The Transcription Factor Network in Embryonic Stem Cells

The virtue of promiscuity?

Debbie van den Berg

The studies presented in this thesis were conducted at the Department of Cell Biology, Erasmus MC, Rotterdam and financially supported by NWO VIDI, ECHO grants and the EuTracc Consortium

Cover: Oct4-centered Protein Interaction Network in Embryonic Stem Cells

Printed by Optima Grafische Communicatie B.V.

The Transcription Factor Network in Embryonic Stem Cells

The virtue of promiscuity?

Het transcriptiefactornetwerk in embryonale stamcellen

Het voordeel van promiscuïteit?

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof.dr. H.G. Schmidt
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 6 oktober 2010 om 15.30 uur

door

Deborah Louisa Carolina van den Berg

geboren te Breda

Promotiecommissie

Promotor:	Prof.dr. F.G. Grosveld
Overige leden:	Prof.dr. J.N.J. Philipsen Prof.dr.ir. D.N. Meijer Dr. J. Gribnau
Copromotor:	Dr. R.A. Poot

Voor ons pa & ons ma

'Het begin is het belangrijkste deel van het werk'

Plato

Contents

Abbreviations

AF	activation function
bp	base pair
CHD	chromodomain helicase DNA-binding
ChIP	chromatin immunoprecipitation
CTD	C-terminal domain
DBD	DNA binding domain
DNA	deoxyribonucleic acid
E	embryonic day
EPI	epiblast
EpiSC	epiblast stem cell
ER	estrogen receptor
ERE	estrogen response element
ERR	estrogen receptor-related receptor
ERRE	ERR response element
ESC	embryonic stem cell
Esrrb	estrogen receptor-related receptor beta
HDAC	histone deacetylase
ICM	inner cell mass
iPS	induced pluripotent stem
kD	kilo Dalton
LBD	ligand binding domain
LIF	leukemia inhibitory factor
mRNA	messenger RNA
MTL	multiple transcription factor-binding locus
NR	nuclear receptor
Oct	octamer binding protein
PcG	Polycomb group
PE	primitive endoderm
PIC	pre-initiation complex
RA	retinoic acid
RNA	ribonucleic acid
RNAi	RNA interference
RNA pol2	RNA polymerase II
TAF	TBP-associated factor
TBP	TATA-binding protein
TE	trophectoderm
TF	transcription factor
TrxG	Trithorax group
TSS	transcription start site

Outline

Embryonic stem (ES) cells are derived from the inner cell mass of blastocyst stage embryos and when cultured *in vitro* can self-renew indefinitely while retaining the capacity to differentiate into derivatives of the three germ layers. These key properties are regulated by a core transcriptional network that revolves around three transcription factors, Oct4, Sox2 and Nanog. Chapter 1 introduces basic aspects of eukaryotic transcription regulation and describes the role of transcription factors in mouse preimplantation development and in maintenance and reinstatement of pluripotency *in vitro*. Whereas for most ES cell transcription factors genomic binding sites and regulated genes have been reported, the scope of their interaction partners remains underinvestigated. Exploring the interactome of ES cell transcription factors, however, can help to elucidate the molecular mechanisms by which they regulate gene expression and potentially leads to identification of novel factors involved in ES cell maintenance. Chapter 3 describes an improved FLAG affinity based protein purification methodology that was used to purify complexes of transcription factors Oct4, Sall4, Dax1, Tcfcp2l1 and Esrrb from relatively small amounts of ES cell nuclear extract. Identification of associated proteins by mass spectrometry analysis resulted in an interactome comprised of 166 proteins, including transcription factors and chromatin modifying complexes with documented roles in pluripotency or self-renewal, but also factors that are novel to the ES cell network. It furthermore demonstrates association of Esrrb with the basal transcription machinery (i.e. Mediator complex, RNA pol2 and TFIID). Chapter 2 reports on the functional implications of the newly identified interaction between Oct4 and Esrrb in regulation of *Nanog* gene expression. Chapter 4 concerns the characterization of Mediator complexes in ES cells and reports reproducible identification of Esrrb in purifications of Mediator complex. Chapter 5 provides a general discussion of the studies presented in this thesis.

1

Introduction

Introduction

Important cell fate decisions during development of an organism are driven by gene expression changes, in which transcription factors play a pivotal role. In this chapter some basic aspects of transcription regulation will be introduced and attention will then be focused on the nature of the earliest fate choices occurring during preimplantation development. The preimplantation embryo is also the source of embryonic stem cells, which, when cultured *in vitro*, can self-renew indefinitely while retaining the ability to differentiate into various tissue types. These hallmark properties of the so called pluripotent cell state rely on signaling and transcription circuits, which will be described in the second part of this chapter. Finally, the role of two key pluripotency transcription factors, Oct4 and Esrrb, in embryonic stem cell maintenance will be discussed in more detail.

Transcription Regulation

The basal transcription machinery

The eukaryotic genome consists of many thousands of genes, which are not all expressed at the same time. Cell type-specific transcriptional programs are set up in a highly controlled manner to support proper cell fate decisions and responses to environmental cues. For a better understanding of the different aspects of gene regulation it is important to take a closer look at some basic principles governing this process. The core promoter directs initiation of transcription and contains several regulative elements that are involved in assembly of the pre-initiation complex (PIC), consisting of basal transcription factors and RNA polymerase II (RNA pol2), which produces all protein coding and most non coding RNAs. These regulative elements are mostly found in focused core promoters, where transcription is initiated from a defined point, as opposed to dispersed promoters that lack a defined transcription start site (TSS)[1]. Only about one-third of vertebrate genes contain focused core promoters, but these genes tend to be highly regulated and tissue-specifically expressed. In contrast, CpG-rich dispersed promoters are mostly associated with housekeeping genes.

The focused core promoter extends from -40 to +40 base pairs relative to the +1 TSS (Figure 1). Binding of basal transcription factor TFIID to the -31/-30 TATA box constitutes the first step in PIC assembly[2]. TFIID is a multi-subunit complex consisting of TATA-binding protein (TBP) and 14 TBP-associated factors (TAFs). TBP binds the TATA box to induce DNA-bending, while several TAFs contact conserved sequences further downstream. TAF1 and TAF2 contact the initiator (Inr), a conserved element surrounding the TSS. The downstream core promoter element (DPE), located from +28 to +33, is recognized by TAF6 and TAF9[1]. Immediately upstream of the DPE, from +18 to +27, the motif ten element (MTE) constitutes another conserved sequence that serves as contact point for the TFIID complex. Binding of TBP to the TATA box is enhanced and stabilized by two additional basal transcription factors, TFIIA and TFIIB. Apart from interacting directly with TBP,

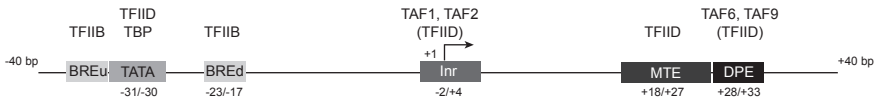


Figure 1. Conserved motifs in focused core promoters
Location of each motif relative to the transcription start site (+1) and motif recognizing factors are indicated. BRE, TFIIB recognition element; u, upstream; d, downstream; Inr, initiator; MTE, motif ten element; DPE, downstream promoter element. Adapted from [1].

TFIIB also binds conserved DNA sequences (BREs for TFIIB recognition elements) found directly upstream and downstream of 10-30% of all TATA boxes. All of these conserved core promoter elements do not merely serve as contact points for the basal transcription machinery, but also influence transcriptional output of sequence specific transcription factors acting upstream of PIC assembly[1].

Once TFIID is engaged with the core promoter and stabilized by TFIIA and TFIIB, RNA pol2 and TFIIF can dock. Association of RNA pol2 with the PIC requires both TFIIB and TFIIF. The final basal transcription factors that join the PIC are TFIIE and TFIIH. The latter contains helicase activity that assists in DNA strand separation and kinase activity that phosphorylates Ser5 of RNA pol2 largest subunit's C-terminal domain (CTD). The CTD of mammalian RNA pol2 contains 52 repeats of a heptapeptide sequence that each harbors 3 possible phosphorylation sites and the extensive possible number of different combinations could play a role in fine tuning RNA pol2 activity. Ser5 phosphorylated RNA pol2 can progress from being bound at the PIC and enter an early phase of elongation. It will, however, pause at a promoter-proximal region until the P-TEFb complex phosphorylates Ser2 of the CTD, allowing transcription elongation through the gene body[3].

The classical view dictates that regulation of gene expression occurs at the step of recruitment of the basal transcription machinery[4]. However, it has now become apparent that a large proportion of eukaryotic genes contain stalled RNA polymerases at promoter proximal regions[5-7]. These polymerases have initiated transcription and produce short 20-50 nucleotide RNA transcripts, but they are incapable of progressing into the elongation phase. Furthermore, in both human fibroblasts and mouse embryonic stem cells (ESCs), stalled RNA polymerases were detected not only downstream, but also upstream of the TSS[6, 8]. Divergent transcription of both the sense and antisense strand was detected, although by an unknown mechanism only RNA polymerase engaged in the sense orientation will proceed towards productive elongation. It thus seems clear that important gene regulatory mechanisms function at post-RNA pol2-recruitment steps.

Chromatin structure

The eukaryotic genome is packaged into chromatin and to gain DNA access, transcription factors and the basal machinery have to overcome this barrier. Chromatin is made up of nucleosomes, consisting of a histone (H2A/H2B/H3/H4) octamer core around which 147 base pairs (bp) of DNA are wrapped. Consecutive nucleosomes are connected via a short stretch of linker DNA, which

can be bound by linker histone H1. These nucleosomal 'beads-on-a-string' can be compacted into higher order chromatin structures, further limiting access of transcription factors and the basal transcription machinery to gene regulatory and promoter elements[9].

Indeed, prior to gene activation promoters of many regulated genes are covered by nucleosomes[10]. Chromatin modifying complexes harboring different enzymatic activities have evolved to overcome these nucleosomal barriers. ATP-dependent chromatin remodelers utilize energy produced by ATP hydrolysis to slide, remove or replace nucleosomes[3]. They can be broadly subdivided into four families: SWI/SNF, ISWI, CHD and INO80[11]. Two main SWI/SNF complexes can be distinguished based on subunit composition, which are conserved from yeast to man. In mouse and human the PBAF complex contains signature subunit Polybromo, whereas BAF is characterized by Arid1a or Arid1b[12]. Exchange of subunits during development creates tissue specific variants of SWI/SNF, which further augments the number of possibilities for combinatorial control of gene expression[11]. ATPase activity is contributed by Brg1 or Brahma subunits and results in nucleosome sliding or ejection, which is predominantly linked to gene activation[10]. SWI/SNF in yeast is found at the -1 nucleosome and therefore may contribute to maintenance of the observed nucleosome free region (NFR) surrounding gene promoters[10].

Mammalian ISWI complexes contain catalytic subunit SNF2H or SNF2L and, similar to SWI/SNF, utilize the enzymatic activity of these subunits to move nucleosomes on the DNA. Unlike SWI/SNF, ISWI complexes are capable of generating regularly spaced nucleosome arrays, which are thought to be required for the formation of higher order chromatin structures[13]. Consequently, ISWI complexes are more frequently linked to negative regulation of transcription[10]. The observed differences in catalytic activity between SWI/SNF and ISWI are thought to reside in a portion of the ATPase domain called helicase-like domain, as exchanging only this part confers remodeling activity of the donor upon the acceptor protein[13].

The CHD family of remodelers consists of nine members that exist as monomers (e.g. Chd1)[14] or as part of multimeric complexes (e.g. Chd3/Chd4 in NuRD)[15]. Besides a SNF2-like helicase-ATPase domain, they contain a chromodomain, which can recognize and bind methylated lysines often found on histone tails. In addition some family members harbor a DNA-binding domain that recognizes A-T rich sequences, a unique feature that sets them apart from other ATP-dependent chromatin remodelers, which generally lack sequence specificity[15].

In the INO80/SWR1 family of remodelers the ATPase domain is split[3]. The most notable activity of both yeast SWR1 and the related mammalian Tip60-p400 complex is their ability to alter nucleosome composition by exchanging histone H2A with the variant H2A.Z[3].

Apart from adjusting nucleosomal positioning, chromatin structure can also be altered by a variety of modifications that target histones. Carried out by a range of different enzymes, these include acetylation, methylation, phosphorylation, ubiquitination and poly-ADP-ribosylation. In most cases enzymes catalyzing removal of these chromatin marks have also been described. Histone modifications mostly take place on histone tails, which protrude from the core histone

octamer. Histone tail modification influences chromatin structure and consequently gene expression in different ways. For one, introduction of a negative charge on a positively charged histone can weaken the interaction with negatively charged DNA and thereby make DNA elements available for transcription factor binding. In addition, specific modifications or combinations thereof can serve as a docking module for secondary effector proteins[16]. For example histone H3K27me3 is specifically targeted by the chromodomain of Polycomb and the H3K4me3 mark found on active promoters is bound by the PHD-finger of TFIID-subunit TAF3[17-18].

Nucleosomal mapping in yeast and human has demonstrated that promoters of actively transcribed genes and, to a minor extent, also non-transcribed genes are devoid of nucleosomes[19-20]. The aforementioned divergent transcription that takes place at most active promoters in both yeast and mammalian cells may be involved in maintaining the nucleosome free region surrounding the TSS, as the length of spacer DNA between the two polymerases equals the size of the NFR[6]. Stalled RNA polymerases are thought to recruit histone modifying enzymes that deposit active histone marks on nucleosomes directly flanking the NFR, which could help to protect against gene silencing mechanisms.

Transcription factors and enhancers

Most chromatin remodeling factors and the basal transcription machinery lack sequence specificity, yet this feature is essential for temporal and spatial control of gene expression patterns. Transcription factors with DNA binding domains (DBD) displaying affinity for particular sequences have evolved to contribute to specificity of gene regulation. The human genome encodes around 2500 DNA-binding domain (DBD) containing proteins, constituting 5-10% of all protein coding genes[21]. Sizes of recognition sequences for these DBDs are typically shorter in eukaryotes than in prokaryotes, as, for example, the average motif length in *Drosophila* is 12.5 bp compared to an average of 24.5 bp in *E.Coli*[22]. This decrease in sequence specificity of individual transcription factors is compensated for by cooperative binding of multiple transcription factors[23]. In addition transcription regulators frequently function in a network of cross-regulatory loops, which increases the complexity of a eukaryotic transcription program[3].

Transcription factor target sites are often located in compacted chromatin and to enable binding, local chromatin structure has to be perturbed. So called 'pioneer' transcription factors, in particular members of the Fox family, are capable of binding compacted nucleosomal arrays *in vitro*[24]. FoxA1 via its C-terminal domain interacts with core histones H3 and H4 to disrupt internucleosomal interactions, resulting in chromatin decompaction[24]. *In vivo* FoxA1 assists in association of estrogen receptor (ER) and glucocorticoid receptor (GR) with their target sites by modulating chromatin structure[25-26].

Transcription factor binding occurs in promoter regions, but also at distal sites upstream, downstream or inside the gene body, contributing to the complexity of mammalian gene regulation. Distal *cis* regulatory elements (or enhancers) can be found as far as 1 mega base away from the corresponding regulated gene[27] and when active are characterized by the presence of

monomethylated H3K4 and binding of histone acetyltransferase p300 and Mediator complex subunit Med1[28]. Furthermore, it was recently demonstrated that bidirectional transcription occurs at active enhancers, similar to divergent transcription observed at promoters[6, 29]. Synthesis of these enhancer RNAs, but not RNA pol2 binding itself, was dependent on presence of the endogenous promoter region, indicative of promoter-enhancer communication. Enhancers can contact promoters directly to regulate gene expression, which is accompanied by formation of chromatin loops, whose presence can be demonstrated using 3C or 4C technology[30-31].

How do transcription factors subsequently induce gene activation? Many gene promoters are covered by nucleosomes and transcription activators may recruit chromatin remodelers to open up promoter chromatin, making it accessible for PIC assembly[32]. Activators can assist in PIC assembly directly by recruiting basal transcription factors or the Mediator complex, whose role in transcription regulation will be discussed in more detail below. Additional steps that can be modulated include promoter escape and initiation of productive elongation[32]. Apart from a role in gene activation, sequence specific transcription factors can function in an analogous manner in gene silencing, by recruitment of repressive chromatin and DNA modifying enzymes or by blocking action of activating complexes.

The Mediator Complex

Structural organization into four submodules

Mediator was originally identified in yeast as a multisubunit complex associated with the CTD of RNA pol2[33-34]. Human Mediator was subsequently isolated as a ligand-dependent co-activator of thyroid hormone receptor (TR)[35] and was independently found associated with liganded vitamin D receptor (VDR) and activation domains of E1A, SREBP, NF-κB[36]. It was shown to be required for activator-dependent transcription on chromatinized templates in an *in vitro* reconstituted system[37-38], but also has an activator-independent role in basal transcription regulation[39-40].

Mammalian Mediator is comprised of 26 subunits that are organized into different modules, making up the head, middle/arm and tail regions of the 1.2 MDa complex (Figure 2). The kinase submodule, consisting of Med12, Med13, Cyclin C and Cdk8, optionally associates via an interaction between Med13 and the tail of Mediator[41]. Individual Mediator subunits are found in most eukaryotes[42]. However, Mediator sequences, in contrast to the basal transcription machinery, have evolved rapidly from yeast to human[43]. Despite these differences in amino acid sequence, there is extensive structural similarity between the yeast and human complex[44] and the four submodule organization additionally is conserved in other eukaryotes[42].

Mediator regulates different steps of the transcription cycle. It plays a role in promoter targeting of the transcription machinery and, although it was initially assumed to form a stable holoenzyme complex with RNA pol2[33], it is now clear that Mediator recruitment can precede RNA pol2 binding[45]. Mediator subsequently enhances targeting of RNA pol2 and basal transcription factors to the promoter and stabilizes the PIC[46-47]. The head module forms an important interaction surface for components of the basal transcription machinery. An RNA pol2-TFIIF complex was

shown to bind recombinant head[48] and interaction of head module subunits with TBP stabilizes an open conformation that would enable subsequent RNA pol2 binding[49]. The critical role of the head module in transcription is substantiated by a temperature sensitive *med17* mutant in yeast, in which expression of virtually all genes was shown to be affected[50].

Kinase submodule

The Med12-Med13-Cdk8-Cyclin C submodule reversibly associates with core Mediator and genetic ablation of individual subunits in *S. cerevisiae* has highly similar effects on gene expression, suggesting that the Cdk8-module forms a distinct functional entity[52]. Whereas Cdk8 is not essential for yeast viability, *Cdk8* null mutation in mice results in early (E2.5) embryonic lethality[53]. Upon association of the Cdk8 module with core Mediator it can induce a structural shift that impedes RNA pol2 recruitment, activation and also reinitiation events[41]. Cdk8-Mediator complex is unable to activate transcription *in vitro*[54]. In addition, both Med12 and Med13 were identified in a genetic screen aimed at isolating novel Polycomb group (PcG) proteins in *Drosophila*, in which they were demonstrated to be required for repression of the HOX gene *Ubx*, albeit independent of Cdk8/cyclin C kinase activity[55]. The kinase submodule additionally plays a role in repression of RE1-silencing transcription factor (REST) target genes in HeLa cells by recruitment of H3K9 histone methyltransferase G9a[56]. Derepression of neuronal genes was specifically observed in Med12-

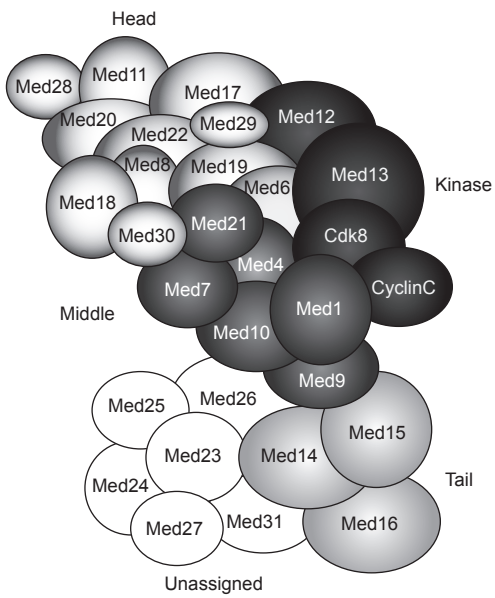


Figure 2. Composition of human Mediator complex.
Subdomains are designated by different shades of grey, unassigned subunits in white. Adapted from [51].

depleted cells and not upon RNAi-mediated reduction of Cdk8. These data indicate that the kinase submodule can exert repressive effects on gene transcription, however, these seem to be mostly independent of its kinase activity incorporated in Cdk8. An exception is Cdk8-mediated phosphorylation of the Cdk7/cyclin H kinase module in TFIIH, which interferes with its ability to phosphorylate RNA pol2 CTD and activate transcription[57].

In contrast, in numerous studies Cdk8 has been implicated in transcription activation[58]. Cdk8-Mediator is recruited upon hormone-induced transcription activation of thyroid hormone receptor (TR) target genes and downregulation of Cdk8 by RNA interference results in a reduction of RNA pol2 promoter occupancy and a concomitant decrease in transcript levels[59]. A role for Cdk8-Mediator post-RNA pol2 recruitment was implied in studies of the serum response in colon cancer cells[60]. Reduction of Cdk8 decreased expression of several serum-responsive immediate-early genes, without affecting RNA pol2 promoter binding. Promoter targeting of Mediator, Cdk7 and Cdk9, however, was reduced, resulting in decreased levels of Ser5 and Ser2 CTD phosphorylated RNA pol2, which ultimately affected the rate of transcription elongation[60].

Apart from components of the basal transcription machinery, Cdk8/cyclin C has also been shown to phosphorylate the linker region of Smads 1, 2 and 3, downstream effectors of BMP (Smad1) and TGFβ signaling (Smad2,3) respectively[61]. Smad linker phosphorylation is dependent on Smad4-mediated incorporation into transcription complexes and is necessary for efficient transcription activation, but simultaneously targets Smads for ubiquitination-mediated degradation[61]. Cdk8/cyclin C plays a related role in limiting the duration of the Notch signal[62]. It is recruited to Notch target genes by coactivator MAM and subsequently phosphorylates Notch intracellular domain (ICD), an event that triggers ICD ubiquitination by E3 ligase Fbw7, followed by degradation[62]. Although isolated Cdk8/cyclin C is capable of phosphorylating Smads and Notch ICD *in vitro*, it remains to be determined whether Med12 and Med13 subunits are required for this activity *in vivo*. Possible functional divergence of the four Cdk8-module subunits is indicated by phenotypical differences observed between Drosophila *Cdk8* and *Cyclin C* mutants compared to *Med12* and *Med13* mutants, in contrast to the largely overlapping phenotypes between *Cdk8* and *Cyclin C*, as well as between *Med12* and *Med13* mutants[63]. The Cdk8-submodule thus exerts both repressive and activating effects on transcription, mediated by structural changes and phosphorylation events, which affect components of the basal transcription machinery, but also secondary factors.

Interaction with transcription factors

Several subunits of core Mediator reportedly interact with specific transcription factors. Med1, at 220 kDa, is the largest subunit of Mediator and provides a general interaction surface for nuclear receptors. It contains two LXXLL motifs that each, in a ligand-dependent manner, can associate with a hydrophobic cleft in the activation function-2 (AF-2) domain of nuclear receptors (NRs), including thyroid hormone and estrogen receptors (ER), peroxisome proliferator-activated receptor (PPAR), retinoic acid (RAR), rentinoic X (RXR) and vitamin D receptor (VDR)[64]. *Med1-/-* embryos die at E8.5-E12.5 with heart failure, abnormal neural development and growth retardation[65]. However,

Med1 is not generally required for cell survival and, in fact, biochemical analysis of Mediator complexes has detected Med1 presence in only a fraction of them[66]. Its core complex-association via interaction with Med7 is enhanced by MAPK-ERK mediated phosphorylation, which in itself is augmented by NR-signaling[67].

The role of nuclear receptor-Mediator interaction was further addressed by mutating both LXXLL motifs in Med1[68]. Mice were viable, but developed mammary gland defects in puberty due to a reduction in ER-target gene expression. The limited effects of LXXLL mutations can be explained by alternative modes of Mediator recruitment to NR-target genes, via LXXLL-independent Med1 interactions[69], via co-activators (e.g. PGC-1α, CCAR1[70-71]) or via interactions with other Mediator subunits[69]. Indeed, LXXLL motifs are also present in Med25 and are required for Med25 recruitment to RAR-target genes[72]. Furthermore, interactions with multiple Mediator subunits have been described for the VP16 activation domain (Med17, Med25)[73] and glucocorticoid receptor (GR) (Med1, Med14)[74].

Activator-dependent recruitment of Mediator can enhance PIC assembly or RNA pol2 CTD phosphorylation, as was described. However, it has also become clear that activator-Mediator interaction can induce structural changes, likely enabled by the propensity of intrinsically disordered regions (IDRs) in several subunits[75]. Interaction of the transcription factor p53 N-terminal activation domain with Med17 induces a structural shift that exposes a binding pocket for RNA pol2[75]. In the absence of the p53 activation domain, Mediator is still contacted at its Med1 subunit by the C-terminal domain of p53, however, this induces a structure that lacks the RNA pol2 binding pocket and therefore is inactive. The active conformation appears to be a common feature of activator-bound Mediator and is not required for RNA pol2 recruitment, but rather stimulates its transition into productive elongation[75].

Patterning in Preimplantation Development

Fertilization of an oocyte results in formation of a zygote that will give rise to all future extraembryonic and embryonic lineages. In the 4.5 days prior to implantation in the uterine wall, the murine zygote undergoes a number of mitotic divisions and cleavages and two important cell fate decisions that set aside embryonic from extra-embryonic lineages (Figure 3). These lineage specifications are presumed to be controlled by cell polarity, cell position and gene expression changes[76]. They result in a blastocyst stage embryo that is comprised of three distinct cell lineages: a surrounding layer of trophectoderm (TE) that will form the placenta, the epiblast (EPI), which will give rise to the embryo proper and a layer of primitive endoderm (PE) that separates the epiblast from a fluid filled cavity and that, upon implantation, will develop into extraembryonic visceral and parietal endoderm[76].

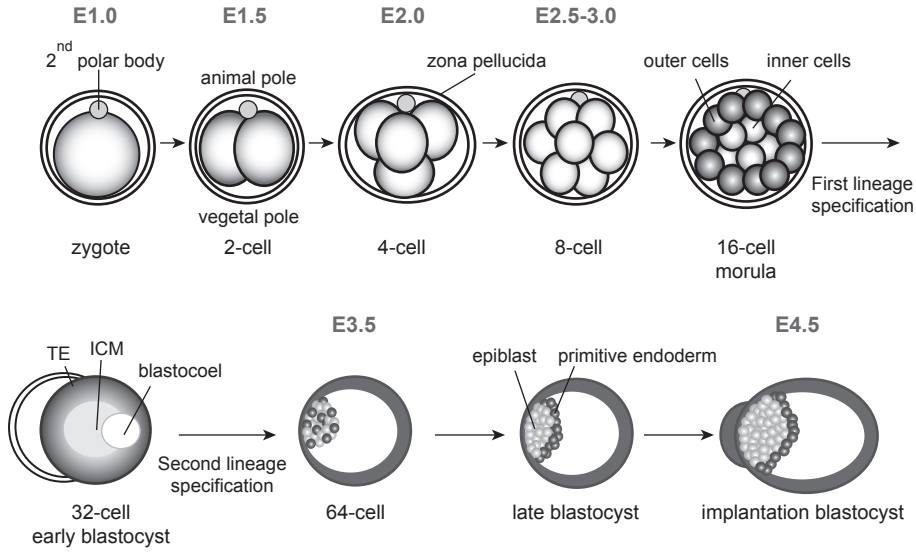


Figure 3. **Mouse preimplantation development**
Successive cleavage stages and lineages choices are outlined.

The first lineage specification

Creating a connection with the bloodstream of the mother to provide nutrients is a vital step in early mammalian development. Therefore the first important cell fate choice to be made is specification into either trophectoderm or inner cell mass (ICM). The TE will be responsible for invading the uterine wall during implantation and will contribute to the embryonic part of the placenta, while the ICM will give rise to the embryo and supportive extraembryonic tissue[77]. Two different models exist to explain how the choice between TE and ICM is made and these consider either cell position or cell polarity to be the fate-determining factor[78]. However, both models would eventually require the induction of lineage specific transcription factors to lock cell fate.

When is the choice between TE and ICM made and what are the molecular mechanisms involved? Lineage tracing studies have suggested that individual blastomeres display a preference towards either TE or ICM as early as the 4 cell stage, that correlates to their position with respect to the animal-vegetal axis, as well as to their histone H3 R2, R17 and R26 methylation status and expression level of the responsible histone methyltransferase CARM1[79-80]. Such a preference, however, is only observed in half of the 4-cell stage embryos and is therefore unlikely to play a decisive role in fate choice.

At the 8-cell stage the individual blastomeres begin to compact and become more adherent. A differential distribution of cellular components toward the apical or basal end of the cell is also detected at this point, resulting in cell polarization[77]. Cell polarity plays a potential role

in determining cell fate, as, upon cell division, it can lead to asymmetric distribution of lineage determining factors among the daughter cells. An example of such a factor is the *Cdx2* mRNA, encoding the trophectoderm reinforcing transcription factor Cdx2, which preferentially localizes to the apical cell membrane and, upon cell division and cleavage perpendicular to the apical-basal axis, will become unevenly distributed among the daughter cells[81].

The fourth round of cleavage gives rise to a pool of inner and outer cells and, although all 16 blastomeres at this stage can still develop into any of the three final blastocyst lineages[82-83], cells positioned on the outside of the morula display a preference towards formation of trophectoderm (TE), while inner cells will more likely contribute to the future inner cell mass (ICM)[84]. A definitive commitment to TE is made at the 16- to 32-cell transition, exemplified by the finding that re-aggregating an embryo at this stage from outer cells only does not result in formation of an ICM[82]. In the same study embryos re-aggregated from inner cells gave rise to morphologically normal blastocysts, however, these did not implant, possibly due to delayed trophectoderm development. In addition, single cell expression analysis on blastomeres revealed a major change in transcript signature occurring from the 16- to the 32-cell stage, enabling classification of individual cells as TE-like or ICM-like[85].

Alterations to the gene expression profile that are not readily reversible would be required to accomplish lineage commitment and, indeed, a number of transcription factors have been implicated in TE fate choice. Tead4 is a member of the TEAD/TEF family of transcription factors and is the most upstream acting factor promoting TE specification identified so far[86]. *Tead4*-/- embryos do not express trophectoderm markers and fail to develop a blastocyst cavity[86]. Nuclear localization of the Tead4 coactivator Yap is regulated by the cell-cell contact inhibition Hippo signaling pathway, resulting in Tead4 activity specifically in the outer cells of the morula stage embryo[87]. Tead4 downstream targets include the trophoblast lineage promoting transcription factors Cdx2 and Gata3[86, 88]. By relaying cell position cues onto downstream effectors Tead4 plays a role in the TE commitment of outer cells in the 32-cell stage early blastocyst.

Lineage bifurcation is further reinforced by reciprocal transcription repression of Cdx2 and Oct4, a POU-domain transcription factor important for development of the ICM[89]. However, mutually exclusive Oct4-Cdx2 expression appears not to be the only determining factor in ICM vs. TE specification, as TE-restricted expression of Cdx2 is initiated in the absence of Oct4, Oct4 protein continues to be present in Cdx2 expressing TE cells until late blastocyst stage and *Cdx2* mutant cells contribute to both TE and ICM in re-aggregated chimeric embryos[88, 90-91]. These observations suggest that other factors have to be involved in setting up the definitive lineage-specific transcription programs. Candidates include the transcription factors Id2 and Sox2, which, at the time of TE specification, were found to be the strongest induced factors out of 48 genes analyzed on a single cell level[85]. An inverse correlation between *Id2* and *Sox2* transcripts can already be detected in 16-cell stage blastomeres, where *Id2*^{high}/*Sox2*^{low} cells are found on the outside of the morula and classify as TE-like, while *Id2*^{low}/*Sox2*^{high} cells represent the inner, ICM-like cells. Sox2

may thus play an important role in ICM specification, although one has to keep in mind that Sox2 function is additionally regulated by nuclear import/export and that maternally contributed Sox2 protein continues to be present until the late blastocyst stage[92].

Expression profiles set up during these initial phases of lineage specification are subsequently locked into place by epigenetic modifications (e.g. histone modifications and DNA methylation), so that future generations remain committed to the fate choice made[93]. For example the *Elf5* gene becomes methylated and thereby stably silenced in the ICM, whereas it remains actively transcribed in the TE, where it upregulates expression of *Cdx2* and *Eomes*[94]. The molecular mechanism of TE specification thus seems to involve positional information forwarded by the Hippo signaling pathway, resulting in activation of Tead4 in outer cells of morula stage embryos. Tead4 downstream targets include TE-promoting transcription factors, which secure the lineage-specific transcription program via a number of positive and negative feedback loops.

The second lineage specification

The second lineage specification occurs within the ICM and separates epiblast (EPI) from primitive endoderm (PE). The epiblast will give rise to the embryo proper and to some extraembryonic tissues, such as the extraembryonic mesoderm. Primitive endoderm at the time of implantation forms a layer on top of the epiblast and will soon thereafter diverge into visceral and parietal endoderm. Visceral endoderm especially holds an important instructive role later in development when it aids in patterning of the embryo[84].

Specification of the two different lineages within the ICM, as with TE commitment, is characterized by mutually exclusive expression of fate-determining transcription factors. In contrast to TE specification, lineage specific expression profiles within the ICM are laid out prior to cell repositioning, resulting in a 'salt-and-pepper' mixed population of PE and epiblast cells in the E3.5 blastocyst[95]. Initial cell position at earlier stages may play a role in segregating epiblast from PE, as was implied by live cell imaging and cell tracking[96]. Inner cells that constitute the ICM are generated during successive waves of cleavage divisions, first in the 8- to 16-cell transition, followed by a second wave in the 16- to 32-cell transition. Tracking cell fates of these successively generated inner cells revealed a preferential contribution to epiblast by ICM cells generated in the first wave and a preference towards PE-fate in cells generated by the second wave. However, similar tracing experiments by an independent laboratory failed to detect any relationship between time of inner cell generation and eventual contribution to epiblast or PE[97].

Initial changes in gene expression profiles reminiscent of lineage segregation can be detected within individual cells of the ICM population at the 32-cell stage[85]. Although at later (64-cell) stages epiblast is distinguished from PE based on expression of either Nanog (EPI) or Gata4/Gata6/Sox17 (PE)[95, 98-99], it is the ligand/receptor encoding *Fgf4/Fgfr2* gene pair that ranks highest for inversely correlated expression within the ICM of early 32-cell blastocysts[85]. *Fgf4* expression will subsequently become restricted to the epiblast, while *Fgf2r* will be exclusively expressed in the PE. The early detectable anti-correlation implies a role for FGF signaling at the onset of

epiblast versus PE specification. Indeed, several lines of evidence support such a role. *Fgf2r*-/- embryos fail to develop a primitive endoderm layer by E4.5 and null mutation of *Grb2*, an adaptor that functions downstream of several receptor kinases, including Fgf2r, results in an E3.5 ICM comprised solely of Nanog-positive epiblast cells[95, 100]. Treatment of *in vitro* cultured 8-cell stage embryos with a combination of an FGF-receptor inhibitor and a Mek-inhibitor similarly results in expansion of the Nanog-positive compartment with a near absence of Gata4-positive PE cells[97, 101]. Comparable effects are also observed when embryos are treated with an FGF receptor inhibitor only, which results in downregulation of PE transcription factors Gata4 and Sox17 and concomitant upregulation of epiblast transcription factors Nanog, Esrrb, Klf2 and others[85]. In reverse, addition of exogenous Fgf4 causes all ICM cells to adopt the PE fate[97]. Differential expression of *Fgf4* by future epiblast cells and its receptor *Fgf2r* by prospective PE cells thus provides an important framework for determining cell fate within the ICM.

Communication between epiblast and PE cells, most likely involving FGF signaling, remains important during later phases of preimplantation development, since *Nanog*-/- embryos that do not establish the epiblast also fail to develop primitive endoderm *in vivo*[102]. Expression of Nanog in the epiblast is necessary for these cells to acquire true pluripotency, but once this pluripotent ground state has been instated, Nanog becomes dispensable and is progressively downregulated at implantation[102-104]. To what extent PE-specific factors Gata4, Gata6 and Sox17 play an essential instructive role in early lineage segregation within the ICM remains unclear. Overexpression of any of these factors in embryonic stem cells induces differentiation towards extraembryonic endoderm, but embryos carrying a null mutation for either of these individual factors can develop until postimplantation stages, indicating possible functional redundancy[105-106].

Pluripotency *in Vitro*

Cytokines and signaling

Mouse embryonic stem cells (mESCs) are derived from the ICM of blastocyst stage embryos and were first isolated almost 30 years ago[107-108]. When cultured under the right conditions they can self-renew indefinitely whilst remaining pluripotent (i.e. retaining the capacity to form any cell type of an organism). To what extent cultured ESCs remain pluripotent can be determined upon *in vitro* differentiation or injection of mESCs into immunodeficient NOD/SCID mice, resulting in formation of teratomas. In the case of mouse ESCs a more stringent assessment of pluripotency can be made following injection into blastocyst stage embryos. In the resulting chimera, ESC-derived cells should contribute to tissues originating from all three germ layers. As an ultimate test of pluripotency mESCs should be capable of tetraploid complementation: when combined with a tetraploid (4n) morula or blastocyst, which can only contribute to extraembryonic tissues, the injected diploid mESCs should give rise to a fully developed organism[109]. Although mESCs merely contribute to embryonic tissues when introduced into blastocyst stage embryos, they do have the ability to form extraembryonic trophectoderm and primitive endoderm *in vitro*[110-111].

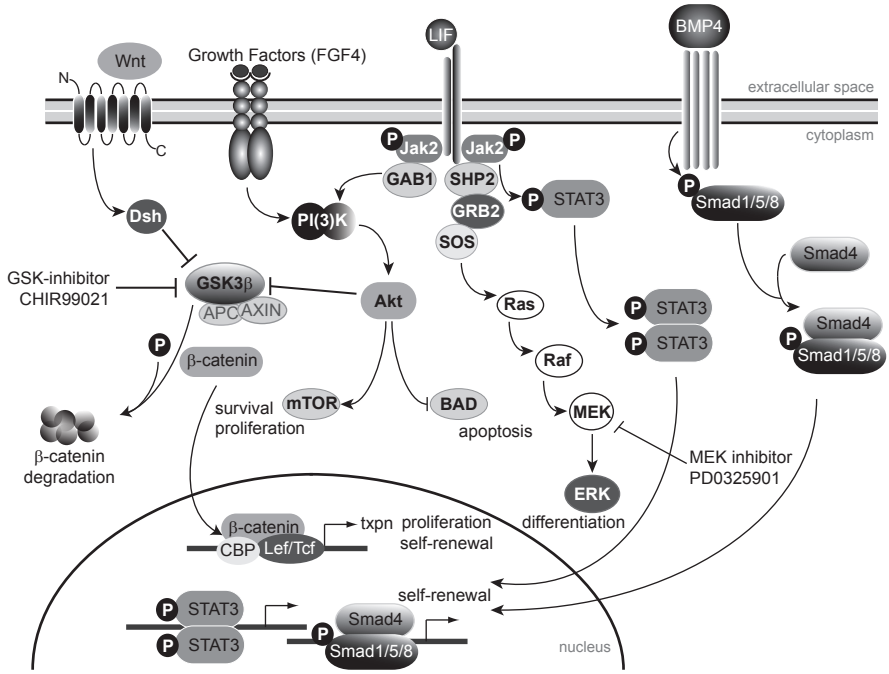


Figure 4. Signalling pathways controlling murine ES cell self-renewal
LIF signal is transduced via PI-3-kinase, Mek and Stat3 pathways, BMP4 acts via Smad1/5/8. Nuclear Stat3 and Smads activate transcription of pluripotency and self-renewal genes. Points and mode of action of 2i components are indicated. Adapted from [128].

Mouse ESCs are cultured in the presence of serum and leukemia inhibitory factor (LIF). Serum can be replaced by bone morphogenetic protein-4 (BMP4), a secreted signaling polypeptide of the TGF-β superfamily[112]. BMP dimers bind to a heterodimeric transmembrane receptor containing a cytoplasmic serine/threonine kinase domain (Figure 4). Ligand binding results in recruitment and phosphorylation of Smad1, -5 or -8, which subsequently associates with Smad4 and translocates to the nucleus[113]. In ESCs Smad4-Smad1/5/8 dimers induce expression of inhibitor of differentiation (*Id*) genes, which in turn act to block differentiation into neuronal lineages[112]. Constitutive expression of Ids alone was sufficient to bypass the demand for BMP signaling.

In addition to BMP4, ESCs require stimulation of several intracellular signaling pathways by LIF, a cytokine first purified from rat liver cells conditioned medium[114]. In preimplantation development LIF is produced by trophectoderm cells, while it receptor is expressed by ICM cells[115]. Deletion of LIF or its receptor, however, does not result in early embryonic lethality, suggesting that other pathways may be involved in maintaining the transient pluripotent state in the developing embryo[116-117]. LIF binds a heterodimer of LIF receptor β and gp130, resulting in the activation of

downstream signal transduction cascades (Figure 4)[109]. Of these the canonical Jak/Stat pathway was shown to be sufficient to mediate the effect of LIF signaling, as activation of Stat3 alone could keep mESCs in an undifferentiated state, whereas expression of a dominant negative form of Stat3 promoted ESC differentiation[118-119]. Stat3 mainly acts to induce expression of Kruppel-like transcription factor Klf4[120]. A parallel circuit of LIF signaling has recently been identified that acts via the phosphatidylinositol-3-OH kinase (PI(3)K)-Akt and MAPK signal transduction cascades to regulate expression and nuclear localization of Tbx3[121]. Both Klf4 and Tbx3 are transcription factors that are integrated with the core transcriptional circuitry of ESCs and they directly regulate expression of the core transcription factors Sox2 and Nanog respectively[121]. Overexpression of Klf4 and Tbx3, similar to Nanog, renders ESCs independent of exogenously added LIF, suggesting these are important downstream targets of the LIF signal[103, 121].

Despite the seeming necessity of BMP and LIF to sustain ES cell self-renewal, it has recently become apparent that ESCs can be maintained in a pluripotent state in the absence of these two cytokines when cultured in the presence of inhibitors targeting FGF receptor kinase activity, its downstream effector Erk and glycogen synthase kinase 3 (GSK3)[122]. These culture conditions are referred to as 3i, or alternatively, when a combination of the Erk and GSK inhibitor is used, 2i. In conventional culture conditions BMPs and, to some extent, LIF revert the inclination of ES cells to differentiate in response to auto- and paracrine FGF4 signaling[123]. The ability of ES cells to self-renew and remain pluripotent in the absence of exogenous cues suggests they can remain in a so called 'ground state' *in vitro*, in which in potential they are fully equivalent to epiblast cells in the preimplantation embryo[124].

ESCs, however, have acquired the ability to self-renew indefinitely, a feature that is obsolete in the developing embryo. They are further characterized by an exceptionally short cell cycle of 8-12 hours and the near complete absence of cell cycle checkpoints[125]. Single cell transcriptome analysis by RNA-sequencing has shown that during establishment of ESCs from ICM outgrowths, expression levels of a number of genes change dramatically[126]. One of the upregulated genes encodes for Eras, a constitutive active form of Ras specifically expressed in ESCs, that activates the PI(3)K-Akt pathway to stimulate proliferation[127]. *Eras*-/- ESCs grow more slowly than their wildtype counterparts, but remain pluripotent, as shown by their ability to contribute to chimeras. Although ESCs to some extent adjust their gene expression profile to adapt to culture conditions, they do retain the full developmental capacity of their *in vivo* equivalents.

Since the first isolation of ES cells from mouse blastocysts, ESC lines have been derived from a number of species, including avian, monkey, human and more recently rat and canine (Table 1)[129]. In most cases establishing an ESC line was not trivial and greatly depended on finding the right culture conditions. Although primate and human ESCs were first isolated in conditions similar to mouse ESCs (i.e. on a feeder cell layer in serum-supplemented medium), it became apparent only later that, for self-renewal, they rely on FGF and TGF-β/Activin/Nodal signaling[130]. These cytokines would induce differentiation of mouse ESCs, while the reverse is true for the effect of BMP signaling

Table 1. Characteristics of different pluripotent stem cell lines

Stem cell line	Growth factors	Cell surface markers	Other markers	Differentiation potential	References
Mouse ES	LIF/BMP4; 3i; 2i	SSEA-1	Oct4, Sox2, Nanog, Klf4, Tbx3, Zfp42	EB, Teratoma, Chimera, GT	107-108
Rat ES	3i/LIF; 2i/LIF	SSEA-1	Oct4, Sox2, Nanog, Klf4, Zfp42, Dppa3	EB, Chimera, GT	139-140
Canine ES	LIF/bFGF	SSEA-1^{low}, SSEA-3, SSEA-4, TRA1-60, TRA1-81	Oct4, Sox2, Nanog, Zfp42, Gbx2	EB, Teratoma	141
Monkey ES	bFGF/Activin	SSEA-3, SSEA-4, TRA1-60, TRA1-81	OCT4, NANOG, ZFP42	EB, Teratoma, Chimera	142-145
Human ES	bFGF/Activin	SSEA-3, SSEA-4, TRA1-60, TRA1-80	OCT4, SOX2, NANOG, ZFP42, TDGF1	EB, Teratoma	146-147
Mouse EpiSC	bFGF/Activin	SSEA-1	Oct4, Sox2, Nanog, Fgf5, T	EB, Teratoma	134-135
Mouse iPS	LIF/BMP4; 2i/LIF	SSEA-1	Oct4, Sox2, Nanog, Klf4, Zfp42, Dppa3	EB, Teratoma, Chimera, GT, TC	148-153
Human iPS	bFGF/Activin	SSEA-3, SSEA-4, TRA1-60, TRA1-80	OCT4, SOX2, NANOG, KLF4, ZFP42, TDGF1	EB, Teratoma	154-155
Monkey iPS	bFGF/Activin	SSEA-4, TRA1-60, TRA1-80	OCT4, SOX2, NANOG, SALL4, TDGF1	EB, Teratoma	156

EB, embryoid body; GT, germline transmission; TC, tetraploid complementation. Modified from [129].

on human ES cells[131]. BMP signaling in hESCs therefore receives repressive input from both FGF and TGF-β/Activin/Nodal signaling pathways[132-133]. In addition, the TGF-β/Activin/Nodal pathway, by means of downstream Smad transcription factors, directly regulates expression of pluripotency transcription factor *NANOG* to stimulate self-renewal[132-133]. Constitutive overexpression of *NANOG* bypasses the need for both FGF and TGF-β/Activin/Nodal in hESC cultures.

In many respects hESCs are more similar to the recently isolated mouse epiblast stem cells (EpiSCs) than to mESCs[134-135]. Compared to the latter both hESCs and mEpiSCs have a more flattened morphology and, like hESCs, mEpiSCs are dependent on FGF and TGF-β/Activin/Nodal signaling for self-renewal *in vitro*. Furthermore, hESCs and mEpiSCs show similarities in epigenetic status at several pluripotency and developmental genes and female cells in most cases have undergone X inactivation, although recently pre-X inactivation hESC lines have been estabilshed under physiological oxygen conditions[135-136]. Unlike hESCs that are derived from pre-implantation blastocysts, mEpiSCs are derived from the E5.5 post-implantation embryo and, even though they are capable of differentiation into the three germ layers *in vitro* and of teratoma formation *in vivo*, they are unable to contribute to chimeras when aggregated with morula stage blastomeres or injected into blastocyst stage embryos[134-135]. mEpiSCs alternatively can be derived relatively easily from mESCs by continuous passaging under EpiSC culture conditions[137]. They are thought to represent a 'primed' state of pluripotency, as they already express a number of lineage specific transcription factors (*Fgf5*, *T*), have lost expression of several key pluripotency factors (e.g. *Klf2*, *Klf4*, *Zfp42*) and female cells have inactivated one X chromosome[124]. Furthermore, it has been

suggested that they cannot be propagated in ground state conditions of 2i/LIF, as these result either in rapid differentiation or cell death, although the Schöler lab has succeeded in reverting EpiSCs into germ-line competent ESC-like cells by continuous culture in the presence of a MEK-inhibitor, GSK3β inhibitor and LIF (essentially 2i/LIF)[137-138].

2i or 3i culture conditions supplemented with LIF have proven to be essential for successful derivation of germ-line competent rat ES cells, an important methodological advance that provides new opportunities for the use of rats as model organism in pharmacological and physiological studies[139-140]. In addition it offers new possibilities for the derivation of ES cell lines from a range of different species for which up till now attempts to establish ES cell cultures have been unsuccessful.

Transcriptional Control of Pluripotency

Apart from extrinsic signals controlling self renewal or differentiation, embryonic stem cells also rely on an intrinsic network of transcription factors to maintain their pluripotent state[157]. Three key players in this network are the transcription factors Oct4, Sox2 and Nanog[158]. Each of these factors is highly expressed in the ICM of the developing embryo and in the absence of Oct4 or Nanog, pluripotency in the ICM does not develop[92, 103, 159]. In mouse ES cells the dose of Oct4 is critical, as more than 50% increase or decrease in Oct4 protein level results in differentiation into primitive endoderm and mesoderm or trophectoderm lineages respectively[111]. Nanog becomes dispensable once the pluripotent ground state has been installed, as deletion of *Nanog* in ESCs does not interfere with their self renewal capacity or ability to differentiate along different lineages upon blastocyst injection[104]. In wild type ES cell colonies *Nanog* is expressed in a mosaic fashion with large variability in expression level between individual cells[104]. Although ESCs in which *Nanog* has been downregulated can revert to a Nanog^{high} state, these cells are more prone to differentiate. *Sox2*-null ESCs, similar to Oct4 depleted ESCs, differentiate into trophectoderm-like cells[160]. This phenotype can be rescued by artificial expression of Oct4, indicating that an important role of Sox2 in ESCs involves the maintenance of Oct4.

There is extensive overlap in genomic binding sites of Oct4, Sox2 and Nanog[161-164]. Motif discovery analysis based on genome wide chromatin immunoprecipitation (ChIP) assays of all three transcription factors identified a common binding motif that strongly resembles a composite Oct/Sox site[162, 164]. Binding of the three transcription factors, depending on the target, correlates to either activation or repression of nearby genes[161, 164-165].

The transcription factor circuitry of ES cells is also closely connected to the several signal transduction pathways that govern pluripotency. *Nanog*, for example, is an important downstream target of the LIF signal and overexpression of Nanog enables ES cells to self renew in the absence of exogenous LIF[103, 121]. Genomic binding sites of Stat3, downstream transcription effector of the Jak/Stat branch of LIF receptor signaling, show a great deal of overlap with Oct4/Sox2/Nanog binding sites and Stat3 occupancy of communal Oct4-Stat3 sites is dependent on Oct4[162]. More strikingly, the same study identified a joint Oct/Sox motif as most prevalent binding motif for Smad1, one of the Smad transcription factors functioning downstream of BMP4 signaling. Oct4 is

required for targeting of Smad1 to Oct4-Smad1 co-occupied binding sites[162]. Nanog on the other hand provides negative feedback to the BMP4/Smad signaling pathway by directly interacting with Smad1, thereby preventing its association with transcription coactivator p300[166].

In recent years several additional transcription factors have been identified as crucial players in the regulation of ES cell self renewal and pluripotency. These include FoxD3[167], Tbx3, Tcl1[168], Klf5[169-170], Sall4[171], Nac1, Zfp281[172] and the nuclear receptors Dax1[173] and Esrrb[168]. Downregulation of these factors in embryonic stem cells by RNAi (Tcl1, Tbx3, Esrrb, Dax1, Sall4, Nac1, Zfp281) or genomic deletion (FoxD3, Dax1) results in overt ESC differentiation. Genome wide binding studies have detected a strong overlap in binding sites of individual factors[162-163, 171, 174]. Both studies by Kim *et al.* and Chen *et al.*, who looked at genomic binding sites of a distinct, but overlapping set of transcription factors, identified sites that were occupied by several of the transcription factors, so called multiple transcription factor-binding loci or MTL[162-163]. These correlate with active transcription of the corresponding gene targets, while genes bound by a single transcription factor are most likely repressed in ES cells and may become activated upon differentiation[163]. A similar clustering of transcription factors was found to occur in human hepatocytes and likewise, binding of multiple factors correlated with active transcription of the associated genes[175]. Computational modeling of multiple transcription factor binding loci in *Drosophila* furthermore was able to accurately predict enhancer activity of these elements during different stages of mesoderm differentiation[176]. In analogy to these predictions, MTL in mESCs containing Nanog, Oct4 and Sox2, but not n-Myc or c-Myc were shown to accommodate ES cell specific enhancer activity in reporter assays, a finding strengthened by observed co-occupancy of Nanog/Oct4/Sox2 MTL by the known enhancer binding histone acetyltransferase p300[28]. A strong correlation between Oct4/Sox2 containing MTL and p300 binding is further confirmed by the identification of a composite Oct/Sox motif as most enriched consensus sequence in the p300 ChIP data set[162].

In addition to co-occupancy of binding sites, interactions in solution between several of these transcription factors have also been described. Oct4, apart from its well described co-operative DNA binding partnership with Sox2[177], was found to interact with Sall4, Zfp42, Zfp219, NF45, Sp1 and Nac1[172]. Transcription factors described to interact with Nanog include Oct4, Sall4, Sall1, Nac1, Dax1, Esrrb, Zfp281, Zfp198, NF45, Sp1 and REST[172]. To what extent these protein-protein interactions play a role in genomic targeting remains under-investigated and is a subject investigated in this thesis.

From the above discussed genome wide binding studies it has become apparent that Oct4 is at the center of the transcriptional circuitry in ES cells, as a general consensus motif deduced from binding site data of multiple transcription factors strongly resembles an extended octamer sequence[163]. Furthermore, combinatorial binding of genomic sites by a range of factors seems to be evolutionary conserved and may provide additional means to regulate or fine-tune target gene transcription.

ES cell chromatin

Open chromatin structure & hypertranscription

In addition to the action of sequence specific transcription factors, gene activation or repression is regulated by local chromatin structure. The unique properties of ES cells (self-renewal and maintenance of pluripotency) require the genes of key lineage regulators to be repressed, but at the same time permissive to activation upon differentiation inducing signals. The chromatin structure may reflect these demands and indeed several lines of evidence indicate that global chromatin structure in ES cells is more decondensed compared to differentiated cell types. For example, mESCs were reported to contain less heterochromatin foci than neural progenitor cells (NPCs) and fluorescence recovery after photobleaching (FRAP) measurements of heterochromatin protein-1α (HP1α) and linker histone H1 demonstrated significantly higher mobility of these proteins in ESCs compared to NPCs[178]. Relative to various tissue-specific cell types, mESCs have a shorter nucleosome repeat length and lower linker histone H1 content, which both are features of decompacted chromatin[179-180]. A difference in chromatin compaction additionally was observed by electron spectroscopic imaging of blastocyst stage embryos, where epiblast cells were characterized by a dispersed and decondensed chromatin structure, while lineage committed cells of the trophectoderm and primitive endoderm contained blocks of compact chromatin[181]. Formation of an open chromatin state in epiblast cells correlates with the establishment of pluripotency, since Oct4-/- ICM cells showed increased chromatin compaction[181].

The relatively decondensed state of ES cell chromatin is accompanied by global hyperactive transcription of exonic, but also intronic and intergenic regions and a concomitant elevated expression of general transcription factors and chromatin remodeling complexes[182]. As will be discussed later, several chromatin remodelers and Polycomb group (PcG) class chromatin modifiers act as repressors of key developmental genes. Chromodomain-helicase-DNA-binding protein 1 (Chd1), however, seems to be required for the maintenance of an open chromatin structure[183]. It binds at or in close proximity of transcription start sites, where it colocalizes with active histone marks (H3K4me3) and RNA pol2. Surprisingly, downregulation of Chd1 by RNAi did not cause major disturbance to the ES cell transcriptome, however, it did result in a significant increase in the number of heterochromatic foci .

Bivalent domains

In the relatively open chromatin environment of ESCs, genes encoding key developmental regulators need to be stably silenced, as premature expression would lead to differentiation. These genes were shown to be bound and repressed by PcG protein complexes, PRC1 and PRC2, in both human and mouse ES cells and a substantial fraction of them is co-occupied by Oct4, Sox2 or Nanog[184-185], of which Oct4 was shown to be important for genomic targeting of PRC1[186]. PcG class proteins were first identified in *Drosophila*, where they are responsible for stable maintenance of transcription repression of *Hox* gene clusters and other important regulators during development[187]. Their role in silencing is counteracted by members of the Trithorax group (TrxG) of proteins, which catalyze

methylation of histone H3 K4 or play a role in nucleosome mobilization[188]. PcG complexes similarly encompass different enzymatic activities to modify chromatin structure and thereby influence transcription. PRC2 subunits Ezh1 and Ezh2 harbor histone H3 K27 methyltransferase activity, creating H3K27me3 epigenetic marks that are subsequently recognized and bound by the PRC1 complex. PRC1-subunit Ring1b displays ubiquitin E3 ligase activity towards histone H2A K119 and mono-ubiquitinated H2A interferes with RNA pol2 activity[189]. In addition, Ring1b was recently shown to affect higher order chromatin structure of the *Hoxb* and *Hoxd* loci in mESCs independent of its enzymatic activity[190].

The majority of PRC1 and PRC2 binding sites in mESCs are found within 1kb of transcription start sites (TSS) of genes encoding factors involved in cell fate specification[184-185]. These genomic regions are further characterized by the presence of so called 'bivalent domains', consisting of simultaneously present repressive histone H3 K27 trimethylation and active histone H3 K4 trimethylation marks[191-192]. Associated genes are thought to be poised for future expression upon induction of differentiation, as they are bound by Ser5 phosphorylated RNA pol2 and transcribed at a low level [189]. Null mutation of *Eed* (PRC2) or conditional deletion of *Ring1b* (PRC1) in ESCs results in loss of dual histone marks and derepression of poised genes[189, 193]. RNA pol2-dependent transcription of intronic, exonic and promoter regions of PcG target genes was recently demonstrated to give rise to short RNAs (50-200 nt) that form stem-loop structures involved in PRC2-targeting[194]. The presence of H3K4me3 marks thus may reflect production of short RNAs rather than RNA pol2 stalling. At many bivalent domains an extra level of repression is provided for by the histone H3 K9 methyltransferease activity of SetDB1 (also known as Eset) and, similar to loss of PRC1 or PRC2, depletion of SetDB1 results in derepression of numerous bivalently marked genes[195].

During the normal course of differentiation a large proportion of bivalent genes loses one of the two histone marks and becomes monovalent upon lineage commitment[196-197]. Besides loss of bivalency, gain of bivalent domains has also been observed and co-occurrence of repressive and active histone marks is therefore not restricted to embryonic stem cells[197]. In addition to the bivalent marks found in promoter regions of key developmental regulators, enhancers of tissue-specific genes are marked by presence of specific transcription factors, such as pioneer factor FoxD3, and absence of CpG methylation[198].

Chromatin remodelers in ES cell maintenance

In addition to the crucial role of PcG complexes and SetDB1 in maintenance of transcriptional silencing at developmental regulator-encoding genes, various other chromatin modifying complexes have been identified to participate in preservation of the pluripotent state. The mammalian SWI/SNF complex, a member of the Trx class , is an ATP-dependent chromatin remodeler whose catalytic subunit Brg1 is essential during preimplantation development[199]. RNAi-mediated depletion of Brg1 in ES cells impairs self-renewal capacity and interferes with proper differentiation into all three germ layers[200]. Genetic ablation of *BAF250a* or *BAF250b* (also called

Arid1a and *Arid1b*), signature subunits of SWI/SNF-A (or BAF for Brg1/Brahma-associated factors) and not found in SWI/SNF-B (or PBAF), similarly inhibits ES cell proliferation and predisposes them to differentiation[201-202]. The subunit composition of BAF in ES cells is somewhat different from that of complexes isolated previously from various other cell types. It exclusively contains Brg1 (and no Brahma) as catalytic subunit, bears BAF155, but not the homologous BAF170 and is enriched for BAF60a and BAF45d[200]. Genome wide ChIP-sequencing reveals preferential localization of esBAF around transcription start sites of genes, many of which are also bound by the key transcription regulators Oct4, Sox2 and Nanog[203]. Comparison of gene expression analysis of Oct4 knockdown or *Sox2* knockout ESCs with that of Brg1-depleted ESCs reveals that, although some common target genes are misregulated in the same direction, the effect on gene expression is predominantly antagonistic[203]. In addition, it was suggested that Brg1 mainly acts as repressor of its ESC target genes[203], which is in stark contrast to the reported essential role for Brg1 in zygotic genome activation[204]. ES cell differentiation induced by the long period during which Brg1 was downregulated[200] may partly explain these contradicting observations.

Subunits of the NuRD chromatin remodeling complex have been identified to interact with Oct4 and Nanog[172, 205]. NuRD contains both histone deacetylase and ATP-dependent nucleosome remodeling activity, contributed by its HDAC and CHD ATPase subunits respectively. Methyl-CpG binding domain protein 3 (Mbd3) is a core subunit of NuRD required for complex assembly. In its absence ESCs can be maintained, albeit at a slower growth rate[206]. Self-renewal of *Mbd3*-/- ESCs occurs even in the absence of LIF, however, they are unable to commit to different lineages when differentiation is induced in the context of embryoid body formation, a phenotype that can be rescued by reintroduction of Mbd3. In contrast, loss of self-renewal capacity was observed upon knockdown of NuRD specific subunit Mta1[205].

An RNAi screen aimed at identifying chromatin modifiers involved in regulation of ESC identity found several subunits of the Tip60-p400 complex to be required for maintenance of normal ESC growth and morphology[207]. Furthermore, individual knockout of two subunits, Tip60 and Trrap, results in pre-implantation lethality[208-209]. The Tip60-p400 complex is involved in a variety of cellular processes, including transcription regulation and the DNA damage response[210]. Similar to the NuRD complex, which contains both histone deacetylase and ATP-dependent nucleosome remodeling activity, the Tip60-p400 complex unifies histone acetyltransferase and ATPase activity in its Tip60 and p400 subunits, respectively. Profiling of p400 binding sites in ESCs detects high confidence binding at over half of all promoter regions, corresponding to both active and silent genes[207]. Targeting to these sites was shown to be regulated independently by both Nanog and H3K4me3 modification. Despite the general assumption that Tip60-p400 mainly functions in gene activation, expression profiling of knockdown ESCs suggests a predominant role in transcriptional repression of genes regulating development and differentiation[207]. These bivalently marked, PcG bound genes are often co-occupied by p400 and H2A.Z, a histone variant whose deposition is catalyzed by p400[207, 211]. Characterization of a Tip60-p400 complex subpopulation, in which p400 represses

Tip60 HAT activity and thereby coactivator function, provides further evidence in support of a role for Tip60-p400 in gene repression[212]. Induction of gene expression is accompanied by a decrease in p400 promoter occupancy and a concomitant increase in histone acetylation, indicating that localization of Tip60-p400 to repressed genes may be a prerequisite for rapid gene induction[212].

It thus seems that in ESCs several chromatin modifying complexes harboring a variety of enzymatic activities functionally converge to repress transcription of developmental genes, while simultaneously priming them for rapid activation upon transduction of differentiation-inducing signals.

Induced Pluripotency

The better understanding of transcriptional networks regulating pluripotency in recent years has greatly aided in development and optimization of strategies to reprogram somatic cells into a pluripotent ESC-like state. Early efforts to reprogram somatic cells relied on nuclear transfer into enucleated oocytes or fusion of somatic cells with pluripotent stem cells[213-215]. Each of these methods has several disadvantages: somatic cell nuclear transfer is a very inefficient and uncontrolled process and relies on availability of oocytes, whereas cell fusion results in tetraploid hybrid cells that do not efficiently contribute to chimeric mice following blastocyst injection[215].

In 2006, ground breaking work from Shinya Yamanaka's group described successful reprogramming of embryonic and adult mouse fibroblasts into a pluripotent state by the introduction of a mere four transcription factors: Oct4, Sox2, Klf4 and c-Myc (OSKM)[150]. Their initial screen consisted of 24 candidate genes that were selected based on their known role in the maintenance of ES cell identity. Via retroviral transduction different pools of factors were introduced into embryonic fibroblasts carrying a neomycin selection cassette in the *Fbx15* gene locus. *Fbx15* is expressed specifically in ES cells and reactivation of transcription from the *Fbx15* promoter thus allows for selection of reprogrammed, ESC-like cells. Transduced fibroblasts were transferred to ES cell culture conditions and indeed, in little over two weeks, neomycin resistant colonies appeared that were morphologically similar to ES cells. The number of required factors could be narrowed down to the four mentioned above. Resulting induced pluripotent stem (iPS) cells contributed to derivatives of all three germ layers in teratoma formation and upon *in vitro* induced differentiation. Furthermore, chimeric contribution was observed in E13.5 embryos, although no live pups were born. Subsequent substitution of *Fbx15* selection with *Oct4* or *Nanog* generated viable chimeras with high iPS contribution to all tissues, including the germ-line[148, 151].

Soon thereafter human iPS cells were successfully derived using a combination of Oct4, Sox2, Klf4 and c-Myc[154] or Oct4, Sox2, Nanog and Lin28[155]. Refinements of the protocol have since resulted in omission of c-Myc, thereby circumventing the potential tumorigenicity caused by c-Myc retroviral reactivation[216-217]. Selection purely on morphology made selectable drug resistance dispensable, allowing reprogramming of genetically unmodified cells from a variety of sources[218-219]. Successfully reprogrammed cells have since been obtained from murine adult neural stem cells, liver and stomach cells, keratinocytes, islet of Langerhans cells and immature

and mature B lymphocytes[220]. Full pluripotency of iPS cells was demonstrated by their capacity to produce viable adult mice upon tetraploid complementation[153].

Efficiency and safety of reprogramming

The potential use of iPS cells in disease modeling or cell replacement therapies has evoked great interest in increasing the efficiency of iPS cell generation and put focus on safety issues related to reprogramming methods. A number of attempts have been made to increase reprogramming efficiency by replacing one of the factors with alternative transcription factors. Omitting c-Myc reduces efficiency and slows down the reprogramming process, but replacing c-Myc with Sall4 overcomes this problem[221].

A screen aimed to identify factors that could replace Klf4 in the reprogramming process found that, besides the related factors Klf2 and Klf5, orphan nuclear receptor Esrrb could reprogram mouse embryonic fibroblasts into iPS cells in conjunction with Oct4 and Sox2, albeit at lower efficiency[222]. Moreover, expression profiling of these OSE-reprogrammed iPS cells revealed they were more dissimilar from wild-type R1 ESCs than OSK-iPS cells, discarding OSE as a suitable alternative for OSK[174]. A more detailed study on the role of nuclear receptors in reprogramming subsequently demonstrated that two nuclear receptors, Nr1i2 and Nr5a2, enhanced efficiency in the context of Oct4, Sox2, Klf4 and c-Myc[223]. In addition, Nr5a2 was capable of replacing Oct4 and generating chimera contributing iPS cells in cooperation with Sox2 and Klf4. Finally, inclusion of Tbx3 in the set of reprogramming factors resulted in increased efficiency of the OSKT-mediated reprogramming process, as well as improved quality of the resulting iPS cells, as demonstrated by enhanced germ-line contribution and transmission[174].

Efficiency of reprogramming seems to be hampered by induction of cellular senescence via the p53-p21 pathway. Inactivation of this pathway by various means, including knockdown or knockout of p53 and downregulation of its target p21, increased reprogramming efficiency nearly 25-fold[224-226]. Furthermore, downregulation of p53 enabled reprogramming of mouse fibroblasts by Oct4 and Sox2 alone. However, p53 being an important tumor suppressor, its permanent inactivation induces genome instability and cancer, as turned out to be the case in several chimeras generated from *p53*-null iPS cells[224]. Therefore ways to transiently suppress p53 function were sought and indeed addition of anti-oxidant vitamin C results in downregulation of p53 and enhanced reprogramming efficiency[227]. Alteration of culture conditions from normoxic (21%) to hypoxic (5%) resulted in a 3- to 7-fold increase in efficiency, possibly mediated by downregulation of the p53 pathway[228].

Apart from efforts to increase the efficiency of reprogramming by adapting culture conditions and using different combinations of transcription factors, research has also focused on diminishing the minimal number of factors required, as accompanying viral integrations increase the risk of oncogene activation. Special attention was drawn to small molecules targeting components of signaling pathways and chromatin modifying enzymes[128]. iPS cells were generated from human fibroblasts following introduction of Oct4 and Sox2 in the presence of valproic acid, a histone

deacetylase (HDAC)-inhibitor[229]. Screening a collection of small molecules for permitting reprogramming by just two factors, Oct4 and Klf4, identified a combination of an L-channel calcium agonist and an inhibitor of G9a histone methyltransferase[230]. The role of Sox2 in reprogramming could likewise be replaced by a small-molecule inhibitor of TGF-β signaling, which capacitates full reprogramming of partially reprogrammed cells via induction of *Nanog* expression[231].

An alternative starting point to increase safety of iPS technology revolves around development of safer methods for gene delivery. To reduce the amount of genetic alteration inflicted by viral integration, *loxP* sites have been introduced in the viral constructs, allowing excision of the transgenes once a fully reprogrammed state has been attained[232]. The number of minimally required viral integrations can be further reduced by making use of a loxP-flanked construct producing a single mRNA, which encodes all four factors, linked by peptides that fail to form peptide bonds, without interfering with ribosomal processivity[233]. The Cre/loxP system, however, is incapable of fully excising all integrated DNA and some remnants are always left behind. The *piggyBac* transposon/transposase system in that respect offers a superior means of introducing and removing reprogramming factors, as no exogenous DNA fragments are left behind following excision[234]. In addition, non-genome integrating methods such as adenovirus transduction, repetitive rounds of transient plasmid transfections and use of episomal (self-replicating, non-integrating) vectors have resulted in fully reprogrammed cells[235]. Non-nucleic acid based iPS cells contributing to various tissues in chimeric embryos, including the germline, were also generated following successive rounds of exposure to purified recombinant reprogramming factors harboring a poly-arginine tag to facilitate passage through the cell membrane[236].

Molecular processes underlying reprogramming

Much effort has been done to develop methods that can enhance the efficiency of somatic cell reprogramming and reduce the genomic invasiveness of the procedure. Gaining a better understanding of the molecular mechanisms underlying a successful reprogramming event may further contribute to improvement of the process. Key events include silencing of lineage specific genes, reactivation of pluripotency genes, setting up an ESC-like chromatin structure and resetting the cell cycle control machinery. To get a better insight into prominent obstacles to reach a pluripotent state, partially reprogrammed cells, which can be stably maintained in culture, were subjected to expression analysis and compared to their origin (e.g. fibroblast or lymphocyte) and to iPS and ES cells[237]. The partially reprogrammed state was characterized by reactivation of self-renewal and ESC maintenance genes, although endogenous expression of key pluripotency factors was not detectable. In addition, lineage-specific factors were not completely silenced and the epigenome, most notably DNA hypermethylation, had not been reset to an ESC-like state. Genomic binding sites of the four factors were mapped in partially reprogrammed cells and clustered binding of multiple factors, like in ESCs, correlated to gene expression[238]. Along these lines, genes bound by more factors in partially reprogrammed iPS (piPS) cells than in ESCs were also expressed at a higher level in piPS. However, not all bonafide ESC target genes were bound

in piPS cells, which seemed to be particularly the case for Oct4/Sox2/Klf4 targets. Site occupancy by these factors may require expression of additional pluripotency factors to facilitate binding. Correct targeting of c-Myc is less dependent on other factors and c-Myc therefore seems to be mostly involved in very early stages of reprogramming.

Activation of pluripotency genes requires, apart from transcription factor binding, extensive epigenetic remodeling of promoter regions, including DNA demethylation. This modification forms an important barrier to acquire a pluripotent state and consequently, inhibition of DNA methyltransferase activity with 5-aza-cytidine in partially reprogrammed cells can suffice to attain full pluripotency[237]. In addition, mechanistic studies on reprogramming in heterokaryons formed by fusion of mESCs with human fibroblasts suggested that active DNA demethylation catalyzed by activation-induced cytidine deaminase (AID) occurs during the reprogramming process[239]. The four factor induced reprogramming process itself can also cause epigenetic alterations, as was recently demonstrated by Stadtfeld *et al.*[240]. Comparison of transcriptomes of genetically identical ES and iPS cells revealed iPS-specific aberrant silencing of an imprinted gene cluster. Silencing had occurred during the four factor-mediated reprogramming process, since it was not present in the somatic starting population and did not establish if reprogramming was mediated by nuclear transfer. The silenced locus was characterized by CpG-hypermethylation and reduction in histone acetylation, which interfered with efficient chimera contribution and ability to generate all-iPS live born animals.

Another prominent feature of murine pluripotent cells relates to (the lack of) cell cycle control. ESCs and iPS cells transit the cell cycle every 8 to 12 hours and spend most (65%) of their time in S-phase. They lack the G1 restriction (R) point, which normally allows cells to exit the cell cycle and enter a quiescence phase in the absence of mitogens. When ESCs differentiate the length of G1 rapidly increases and R point control is instated[241]. On the contrary, during somatic cell reprogramming the R point has to be alleviated, a step in which partially reprogrammed cells have not succeeded[242]. c-Myc is potentially involved in this process by inactivation of senescence factors retinoblastoma (pRb) and p53 via modulation of cyclin-dependent kinase activity[242].

The rapid advances in iPS technology have also stimulated interest in trans-differentiation, in which one differentiated cell type is reprogrammed directly into another terminally differentiated type by introduction of a few key factors. This avoids passage through a pluripotent state, which is an important advantage if reprogrammed cells are to be used in patients, since failure to remove all residual pluripotent cells entails the possibility of teratoma formation. Successful trans-differentiation has been reported for conversion of exocrine pancreatic cells into insulin-producing endocrine cells, as well as for direct reprogramming of fibroblasts into functional neurons[243-244].

Octamer-binding protein 4 (Oct4)

Oct4 (also known as Oct3 or Pou5f1) was first identified as an octamer binding protein specifically expressed in pre-implantation embryos and the germline[245]. It is a member of the POU family of transcription factors that adopted its name from observed homology in mammalian *Pit1*, *Oct1*,

Oct2 and *C. elegans Unc86* genes, resulting in identification of the metazoan-specific POU domain. This domain partially resembles the previously characterized and highly conserved DNA binding homeodomain. The POU domain however contains an N-terminal extension termed POU-specific domain (POU_S) which is separated from the POU-homeodomain (POU_H) by a variable linker[246]. The two subdomains cooperatively bind DNA, even when not connected via linker sequence and the linker likely functions to locally increase concentration of both POU domains[247]. The smaller POU_H forms three α-helices that are arranged into a helix-turn-helix structure. It strongly resembles classical homeodomains that utilize helix 3 to bind a core ATTA sequence. The POU-specific domain is comprised of 75-82 amino acids forming four α-helices, of which helices 2 and 3 are arranged into a helix-turn-helix motif that structurally resembles several prokaryotic repressors[248]. Each POU-subdomain recognizes half of the octameric ATGCAAAT site bound by Oct proteins. Helix 3 of POU_S forms extensive hydrogen bonds and one hydrophobic interaction with the ATGC half of the octamer element, while all four helices contact the DNA phosphate backbone. The third helix of POU_H makes base-specific contacts with the second half of the octamer[247]. In the most frequently adopted DNA-binding conformation, the two individual subdomains localize to opposite sides of the DNA double helix, with the connecting linker stretch tracking along the minor groove. Alternative ways of DNA binding are possible and are influenced by phosphorylation and multimerization events. This can result in conformations where POU_S and POU_H are found on the same side of the helix and recognize a non-canonical sequence element termed PORE (Palindromic Octamer Related Element) or MORE (More PORE)[249].

The 15 POU-domain proteins that have been identified in mice can be classified into six groups based on sequence conservation of their subdomains and linker region[246]. Looking at interrelatedness based on full length sequence of human POU-proteins reinforces these different classes (Figure 5A), whereas comparison of DBD regions only accurately separates octameric from non-octamer binding POU domain proteins (Figure 5B)[249].

Oct-Sox partnership

Classical High Mobility Group (HMG) proteins (e.g. HMG1, HMG2) constitute a group of chromatin-associated proteins lacking DNA sequence specificity, that are found in all eukaryotes. They can bind DNA cooperatively with classical homeodomain proteins, but also with POU-domain factors[248]. Sox (SRY HMG box) proteins arose in metazoans where they seem to have co-evolved with POU-domain factors. In contrast to the two or more HMG boxes found in classical HMG proteins, Sox factors have one HMG box and display binding preference to (A/T)(A/T)CAA(A/T)G sequences. Sox and POU factors in metazoans have developed similar partnerships as were observed between classical HMG proteins and classical homeodomains[248].

The POU/Sox relationship of ESC transcription factors Oct4 and Sox2 has been intensively studied in relation to their ability to transactivate transcription from the *Fgf4* enhancer[177]. Oct4 and Sox2 interact through their DNA-binding domains and can do so in the absence of DNA[177]. They cooperatively bind to form ternary (Oct4-Sox2-DNA) complexes on the *Fgf4* enhancer and

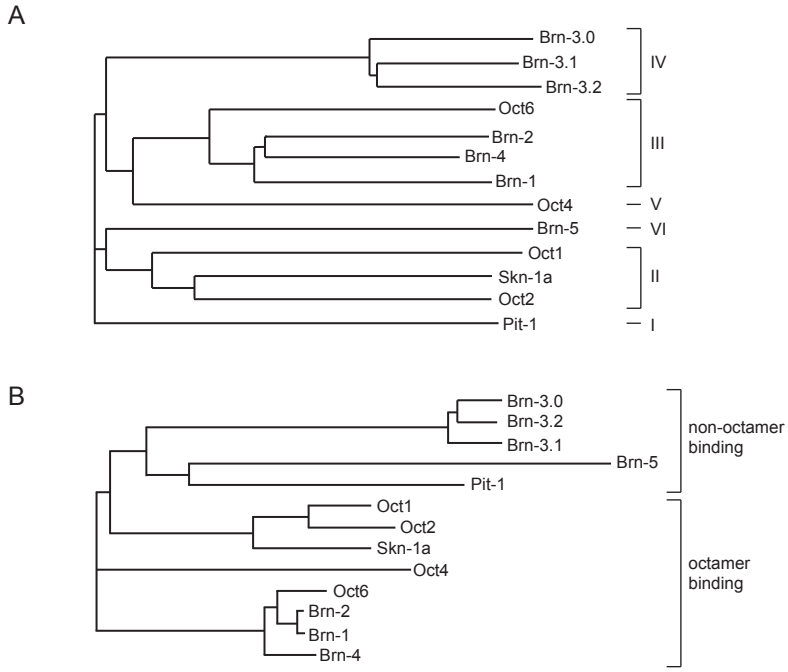


Figure 5 Dendrogram displaying relatedness between different human POU-domain transcription factors.
(A) Clustering based on kinship between full length proteins. (B) Clustering of DNA binding domains. Adapted from [249].

this cooperativity critically depends on spacing between the individual Oct and Sox binding elements[177]. Whereas DBDs of Oct4 and Sox2 alone are sufficient to mediate cooperative binding, additional domains in both proteins are required to activate transcription[250]. Curiously, a chimeric protein composed of Oct1 DBD and Oct4 N- and C-terminal parts is unable to activate transcription in conjunction with Sox2, despite their ability to bind cooperatively[250]. Transactivation function is partially restored by reintroduction of POU_H, suggesting this subdomain plays an essential role in capacitating the activator function, either by inducing structural changes or by directly mediating interactions with co-activators. These findings additionally can provide an explanation for the inaptitude of other ESC-expressed Oct proteins (i.e. Oct1, Oct6[251]) to maintain ESC identity[111, 252] or to replace Oct4 in somatic cell reprogramming[216]. Sox2 does not appear to be absolutely required for the Oct-Sox partnership in ESCs, as *Sox2* null cells can be kept in a pluripotent state if Oct4 expression is artificially maintained[160]. Sox4, Sox11 and Sox15 are additional Sox factors expressed in ESCs and are capable of binding Oct-Sox enhancers[160]. Of these, at least Sox15 can substitute for Sox2 in reprogramming[216]. These data thus suggest Oct4 can form functional partnerships with other Sox proteins that are sufficient to sustain the pluripotent state.

Oct4 is a dose-dependent pluripotency factor

As mentioned previously, Oct4 is a dose-dependent regulator of the ES cell state, since a 50% increase or decrease in expression level induces differentiation[111]. Overexpression of a single point mutant Oct4 that is unable to bind DNA exhibits the same dose-response, suggesting it is independent of DNA binding capacity[252]. Overexpression of Oct4 DBD alone, however, did not interfere with ESC maintenance and so the possible involvement of N- and C-terminal activation domains was assessed. Presence of either activation domain sufficed to complement for loss of full length Oct4 and, in accordance with that, overexpression of Oct4 POU domain fused with either N- or C-terminal domain or even Oct2 CTD had the same dose-dependent effect as overexpression of the full length protein[252]. Titration of a critical binding partner can be the underlying cause for the observed overexpression-induced differentiation into primitive endoderm and mesoderm lineages. Interaction with this binding partner would be mediated by the Oct4 POU-domain and an unspecific activation domain.

Regulation of Oct4 expression and activity

Transcription of the murine *Pou5f1* gene initiates at multiple sites in a GC-rich region that lacks a TATA box[253]. Comparison of human, bovine and murine promoter sequences has led to the identification of four conserved regions (CR1-4) located within 2 kb upstream of the TSS[254]. A conserved hormone responsive element (HRE) and putative Sp1/Sp3 site are found in the promoter proximal CR1 region. Several nuclear hormone receptors were demonstrated to bind HREs in the Pou5f1 promoter and have been implicated in either positive or negative regulation of Oct4 expression. These include positive regulators RXRβ[255], SF-1 (Nr5a1)[256], LRH-1 (Nr5a2)[257], Esrrb (Nr3b2)[258] and negative regulators GCNF (Nr6a1)[259], COUP-TF1 (Nr2f1) and COUP-TF2 (Nr2f2)[255].

Expression of *Pou5f1* is gradually downregulated within 5-8 days following retinoic acid (RA) induced embryoid body differentiation[259]. Stable silencing involves promoter CpG methylation by Dnmt3a and Dnmt3b, possibly targeted to *Pou5f1* by GCNF[259-260], as well as PRC2-mediated H3K27 methylation[261-262]. Expression of Oct4 is positively regulated by numerous transcription factors that mostly bind to an enhancer region located 2 kb upstream of the TSS. Positive regulators include Nanog[164], Klf5[170], Sall4[171] and the Oct4-Sox2 pair itself[263].

Post-transcriptional regulation of Oct4 levels is mediated by coding sequence-targeting miRNAs in mESCs[264] and 3'UTR-targeting miRNAs in hESCs[265]. When mESCs are induced to differentiate by addition of RA, these miRNAs are upregulated to lower Oct4 protein level and blocking miRNA action at this stage transiently stabilizes Oct4, causing a delay in ESC differentiation[264]. Furthermore, several post-translational modifications have been implicated in regulation of Oct4 action. Phosphorylation of residues in the C-terminal and POU_H domains has been reported[252, 266] and phosphorylation of the latter was shown to negatively affect Oct4 transactivation from PORE sites specifically, which require cooperative binding of two Oct4 molecules[266]. Two papers have reported sumoylation of Oct4 at K118 in the N-terminal domain and mutation of this sumoylation site was shown to destabilize protein levels, although it did not have major consequences for ESC

maintenance[267-268]. Human Oct4 finally was shown to be modified by addition of a monosaccharide, O-linked beta-N-acetylglucosamine (O-GlcNAc)[269], which has been demonstrated to modulate the activity of several transcription factors and whose addition is catalyzed by the PcG class protein Ogt/Sxc[270-271].

Oct4 target gene regulation

Genome-wide binding sites and putative target genes of Oct4 in ESCs have been mapped[164, 272]. The most detailed integrated analysis of these different data sets has been conducted by Sharov *et al*.[273]. Expression profiling was performed at different time points (0, 3, 6, 12 and 24 hours) following downregulation of a tetracycline-responsive *Oct4* transgene in ZHBTc4 ES cells and responding genes were subsequently classified according to direction and time of their response. Genomic binding sites of Oct4 were assigned a qualitative value (score of potential function or SPF) that is positively affected by number of reads in the chromatin immunoprecipitation (ChIP) and negatively by CpG richness and distance between the binding site and the nearest gene. This culminated in the identification of 420 tentative target genes (TTGs), 85% of which are activated by Oct4[273]. Early response TTGs include a number of TFs that are important to the ESC-state, such as *FoxD3*[274], *Klf2*[120], *Dax1* (Nr0b1)[172] and *Zic3*[275]. Expression of key pluripotency factors *Sox2* and *Nanog* only became affected at much later timepoints, suggesting that Oct4 is not the main activator of these genes. TTGs that are repressed by Oct4 include trophectoderm transcription factors *Cdx2* and *Eomes*[276] and, indeed, mechanistic studies have since demonstrated that Oct4 recruits the histone H3K9 methyltransferase Eset to repress Cdx2 expression in ES cells[277-278].

Use of SPF to identify TTGs may introduce bias towards identification of genes where Oct4 functions as activator if, for example, Oct4 preferentially acts as transcription repressor from distal enhancers. Independent analysis of ZHBTc4 transcriptomes upon Oct4-depletion, however, showed a similar trend towards a predominantly activating role for Oct4 in transcription, although this alternatively could reflect mechanistic differences between retaining an active or repressed state[120].

To better understand correlations between transcription factor binding and transcription output, ChIP-Seq data of 12 ESC transcription factors was combined with RNA-Seq expression data to devise a model that could accurately predict gene expression outcome[165]. This revealed that Oct4 and also Sox2, Nanog, Esrrb, Smad1, Stat3 and Tcfcp2l1, can be either an activator or repressor. Affinity purification of Oct4 followed by mass spectrometry analysis has identified, apart from the aforementioned transcription factors, members of the NuRD chromatin remodeling complex as Oct4-interacting proteins[172, 205], further suggestive of a repressive role in transcription regulation.

Estrogen receptor-related receptor β (Esrrb)

The estrogen receptor-related receptors were among the first orphan nuclear receptors identified. Their subfamily is comprised of three members: Esrra (Nr3B1 or ERRα), Esrrb (Nr3B2 or ERRβ) and Esrrg (Nr3B3 or ERRγ). They belong to the superfamily of nuclear receptors, which in mice consists

of 49 members that can be further divided into six subfamilies[279]. The ERR subfamily (group III) additionally contains estrogen receptor α and β, androgen (AR), progesterone (PR), glucocorticoid (GR) and mineralocorticoid (MR) receptors (Figure 6).

All nuclear receptors have a similar domain structure that consists of a non-conserved N-terminal activation domain (AF-1), a DNA-binding domain comprised of two zinc fingers, followed by a ligand binding domain (LBD or AF-2). Both activation domains can recruit coactivators or corepressors and AF-2 interactors usually contain an LXXLL motif that mediates the interaction[280]. Sequence conservation between hERRs and hERα is highest in the DBD (68-73%) and around 37% in the LBD[281]. In contrast to ER, the LBD of ERRs appears to always be in an active conformation, irrespective of ligand and, as no endogenous ligand for ERRs has been described, these receptors are classified as orphan[282].

Most group III nuclear receptors recognize a consensus sequence of two inverted repeat half sites (e.g. AGGTCA for ER) and bind as dimers. Genome wide mapping of binding sites for Esrra and Esrrg in cardiac tissue[283] and for Esrrb in ES cells[162] has identified a single extended half site (TCAAGGTCA) as most prevalent recognition sequence. This so called ERRE (for ERR response element) can be bound by monomeric ERRs, in which case interaction with DNA is further stabilized by contacts between a C-terminal extension of the DBD and the extended half site[284]. Alternatively, homo- or heterodimers and even a homodimer of ERα can bind to ERRE motifs[285]. The reverse, binding of estrogen-response elements (EREs) by homodimers of Esrra or Esrrb, has also been reported[286].

ERRs are present in most mouse tissues and especially Esrra shows widespread expression at high levels. Esrrb expression patterns are more restricted, with high levels detected in the eye, thyroid, kidney, heart and testis but near absence in a range of other tissues[287]. Furthermore, NR expression analysis in different tissues during a 24 hour period revealed a strong rhythmic expression of Esrrb in all tissues tested, suggestive of circadian clock control[288]. *Esrrb* null mice die at E10.5 with placental defects characterized by an overabundance of trophoblast giant cells and near absence of diploid trophoblast cells[289].This phenotype could be rescued by tetraploid complementation, in which *Esrrb*-/- embryos at E12.5 were indistinguishable from their wildtype littermates. Interestingly, whereas no ERR ligand has been described, the synthetic estrogen diethylstilbestrol (DES) was shown to act as an ERR antagonist and, when administered to pregnant mice, reproduces the *Esrrb*-/- phenotype[290]. In human, point mutations in the DBD and LBD of Esrrb have been linked to non-syndromic hearing loss[291]. A role for Esrrb in inner ear development is further supported by observed misspecification of epithelial cells in the inner ear, accompanied by a reduction in endolymph production in mice where Esrrb is deleted in the embryonic lineage[292]. Incidentally, this study also demonstrated that *Esrrb*-/- mice are viable and, apart from the inner ear defect, a reduction in primordial germ cell number was independently reported[293].

Esrrb was identified as a novel regulator of mES cell maintenance in an RNAi-screen conducted by Ivanova *et al*.[168]. Reduction of Esrrb resulted in ESC differentiation as assessed by morphological

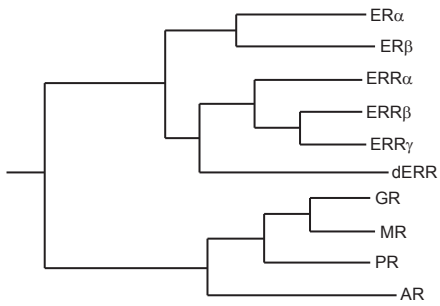


Figure 6 Dendrogram displaying interrelatedness of class III nuclear hormone receptors
ER, estrogen receptor; ERR, estrogen receptor-related receptor; d, *Drosophila*; GR, glucocorticoid receptor; MR, mineralocorticoid receptor; PR, progesterone receptor; AR, androgen receptor. Adapted from [285].

changes, loss of alkaline phosphatase activity and upregulation of differentiation markers. This phenotype could be partially rescued by overexpression of Nanog. The other ERR family member that is expressed in ESCs, Esrra, cannot substitute for Esrrb neither in ES cell maintenance nor in induction of pluripotency[222]. Esrrg, which is more closely related to Esrrb than Esrra, can replace Esrrb in somatic cell reprogramming, but is normally near absent from ESCs[222]. Esrrb is not detectable in human ES cells, which may reflect the suggestion that mouse and human ESCs correspond to different stages of development[294]. Indeed mEpiSCs, which are thought to be equivalent to hESCs, express Esrrb at much lower levels than mESCs[138].

Esrrb target gene regulation

Genome wide binding sites of Esrrb in ESCs overlap with those of Nanog, Tcfcp2l1 and, to a lesser extent Klf4, Sox2 and Oct4[162]. Computational modeling predicted that Esrrb, depending on the target gene, acts either as activator or repressor and this prediction could be independently verified for several target genes upon Esrrb knockdown[165]. Apart from its reported co-purification with Nanog[172] little is known about Esrrb-interacting proteins in ESCs that could potentially function as coactivator or corepressor. However, interactions with co-regulators have been reported in other cell types. Human ERRβ binds co-activators of the NCoA/SRC/p160 family (NCoA2 and NCoA3) via its LBD in a ligand-independent manner, resulting in transactivation of a reporter gene[295]. SRC-family coactivators interact with several NRs via any of their three central LXXLL motifs and can recruit chromatin modifying enzymes such as CBP, p300, CARM1 and PRMT1 via their C-terminal activation domains[296]. Transcription activation by hERRγ is enhanced by AF-1 associating factors PNRC2 and TLE1[297] and also by PGC-1α and PGC1-β[298]. Of the latter, PGC-1α was shown to interact with p300 and Mediator-subunit Med1 to stimulate nuclear receptor PPAR-mediated transcription[299].

Due to its close kinship, mechanistic analysis of ER-function may additionally be informative on a possible mode of action for Esrrb. ER-mediated transcription activation however requires the presence of its ligand, estrogen (E_2). The Med1-containing Mediator fraction, which is also enriched for RNA pol2, was demonstrated to be recruited to ER-target genes in a ligand-dependent

manner and was shown to be required for ER-mediated transcription[66]. Analysis of the ER-target gene pS2 following E_2-stimulation of MCF-7 cells revealed a cyclic recruitment pattern of ERα and a variety of cofactors, including NCoA1, histone methyltransferases, histone acetylases, RNA pol2 and Mediator complex[300]. At the onset of this cyclic process recruitment of chromatin remodeling or modifying enzymes, such as Brg1, p300 and Tip60, precedes binding of the basal transcription machinery. Once RNA pol2 enters a productive elongation phase a number of repressive activities, such as histone deacetylases (Hdacs) and NuRD complex, are recruited to the E_2-responsive gene promoter to limit the transcriptional response. In addition, expression analysis more than 12 hours after a short E_2-pulse demonstrated that ER induces an NR-interacting protein, Nrip1, which acts as NR-corepressor to restrain the response[301].

Motif discovery analysis on genome wide binding data exposed a FoxA1 motif as second most prevalent among ER-bound sites, indicating pioneer factor FoxA1 may facilitate initial ER-binding[301]. Chromatin interactions of ERα-bound sites as determined by ChIA-PET predominantly occur between distant sites on the same chromosome and in most cases span regions of less than 100 kb[302]. The majority of ERα binding sites are found distal to transcription start sites and coincide with FoxA1 binding. These findings, combined with the notion that many distal sites are engaged in complex chromatin interactions with promoter proximal sites, led to the postulation that ERα-binding and recruitment of cofactors may result in formation of chromatin loop structures around target genes, which could positively affect transcription by increasing local concentrations of cofactors. Esrrb could function in a similar manner, although it is unclear whether a ligand is required.

REFERENCES

1. Juven-Gershon, T. and J.T. Kadonaga, *Regulation of gene expression via the core promoter and the basal transcriptional machinery.* Dev Biol, 2010. **339**(2): p. 225-9.
2. Thomas, M.C. and C.M. Chiang, *The general transcription machinery and general cofactors.* Crit Rev Biochem Mol Biol, 2006. **41**(3): p. 105-78.
3. Venters, B.J. and B.F. Pugh, *How eukaryotic genes are transcribed.* Crit Rev Biochem Mol Biol, 2009. **44**(2-3): p. 117-41.
4. Ptashne, M. and A. Gann, *Transcriptional activation by recruitment.* Nature, 1997. **386**(6625): p. 569-77.
5. Muse, G.W., D.A. Gilchrist, S. Nechaev, R. Shah, J.S. Parker, S.F. Grissom, J. Zeitlinger, and K. Adelman, *RNA polymerase is poised for activation across the genome.* Nat Genet, 2007. **39**(12): p. 1507-11.
6. Seila, A.C., J.M. Calabrese, S.S. Levine, G.W. Yeo, P.B. Rahl, R.A. Flynn, R.A. Young, and P.A. Sharp, *Divergent transcription from active promoters.* Science, 2008. **322**(5909): p. 1849-51.
7. Zeitlinger, J., A. Stark, M. Kellis, J.W. Hong, S. Nechaev, K. Adelman, M. Levine, and R.A. Young, *RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo.* Nat Genet, 2007. **39**(12): p. 1512-6.
8. Core, L.J., J.J. Waterfall, and J.T. Lis, *Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.* Science, 2008. **322**(5909): p. 1845-8.
9. Campos, E.I. and D. Reinberg, *Histones: annotating chromatin.* Annu Rev Genet, 2009. **43**: p. 559-99.
10. Cairns, B.R., *The logic of chromatin architecture and remodelling at promoters.* Nature, 2009. **461**(7261): p. 193-8.
11. Ho, L. and G.R. Crabtree, *Chromatin remodelling during development.* Nature, 2010. **463**(7280): p. 474-84.

12. Mohrmann, L. and C.P. Verrijzer, *Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes.* Biochim Biophys Acta, 2005. **1681**(2-3): p. 59-73.
13. Racki, L.R. and G.J. Narlikar, *ATP-dependent chromatin remodeling enzymes: two heads are not better, just different.* Curr Opin Genet Dev, 2008. **18**(2): p. 137-44.
14. Lusser, A., D.L. Urwin, and J.T. Kadonaga, *Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly.* Nat Struct Mol Biol, 2005. **12**(2): p. 160-6.
15. Hall, J.A. and P.T. Georgel, *CHD proteins: a diverse family with strong ties.* Biochem Cell Biol, 2007. **85**(4): p. 463-76.
16. Taverna, S.D., H. Li, A.J. Ruthenburg, C.D. Allis, and D.J. Patel, *How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers.* Nat Struct Mol Biol, 2007. **14**(11): p. 1025-40.
17. Min, J., Y. Zhang, and R.M. Xu, *Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27.* Genes Dev, 2003. **17**(15): p. 1823-8.
18. Vermeulen, M., K.W. Mulder, S. Denissov, W.W. Pijnappel, F.M. van Schaik, R.A. Varier, M.P. Baltissen, H.G. Stunnenberg, M. Mann, and H.T. Timmers, *Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4.* Cell, 2007. **131**(1): p. 58-69.
19. Mavrich, T.N., I.P. Ioshikhes, B.J. Venters, C. Jiang, L.P. Tomsho, J. Qi, S.C. Schuster, I. Albert, and B.F. Pugh, *A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome.* Genome Res, 2008. **18**(7): p. 1073-83.
20. Schones, D.E., K. Cui, S. Cuddapah, T.Y. Roh, A. Barski, Z. Wang, G. Wei, and K. Zhao, *Dynamic regulation of nucleosome positioning in the human genome.* Cell, 2008. **132**(5): p. 887-98.
21. Babu, M.M., N.M. Luscombe, L. Aravind, M. Gerstein, and S.A. Teichmann, *Structure and evolution of transcriptional regulatory networks.* Curr Opin Struct Biol, 2004. **14**(3): p. 283-91.
22. Georges, A.B., B.A. Benayoun, S. Caburet, and R.A. Veitia, *Generic binding sites, generic DNA-binding domains: where does specific promoter recognition come from?* FASEB J, 2010. **24**(2): p. 346-56.
23. Harbison, C.T., D.B. Gordon, T.I. Lee, N.J. Rinaldi, K.D. Macisaac, T.W. Danford, N.M. Hannett, J.B. Tagne, D.B. Reynolds, J. Yoo, E.G. Jennings, J. Zeitlinger, D.K. Pokholok, M. Kellis, P.A. Rolfe*, et al.*, *Transcriptional regulatory code of a eukaryotic genome.* Nature, 2004. **431**(7004): p. 99-104.
24. Cirillo, L.A., F.R. Lin, I. Cuesta, D. Friedman, M. Jarnik, and K.S. Zaret, *Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4.* Mol Cell, 2002. **9**(2): p. 279-89.
25. Carroll, J.S., X.S. Liu, A.S. Brodsky, W. Li, C.A. Meyer, A.J. Szary, J. Eeckhoute, W. Shao, E.V. Hestermann, T.R. Geistlinger, E.A. Fox, P.A. Silver, and M. Brown, *Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1.* Cell, 2005. **122**(1): p. 33-43.
26. Holmqvist, P.H., S. Belikov, K.S. Zaret, and O. Wrange, *FoxA1 binding to the MMTV LTR modulates chromatin structure and transcription.* Exp Cell Res, 2005. **304**(2): p. 593-603.
27. Sagai, T., M. Hosoya, Y. Mizushina, M. Tamura, and T. Shiroishi, *Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb.* Development, 2005. **132**(4): p. 797-803.
28. Heintzman, N.D., R.K. Stuart, G. Hon, Y. Fu, C.W. Ching, R.D. Hawkins, L.O. Barrera, S. Van Calcar, C. Qu, K.A. Ching, W. Wang, Z. Weng, R.D. Green, G.E. Crawford, and B. Ren, *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.* Nat Genet, 2007. **39**(3): p. 311-8.
29. Kim, T.K., M. Hemberg, J.M. Gray, A.M. Costa, D.M. Bear, J. Wu, D.A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P.F. Worley, G. Kreiman*, et al.*, *Widespread transcription at neuronal activity-regulated enhancers.* Nature, 2010. **465**(7295): p. 182-7.
30. Simonis, M., P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. de Wit, B. van Steensel, and W. de Laat, *Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C).* Nat Genet, 2006. **38**(11): p. 1348-54.
31. Tolhuis, B., R.J. Palstra, E. Splinter, F. Grosveld, and W. de Laat, *Looping and interaction between hypersensitive sites in the active beta-globin locus.* Mol Cell, 2002. **10**(6): p. 1453-65.
32. Fuda, N.J., M.B. Ardehali, and J.T. Lis, *Defining mechanisms that regulate RNA polymerase II transcription in vivo.* Nature, 2009. **461**(7261): p. 186-92.
33. Kim, Y.J., S. Bjorklund, Y. Li, M.H. Sayre, and R.D. Kornberg, *A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II.* Cell, 1994. **77**(4): p. 599-608.
34. Thompson, C.M., A.J. Koleske, D.M. Chao, and R.A. Young, *A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast.* Cell, 1993. **73**(7): p. 1361-75.

35. Fondell, J.D., H. Ge, and R.G. Roeder, *Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex.* Proc Natl Acad Sci U S A, 1996. **93**(16): p. 8329-33.
36. Conaway, R.C., S. Sato, C. Tomomori-Sato, T. Yao, and J.W. Conaway, *The mammalian Mediator complex and its role in transcriptional regulation.* Trends Biochem Sci, 2005. **30**(5): p. 250-5.
37. Naar, A.M., P.A. Beaurang, S. Zhou, S. Abraham, W. Solomon, and R. Tjian, *Composite co-activator ARC mediates chromatin-directed transcriptional activation.* Nature, 1999. **398**(6730): p. 828-32.
38. Rachez, C., B.D. Lemon, Z. Suldan, V. Bromleigh, M. Gamble, A.M. Naar, H. Erdjument-Bromage, P. Tempst, and L.P. Freedman, *Ligand-dependent transcription activation by nuclear receptors requires the DRIP complex.* Nature, 1999. **398**(6730): p. 824-8.
39. Baek, H.J., S. Malik, J. Qin, and R.G. Roeder, *Requirement of TRAP/mediator for both activator-independent and activator-dependent transcription in conjunction with TFIID-associated TAF(II)s.* Mol Cell Biol, 2002. **22**(8): p. 2842-52.
40. Mittler, G., E. Kremmer, H.T. Timmers, and M. Meisterernst, *Novel critical role of a human Mediator complex for basal RNA polymerase II transcription.* EMBO Rep, 2001. **2**(9): p. 808-13.
41. Knuesel, M.T., K.D. Meyer, C. Bernecky, and D.J. Taatjes, *The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function.* Genes Dev, 2009. **23**(4): p. 439-51.
42. Bourbon, H.M., *Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex.* Nucleic Acids Res, 2008. **36**(12): p. 3993-4008.
43. Levine, M. and R. Tjian, *Transcription regulation and animal diversity.* Nature, 2003. **424**(6945): p. 147-51.
44. Cai, G., T. Imasaki, Y. Takagi, and F.J. Asturias, *Mediator structural conservation and implications for the regulation mechanism.* Structure, 2009. **17**(4): p. 559-67.
45. Park, J.M., J. Werner, J.M. Kim, J.T. Lis, and Y.J. Kim, *Mediator, not holoenzyme, is directly recruited to the heat shock promoter by HSF upon heat shock.* Mol Cell, 2001. **8**(1): p. 9-19.
46. Cantin, G.T., J.L. Stevens, and A.J. Berk, *Activation domain-mediator interactions promote transcription preinitiation complex assembly on promoter DNA.* Proc Natl Acad Sci U S A, 2003. **100**(21): p. 12003-8.
47. Wu, S.Y., T. Zhou, and C.M. Chiang, *Human mediator enhances activator-facilitated recruitment of RNA polymerase II and promoter recognition by TATA-binding protein (TBP) independently of TBP-associated factors.* Mol Cell Biol, 2003. **23**(17): p. 6229-42.
48. Takagi, Y., G. Calero, H. Komori, J.A. Brown, A.H. Ehrensberger, A. Hudmon, F. Asturias, and R.D. Kornberg, *Head module control of mediator interactions.* Mol Cell, 2006. **23**(3): p. 355-64.
49. Cai, G., T. Imasaki, K. Yamada, F. Cardelli, Y. Takagi, and F.J. Asturias, *Mediator head module structure and functional interactions.* Nat Struct Mol Biol, 2010. **17**(3): p. 273-9.
50. Holstege, F.C., E.G. Jennings, J.J. Wyrick, T.I. Lee, C.J. Hengartner, M.R. Green, T.R. Golub, E.S. Lander, and R.A. Young, *Dissecting the regulatory circuitry of a eukaryotic genome.* Cell, 1998. **95**(5): p. 717-28.
51. Casamassimi, A. and C. Napoli, *Mediator complexes and eukaryotic transcription regulation: an overview.* Biochimie, 2007. **89**(12): p. 1439-46.
52. van de Peppel, J., N. Kettelarij, H. van Bakel, T.T. Kockelkorn, D. van Leenen, and F.C. Holstege, *Mediator expression profiling epistasis reveals a signal transduction pathway with antagonistic submodules and highly specific downstream targets.* Mol Cell, 2005. **19**(4): p. 511-22.
53. Westerling, T., E. Kuuluvainen, and T.P. Makela, *Cdk8 is essential for preimplantation mouse development.* Mol Cell Biol, 2007. **27**(17): p. 6177-82.
54. Taatjes, D.J., A.M. Naar, F. Andel, 3rd, E. Nogales, and R. Tjian, *Structure, function, and activator-induced conformations of the CRSP coactivator.* Science, 2002. **295**(5557): p. 1058-62.
55. Gaytan de Ayala Alonso, A., L. Gutierrez, C. Fritsch, B. Papp, D. Beuchle, and J. Muller, *A genetic screen identifies novel polycomb group genes in Drosophila.* Genetics, 2007. **176**(4): p. 2099-108.
56. Ding, N., H. Zhou, P.O. Esteve, H.G. Chin, S. Kim, X. Xu, S.M. Joseph, M.J. Friez, C.E. Schwartz, S. Pradhan, and T.G. Boyer, *Mediator links epigenetic silencing of neuronal gene expression with x-linked mental retardation.* Mol Cell, 2008. **31**(3): p. 347-59.
57. Akoulitchev, S., S. Chuikov, and D. Reinberg, *TFIIH is negatively regulated by cdk8-containing mediator complexes.* Nature, 2000. **407**(6800): p. 102-6.
58. Taatjes, D.J., *The human Mediator complex: a versatile, genome-wide regulator of transcription.* Trends Biochem Sci, 2010.
59. Belakavadi, M. and J.D. Fondell, *Cyclin-dependent kinase 8 positively cooperates with Mediator to promote thyroid hormone receptor-dependent transcriptional activation.* Mol Cell Biol, 2010. **30**(10): p. 2437-48.
60. Donner, A.J., C.C. Ebmeier, D.J. Taatjes, and J.M. Espinosa, *CDK8 is a positive regulator of transcriptional elongation within the serum response network.* Nat Struct Mol Biol, 2010. **17**(2): p. 194-201.

61. Alarcon, C., A.I. Zaromytidou, Q. Xi, S. Gao, J. Yu, S. Fujisawa, A. Barlas, A.N. Miller, K. Manova-Todorova, M.J. Macias, G. Sapkota, D. Pan, and J. Massague, *Nuclear CDKs drive Smad transcriptional activation and turnover in BMP and TGF-beta pathways.* Cell, 2009. **139**(4): p. 757-69.
62. Fryer, C.J., J.B. White, and K.A. Jones, *Mastermind recruits CycC:CDK8 to phosphorylate the Notch ICD and coordinate activation with turnover.* Mol Cell, 2004. **16**(4): p. 509-20.
63. Loncle, N., M. Boube, L. Joulia, C. Boschiero, M. Werner, D.L. Cribbs, and H.M. Bourbon, *Distinct roles for Mediator Cdk8 module subunits in Drosophila development.* EMBO J, 2007. **26**(4): p. 1045-54.
64. Yuan, C.X., M. Ito, J.D. Fondell, Z.Y. Fu, and R.G. Roeder, *The TRAP220 component of a thyroid hormone receptor- associated protein (TRAP) coactivator complex interacts directly with nuclear receptors in a ligand-dependent fashion.* Proc Natl Acad Sci U S A, 1998. **95**(14): p. 7939-44.
65. Ito, M., C.X. Yuan, H.J. Okano, R.B. Darnell, and R.G. Roeder, *Involvement of the TRAP220 component of the TRAP/SMCC coactivator complex in embryonic development and thyroid hormone action.* Mol Cell, 2000. **5**(4): p. 683-93.
66. Zhang, X., A. Krutchinsky, A. Fukuda, W. Chen, S. Yamamura, B.T. Chait, and R.G. Roeder, *MED1/TRAP220 exists predominantly in a TRAP/ Mediator subpopulation enriched in RNA polymerase II and is required for ER-mediated transcription.* Mol Cell, 2005. **19**(1): p. 89-100.
67. Belakavadi, M., P.K. Pandey, R. Vijayvargia, and J.D. Fondell, *MED1 phosphorylation promotes its association with mediator: implications for nuclear receptor signaling.* Mol Cell Biol, 2008. **28**(12): p. 3932-42.
68. Jiang, P., Q. Hu, M. Ito, S. Meyer, S. Waltz, S. Khan, R.G. Roeder, and X. Zhang, *Key roles for MED1 LxxLL motifs in pubertal mammary gland development and luminal-cell differentiation.* Proc Natl Acad Sci U S A, 2010. **107**(15): p. 6765-70.
69. Ge, K., Y.W. Cho, H. Guo, T.B. Hong, M. Guermah, M. Ito, H. Yu, M. Kalkum, and R.G. Roeder, *Alternative mechanisms by which mediator subunit MED1/TRAP220 regulates peroxisome proliferator-activated receptor gamma-stimulated adipogenesis and target gene expression.* Mol Cell Biol, 2008. **28**(3): p. 1081-91.
70. Chen, W., Q. Yang, and R.G. Roeder, *Dynamic interactions and cooperative functions of PGC-1alpha and MED1 in TRalpha-mediated activation of the brown-fat-specific UCP-1 gene.* Mol Cell, 2009. **35**(6): p. 755-68.
71. Kim, J.H., C.K. Yang, K. Heo, R.G. Roeder, W. An, and M.R. Stallcup, *CCAR1, a key regulator of mediator complex recruitment to nuclear receptor transcription complexes.* Mol Cell, 2008. **31**(4): p. 510-9.
72. Lee, H.K., U.H. Park, E.J. Kim, and S.J. Um, *MED25 is distinct from TRAP220/MED1 in cooperating with CBP for retinoid receptor activation.* EMBO J, 2007. **26**(15): p. 3545-57.
73. Malik, S. and R.G. Roeder, *Dynamic regulation of pol II transcription by the mammalian Mediator complex.* Trends Biochem Sci, 2005. **30**(5): p. 256-63.
74. Hittelman, A.B., D. Burakov, J.A. Iniguez-Lluhi, L.P. Freedman, and M.J. Garabedian, *Differential regulation of glucocorticoid receptor transcriptional activation via AF-1-associated proteins.* EMBO J, 1999. **18**(19): p. 5380-8.
75. Meyer, K.D., S.C. Lin, C. Bernecky, Y. Gao, and D.J. Taatjes, *p53 activates transcription by directing structural shifts in Mediator.* Nat Struct Mol Biol, 2010. **17**(6): p. 753-60.
76. Zernicka-Goetz, M., S.A. Morris, and A.W. Bruce, *Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo.* Nat Rev Genet, 2009. **10**(7): p. 467-77.
77. Marikawa, Y. and V.B. Alarcon, *Establishment of trophectoderm and inner cell mass lineages in the mouse embryo.* Mol Reprod Dev, 2009. **76**(11): p. 1019-32.
78. Sasaki, H., *Mechanisms of trophectoderm fate specification in preimplantation mouse development.* Dev Growth Differ, 2010. **52**(3): p. 263-73.
79. Piotrowska-Nitsche, K., A. Perea-Gomez, S. Haraguchi, and M. Zernicka-Goetz, *Four-cell stage mouse blastomeres have different developmental properties.* Development, 2005. **132**(3): p. 479-90.
80. Torres-Padilla, M.E., D.E. Parfitt, T. Kouzarides, and M. Zernicka-Goetz, *Histone arginine methylation regulates pluripotency in the early mouse embryo.* Nature, 2007. **445**(7124): p. 214-8.
81. Jedrusik, A., D.E. Parfitt, G. Guo, M. Skamagki, J.B. Grabarek, M.H. Johnson, P. Robson, and M. Zernicka-Goetz, *Role of Cdx2 and cell polarity in cell allocation and specification of trophectoderm and inner cell mass in the mouse embryo.* Genes Dev, 2008. **22**(19): p. 2692-706.
82. Suwinska, A., R. Czolowska, W. Ozdzenski, and A.K. Tarkowski, *Blastomeres of the mouse embryo lose totipotency after the fifth cleavage division: expression of Cdx2 and Oct4 and developmental potential of inner and outer blastomeres of 16- and 32-cell embryos.* Dev Biol, 2008. **322**(1): p. 133-44.
83. Ziomek, C.A., M.H. Johnson, and A.H. Handyside, *The developmental potential of mouse 16-cell*

blastomeres. J Exp Zool, 1982. **221**(3): p. 345-55.
84. Rossant, J. and P.P. Tam, *Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse.* Development, 2009. **136**(5): p. 701-13.
85. Guo, G., M. Huss, G.Q. Tong, C. Wang, L. Li Sun, N.D. Clarke, and P. Robson, *Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst.* Dev Cell, 2010. **18**(4): p. 675-85.
86. Nishioka, N., S. Yamamoto, H. Kiyonari, H. Sato, A. Sawada, M. Ota, K. Nakao, and H. Sasaki, *Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos.* Mech Dev, 2008. **125**(3-4): p. 270-83.
87. Nishioka, N., K. Inoue, K. Adachi, H. Kiyonari, M. Ota, A. Ralston, N. Yabuta, S. Hirahara, R.O. Stephenson, N. Ogonuki, R. Makita, H. Kurihara, E.M. Morin-Kensicki, H. Nojima, J. Rossant*, et al.*, *The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass.* Dev Cell, 2009. **16**(3): p. 398-410.
88. Ralston, A., B.J. Cox, N. Nishioka, H. Sasaki, E. Chea, P. Rugg-Gunn, G. Guo, P. Robson, J.S. Draper, and J. Rossant, *Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2.* Development, 2010. **137**(3): p. 395-403.
89. Niwa, H., Y. Toyooka, D. Shimosato, D. Strumpf, K. Takahashi, R. Yagi, and J. Rossant, *Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation.* Cell, 2005. **123**(5): p. 917-29.
90. Dietrich, J.E. and T. Hiiragi, *Stochastic patterning in the mouse pre-implantation embryo.* Development, 2007. **134**(23): p. 4219-31.
91. Ralston, A. and J. Rossant, *Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo.* Dev Biol, 2008. **313**(2): p. 614-29.
92. Avilion, A.A., S.K. Nicolis, L.H. Pevny, L. Perez, N. Vivian, and R. Lovell-Badge, *Multipotent cell lineages in early mouse development depend on SOX2 function.* Genes Dev, 2003. **17**(1): p. 126-40.
93. Hemberger, M., W. Dean, and W. Reik, *Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal.* Nat Rev Mol Cell Biol, 2009. **10**(8): p. 526-37.
94. Ng, R.K., W. Dean, C. Dawson, D. Lucifero, Z. Madeja, W. Reik, and M. Hemberger, *Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5.* Nat Cell Biol, 2008. **10**(11): p. 1280-90.
95. Chazaud, C., Y. Yamanaka, T. Pawson, and J. Rossant, *Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway.* Developmental Cell, 2006. **10**(5): p. 615-624.
96. Morris, S.A., R.T. Teo, H. Li, P. Robson, D.M. Glover, and M. Zernicka-Goetz, *Origin and formation of the first two distinct cell types of the inner cell mass in the mouse embryo.* Proc Natl Acad Sci U S A, 2010. **107**(14): p. 6364-9.
97. Yamanaka, Y., F. Lanner, and J. Rossant, *FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst.* Development, 2010. **137**(5): p. 715-24.
98. Kurimoto, K., Y. Yabuta, Y. Ohinata, Y. Ono, K.D. Uno, R.G. Yamada, H.R. Ueda, and M. Saitou, *An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis.* Nucleic Acids Res, 2006. **34**(5): p. e42.
99. Plusa, B., A. Piliszek, S. Frankenberg, J. Artus, and A.K. Hadjantonakis, *Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst.* Development, 2008. **135**(18): p. 3081-91.
100. Arman, E., R. Haffner-Krausz, Y. Chen, J.K. Heath, and P. Lonai, *Targeted disruption of fibroblast growth factor (FGF) receptor 2 suggests a role for FGF signaling in pregastrulation mammalian development.* Proc Natl Acad Sci U S A, 1998. **95**(9): p. 5082-7.
101. Nichols, J., J. Silva, M. Roode, and A. Smith, *Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo.* Development, 2009. **136**(19): p. 3215-22.
102. Silva, J., J. Nichols, T.W. Theunissen, G. Guo, A.L. van Oosten, O. Barrandon, J. Wray, S. Yamanaka, I. Chambers, and A. Smith, *Nanog is the gateway to the pluripotent ground state.* Cell, 2009. **138**(4): p. 722-37.
103. Chambers, I., D. Colby, M. Robertson, J. Nichols, S. Lee, S. Tweedie, and A. Smith, *Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells.* Cell, 2003. **113**(5): p. 643-55.
104. Chambers, I., J. Silva, D. Colby, J. Nichols, B. Nijmeijer, M. Robertson, J. Vrana, K. Jones, L. Grotewold, and A. Smith, *Nanog safeguards pluripotency and mediates germline development.* Nature, 2007. **450**(7173): p. 1230-4.
105. Fujikura, J., E. Yamato, S. Yonemura, K. Hosoda, S. Masui, K. Nakao, J. Miyazaki Ji, and H. Niwa, *Differentiation of embryonic stem cells is induced by GATA factors.* Genes Dev, 2002. **16**(7): p. 784-9.
106. Niakan, K.K., H. Ji, R. Maehr, S.A. Vokes, K.T. Rodolfa, R.I. Sherwood, M. Yamaki, J.T. Dimos, A.E. Chen, D.A. Melton, A.P. McMahon, and K. Eggan, *Sox17 promotes differentiation in mouse embryonic*

stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. Genes Dev, 2010. **24**(3): p. 312-26.
107. Evans, M.J. and M.H. Kaufman, *Establishment in culture of pluripotential cells from mouse embryos.* Nature, 1981. **292**(5819): p. 154-6.
108. Martin, G.R., *Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells.* Proc Natl Acad Sci U S A, 1981. **78**(12): p. 7634-8.
109. Niwa, H., *Mouse ES cell culture system as a model of development.* Dev Growth Differ, 2010. **52**(3): p. 275-83.
110. Doetschman, T.C., H. Eistetter, M. Katz, W. Schmidt, and R. Kemler, *The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium.* J Embryol Exp Morphol, 1985. **87**: p. 27-45.
111. Niwa, H., J. Miyazaki, and A.G. Smith, *Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells.* Nat Genet, 2000. **24**(4): p. 372-6.
112. Ying, Q.L., J. Nichols, I. Chambers, and A. Smith, *BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3.* Cell, 2003. **115**(3): p. 281-92.
113. von Bubnoff, A. and K.W. Cho, *Intracellular BMP signaling regulation in vertebrates: pathway or network?* Dev Biol, 2001. **239**(1): p. 1-14.
114. Smith, A.G., J.K. Heath, D.D. Donaldson, G.G. Wong, J. Moreau, M. Stahl, and D. Rogers, *Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides.* Nature, 1988. **336**(6200): p. 688-90.
115. Nichols, J., D. Davidson, T. Taga, K. Yoshida, I. Chambers, and A. Smith, *Complementary tissue-specific expression of LIF and LIF-receptor mRNAs in early mouse embryogenesis.* Mech Dev, 1996. **57**(2): p. 123-31.
116. Escary, J.L., J. Perreau, D. Dumenil, S. Ezine, and P. Brulet, *Leukaemia inhibitory factor is necessary for maintenance of haematopoietic stem cells and thymocyte stimulation.* Nature, 1993. **363**(6427): p. 361-4.
117. Yoshida, K., T. Taga, M. Saito, S. Suematsu, A. Kumanogoh, T. Tanaka, H. Fujiwara, M. Hirata, T. Yamagami, T. Nakahata, T. Hirabayashi, Y. Yoneda, K. Tanaka, W.Z. Wang, C. Mori*, et al.*, *Targeted disruption of gp130, a common signal transducer for the interleukin 6 family of cytokines, leads to myocardial and hematological disorders.* Proc Natl Acad Sci U S A, 1996. **93**(1): p. 407-11.
118. Matsuda, T., T. Nakamura, K. Nakao, T. Arai, M. Katsuki, T. Heike, and T. Yokota, *STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells.* EMBO J, 1999. **18**(15): p. 4261-9.
119. Niwa, H., T. Burdon, I. Chambers, and A. Smith, *Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3.* Genes Dev, 1998. **12**(13): p. 2048-60.
120. Hall, J., G. Guo, J. Wray, I. Eyres, J. Nichols, L. Grotewold, S. Morfopoulou, P. Humphreys, W. Mansfield, R. Walker, S. Tomlinson, and A. Smith, *Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem cell self-renewal.* Cell Stem Cell, 2009. **5**(6): p. 597-609.
121. Niwa, H., K. Ogawa, D. Shimosato, and K. Adachi, *A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells.* Nature, 2009. **460**(7251): p. 118-22.
122. Ying, Q.L., J. Wray, J. Nichols, L. Batlle-Morera, B. Doble, J. Woodgett, P. Cohen, and A. Smith, *The ground state of embryonic stem cell self-renewal.* Nature, 2008. **453**(7194): p. 519-23.
123. Silva, J. and A. Smith, *Capturing pluripotency.* Cell, 2008. **132**(4): p. 532-6.
124. Nichols, J. and A. Smith, *Naive and primed pluripotent states.* Cell Stem Cell, 2009. **4**(6): p. 487-92.
125. Boheler, K.R., *Stem cell pluripotency: a cellular trait that depends on transcription factors, chromatin state and a checkpoint deficient cell cycle.* J Cell Physiol, 2009. **221**(1): p. 10-7.
126. Tang, F., C. Barbacioru, S. Bao, C. Lee, E. Nordman, X. Wang, K. Lao, and M.A. Surani, *Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis.* Cell Stem Cell, 2010. **6**(5): p. 468-78.
127. Takahashi, K., K. Mitsui, and S. Yamanaka, *Role of ERas in promoting tumour-like properties in mouse embryonic stem cells.* Nature, 2003. **423**(6939): p. 541-5.
128. Li, W. and S. Ding, *Small molecules that modulate embryonic stem cell fate and somatic cell reprogramming.* Trends Pharmacol Sci, 2010. **31**(1): p. 36-45.
129. Martins-Taylor, K. and R.H. Xu, *Determinants of pluripotency: from avian, rodents, to primates.* J Cell Biochem, 2010. **109**(1): p. 16-25.
130. Yu, J. and J.A. Thomson, *Pluripotent stem cell lines.* Genes Dev, 2008. **22**(15): p. 1987-97.
131. Xu, R.H., X. Chen, D.S. Li, R. Li, G.C. Addicks, C. Glennon, T.P. Zwaka, and J.A. Thomson, *BMP4 initiates human embryonic stem cell differentiation to trophoblast.* Nat Biotechnol, 2002. **20**(12): p. 1261-4.
132. Vallier, L., S. Mendjan, S. Brown, Z. Chng, A. Teo, L.E. Smithers, M.W. Trotter, C.H. Cho, A. Martinez,

P. Rugg-Gunn, G. Brons, and R.A. Pedersen, *Activin/Nodal signalling maintains pluripotency by controlling Nanog expression.* Development, 2009. **136**(8): p. 1339-49.
133. Xu, R.H., T.L. Sampsell-Barron, F. Gu, S. Root, R.M. Peck, G. Pan, J. Yu, J. Antosiewicz-Bourget, S. Tian, R. Stewart, and J.A. Thomson, *NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs.* Cell Stem Cell, 2008. **3**(2): p. 196-206.
134. Brons, I.G., L.E. Smithers, M.W. Trotter, P. Rugg-Gunn, B. Sun, S.M. Chuva de Sousa Lopes, S.K. Howlett, A. Clarkson, L. Ahrlund-Richter, R.A. Pedersen, and L. Vallier, *Derivation of pluripotent epiblast stem cells from mammalian embryos.* Nature, 2007. **448**(7150): p. 191-5.
135. Tesar, P.J., J.G. Chenoweth, F.A. Brook, T.J. Davies, E.P. Evans, D.L. Mack, R.L. Gardner, and R.D. McKay, *New cell lines from mouse epiblast share defining features with human embryonic stem cells.* Nature, 2007. **448**(7150): p. 196-9.
136. Lengner, C.J., A.A. Gimelbrant, J.A. Erwin, A.W. Cheng, M.G. Guenther, G.G. Welstead, R. Alagappan, G.M. Frampton, P. Xu, J. Muffat, S. Santagata, D. Powers, C.B. Barrett, R.A. Young, J.T. Lee, *et al.*, *Derivation of Pre-X Inactivation Human Embryonic Stem Cells under Physiological Oxygen Concentrations.* Cell, 2010.
137. Guo, G., J. Yang, J. Nichols, J.S. Hall, I. Eyres, W. Mansfield, and A. Smith, *Klf4 reverts developmentally programmed restriction of ground state pluripotency.* Development, 2009. **136**(7): p. 1063-9.
138. Greber, B., G. Wu, C. Bernemann, J.Y. Joo, D.W. Han, K. Ko, N. Tapia, D. Sabour, J. Sterneckert, P. Tesar, and H.R. Scholer, *Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells.* Cell Stem Cell, 2010. **6**(3): p. 215-26.
139. Buehr, M., S. Meek, K. Blair, J. Yang, J. Ure, J. Silva, R. McLay, J. Hall, Q.L. Ying, and A. Smith, *Capture of authentic embryonic stem cells from rat blastocysts.* Cell, 2008. **135**(7): p. 1287-98.
140. Li, P., C. Tong, R. Mehrian-Shai, L. Jia, N. Wu, Y. Yan, R.E. Maxson, E.N. Schulze, H. Song, C.L. Hsieh, M.F. Pera, and Q.L. Ying, *Germline competent embryonic stem cells derived from rat blastocysts.* Cell, 2008. **135**(7): p. 1299-310.
141. Vaags, A.K., S. Rosic-Kablar, C.J. Gartley, Y.Z. Zheng, A. Chesney, D.A. Villagomez, S.A. Kruth, and M.R. Hough, *Derivation and characterization of canine embryonic stem cell lines with in vitro and in vivo differentiation potential.* Stem Cells, 2009. **27**(2): p. 329-40.
142. Takada, T., Y. Suzuki, Y. Kondo, N. Kadota, K. Kobayashi, S. Nito, H. Kimura, and R. Torii, *Monkey embryonic stem cell lines expressing green fluorescent protein.* Cell Transplant, 2002. **11**(7): p. 631-5.
143. Thomson, J.A., J. Kalishman, T.G. Golos, M. Durning, C.P. Harris, R.A. Becker, and J.P. Hearn, *Isolation of a primate embryonic stem cell line.* Proc Natl Acad Sci U S A, 1995. **92**(17): p. 7844-8.
144. Thomson, J.A., J. Kalishman, T.G. Golos, M. Durning, C.P. Harris, and J.P. Hearn, *Pluripotent cell lines derived from common marmoset (Callithrix jacchus) blastocysts.* Biol Reprod, 1996. **55**(2): p. 254-9.
145. Ueda, S., M. Yoshikawa, Y. Ouji, K. Saito, K. Moriya, M. Nishiofuku, N. Hayashi, S. Ishizaka, K. Shimada, N. Konishi, and H. Fukui, *Cynomolgus monkey embryonic stem cell lines express green fluorescent protein.* J Biosci Bioeng, 2006. **102**(1): p. 14-20.
146. Adewumi, O., B. Aflatoonian, L. Ahrlund-Richter, M. Amit, P.W. Andrews, G. Beighton, P.A. Bello, N. Benvenisty, L.S. Berry, S. Bevan, B. Blum, J. Brooking, K.G. Chen, A.B. Choo, G.A. Churchill, *et al.*, *Characterization of human embryonic stem cell lines by the International Stem Cell Initiative.* Nat Biotechnol, 2007. **25**(7): p. 803-16.
147. Thomson, J.A., J. Itskovitz-Eldor, S.S. Shapiro, M.A. Waknitz, J.J. Swiergiel, V.S. Marshall, and J.M. Jones, *Embryonic stem cell lines derived from human blastocysts.* Science, 1998. **282**(5391): p. 1145-7.
148. Okita, K., T. Ichisaka, and S. Yamanaka, *Generation of germline-competent induced pluripotent stem cells.* Nature, 2007. **448**(7151): p. 313-7.
149. Silva, J., O. Barrandon, J. Nichols, J. Kawaguchi, T.W. Theunissen, and A. Smith, *Promotion of reprogramming to ground state pluripotency by signal inhibition.* PLoS Biol, 2008. **6**(10): p. e253.
150. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.
151. Wernig, M., A. Meissner, R. Foreman, T. Brambrink, M. Ku, K. Hochedlinger, B.E. Bernstein, and R. Jaenisch, *In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state.* Nature, 2007. **448**(7151): p. 318-24.
152. Zhao, X.Y., W. Li, Z. Lv, L. Liu, M. Tong, T. Hai, J. Hao, C.L. Guo, Q.W. Ma, L. Wang, F. Zeng, and Q. Zhou, *iPS cells produce viable mice through tetraploid complementation.* Nature, 2009. **461**(7260): p. 86-90.
153. Boland, M.J., J.L. Hazen, K.L. Nazor, A.R. Rodriguez, W. Gifford, G. Martin, S. Kupriyanov, and K.K. Baldwin, *Adult mice generated from induced pluripotent stem cells.* Nature, 2009. **461**(7260): p. 91-4.
154. Takahashi, K., K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, and S. Yamanaka, *Induction of*

pluripotent stem cells from adult human fibroblasts by defined factors. Cell, 2007. **131**(5): p. 861-72.
155. Yu, J., M.A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J.L. Frane, S. Tian, J. Nie, G.A. Jonsdottir, V. Ruotti, R. Stewart, Slukvin, II, and J.A. Thomson, *Induced pluripotent stem cell lines derived from human somatic cells.* Science, 2007. **318**(5858): p. 1917-20.
156. Liu, H., F. Zhu, J. Yong, P. Zhang, P. Hou, H. Li, W. Jiang, J. Cai, M. Liu, K. Cui, X. Qu, T. Xiang, D. Lu, X. Chi, G. Gao*, et al.*, *Generation of induced pluripotent stem cells from adult rhesus monkey fibroblasts.* Cell Stem Cell, 2008. **3**(6): p. 587-90.
157. Chambers, I. and S.R. Tomlinson, *The transcriptional foundation of pluripotency.* Development, 2009. **136**(14): p. 2311-22.
158. Rao, S. and S.H. Orkin, *Unraveling the transcriptional network controlling ES cell pluripotency.* Genome Biol, 2006. **7**(8): p. 230.
159. Nichols, J., B. Zevnik, K. Anastassiadis, H. Niwa, D. Klewe-Nebenius, I. Chambers, H. Scholer, and A. Smith, *Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4.* Cell, 1998. **95**(3): p. 379-91.
160. Masui, S., Y. Nakatake, Y. Toyooka, D. Shimosato, R. Yagi, K. Takahashi, H. Okochi, A. Okuda, R. Matoba, A.A. Sharov, M.S. Ko, and H. Niwa, *Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells.* Nat Cell Biol, 2007. **9**(6): p. 625-35.
161. Boyer, L.A., T.I. Lee, M.F. Cole, S.E. Johnstone, S.S. Levine, J.P. Zucker, M.G. Guenther, R.M. Kumar, H.L. Murray, R.G. Jenner, D.K. Gifford, D.A. Melton, R. Jaenisch, and R.A. Young, *Core transcriptional regulatory circuitry in human embryonic stem cells.* Cell, 2005. **122**(6): p. 947-56.
162. Chen, X., H. Xu, P. Yuan, F. Fang, M. Huss, V.B. Vega, E. Wong, Y.L. Orlov, W. Zhang, J. Jiang, Y.H. Loh, H.C. Yeo, Z.X. Yeo, V. Narang, K.R. Govindarajan*, et al.*, *Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.* Cell, 2008. **133**(6): p. 1106-17.
163. Kim, J., J. Chu, X. Shen, J. Wang, and S.H. Orkin, *An extended transcriptional network for pluripotency of embryonic stem cells.* Cell, 2008. **132**(6): p. 1049-61.
164. Loh, Y.H., Q. Wu, J.L. Chew, V.B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K.Y. Wong, K.W. Sung, C.W. Lee, X.D. Zhao, K.P. Chiu*, et al.*, *The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells.* Nat Genet, 2006. **38**(4): p. 431-40.
165. Ouyang, Z., Q. Zhou, and W.H. Wong, *ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells.* Proc Natl Acad Sci U S A, 2009. **106**(51): p. 21521-6.
166. Suzuki, A., A. Raya, Y. Kawakami, M. Morita, T. Matsui, K. Nakashima, F.H. Gage, C. Rodriguez-Esteban, and J.C. Izpisua Belmonte, *Nanog binds to Smad1 and blocks bone morphogenetic protein-induced differentiation of embryonic stem cells.* Proc Natl Acad Sci U S A, 2006. **103**(27): p. 10294-9.
167. Hanna, L.A., R.K. Foreman, I.A. Tarasenko, D.S. Kessler, and P.A. Labosky, *Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo.* Genes Dev, 2002. **16**(20): p. 2650-61.
168. Ivanova, N., R. Dobrin, R. Lu, I. Kotenko, J. Levorse, C. DeCoste, X. Schafer, Y. Lun, and I.R. Lemischka, *Dissecting self-renewal in stem cells with RNA interference.* Nature, 2006. **442**(7102): p. 533-8.
169. Ema, M., D. Mori, H. Niwa, Y. Hasegawa, Y. Yamanaka, S. Hitoshi, J. Mimura, Y. Kawabe, T. Hosoya, M. Morita, D. Shimosato, K. Uchida, N. Suzuki, J. Yanagisawa, K. Sogawa*, et al.*, *Kruppel-like factor 5 is essential for blastocyst development and the normal self-renewal of mouse ESCs.* Cell Stem Cell, 2008. **3**(5): p. 555-67.
170. Parisi, S., F. Passaro, L. Aloia, I. Manabe, R. Nagai, L. Pastore, and T. Russo, *Klf5 is involved in self-renewal of mouse embryonic stem cells.* J Cell Sci, 2008. **121**(Pt 16): p. 2629-34.
171. Zhang, J., W.L. Tam, G.Q. Tong, Q. Wu, H.Y. Chan, B.S. Soh, Y. Lou, J. Yang, Y. Ma, L. Chai, H.H. Ng, T. Lufkin, P. Robson, and B. Lim, *Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1.* Nat Cell Biol, 2006. **8**(10): p. 1114-23.
172. Wang, J., S. Rao, J. Chu, X. Shen, D.N. Levasseur, T.W. Theunissen, and S.H. Orkin, *A protein interaction network for pluripotency of embryonic stem cells.* Nature, 2006. **444**(7117): p. 364-8.
173. Niakan, K.K., E.C. Davis, R.C. Clipsham, M. Jiang, D.B. Dehart, K.K. Sulik, and E.R. McCabe, *Novel role for the orphan nuclear receptor Dax1 in embryogenesis, different from steroidogenesis.* Mol Genet Metab, 2006. **88**(3): p. 261-71.
174. Han, J., P. Yuan, H. Yang, J. Zhang, B.S. Soh, P. Li, S.L. Lim, S. Cao, J. Tay, Y.L. Orlov, T. Lufkin, H.H. Ng, W.L. Tam, and B. Lim, *Tbx3 improves the germ-line competency of induced pluripotent stem cells.* Nature, 2010. **463**(7284): p. 1096-100.
175. Odom, D.T., R.D. Dowell, E.S. Jacobsen, L. Nekludova, P.A. Rolfe, T.W. Danford, D.K. Gifford, E. Fraenkel, G.I. Bell, and R.A. Young, *Core transcriptional regulatory circuitry in human hepatocytes.* Mol Syst Biol, 2006. **2**: p. 2006 0017.
176. Zinzen, R.P., C. Girardot, J. Gagneur, M. Braun, and E.E. Furlong, *Combinatorial binding predicts spatio-temporal cis-regulatory activity.* Nature, 2009. **462**(7269): p. 65-70.
177. Ambrosetti, D.C., C. Basilico, and L. Dailey, *Synergistic activation of the fibroblast growth factor 4*

enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. Mol Cell Biol, 1997. **17**(11): p. 6321-9.
178. Meshorer, E., D. Yellajoshula, E. George, P.J. Scambler, D.T. Brown, and T. Misteli, *Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells.* Dev Cell, 2006. **10**(1): p. 105-16.
179. Fan, Y., T. Nikitina, E.M. Morin-Kensicki, J. Zhao, T.R. Magnuson, C.L. Woodcock, and A.I. Skoultchi, *H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo.* Mol Cell Biol, 2003. **23**(13): p. 4559-72.
180. Fan, Y., T. Nikitina, J. Zhao, T.J. Fleury, R. Bhattacharyya, E.E. Bouhassira, A. Stein, C.L. Woodcock, and A.I. Skoultchi, *Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation.* Cell, 2005. **123**(7): p. 1199-212.
181. Ahmed, K., H. Dehghani, P. Rugg-Gunn, E. Fussner, J. Rossant, and D.P. Bazett-Jones, *Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo.* PLoS One, 2010. **5**(5): p. e10531.
182. Efroni, S., R. Duttagupta, J. Cheng, H. Dehghani, D.J. Hoeppner, C. Dash, D.P. Bazett-Jones, S. Le Grice, R.D. McKay, K.H. Buetow, T.R. Gingeras, T. Misteli, and E. Meshorer, *Global transcription in pluripotent embryonic stem cells.* Cell Stem Cell, 2008. **2**(5): p. 437-47.
183. Gaspar-Maia, A., A. Alajem, F. Polesso, R. Sridharan, M.J. Mason, A. Heidersbach, J. Ramalho-Santos, M.T. McManus, K. Plath, E. Meshorer, and M. Ramalho-Santos, *Chd1 regulates open chromatin and pluripotency of embryonic stem cells.* Nature, 2009. **460**(7257): p. 863-8.
184. Boyer, L.A., K. Plath, J. Zeitlinger, T. Brambrink, L.A. Medeiros, T.I. Lee, S.S. Levine, M. Wernig, A. Tajonar, M.K. Ray, G.W. Bell, A.P. Otte, M. Vidal, D.K. Gifford, R.A. Young*, et al.*, *Polycomb complexes repress developmental regulators in murine embryonic stem cells.* Nature, 2006. **441**(7091): p. 349-53.
185. Lee, T.I., R.G. Jenner, L.A. Boyer, M.G. Guenther, S.S. Levine, R.M. Kumar, B. Chevalier, S.E. Johnstone, M.F. Cole, K. Isono, H. Koseki, T. Fuchikami, K. Abe, H.L. Murray, J.P. Zucker*, et al.*, *Control of developmental regulators by Polycomb in human embryonic stem cells.* Cell, 2006. **125**(2): p. 301-13.
186. Endoh, M., T.A. Endo, T. Endoh, Y. Fujimura, O. Ohara, T. Toyoda, A.P. Otte, M. Okano, N. Brockdorff, M. Vidal, and H. Koseki, *Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity.* Development, 2008. **135**(8): p. 1513-24.
187. Ringrose, L. and R. Paro, *Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins.* Annu Rev Genet, 2004. **38**: p. 413-43.
188. Schuettengruber, B., D. Chourrout, M. Vervoort, B. Leblanc, and G. Cavalli, *Genome regulation by polycomb and trithorax proteins.* Cell, 2007. **128**(4): p. 735-45.
189. Stock, J.K., S. Giadrossi, M. Casanova, E. Brookes, M. Vidal, H. Koseki, N. Brockdorff, A.G. Fisher, and A. Pombo, *Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells.* Nat Cell Biol, 2007. **9**(12): p. 1428-35.
190. Eskeland, R., M. Leeb, G.R. Grimes, C. Kress, S. Boyle, D. Sproul, N. Gilbert, Y. Fan, A.I. Skoultchi, A. Wutz, and W.A. Bickmore, *Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination.* Mol Cell, 2010. **38**(3): p. 452-64.
191. Bernstein, B.E., T.S. Mikkelsen, X. Xie, M. Kamal, D.J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S.L. Schreiber, and E.S. Lander, *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-26.
192. Pan, G., S. Tian, J. Nie, C. Yang, V. Ruotti, H. Wei, G.A. Jonsdottir, R. Stewart, and J.A. Thomson, *Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells.* Cell Stem Cell, 2007. **1**(3): p. 299-312.
193. Azuara, V., P. Perry, S. Sauer, M. Spivakov, H.F. Jorgensen, R.M. John, M. Gouti, M. Casanova, G. Warnes, M. Merkenschlager, and A.G. Fisher, *Chromatin signatures of pluripotent cell lines.* Nat Cell Biol, 2006. **8**(5): p. 532-8.
194. Kanhere, A., K. Viiri, C.C. Araujo, J. Rasaiyaah, R.D. Bouwman, W.A. Whyte, C.F. Pereira, E. Brookes, K. Walker, G.W. Bell, A. Pombo, A.G. Fisher, R.A. Young, and R.G. Jenner, *Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2.* Mol Cell, 2010. **38**(5): p. 675-88.
195. Bilodeau, S., M.H. Kagey, G.M. Frampton, P.B. Rahl, and R.A. Young, *SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state.* Genes Dev, 2009. **23**(21): p. 2484-9.
196. Mikkelsen, T.S., M. Ku, D.B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T.K. Kim, R.P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ*, et al.*, *Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.* Nature, 2007. **448**(7153): p. 553-60.

197. Mohn, F., M. Weber, M. Rebhan, T.C. Roloff, J. Richter, M.B. Stadler, M. Bibel, and D. Schubeler, *Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors.* Mol Cell, 2008. **30**(6): p. 755-66.
198. Xu, J., J.A. Watts, S.D. Pope, P. Gadue, M. Kamps, K. Plath, K.S. Zaret, and S.T. Smale, *Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells.* Genes Dev, 2009. **23**(24): p. 2824-38.
199. Bultman, S., T. Gebuhr, D. Yee, C. La Mantia, J. Nicholson, A. Gilliam, F. Randazzo, D. Metzger, P. Chambon, G. Crabtree, and T. Magnuson, *A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes.* Mol Cell, 2000. **6**(6): p. 1287-95.
200. Ho, L., J.L. Ronan, J. Wu, B.T. Staahl, L. Chen, A. Kuo, J. Lessard, A.I. Nesvizhskii, J. Ranish, and G.R. Crabtree, *An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency.* Proc Natl Acad Sci U S A, 2009. **106**(13): p. 5181-6.
201. Gao, X., P. Tate, P. Hu, R. Tjian, W.C. Skarnes, and Z. Wang, *ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a.* Proc Natl Acad Sci U S A, 2008. **105**(18): p. 6656-61.
202. Yan, Z., Z. Wang, L. Sharova, A.A. Sharov, C. Ling, Y. Piao, K. Aiba, R. Matoba, W. Wang, and M.S. Ko, *BAF250B-associated SWI/SNF chromatin-remodeling complex is required to maintain undifferentiated mouse embryonic stem cells.* Stem Cells, 2008. **26**(5): p. 1155-65.
203. Ho, L., R. Jothi, J.L. Ronan, K. Cui, K. Zhao, and G.R. Crabtree, *An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network.* Proc Natl Acad Sci U S A, 2009. **106**(13): p. 5187-91.
204. Bultman, S.J., T.C. Gebuhr, H. Pan, P. Svoboda, R.M. Schultz, and T. Magnuson, *Maternal BRG1 regulates zygotic genome activation in the mouse.* Genes Dev, 2006. **20**(13): p. 1744-54.
205. Liang, J., M. Wan, Y. Zhang, P. Gu, H. Xin, S.Y. Jung, J. Qin, J. Wong, A.J. Cooney, D. Liu, and Z. Songyang, *Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells.* Nat Cell Biol, 2008. **10**(6): p. 731-9.
206. Kaji, K., I.M. Caballero, R. MacLeod, J. Nichols, V.A. Wilson, and B. Hendrich, *The NuRD component Mbd3 is required for pluripotency of embryonic stem cells.* Nat Cell Biol, 2006. **8**(3): p. 285-92.
207. Fazzio, T.G., J.T. Huff, and B. Panning, *An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity.* Cell, 2008. **134**(1): p. 162-74.
208. Herceg, Z., W. Hulla, D. Gell, C. Cuenin, M. Lleonart, S. Jackson, and Z.Q. Wang, *Disruption of Trrap causes early embryonic lethality and defects in cell cycle progression.* Nat Genet, 2001. **29**(2): p. 206-11.
209. Hu, Y., J.B. Fisher, S. Koprowski, D. McAllister, M.S. Kim, and J. Lough, *Homozygous disruption of the Tip60 gene causes early embryonic lethality.* Dev Dyn, 2009. **238**(11): p. 2912-21.
210. Sapountzi, V., I.R. Logan, and C.N. Robson, *Cellular functions of TIP60.* Int J Biochem Cell Biol, 2006. **38**(9): p. 1496-509.
211. Creyghton, M.P., S. Markoulaki, S.S. Levine, J. Hanna, M.A. Lodato, K. Sha, R.A. Young, R. Jaenisch, and L.A. Boyer, *H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment.* Cell, 2008. **135**(4): p. 649-61.
212. Park, J.H., X.J. Sun, and R.G. Roeder, *The SANT domain of p400 ATPase represses acetyltransferase activity and coactivator function of TIP60 in basal p21 gene expression.* Mol Cell Biol, 2010. **30**(11): p. 2750-61.
213. Campbell, K.H., J. McWhir, W.A. Ritchie, and I. Wilmut, *Sheep cloned by nuclear transfer from a cultured cell line.* Nature, 1996. **380**(6569): p. 64-6.
214. Cowan, C.A., J. Atienza, D.A. Melton, and K. Eggan, *Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells.* Science, 2005. **309**(5739): p. 1369-73.
215. Tada, M., Y. Takahama, K. Abe, N. Nakatsuji, and T. Tada, *Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells.* Curr Biol, 2001. **11**(19): p. 1553-8.
216. Nakagawa, M., M. Koyanagi, K. Tanabe, K. Takahashi, T. Ichisaka, T. Aoi, K. Okita, Y. Mochiduki, N. Takizawa, and S. Yamanaka, *Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts.* Nat Biotechnol, 2008. **26**(1): p. 101-6.
217. Wernig, M., A. Meissner, J.P. Cassady, and R. Jaenisch, *c-Myc is dispensable for direct reprogramming of mouse fibroblasts.* Cell Stem Cell, 2008. **2**(1): p. 10-2.
218. Meissner, A., M. Wernig, and R. Jaenisch, *Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells.* Nat Biotechnol, 2007. **25**(10): p. 1177-81.
219. Takahashi, K., K. Okita, M. Nakagawa, and S. Yamanaka, *Induction of pluripotent stem cells from fibroblast cultures.* Nat Protoc, 2007. **2**(12): p. 3081-9.
220. Robbins, R.D., N. Prasain, B.F. Maier, M.C. Yoder, and R.G. Mirmira, *Inducible pluripotent stem cells: not quite ready for prime time?* Curr Opin Organ Transplant, 2010. **15**(1): p. 61-7.
221. Tsubooka, N., T. Ichisaka, K. Okita, K. Takahashi, M. Nakagawa, and S. Yamanaka, *Roles of Sall4 in the*

generation of pluripotent stem cells from blastocysts and fibroblasts. Genes Cells, 2009. **14**(6): p. 683-94.
222. Feng, B., J. Jiang, P. Kraus, J.H. Ng, J.C. Heng, Y.S. Chan, L.P. Yaw, W. Zhang, Y.H. Loh, J. Han, V.B. Vega, V. Cacheux-Rataboul, B. Lim, T. Lufkin, and H.H. Ng, *Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb.* Nat Cell Biol, 2009. **11**(2): p. 197-203.
223. Heng, J.C., B. Feng, J. Han, J. Jiang, P. Kraus, J.H. Ng, Y.L. Orlov, M. Huss, L. Yang, T. Lufkin, B. Lim, and H.H. Ng, *The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells.* Cell Stem Cell, 2010. **6**(2): p. 167-74.
224. Hong, H., K. Takahashi, T. Ichisaka, T. Aoi, O. Kanagawa, M. Nakagawa, K. Okita, and S. Yamanaka, *Suppression of induced pluripotent stem cell generation by the p53-p21 pathway.* Nature, 2009. **460**(7259): p. 1132-5.
225. Kawamura, T., J. Suzuki, Y.V. Wang, S. Menendez, L.B. Morera, A. Raya, G.M. Wahl, and J.C. Belmonte, *Linking the p53 tumour suppressor pathway to somatic cell reprogramming.* Nature, 2009. **460**(7259): p. 1140-4.
226. Zhao, Y., X. Yin, H. Qin, F. Zhu, H. Liu, W. Yang, Q. Zhang, C. Xiang, P. Hou, Z. Song, Y. Liu, J. Yong, P. Zhang, J. Cai, M. Liu*, et al.*, *Two supporting factors greatly improve the efficiency of human iPSC generation.* Cell Stem Cell, 2008. **3**(5): p. 475-9.
227. Esteban, M.A., T. Wang, B. Qin, J. Yang, D. Qin, J. Cai, W. Li, Z. Weng, J. Chen, S. Ni, K. Chen, Y. Li, X. Liu, J. Xu, S. Zhang*, et al.*, *Vitamin C enhances the generation of mouse and human induced pluripotent stem cells.* Cell Stem Cell, 2010. **6**(1): p. 71-9.
228. Yoshida, Y., K. Takahashi, K. Okita, T. Ichisaka, and S. Yamanaka, *Hypoxia enhances the generation of induced pluripotent stem cells.* Cell Stem Cell, 2009. **5**(3): p. 237-41.
229. Huangfu, D., K. Osafune, R. Maehr, W. Guo, A. Eijkelenboom, S. Chen, W. Muhlestein, and D.A. Melton, *Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2.* Nat Biotechnol, 2008. **26**(11): p. 1269-75.
230. Shi, Y., C. Desponts, J.T. Do, H.S. Hahm, H.R. Scholer, and S. Ding, *Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds.* Cell Stem Cell, 2008. **3**(5): p. 568-74.
231. Ichida, J.K., J. Blanchard, K. Lam, E.Y. Son, J.E. Chung, D. Egli, K.M. Loh, A.C. Carter, F.P. Di Giorgio, K. Koszka, D. Huangfu, H. Akutsu, D.R. Liu, L.L. Rubin, and K. Eggan, *A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog.* Cell Stem Cell, 2009. **5**(5): p. 491-503.
232. Soldner, F., D. Hockemeyer, C. Beard, Q. Gao, G.W. Bell, E.G. Cook, G. Hargus, A. Blak, O. Cooper, M. Mitalipova, O. Isacson, and R. Jaenisch, *Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors.* Cell, 2009. **136**(5): p. 964-77.
233. Kaji, K., K. Norrby, A. Paca, M. Mileikovsky, P. Mohseni, and K. Woltjen, *Virus-free induction of pluripotency and subsequent excision of reprogramming factors.* Nature, 2009. **458**(7239): p. 771-5.
234. Woltjen, K., I.P. Michael, P. Mohseni, R. Desai, M. Mileikovsky, R. Hamalainen, R. Cowling, W. Wang, P. Liu, M. Gertsenstein, K. Kaji, H.K. Sung, and A. Nagy, *piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells.* Nature, 2009. **458**(7239): p. 766-70.
235. O'Malley, J., K. Woltjen, and K. Kaji, *New strategies to generate induced pluripotent stem cells.* Curr Opin Biotechnol, 2009. **20**(5): p. 516-21.
236. Zhou, H., S. Wu, J.Y. Joo, S. Zhu, D.W. Han, T. Lin, S. Trauger, G. Bien, S. Yao, Y. Zhu, G. Siuzdak, H.R. Scholer, L. Duan, and S. Ding, *Generation of induced pluripotent stem cells using recombinant proteins.* Cell Stem Cell, 2009. **4**(5): p. 381-4.
237. Mikkelsen, T.S., J. Hanna, X. Zhang, M. Ku, M. Wernig, P. Schorderet, B.E. Bernstein, R. Jaenisch, E.S. Lander, and A. Meissner, *Dissecting direct reprogramming through integrative genomic analysis.* Nature, 2008. **454**(7200): p. 49-55.
238. Sridharan, R., J. Tchieu, M.J. Mason, R. Yachechko, E. Kuoy, S. Horvath, Q. Zhou, and K. Plath, *Role of the murine reprogramming factors in the induction of pluripotency.* Cell, 2009. **136**(2): p. 364-77.
239. Bhutani, N., J.J. Brady, M. Damian, A. Sacco, S.Y. Corbel, and H.M. Blau, *Reprogramming towards pluripotency requires AID-dependent DNA demethylation.* Nature, 2010. **463**(7284): p. 1042-7.
240. Stadtfeld, M., E. Apostolou, H. Akutsu, A. Fukuda, P. Follett, S. Natesan, T. Kono, T. Shioda, and K. Hochedlinger, *Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells.* Nature, 2010. **465**(7295): p. 175-81.
241. White, J., E. Stead, R. Faast, S. Conn, P. Cartwright, and S. Dalton, *Developmental activation of the Rb-E2F pathway and establishment of cell cycle-regulated cyclin-dependent kinase activity during embryonic stem cell differentiation.* Mol Biol Cell, 2005. **16**(4): p. 2018-27.
242. Singh, A.M. and S. Dalton, *The cell cycle and Myc intersect with mechanisms that regulate pluripotency and reprogramming.* Cell Stem Cell, 2009. **5**(2): p. 141-9.
243. Vierbuchen, T., A. Ostermeier, Z.P. Pang, Y. Kokubu, T.C. Sudhof, and M. Wernig, *Direct conversion of*

fibroblasts to functional neurons by defined factors. Nature, 2010. **463**(7284): p. 1035-41.
244. Zhou, Q., J. Brown, A. Kanarek, J. Rajagopal, and D.A. Melton, *In vivo reprogramming of adult pancreatic exocrine cells to beta-cells.* Nature, 2008. **455**(7213): p. 627-32.
245. Scholer, H.R., S. Ruppert, N. Suzuki, K. Chowdhury, and P. Gruss, *New type of POU domain in germ line-specific protein Oct-4.* Nature, 1990. **344**(6265): p. 435-9.
246. Ryan, A.K. and M.G. Rosenfeld, *POU domain family values: flexibility, partnerships, and developmental codes.* Genes Dev, 1997. **11**(10): p. 1207-25.
247. Phillips, K. and B. Luisi, *The virtuoso of versatility: POU proteins that flex to fit.* J Mol Biol, 2000. **302**(5): p. 1023-39.
248. Dailey, L. and C. Basilico, *Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes.* J Cell Physiol, 2001. **186**(3): p. 315-28.
249. Kang, J., A. Shakya, and D. Tantin, *Stem cells, stress, metabolism and cancer: a drama in two Octs.* Trends Biochem Sci, 2009. **34**(10): p. 491-9.
250. Ambrosetti, D.C., H.R. Scholer, L. Dailey, and C. Basilico, *Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer.* J Biol Chem, 2000. **275**(30): p. 23387-97.
251. Scholer, H.R., A.K. Hatzopoulos, R. Balling, N. Suzuki, and P. Gruss, *A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor.* EMBO J, 1989. **8**(9): p. 2543-50.
252. Niwa, H., S. Masui, I. Chambers, A.G. Smith, and J. Miyazaki, *Phenotypic complementation establishes requirements for specific POU domain and generic transactivation function of Oct-3/4 in embryonic stem cells.* Mol Cell Biol, 2002. **22**(5): p. 1526-36.
253. Okazawa, H., K. Okamoto, F. Ishino, T. Ishino-Kaneko, S. Takeda, Y. Toyoda, M. Muramatsu, and H. Hamada, *The oct3 gene, a gene for an embryonic transcription factor, is controlled by a retinoic acid repressible enhancer.* EMBO J, 1991. **10**(10): p. 2997-3005.
254. Nordhoff, V., K. Hubner, A. Bauer, I. Orlova, A. Malapetsa, and H.R. Scholer, *Comparative analysis of human, bovine, and murine Oct-4 upstream promoter sequences.* Mamm Genome, 2001. **12**(4): p. 309-17.
255. Ben-Shushan, E., H. Sharir, E. Pikarsky, and Y. Bergman, *A dynamic balance between ARP-1/COUP-TFII, EAR-3/COUP-TFI, and retinoic acid receptor:retinoid X receptor heterodimers regulates Oct-3/4 expression in embryonal carcinoma cells.* Mol Cell Biol, 1995. **15**(2): p. 1034-48.
256. Yang, H.M., H.J. Do, D.K. Kim, J.K. Park, W.K. Chang, H.M. Chung, S.Y. Choi, and J.H. Kim, *Transcriptional regulation of human Oct4 by steroidogenic factor-1.* J Cell Biochem, 2007. **101**(5): p. 1198-209.
257. Gu, P., B. Goodwin, A.C. Chung, X. Xu, D.A. Wheeler, R.R. Price, C. Galardi, L. Peng, A.M. Latour, B.H. Koller, J. Gossen, S.A. Kliewer, and A.J. Cooney, *Orphan nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development.* Mol Cell Biol, 2005. **25**(9): p. 3492-505.
258. Zhang, X., J. Zhang, T. Wang, M.A. Esteban, and D. Pei, *Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells.* J Biol Chem, 2008. **283**(51): p. 35825-33.
259. Sato, N., M. Kondo, and K. Arai, *The orphan nuclear receptor GCNF recruits DNA methyltransferase for Oct-3/4 silencing.* Biochem Biophys Res Commun, 2006. **344**(3): p. 845-51.
260. Li, J.Y., M.T. Pu, R. Hirasawa, B.Z. Li, Y.N. Huang, R. Zeng, N.H. Jing, T. Chen, E. Li, H. Sasaki, and G.L. Xu, *Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog.* Mol Cell Biol, 2007. **27**(24): p. 8748-59.
261. Li, G., R. Margueron, M. Ku, P. Chambon, B.E. Bernstein, and D. Reinberg, *Jarid2 and PRC2, partners in regulating gene expression.* Genes Dev, 2010. **24**(4): p. 368-80.
262. Pasini, D., A.P. Bracken, J.B. Hansen, M. Capillo, and K. Helin, *The polycomb group protein Suz12 is required for embryonic stem cell differentiation.* Mol Cell Biol, 2007. **27**(10): p. 3769-79.
263. Chew, J.L., Y.H. Loh, W. Zhang, X. Chen, W.L. Tam, L.S. Yeap, P. Li, Y.S. Ang, B. Lim, P. Robson, and H.H. Ng, *Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells.* Mol Cell Biol, 2005. **25**(14): p. 6031-46.
264. Tay, Y., J. Zhang, A.M. Thomson, B. Lim, and I. Rigoutsos, *MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation.* Nature, 2008. **455**(7216): p. 1124-8.
265. Xu, N., T. Papagiannakopoulos, G. Pan, J.A. Thomson, and K.S. Kosik, *MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells.* Cell, 2009. **137**(4): p. 647-58.
266. Saxe, J.P., A. Tomilin, H.R. Scholer, K. Plath, and J. Huang, *Post-translational regulation of Oct4 transcriptional activity.* PLoS One, 2009. **4**(2): p. e4467.
267. Tsuruzoe, S., K. Ishihara, Y. Uchimura, S. Watanabe, Y. Sekita, T. Aoto, H. Saitoh, Y. Yuasa, H. Niwa, M. Kawasuji, H. Baba, and M. Nakao, *Inhibition of DNA binding of Sox2 by the SUMO conjugation.* Biochem Biophys Res Commun, 2006. **351**(4): p. 920-6.

268. Zhang, Z., B. Liao, M. Xu, and Y. Jin, *Post-translational modification of POU domain transcription factor Oct-4 by SUMO-1.* FASEB J, 2007. **21**(12): p. 3042-51.
269. Webster, D.M., C.F. Teo, Y. Sun, D. Wloga, S. Gay, K.D. Klonowski, L. Wells, and S.T. Dougan, *O-GlcNAc modifications regulate cell survival and epiboly during zebrafish development.* BMC Dev Biol, 2009. **9**: p. 28.
270. Gambetta, M.C., K. Oktaba, and J. Muller, *Essential role of the glycosyltransferase sxc/Ogt in polycomb repression.* Science, 2009. **325**(5936): p. 93-6.
271. Ozcan, S., S.S. Andrali, and J.E. Cantrell, *Modulation of transcription factor function by O-GlcNAc modification.* Biochim Biophys Acta, 2010. **1799**(5-6): p. 353-64.
272. Matoba, R., H. Niwa, S. Masui, S. Ohtsuka, M.G. Carter, A.A. Sharov, and M.S. Ko, *Dissecting Oct3/4-regulated gene networks in embryonic stem cells by expression profiling.* PLoS One, 2006. **1**: p. e26.
273. Sharov, A.A., S. Masui, L.V. Sharova, Y. Piao, K. Aiba, R. Matoba, L. Xin, H. Niwa, and M.S. Ko, *Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data.* BMC Genomics, 2008. **9**: p. 269.
274. Liu, Y. and P.A. Labosky, *Regulation of embryonic stem cell self-renewal and pluripotency by Foxd3.* Stem Cells, 2008. **26**(10): p. 2475-84.
275. Lim, L.S., Y.H. Loh, W. Zhang, Y. Li, X. Chen, Y. Wang, M. Bakre, H.H. Ng, and L.W. Stanton, *Zic3 is required for maintenance of pluripotency in embryonic stem cells.* Mol Biol Cell, 2007. **18**(4): p. 1348-58.
276. Strumpf, D., C.A. Mao, Y. Yamanaka, A. Ralston, K. Chawengsaksophak, F. Beck, and J. Rossant, *Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst.* Development, 2005. **132**(9): p. 2093-102.
277. Yeap, L.S., K. Hayashi, and M.A. Surani, *ERG-associated protein with SET domain (ESET)-Oct4 interaction regulates pluripotency and represses the trophectoderm lineage.* Epigenetics Chromatin, 2009. **2**(1): p. 12.
278. Yuan, P., J. Han, G. Guo, Y.L. Orlov, M. Huss, Y.H. Loh, L.P. Yaw, P. Robson, B. Lim, and H.H. Ng, *Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells.* Genes Dev, 2009. **23**(21): p. 2507-20.
279. Robinson-Rechavi, M., H. Escriva Garcia, and V. Laudet, *The nuclear receptor superfamily.* J Cell Sci, 2003. **116**(Pt 4): p. 585-6.
280. Warnmark, A., E. Treuter, A.P. Wright, and J.A. Gustafsson, *Activation functions 1 and 2 of nuclear receptors: molecular strategies for transcriptional activation.* Mol Endocrinol, 2003. **17**(10): p. 1901-9.
281. Rollerova, E. and M. Urbancikova, *Intracellular estrogen receptors, their characterization and function (Review).* Endocr Regul, 2000. **34**(4): p. 203-18.
282. Kallen, J., J.M. Schlaeppi, F. Bitsch, I. Filipuzzi, A. Schilb, V. Riou, A. Graham, A. Strauss, M. Geiser, and B. Fournier, *Evidence for ligand-independent transcriptional activation of the human estrogen-related receptor alpha (ERRalpha): crystal structure of ERRalpha ligand binding domain in complex with peroxisome proliferator-activated receptor coactivator-1alpha.* J Biol Chem, 2004. **279**(47): p. 49330-7.
283. Dufour, C.R., B.J. Wilson, J.M. Huss, D.P. Kelly, W.A. Alaynick, M. Downes, R.M. Evans, M. Blanchette, and V. Giguere, *Genome-wide orchestration of cardiac functions by the orphan nuclear receptors ERRalpha and gamma.* Cell Metab, 2007. **5**(5): p. 345-56.
284. Gearhart, M.D., S.M. Holmbeck, R.M. Evans, H.J. Dyson, and P.E. Wright, *Monomeric complex of human orphan estrogen related receptor-2 with DNA: a pseudo-dimer interface mediates extended half-site recognition.* J Mol Biol, 2003. **327**(4): p. 819-32.
285. Giguere, V., *To ERR in the estrogen pathway.* Trends Endocrinol Metab, 2002. **13**(5): p. 220-5.
286. Tremblay, A.M. and V. Giguere, *The NR3B subgroup: an ovERRview.* Nucl Recept Signal, 2007. **5**: p. e009.
287. Bookout, A.L., Y. Jeong, M. Downes, R.T. Yu, R.M. Evans, and D.J. Mangelsdorf, *Anatomical profiling of nuclear receptor expression reveals a hierarchical transcriptional network.* Cell, 2006. **126**(4): p. 789-99.
288. Yang, X., M. Downes, R.T. Yu, A.L. Bookout, W. He, M. Straume, D.J. Mangelsdorf, and R.M. Evans, *Nuclear receptor expression links the circadian clock to metabolism.* Cell, 2006. **126**(4): p. 801-10.
289. Luo, J., R. Sladek, J.A. Bader, A. Matthyssen, J. Rossant, and V. Giguere, *Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-beta.* Nature, 1997. **388**(6644): p. 778-82.
290. Tremblay, G.B., T. Kunath, D. Bergeron, L. Lapointe, C. Champigny, J.A. Bader, J. Rossant, and V. Giguere, *Diethylstilbestrol regulates trophoblast stem cell differentiation as a ligand of orphan nuclear receptor ERR beta.* Genes Dev, 2001. **15**(7): p. 833-8.
291. Collin, R.W., E. Kalay, M. Tariq, T. Peters, B. van der Zwaag, H. Venselaar, J. Oostrik, K. Lee, Z.M.

Ahmed, R. Caylan, Y. Li, H.A. Spierenburg, E. Eyupoglu, A. Heister, S. Riazuddin*, et al.*, *Mutations of ESRRB encoding estrogen-related receptor beta cause autosomal-recessive nonsyndromic hearing impairment DFNB35.* Am J Hum Genet, 2008. **82**(1): p. 125-38.

292. Chen, J. and J. Nathans, *Estrogen-related receptor beta/NR3B2 controls epithelial cell fate and endolymph production by the stria vascularis.* Dev Cell, 2007. **13**(3): p. 325-37.

293. Mitsunaga, K., K. Araki, H. Mizusaki, K. Morohashi, K. Haruna, N. Nakagata, V. Giguere, K. Yamamura, and K. Abe, *Loss of PGC-specific expression of the orphan nuclear receptor ERR-beta results in reduction of germ cell number in mouse embryos.* Mech Dev, 2004. **121**(3): p. 237-46.

294. Xie, C.Q., Y. Jeong, M. Fu, A.L. Bookout, M.T. Garcia-Barrio, T. Sun, B.H. Kim, Y. Xie, S. Root, J. Zhang, R.H. Xu, Y.E. Chen, and D.J. Mangelsdorf, *Expression profiling of nuclear receptors in human and mouse embryonic stem cells.* Mol Endocrinol, 2009. **23**(5): p. 724-33.

295. Xie, W., H. Hong, N.N. Yang, R.J. Lin, C.M. Simon, M.R. Stallcup, and R.M. Evans, *Constitutive activation of transcription and binding of coactivator by estrogen-related receptors 1 and 2.* Mol Endocrinol, 1999. **13**(12): p. 2151-62.

296. Xu, J. and Q. Li, *Review of the in vivo functions of the p160 steroid receptor coactivator family.* Mol Endocrinol, 2003. **17**(9): p. 1681-92.

297. Hentschke, M. and U. Borgmeyer, *Identification of PNRC2 and TLE1 as activation function-1 cofactors of the orphan nuclear receptor ERRgamma.* Biochem Biophys Res Commun, 2003. **312**(4): p. 975-82.

298. Hentschke, M., U. Susens, and U. Borgmeyer, *PGC-1 and PERC, coactivators of the estrogen receptor-related receptor gamma.* Biochem Biophys Res Commun, 2002. **299**(5): p. 872-9.

299. Wallberg, A.E., S. Yamamura, S. Malik, B.M. Spiegelman, and R.G. Roeder, *Coordination of p300-mediated chromatin remodeling and TRAP/mediator function through coactivator PGC-1alpha.* Mol Cell, 2003. **12**(5): p. 1137-49.

300. Metivier, R., G. Penot, M.R. Hubner, G. Reid, H. Brand, M. Kos, and F. Gannon, *Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter.* Cell, 2003. **115**(6): p. 751-63.

301. Carroll, J.S., C.A. Meyer, J. Song, W. Li, T.R. Geistlinger, J. Eeckhoute, A.S. Brodsky, E.K. Keeton, K.C. Fertuck, G.F. Hall, Q. Wang, S. Bekiranov, V. Sementchenko, E.A. Fox, P.A. Silver*, et al.*, *Genome-wide analysis of estrogen receptor binding sites.* Nat Genet, 2006. **38**(11): p. 1289-97.

302. Fullwood, M.J., M.H. Liu, Y.F. Pan, J. Liu, H. Xu, Y.B. Mohamed, Y.L. Orlov, S. Velkov, A. Ho, P.H. Mei, E.G. Chew, P.Y. Huang, W.J. Welboren, Y. Han, H.S. Ooi*, et al.*, *An oestrogen-receptor-alpha-bound human chromatin interactome.* Nature, 2009. **462**(7269): p. 58-64.

2

Estrogen-Related Receptor Beta Interacts with Oct4 to Positively Regulate *Nanog* Gene Expression

Mol Cell Biol. 2008 Oct;28(19):5986-95.

Estrogen-Related Receptor Beta Interacts with Oct4 to Positively Regulate *Nanog* Gene Expression

Debbie L.C. van den Berg,[1] Wensheng Zhang,[2] Adam Yates,[2] Erik Engelen,[1] Katalin Takacs,[3] Karel Bezstarosti,[4] Jeroen Demmers,[4] Ian Chambers,[2] and Raymond A. Poot[1,*]

Department of Cell Biology, Erasmus MC, Dr.Molewaterplein 50, 3015GE Rotterdam, The Netherlands[1]; MRC Centre for Regenerative Medicine, Institute for Stem Cell Research, School of Biological Sciences, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JQ, Scotland[2]; MRC Clinical Sciences Centre, Du Cane Road, London W12 0NN, United Kingdom[3]; Proteomics Center, Erasmus MC, Rotterdam, The Netherlands[4]
*Corresponding author

ABSTRACT

Embryonic stem (ES) cell self-renewal is regulated by transcription factors including Oct4, Sox2 and Nanog. A number of additional transcriptional regulators of ES cell self-renewal have recently been identified including the orphan nuclear receptor Estrogen Related Receptor Beta (Esrrb). However, the mode of action of Esrrb in ES cells is unknown. Here, using an Oct4 affinity screen, we identify Esrrb as an Oct4 partner protein. Esrrb can interact with Oct4 independent of DNA. Esrrb is recruited near the Oct-Sox element in the *Nanog* proximal promoter, where it positively regulates *Nanog* expression. Esrrb recruitment to the *Nanog* promoter requires both the presence of Oct4 and a degenerate Estrogen Related Receptor DNA Element (ERRE). Consistent with its role in *Nanog* regulation, expression of Esrrb protein within the Oct4^{+} ES cell population is mosaic and correlates with the mosaic expression of Nanog protein. Together with previous reports that Nanog may regulate *Esrrb* gene expression, our results suggests that Esrrb and Nanog act as part of a feedback regulatory circuit that modulates the fluctuating self-renewal capacity of ES cell populations.

INTRODUCTION

Self-renewal of mouse ES cells is regulated by a network of transcription factors that includes Oct4, Nanog and Sox2[1]. The expression level of Oct4 protein needs to be kept within a tight range in order to maintain ES cell self-renewal[2]. Decreasing Oct4 levels below 50% induces differentiation into trophectoderm, whereas a two-fold increase causes differentiation into cells expressing markers of endoderm and mesoderm[2]. In contrast, overexpression of *Nanog* allows mouse ES cells to remain undifferentiated in the absence of the otherwise requisite stimulation by LIF and BMP[3-5]. Oct4 is thought to act together with Sox2 by binding to adjacent cognate DNA sequences in many genes[6], including *Nanog*[7-8]. Genome-wide chromatin-immunoprecipitation studies have suggested that composite Oct-Sox motifs regulate the expression of many genes in mouse and human ES cells[9-10]. Recent evidence has shown that the critical role of Sox2 in maintaining ES cell self-renewal is to regulate Oct4 expression, suggesting that the secondary role of gene regulation via Oct-Sox motifs is performed redundantly with Sox4, Sox11 and Sox15[11]. Recent reports have expanded the

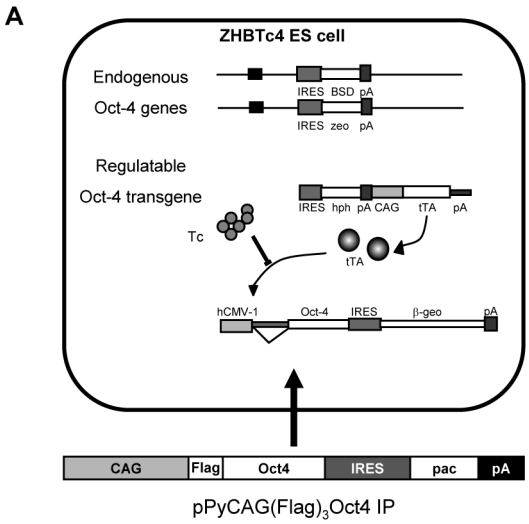
list of factors that contribute to ES cell self-renewal. Wang et al.[12] reported a proteomic analysis of interactors of Oct4 and Nanog and suggested that some Nanog interactors may assist in Nanog-mediated gene regulation. A separate study using a RNA interference screen found that depletion of Esrrb, Tbx3 or Tcl1 resulted in ES cell differentiation but that this differentiation could be attenuated by overexpression of *Nanog*[13]. However, it is unclear how any of these novel regulatory factors mediate their function. Here we use an unbiased analysis of Oct4 binding proteins to identify Esrrb as an Oct4 interacting partner protein. We show that Esrrb is recruited to the Oct4 responsive element within the proximal *Nanog* promoter where it is responsible for mediating the positive regulatory effect of Oct4.

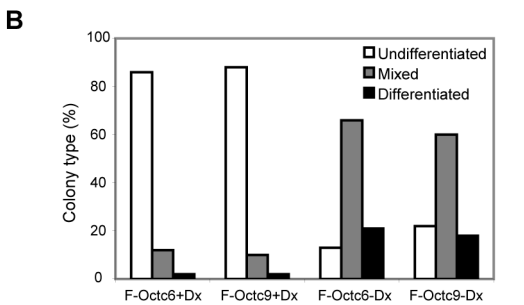
RESULTS

The Oct4 protein interacts with Esrrb

To identify interaction partners of the Oct4 protein in ES cells, we constructed an ES cell line where, under self-renewing conditions, all Oct4 in the cell has an N-terminal triple FLAG-tag (FLAG-Oct4). The parental ZHBTc4 ES cell line[2] has both Oct4 alleles disrupted and the only Oct4 protein in the cell is transcribed from a doxycycline-suppressible transgene (Figure 1A). ZHBTc4 cells were transfected with a construct in which constitutive expression of FLAG-Oct4 is linked through an IRES to puromycin resistance. Simultaneously, doxycycline was added to the medium to repress the inducible Oct4 transgene expression[2]. After 12 days growth, colonies were picked and expanded into cell lines. All cell lines expressed a protein of the same relative molecular weight that reacted with an anti-flag antibody on western blots (data not shown). As the Oct4 level must be tightly regulated to allow continued self-renewal[2], the survival of puromycin resistant colonies indicates that the FLAG-Oct4 protein is functional. Two of these lines were further tested for their response to doxycycline treatment and both underwent efficient differentiation at clonal density (Figures 1B and 1C). One of these lines (c6), called F-Oct4 ES cells from here onwards, was taken forward for biochemical analysis.

F-Oct4 ES cells were expanded, nuclear extracts prepared and the FLAG-Oct4 protein purified using FLAG-affinity technology (see material and methods). Silverstain analysis of a polyacrylamide gel containing the proteins purified from F-Oct4 extracts identified a major and a minor band running just above the 54 kD marker (Figure 2A). Both bands are recognised by a FLAG antibody and are not present in purifications from control extracts, suggesting they represent the FLAG-Oct4 protein (Figure 2B). No other major bands were observed, indicating that Oct4 does not purify as part of a major, stoichiometric, complex, despite the mild purification conditions used. Mass spectrometry analysis of two independent purifications identified the presence of Oct4 (10 unique peptides) and Estrogen-related receptor beta (5 unique peptides) in F-Oct4 samples and not in control samples. Indeed, Esrrb could be detected by western in the F-Oct4 sample and not in the control (Figure 2C). The interaction between Esrrb and Oct4 was independently verified by co-immunoprecipitation from extracts of a different ES cell line, 46C, using antibodies





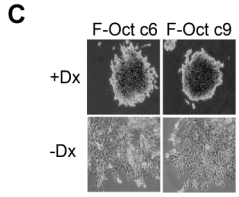


Figure 1. Construction and characterization of F-Oct ES cell lines

(A) In ZHBTc4.1 ES cells both the Oct4 alleles have been replaced and Oct4 expression is directed from a doxycycline-suppressible transgene. F-Oct ES cell lines were derived from ZHBTc4.1 cells by transfection with linearized pPyCAG(Flag)3Oct4IP and concomitant addition of doxycycline. (B) Clonal assays on two clones demonstrate that following withdrawal of doxycycline, Oct4-induced differentiation occurs efficiently in representative F-Oct cell lines. Cells were plated at 600 per 10 cm dish in the presence or absence of 1 mM doxycycline, cultured for 6 days, stained for alkaline phosphatase activity and differentiation status counted. (C) Examples of colony morphologies in the presence and absence of doxycycline are shown.

against endogenous Oct4 and Esrrb (Figures 2D and 2E). Treatment of the extract with benzonase nuclease or ethidium bromide did not affect the interaction (Figures 2D and 2E), indicating that it is not mediated indirectly through DNA. Moreover, the ability of bacterially expressed GST-Oct4 to pull down FLAG-Esrrb from transfected ES cells (Figure 2F) indicates that post-translational modification of Oct4 is not required for interaction between Oct4 and Esrrb.

Esrrb regulates expression of the Oct4 target gene *Nanog*

Oct4 regulates expression of a cohort of target genes in ES cells, often acting in concert with Sox proteins[10]. To determine whether the binding of Esrrb to Oct4 affected the regulation of gene expression by Oct4, we first examined expression of the Oct4 target gene *Nanog*. We depleted Esrrb by RNA interference using two vectors that express different, previously reported, Esrrb shRNAs[10] and harbour a puromycin selection marker. After two days of puromycin selection,

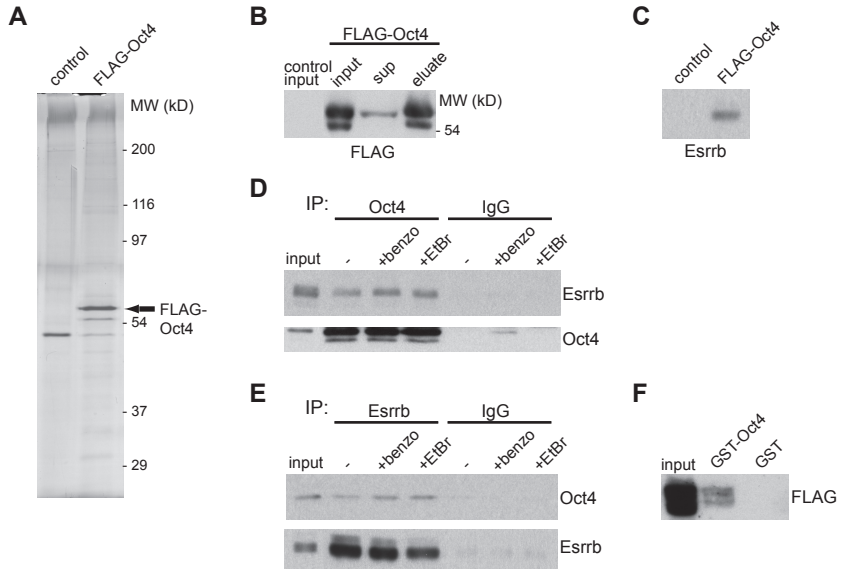


Figure 2. Oct4 interacts with Esrrb
(A) Silver-stained SDS-PAA gel of peptide-eluted FLAG-Oct4 versus a control purification. Protein marker in kilo Dalton (kD) is shown. A band around 54 kD represents FLAG-Oct4, indicated by the arrow. (B) Western blot with anti-Flag antibody on input, supernatant and elution of FLAG-Oct4 purification. (C) Western blot with anti-Esrrb antibody on the eluted FLAG-Oct4 or control sample. (D, E) Co-immunoprecipitation experiments using antibodies against Oct4, Esrrb, or control IgGs confirm the Oct4-Esrrb interaction. DNA-independency of the interaction was shown by its insensitivity to benzonase (*Benzo)* or ethidiumbromide *(EtBr)*.

the levels of *Nanog* mRNA (Figure 3A) and Nanog protein (Figure 3B) were reduced in 46C ES cells treated with either Esrrb shRNA vector compared to the control. Importantly, this specific depletion of Nanog occurred prior to any reduction in Oct4 expression (Figures 3A and 3B) and prior to any morphological evidence of ES cell differentiation (Figure 3C). This indicates that Esrrb shRNA-induced Nanog depletion is not a consequence of differentiation but occurs prior to differentiation. We also tested the effect of Esrrb depletion on TNG-PS ES cells which have GFP inserted at the start codon of one of the endogenous *Nanog* alleles[14]. Transfection of either Esrrb shRNA vector decreased the mean GFP fluorescence of the population (Figure 3D) after 2 days of puromycin selection, suggesting that depletion of Esrrb reduces transcription from the *Nanog* locus. To determine whether Esrrb affects the activity of the *Nanog* promoter, similar shRNA experiments were performed using a luciferase reporter under the control of a *Nanog* promoter fragment extending from –2.5 kb to +50 bp compared to the transcription start site. Esrrb shRNA vectors were co-transfected with the luciferase reporters and luciferase activity was measured 2 days post-transfection. Figure 3E shows that *Nanog* promoter activity is strongly reduced with either Esrrb shRNA vector, compared to a control shRNA vector.

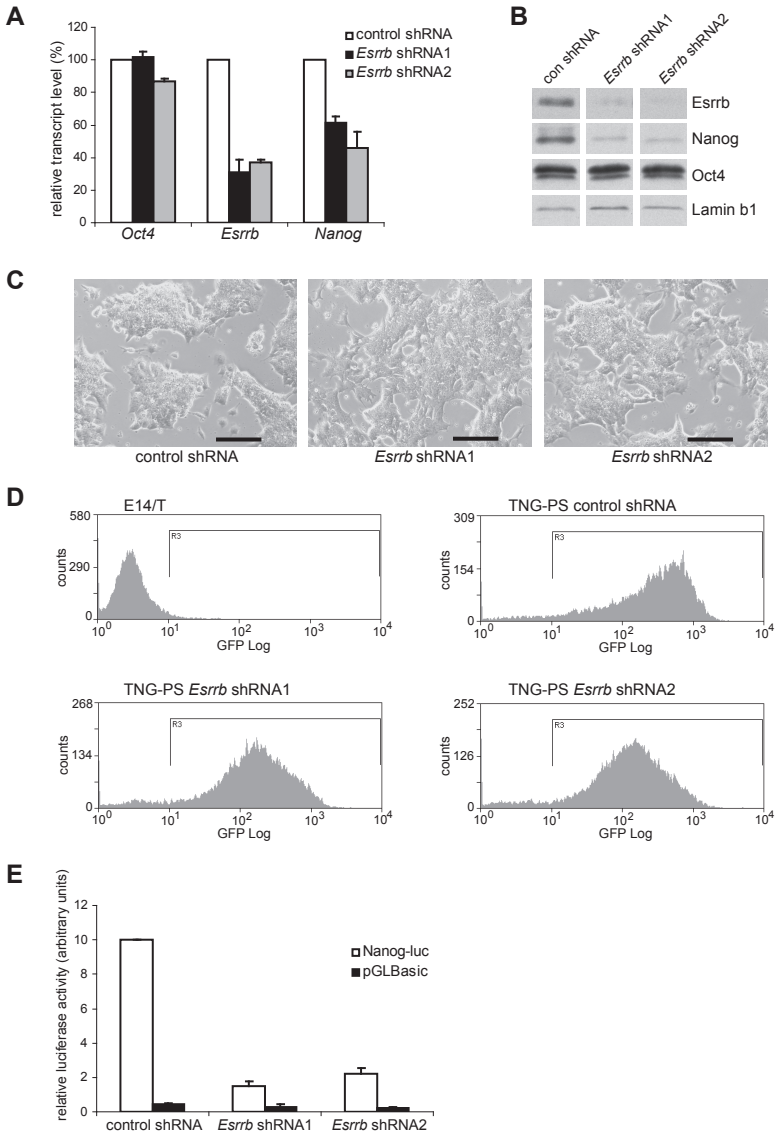


Figure 3. Esrrb regulates Nanog expression
(A) Quantitative RT-PCR analysis shows downregulation of the mRNA levels of Nanog but not Oct4, following shRNA-mediated knock down of Esrrb in 46C ES cells by transfection of pSUPER plasmids expressing Esrrb shRNA1 or Esrrb shRNA2. RNA levels are compared to cells tranfected with pSUPER expressing a control shRNA. Error bars represent the SEM of three independent experiments. (B) Western blot on total cell lysates confirms depletion at the protein level of Esrrb and Nanog, but not Oct4, upon shRNA-mediated knock down of Esrrb. One of two independent experiments is shown, Lamin b1 is used as loading control. (C) Phase-contrast images of live 46C ES cell cultures 3 days post transfection of the indicated shRNA constructs. Scale bars represent 200 μm. (D) FACS profiles of TNG-PS ES cells, in which the GFP ORF has been placed at the *Nanog* start codon. Knock down of Esrrb with either of two shRNA constructs reduces GFP expression compared to the control construct. E14/T is an ES cell line lacking a GFP gene. One of two independent experiments is shown. (E) Luciferase reporter assays with a *Nanog*-promoter construct. The luciferase activity of the *Nanog* promoter (-2.5kb/+50bp), co-transfected with control-shRNA plasmid is arbitrarily set at 10 and compared to the luciferase activities in the presence of either of two Esrrb-shRNA plasmids, or of pGL-Basic control vector. Error bars represent the SEM of three independent experiments.

Essrb regulates *Nanog* expression using contacts with both Oct4 and a degenerate ERRE

Oct4 contributes to the regulation of *Nanog* by binding to an Oct-Sox site[7-8], located 166 to 180 bp upstream of the mapped transcription start site of the *Nanog* gene[15-16]. To investigate the relationship between regulation of *Nanog* by Esrrb and Oct4, we performed chromatin immunoprecipitation (ChIP) experiments with antibodies for Oct4 and Esrrb in 46C ES cells and examined the precipitates for the presence of *Nanog* promoter. A standard ChIP protocol using only formaldehyde as a crosslinking agent confirms that Oct4 binds in the vicinity of the Oct-Sox site (Figure 4B). Using standard ChIP, we found no enrichment of Esrrb at the *Nanog* promoter (Figure 4B), although the Esrrb protein is immunoprecipitated during the ChIP procedure (data not shown). Conventional formaldehyde-based ChIP methods efficiently detect protein-DNA interactions but may not detect the binding of Esrrb to the *Nanog* promoter, if it is stabilized by protein-protein interactions. We therefore used a dual-crosslinking ChIP method (XX-ChIP) that uses Di(N-succinimidyl) glutarate (DSG) prior to formaldehyde crosslinking[17]. DSG has a longer spacer-arm than formaldehyde and has been used to crosslink transcription factor protein-protein interactions on DNA[18]. XX-ChIP indeed detects Esrrb at the Oct-Sox site within the proximal *Nanog* promoter (Figure 4C). XX-ChIP using the Oct4 antibody also gives a specific enrichment of Oct4 at the Oct-Sox motif (Figure 4C). Oct4 and Esrrb were also specifically enriched on the Oct-Sox site, when compared to input and an IgG control (see Figures S1A and S1B in the supplementary material). We conclude that Esrrb binds to the *Nanog* promoter, in the vicinity of the Oct-Sox motif.

To assess whether Esrrb binding to the *Nanog* promoter is dependent on Oct4, we made use of the ES cell line ZHBTc4 in which the endogenous Oct4 alleles are disrupted and in which Oct4 is expressed from a doxycycline-suppressible promoter. Addition of doxycycline for 12 hrs removes all Oct4 protein from the cell (Figure 4D). At this time point, the level of Esrrb is unaffected (Figure 4D). Esrrb XX ChIP shows that Esrrb is no longer enriched at the *Nanog* promoter in the absence of Oct4 (Figure 4E). To test functionally whether maintenance of *Nanog* promoter activity by Esrrb is via Oct4, we tested *Nanog*-luc promoter constructs where the Oct4 binding site is mutated (*Nanog* mOS)[8]. As expected, the *Nanog* mOS reporter is less active than the wt construct, although still clearly above background (Figure 4F)[8]. Depletion of Esrrb does not further reduce the activity of *Nanog* mOS (Figure 4F), suggesting that the effect of Esrrb on *Nanog* expression requires Oct4 binding to the oct-sox site in the *Nanog* promoter.

Estrogen related receptors are thought to act via an Estrogen Related Receptor response element (ERRE). A consensus ERRE of tcaaGGttca (invariant positions in upper case) was determined by SELEX and confirmed by *in vivo* studies[19-20]. Visual inspection of the *Nanog* promoter identified a degenerate ERRE (sequence TCTGGGTCA) 12bp upstream of the Oct-Sox motif (Figure 5A). This sequence is largely conserved in many mammalian species[8]. To test the contribution of this putative ERRE to *Nanog* promoter activity, we mutated the core GGT in this motif into AAC (Figure 5A) in the context of a *Nanog*-luc promoter construct. NMR structural analysis of the Esrrb DNA binding domain in complex with DNA shows that these bases make a major contribution to DNA

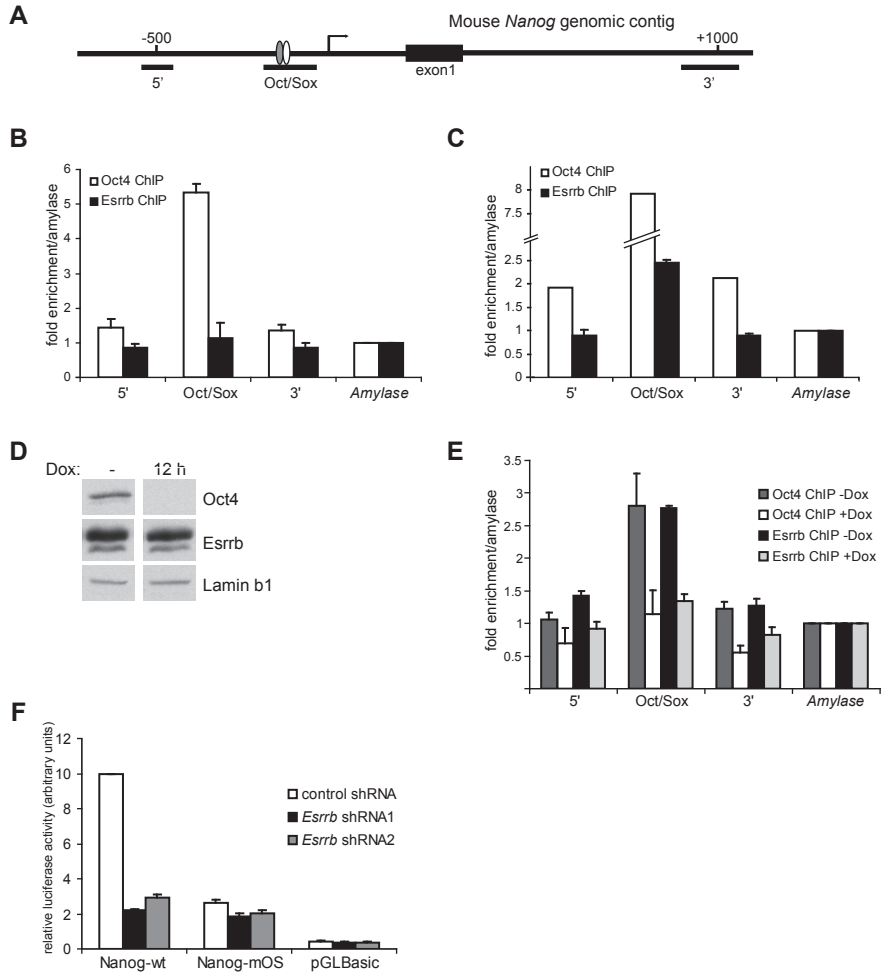


Figure 4. Esrrb binds to the Nanog promoter in an Oct4-dependent manner.
(A) Outline of the *Nanog* genomic contig showing the amplicons (5', Oct-Sox and 3') used in chromatin immunoprecipitation (ChIP) analysis, size markers in basepairs. (B) ChIP analysis of formaldehyde cross-linked 46C ES cells using antibodies against Oct4 and Esrrb. Relative enrichments of the *Nanog* amplicons are depicted as fold enrichments over an unrelated region (*Amylase*). Oct4 binds to the Oct-Sox element, but no significant enrichment for Esrrb can be detected at this region. Error bars represent the SEM of two independent experiments. (C) ChIP analysis of dual-cross-linked 46C ES cell chromatin shows enrichment of both Oct4 and Esrrb on the Oct-Sox element of the *Nanog* promoter. Error bars for Esrrb enrichments represent SEM of two independent experiments. (D) Western blot showing the complete depletion of Oct4 protein in ZHBTc4 ES cells after 12 hours treatment with 1 µg/ml doxycycline, compared to untreated cells. Levels of Esrrb are not affected. Lamin b1 is used as loading control. (E) Dual-cross-linked chromatin from ZHBTc4 ES cells that were untreated (-Dox) or treated for twelve hours with 1 µg/ml doxycycline (+Dox). Binding of Esrrb to the Oct-Sox element is no longer detected when Oct4 is absent. Error bars represent SEM of two independent experiments. (F) Luciferase assays with wild-type *Nanog* promoter construct (*Nanog*-wt), with mutated Oct binding site (*Nanog*-mOS) or the empty vector pGLBasic. Co-transfected shRNA plasmids are indicated. Error bars represent the SEM of three independent experiments.

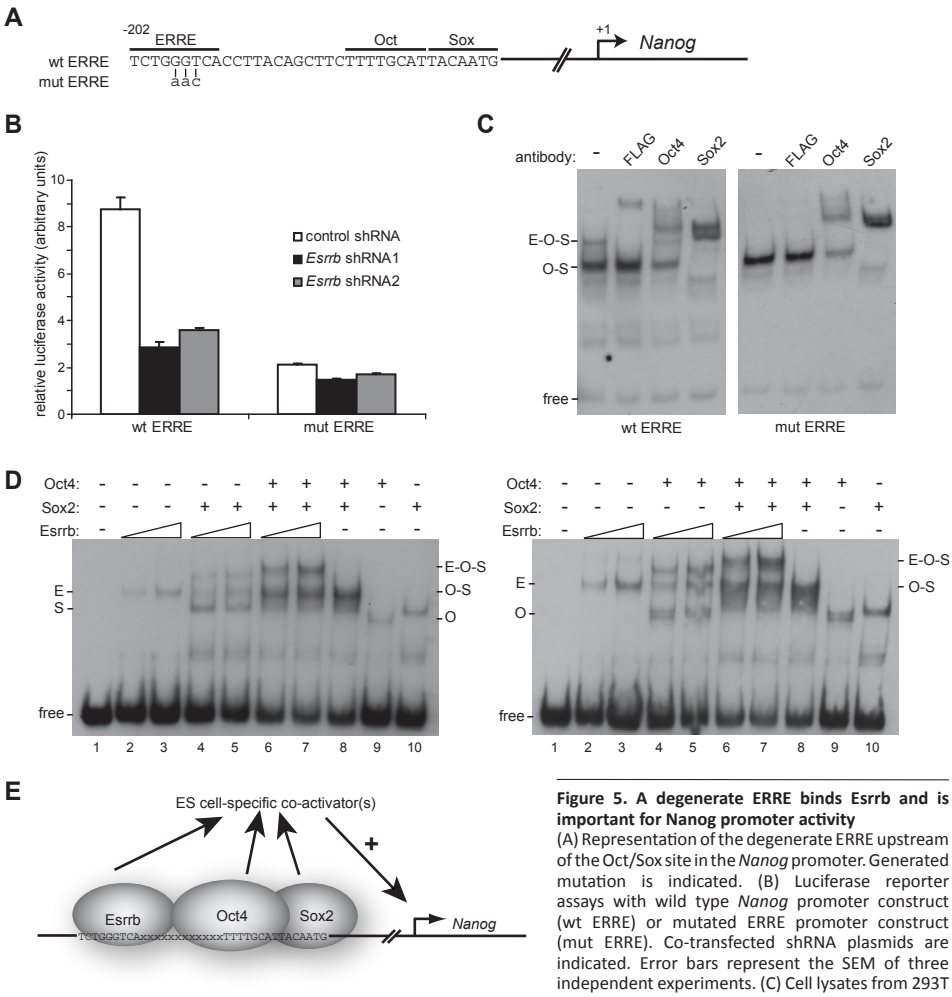


Figure 5. A degenerate ERRE binds Esrrb and is important for Nanog promoter activity
(A) Representation of the degenerate ERRE upstream of the Oct/Sox site in the *Nanog* promoter. Generated mutation is indicated. (B) Luciferase reporter assays with wild type *Nanog* promoter construct (wt ERRE) or mutated ERRE promoter construct (mut ERRE). Co-transfected shRNA plasmids are indicated. Error bars represent the SEM of three independent experiments. (C) Cell lysates from 293T cells co-transfected with Flag-Esrrb, Oct4 and Sox2 expression plasmids were used in EMSA with a 57 nucleotide *Nanog* promoter probe, containing the Oct-Sox sequence and a wild type or mutated ERRE. Antibodies were added as indicated. (D) Cell lysates from 293T cells transfected with Flag-Esrrb, Oct4 or Sox2 expression plasmids were subjected to EMSA using the wild type *Nanog* promoter probe. (E) Model showing the enhancement of Esrrb binding to the *Nanog* promoter by Oct4 and Sox2 which positively affects *Nanog* promoter activity.

binding by Esrrb[21]. This mutation strongly reduced *Nanog* promoter activity (Fig 5B). Moreover, in contrast to the situation with the unmutated construct, the activity of this mutant could not be further reduced by Esrrb shRNA expression (Figure 5B).

EMSA was employed to investigate the potential binding of Esrrb to the *Nanog* promoter. A 57-mer oligonucleotide corresponding to the sequence of the *Nanog* promoter was used that includes the putative ERRE and the Oct-Sox site. Lysate prepared from 293T cells co-transfected with FLAG-Esrrb, Oct4 and Sox2 expression plasmids caused a shift in migration of the probe

into two complexes (Figure 5C, left panel). The faster migrating complex could be supershifted by antibodies against Oct4 and Sox2 but not by an anti-FLAG antibody. In contrast, the slower migrating band was supershifted by all three antibodies, suggesting that the slower migrating complex is formed by binding of Esrrb, Sox 2 and Oct4 (Figure 5C, left panel, lanes 2-4). Using a second 57-mer oligonucleotide that has the GGT to AAC mutation in the putative ERRE, only the faster of the two complexes was observed and this could be shifted by antibodies against Oct4 and Sox2 but not by anti-FLAG antibody (Figure 5C, right panel). Therefore, we conclude that there is an ERRE upstream of the Oct-Sox site in the *Nanog* promoter that is essential for Esrrb binding and optimal *Nanog* promoter activity.

To determine whether binding of Esrrb to the *Nanog* promoter *in vitro* is dependent upon binding of Oct4-Sox2, EMSAs were performed by mixing cell lysates prepared from individual transfections of Oct4, Sox2 and Esrrb into 293T cells. Addition of Esrrb lysate caused a weak probe shift (Figure 5D, lanes 2 and 3), showing that Esrrb alone can bind the *Nanog* promoter probe. However, in the presence of Oct4 and Sox2, DNA complex formation by Esrrb was enhanced compared to DNA complex formation by Esrrb alone (Figure 5D, compare lanes 2 and 3 to lanes 6 and 7). This effect was greatest with both Oct4 and Sox2 present suggesting that an Oct4-Sox2 complex is required

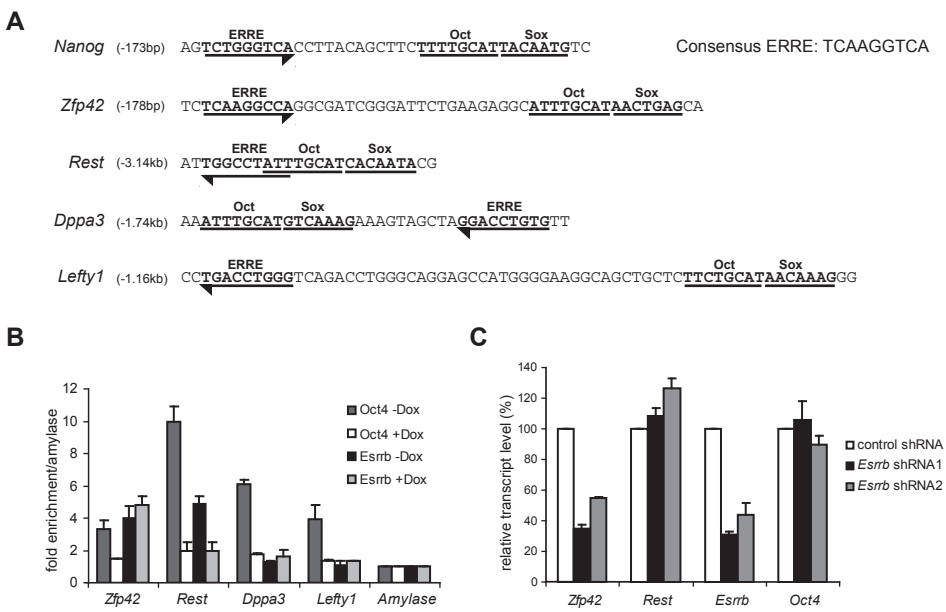


Figure 6. Esrrb binds and regulates other Oct4 target genes
(A) The sequences surrounding the Oct-Sox motifs in the regulatory elements of the *Nanog*, *Zfp42*, *Rest*, *Dppa3* and *Lefty1* genes are shown with the position and orientation of the putative ERREs indicated. Distance from the 3' nucleotide of the shown Oct motif to the transcription start site is indicated. (B) Chromatin immunoprecipitation with anti-Oct4 or anti-Esrrb antibodies on dual cross-linked chromatin isolated from ZHBTc4 cells that were either non-treated or treated for 12 hours with 1 µg/ml doxycycline to downregulate Oct4 expression. Enrichments over a negative control region (*Amylase*) are depicted; error bars represent SEM of two independent experiments. (C) Quantitative PCR analysis of transcript levels in cells transfected with control or Esrrb shRNA. Error bars represent SEM of two independent experiments.

for this co-operative effect. Using different combinations of Esrrb, Oct4, Sox2 lysates leads to the same conclusions (Figure 5D, lower panel). These data, together with the Esrrb ChIP experiments (Figures 4C and 4E), suggest a model (Figure 5E) in which the presence of Oct4-Sox2 complex bound to the Oct-Sox site in the *Nanog* promoter strongly enhances the intrinsic capacity of Esrrb to bind to an ERRE located upstream of the Oct-Sox site.

Esrrb binds and regulates other Oct4 target genes

To determine the generality of the association of Esrrb with Oct4 on Oct4 target genes, we investigate a number of Oct4 target genes with a characterized Oct-Sox DNA element (Figure 6A), using the ES cell line ZHBTc4. These genes all had a putative ERRE at different distances from the Oct-Sox site and in different orientations (Figure 6A). XX-ChIP analysis showed that Esrrb binds near the Oct-Sox element in the promoters of *Zfp42* (*Rex1*) and *Rest* but not near the Oct-Sox site of *Dppa3* and *Lefty1* (Figure 6B). Interestingly, removing Oct4 from the promoters by doxycycline treatment (see Figure 4D) prevented detection of Esrrb at the *Rest* promoter, whereas Esrrb binding to the *Zfp42* promoter was unaffected.

We next determined the contribution of Esrrb to the expression of *Zfp42* and *Rest* by Esrrb

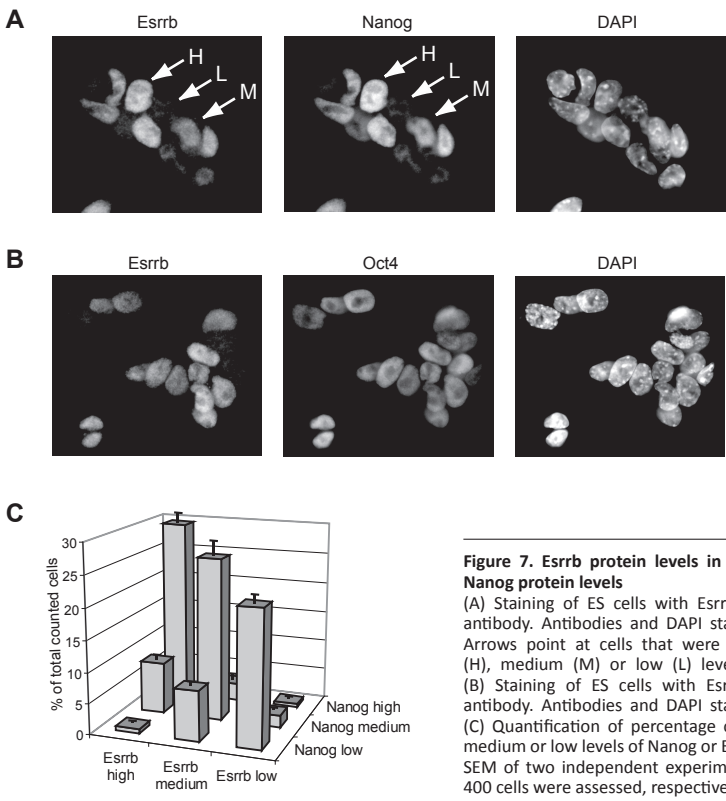


Figure 7. Esrrb protein levels in cells co-fluctuate with Nanog protein levels

(A) Staining of ES cells with Esrrb antibody and Nanog antibody. Antibodies and DAPI staining are as indicated. Arrows point at cells that were assigned to have high (H), medium (M) or low (L) levels of Esrrb or Nanog. (B) Staining of ES cells with Esrrb antibody and Oct4 antibody. Antibodies and DAPI staining are as indicated. (C) Quantification of percentage of cells that have high, medium or low levels of Nanog or Esrrb. Error bars indicate SEM of two independent experiments, in which 350 and 400 cells were assessed, respectively.

knock-down. Expression of Esrrb shRNAs caused a reduction in *Zfp42* mRNA expression, whereas *Rest* mRNA expression was unaffected (Fig 6C). We conclude that Esrrb binds near the Oct-Sox sites of two other Oct4 targets, *Zfp42* and *Rest*, but that only in the case of *Zfp42* does this binding contribute to its expression in ES cells.

Esrrb protein levels correlate with Nanog levels in ES cell colonies

Oct4$^+$ ES cell colonies express *Nanog* in a mosaic fashion[14, 22-23]. The cellular expression of Esrrb was examined by immunostaining of ES cells with antibodies against Nanog or Esrrb. Interestingly, 46C ES cell colonies show a mosaic pattern of Esrrb expression within the Oct4$^+$ population (Figures 7A and 7B). Moreover, when cells are classified according to their relative high, medium, or low levels of Esrrb staining within ES cell colonies (Figure 7A), there is a good correlation between the levels of expression of Nanog and that of Esrrb (Figure 7C). This correlation between the cellular levels of Esrrb and Nanog, in combination with our *Nanog* gene regulation data, suggests high *Nanog* expression in the cell may be facilitated by high *Esrrb* expression.

DISCUSSION

Using an Oct4 protein-affinity strategy, we have identified Esrrb as a binding partner of the Oct4 protein. A previous report identified a number of other putative Oct4 interacting proteins, but did not detect Esrrb[12]. Our approach also identified a number of the reported interactors in our Oct4 sample, but these were often also present in the control sample. Our milder purifications conditions may account for both the reproducible and verified identification of Esrrb as a specific Oct4 binding partner, as well as the non-specific binding of other reported interactors. Indeed, we find that the Oct4-Esrrb interaction is sensitive to higher salt conditions (data not shown).

Members of the estrogen receptor related family can bind to the palindromic 12bp Estrogen Response Element (ERE)[24] or to the “extended half-site” 9bp Estrogen Related Receptor response element (ERRE)[19-20]. Either element can support Esrrb-mediated transcription[25-26]. Esrrb was also suggested to activate targets genes, independent of a DNA element by binding to transcription factors, such as Sp1[27]. We show here, using ChIP and EMSA experiments, that Esrrb recruitment to the *Nanog* promoter requires both a degenerate, but conserved, ERRE and binding of Oct4 to the downstream Oct-Sox element. EMSA experiments (Figure 5) confirm the previously reported synergistic binding of Oct4 and Sox2 to the Oct-Sox site[7], but also indicate that binding of Esrrb to the *Nanog* promoter oligonucleotides occurs co-operatively with binding of Oct4 and Sox2. This effect required binding of both Oct4 and Sox2.

We also provide evidence that recruitment of Esrrb to the *Nanog* promoter by Oct4 and the ERRE positively regulates *Nanog* expression. Depletion of Esrrb with shRNAs caused transcriptional down regulation of the endogenous *Nanog* gene and a *Nanog* promoter-reporter. Mutation of the ERRE also had a negative effect on reporter expression, underscoring the importance of the ERRE for the functional recruitment of Esrrb to the *Nanog* promoter.

Based upon the effect of mutations within the composite Oct-Sox site upon reporter gene

expression directed by the proximal *Nanog* promoter, Oct4 has been suggested to be important for maintaining *Nanog* expression in ES cells[7-8]. However, Oct4 is not required to initiate *Nanog* expression in the pre-implantation embryo, since *Nanog* transcripts[3] are present in *Oct4*$^{-/-}$ morulae and early blastocysts. This apparent discrepancy could be due, in part, to different requirements for establishment versus maintenance of *Nanog* expression. Our data on the Esrrb requirement for *Nanog* expression in ES cells suggest that one function of Oct4 binding to the Oct-Sox motif is that it facilitates Esrrb binding to the *Nanog* promoter, which in turn promotes *Nanog* transcription in ES cells.

Transcriptional regulation via transcription factor interactions in ES cell self-renewal

Here we have provided evidence of a stem cell factor, Oct4, directing its physical interactor, Esrrb, to a target gene, *Nanog*, to positively regulate transcription. Gene regulation facilitated by complexes of individual transcription factors, like the Oct4-Esrrb complex may be widespread in ES cells. Visual inspection of the sequences around the Oct-Sox sites of a number of Oct-Sox target genes for homology to the ERRE identified several potential Esrrb targets that were tested for regulation by Esrrb. Of these, ChIP analysis showed Esrrb to be detectable on *Zfp42* and *Rest* but not *Lefty1* or *Dppa3*. The Oct4-independent binding of Esrrb to the *Zfp42* promoter may be due to the high match (8/9) of the *Zfp42* ERRE sequence to the consensus (Figure 6A) which may provide sufficient DNA binding affinity. The extreme proximity of the ERRE and the Oct site in *Rest* coupled with the low match to the ERRE consensus (6/9) could underlie the Oct4-dependent ChIP of Esrrb at *Rest*. However, the ERRE in *Nanog* is further removed form the Oct-Sox site and is a better match to the consensus (7/9) than that in *Rest* so there is no obvious common feature that can explain the Oct4-dependent binding of Esrrb to each of these sequences. There is also no clear reason why the remaining two Oct-Sox targets do not bind Esrrb. A low match to the consensus could explain the lack of binding to Dppa3 (6/9). However, the consensus match for Lefty1 is the same as that for *Nanog* (7/9) suggesting that relative spatial disposition and /or distance could play a role. Further experimentation will be required to more deeply understand the relationship of Oct4 and Esrrb binding to DNA and how this affects gene regulation.

Nanog is expressed mosaically within the Oct4^{+} populations in ES cell cultures[14, 22-23]. Moreover, Nanog levels fluctuate in ES cell cultures such that cells expressing low or no Nanog can re-express a high level of Nanog. However, lowered *Nanog* expression pre-disposes cells towards differentiation, without marking a commitment event[14]. Here we show that the mosaic patterns of Esrrb and Nanog expression in ES cell colonies largely overlap. We also show that Esrrb positively regulates Nanog expression. As Nanog has been reported to positively regulate Esrrb expression[10], Esrrb and Nanog may both act to reinforce expression of the reciprocal gene through a positive feedback loop. How this leads to mosaic and co-fluctuating levels of both proteins remains to be determined. Oct4 is not obviously mosaic and appears not to fluctuate, suggesting it is not a determining factor of fluctuations in Nanog and Esrrb in ES cells. As Nanog levels, and, by implication, Esrrb levels, regulate the self-renewal efficiency of ES cells, unravelling this regulatory

mechanism will be important for a fuller understanding of ES cell self-renewal and the maintenance of pluripotency.

EXPERIMENTAL PROCEDURES

Plasmids, cell culture

The RNAi constructs pSuper-Esrrb-sh1 and pSuper-Esrrb-sh2 were constructed by cloning Esrrb RNAi1 and RNAi2[10] into pSuper-puro (Oligoengine). pSuper-control contains an oligo without complementarity to any known mammalian sequence (Dharmacon). Mouse ES cell lines 46C[28], ZHBTc4[2] and its derivatives were grown on gelatin-coated dishes without feeders on GMEM supplemented with leukaemia inhibitory factor (LIF), 15% FBS, 0.25% sodium bicarbonate, 1mM glutamine, 1 mM sodium pyruvate, non-essential amino acids, 50 µM beta-mercaptoethanol and penicillin/streptomycin. F-Oct4 ES cells were created by electroporating ZHBTc4 cells with linearized pPyCAG $(FLAG)_3$ Oct4IP, a plasmid in which the Oct4 ORF was placed between the N-terminal triple FLAG tag and the IRES-puromycin resistance cassette of pPyCAG $(FLAG)_3$IP[29]. Electroporated cells were plated in ES cell medium and after 24 hours 1 µg/ml puromycin and 1 µg/ml doxycycline were added. After 12 days selection puromycin resistant colonies were picked and tested for FLAG-Oct4 expression by anti-FLAG western. TNG cells have been described[14]. Puromycin-sensitive TNG-PS cells were derived from TNG cells by excision of the frt-IRES-pac-frt cassette by transient expression of FLPe. Brightfield pictures of ES cell cultures were taken using the IX70 inverted microscope (Olympus).

Oct4 purification and mass spectrometry

F-Oct4 ES cells and control cells (ZHBTc4) were expanded to 50 14-cm dishes, plates were washed once with PBS, cells scraped off, nuclear extracts prepared[30], and dialysed to 100 mM KCl[30]. 100 µl of anti-FLAG M2 agarose beads (Sigma), equilibrated in buffer C-100 (20 mM Hepes pH 7.6, 10% glycerol, 100 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 1x *Complete EDTA-free* protease inhibitor, Roche) were added to 10 ml of nuclear extract, incubated for 3 hrs at 4°C, transferred into an eppendorf tube and washed five times with 1 ml of C-100 buffer + 0.02% NP40 (C-100*) and four times eluted with C-100* containing 0.2 mg/ml FLAG tripeptide (Sigma) for 15 minutes at 4°C. Fractions were loaded onto a 10% SDS PAA gel and silver-stained. Elutions 1 and 2, containing the majority of FLAG-Oct4 in purification from the F-Oct4 extract, were concentrated by speedvac condensation, loaded onto a 10% SDS PAA gel, and stained with colloidal coomassie. Gel lanes were cut and subjected to in-gel digestion with trypsin (Promega), essentially as described previously[31]. NanoflowLC-MS/MS was performed on an 1100 series capillary LC system (Agilent Technologies) coupled to an LTQ-Orbitrap mass spectrometer (Thermo), as described[32]. Database searches to assign proteins to the found peptide fragmentation spectra were performed using MASCOT, as described[32].

Immunoprecipitation

For immunoprecipitation, 2.5 µg of Oct4 antibody (N19, Santa Cruz) or Esrrb antibody (R&D systems) was added to 200 µl of 46C ES cell nuclear extract and incubated under rotation for 2 hrs at 4°C. 1U Benzonase (Novagen) or 25 µg/ml ethidium bromide was added where indicated. The antibody-extract mixture was added to 20 µl of protein G sepharose beads (Amersham) blocked with 1% fish-skin gelatin (Sigma) and 0.2 mg/ml chicken egg albumin (Sigma), and rotated for another 90 minutes. Beads were washed four times with 100 µl of C-100* buffer and boiled in SDS-loading dye.

RNA interference and assays for *Nanog* expression

46C ES cells were transfected with pSuper-Esrrb-sh1, pSuper-Esrrb-sh2 or pSuper-control using Lipofectamine 2000 transfection reagent (Invitrogen). For measuring the effect of Esrrb RNAi on endogenous mRNA and protein levels of Nanog, Oct4 and Esrrb, transfected cells were selected with 1 µg/ml puromycin for 48 hrs, starting 24 hrs post-transfection. RNA was isolated from these samples using Trizol (Invitrogen) and mRNA levels measured by performing RT-qPCR on an Opticon Real Time PCR machine. Protein levels were measured by Westerns using antibodies against Nanog[14], Oct4, Esrrb and Lamin B1 (Santa Cruz). TNG-PS cells were transfected with pSuper-Esrrb-sh1, pSuper-Esrrb-sh2 or pSuper-control and transfected cells selected with puromycin for 48 hrs, starting 24 hrs post-transfection. GFP fluorescence of the TNG cells was measured using a FACSCalibur flow cytometer (Becton Dickinson), as described[14]. For measuring the effect of Esrrb RNAi on expression from the *Nanog* promoter, pNanog-Luc containing a –2.5kb until +50 bp Nanog promoter fragment[33] and pRenilla-TK (Promega) were co-transfected with the pSuper constructs. Luciferase/Renilla assays were done 48 hours post-transfection using the dual-luciferase reporter system (Promega). pNanog-Luc mOS and its control were described[8]. pNanog-Luc constructs with a mutant ERRE contained the mutation GGT to AAC in the ERRE sequence TCTGGGTCA in the Nanog proximal promoter from -230 to +106, compared to the Nanog transcriptional start, and were tested 24 hrs post-transfection.

Chromatin immunoprecipitations

Chromatin immunoprecipitation (ChIP) using formaldehyde cross-linking and Oct4 antibodies (sc-8628, Santa Cruz) or Esrrb antibodies (R&D systems) was done on 46C ES cells, as described[9]. Dual cross-linking ChIP using formaldehyde and di(N-succimidyl) glutarate was performed on 46C and ZBHTc4 ES cells, as described[17]. Quantitative PCR analysis was performed using the DNA Engine Opticon 2. Relative enrichments were calculated by comparing immunoprecipitation efficiency of the region of interest to that of an unrelated region (*Amylase*). Primers used to amplify *Nanog* genomic region: 5'(-550/-462): CACAGGCTCTTTCTTCAGACTTG and TCTTGCTTGCTCTTCACATTGG, Oct-Sox (-215/-60): TCCCTCCCTCCCAGTCTG and CCTCCTACCCTACCCACCC, 3' (+929/+988): GGTAGAACCAAGAGGCTGCT and CATCACAACACGCACCTGA. Primers used for *Zfp42* (-283/-117): TGCATCCTCTGCTTGTGTAA and CAGAGCTGTCCCCTTGTCT; *Rest* (-3216/-3071): CTCCCCTGGACAATAGCTTC and CGTCCTTCATTTCCTCAGTG; *Dppa3* (-1770/-1550):

GATCCAGCTGGTCTGAGCTA and GTGCAGGGATCATAGGAGTG; *Lefty1* (-1264/-1060): AAGCTGCAGACTTCATTCCA and CGGGGGATAGATGAAGAAAC ([34]). Primers used for *Amylase*: CTCCTTGTACGGGTTGGT and AATGATGTGCACAGCTGAA.

Electrophoretic Mobility Shift Assays (EMSA)

293T cells were transfected with pPyCAGIP derivatives[29] expressing the cDNAs of FLAG-Esrrb, Oct4 or Sox2. After 24 hrs, cells were washed and harvested in phosphate-buffered saline, resuspended in lysis buffer (50 mM Hepes, pH 7.9, 150 mM NaCl, 1 mM EDTA, 0.5% NP-40, 1x *Complete EDTA-free* protease inhibitor, Roche) and rotated at 4 °C for 20 min. After microcentrifuge at 13,000 rpm for 10 min, supernatants were aliquoted and stored at -80 °C. EMSA was carried out as described[35]. The antibodies for supershift were added after the initial incubation for a further 10 min as follows: 2 μg of anti-Oct4 (sc-9081), 2 μg of anti-Sox2 (sc-17320) or 1 μg of FLAG M2 antibody (Sigma).

Immunofluorescence

46C ES cells were grown on coverslips coated with 0.1% gelatin and stained with antibodies using a standard protocol. In short, cells were fixed in 4% paraformaldehyde and incubated with antibodies against Nanog[14], Oct4 (N19, Santa Cruz) and Esrrb (R&D systems). Secondary antibodies are from DAKO laboratories (FITC swine anti-rabbit), Molecular probes (Alexa Fluor 594 goat anti-mouse) and Jackson Immunoresearch Laboratories (FITC rabbit anti-goat). Images were taken with Axio Imager (Zeiss).

ACKNOWLEDGEMENTS

We thank Dr. Hitoshi Niwa for ZHBTc4 cells, Dr. Austin Cooney for the 2.5 kb *Nanog* promoter-luciferase construct, Dr. Paul Robson for the *Nanog* mOS luciferase and control *Nanog* luciferase constructs, Dr. Austin Smith for 46C ES cells, Rodrigo Osorno for technical assistance and Dr. Frank Grosveld for critically reading the manuscript and helpful suggestions, This work was supported by an NWO Vidi grant to the R.P. lab and by the Medical Research Council and the Biotechnological and Biological Sciences Research Councils of the UK, the Wellcome Trust and the Juvenile Diabetes Research Foundation.

REFERENCES

1. Niwa, H., *How is pluripotency determined and maintained?* Development, 2007. **134**(4): p. 635-46.
2. Niwa, H., J. Miyazaki, and A.G. Smith, *Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells.* Nat Genet, 2000. **24**(4): p. 372-6.
3. Chambers, I., D. Colby, M. Robertson, J. Nichols, S. Lee, S. Tweedie, and A. Smith, *Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells.* Cell, 2003. **113**(5): p. 643-55.
4. Mitsui, K., Y. Tokuzawa, H. Itoh, K. Segawa, M. Murakami, K. Takahashi, M. Maruyama, M. Maeda, and S. Yamanaka, *The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells.* Cell, 2003. **113**(5): p. 631-642.
5. Ying, Q.L., J. Nichols, I. Chambers, and A. Smith, *BMP induction of Id proteins suppresses*

differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell, 2003. **115**(3): p. 281-92.

6. Ambrosetti, D.C., C. Basilico, and L. Dailey, *Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites.* Mol Cell Biol, 1997. **17**(11): p. 6321-9.
7. Kuroda, T., M. Tada, H. Kubota, H. Kimura, S.Y. Hatano, H. Suemori, N. Nakatsuji, and T. Tada, *Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression.* Mol Cell Biol, 2005. **25**(6): p. 2475-85.
8. Rodda, D.J., J.L. Chew, L.H. Lim, Y.H. Loh, B. Wang, H.H. Ng, and P. Robson, *Transcriptional regulation of nanog by OCT4 and SOX2.* J Biol Chem, 2005. **280**(26): p. 24731-7.
9. Boyer, L.A., T.I. Lee, M.F. Cole, S.E. Johnstone, S.S. Levine, J.P. Zucker, M.G. Guenther, R.M. Kumar, H.L. Murray, R.G. Jenner, D.K. Gifford, D.A. Melton, R. Jaenisch, and R.A. Young, *Core transcriptional regulatory circuitry in human embryonic stem cells.* Cell, 2005. **122**(6): p. 947-56.
10. Loh, Y.H., Q. Wu, J.L. Chew, V.B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K.Y. Wong, K.W. Sung, C.W. Lee, X.D. Zhao, K.P. Chiu*, et al.*, *The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells.* Nat Genet, 2006. **38**(4): p. 431-40.
11. Masui, S., Y. Nakatake, Y. Toyooka, D. Shimosato, R. Yagi, K. Takahashi, H. Okochi, A. Okuda, R. Matoba, A.A. Sharov, M.S. Ko, and H. Niwa, *Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells.* Nat Cell Biol, 2007. **9**(6): p. 625-35.
12. Wang, J., S. Rao, J. Chu, X. Shen, D.N. Levasseur, T.W. Theunissen, and S.H. Orkin, *A protein interaction network for pluripotency of embryonic stem cells.* Nature, 2006. **444**(7117): p. 364-8.
13. Ivanova, N., R. Dobrin, R. Lu, I. Kotenko, J. Levorse, C. DeCoste, X. Schafer, Y. Lun, and I.R. Lemischka, *Dissecting self-renewal in stem cells with RNA interference.* Nature, 2006. **442**(7102): p. 533-8.
14. Chambers, I., J. Silva, D. Colby, J. Nichols, B. Nijmeijer, M. Robertson, J. Vrana, K. Jones, L. Grotewold, and A. Smith, *Nanog safeguards pluripotency and mediates germline development.* Nature, 2007. **450**(7173): p. 1230-4.
15. Chambers, I., *Mechanisms and factors in embryonic stem cell renewal.* Rend. Fis. Acc. Lincei, 2005. **16**: p. 83-97.
16. Wu da, Y. and Z. Yao, *Isolation and characterization of the murine Nanog gene promoter.* Cell Res, 2005. **15**(5): p. 317-24.
17. Nowak, D.E., B. Tian, and A.R. Brasier, *Two-step cross-linking method for identification of NF-kappaB gene network by chromatin immunoprecipitation.* Biotechniques, 2005. **39**(5): p. 715-25.
18. Liu, T., A.E. Tee, A. Porro, S.A. Smith, T. Dwarte, P.Y. Liu, N. Iraci, E. Sekyere, M. Haber, M.D. Norris, D. Diolaiti, G. Della Valle, G. Perini, and G.M. Marshall, *Activation of tissue transglutaminase transcription by histone deacetylase inhibition as a therapeutic approach for Myc oncogenesis.* Proc Natl Acad Sci U S A, 2007. **104**(47): p. 18682-7.
19. Dufour, C.R., B.J. Wilson, J.M. Huss, D.P. Kelly, W.A. Alaynick, M. Downes, R.M. Evans, M. Blanchette, and V. Giguere, *Genome-wide orchestration of cardiac functions by the orphan nuclear receptors ERRalpha and gamma.* Cell Metab, 2007. **5**(5): p. 345-56.
20. Sladek, R., J.A. Bader, and V. Giguere, *The orphan nuclear receptor estrogen-related receptor alpha is a transcriptional regulator of the human medium-chain acyl coenzyme A dehydrogenase gene.* Mol Cell Biol, 1997. **17**(9): p. 5400-9.
21. Gearhart, M.D., S.M. Holmbeck, R.M. Evans, H.J. Dyson, and P.E. Wright, *Monomeric complex of human orphan estrogen related receptor-2 with DNA: a pseudo-dimer interface mediates extended half-site recognition.* J Mol Biol, 2003. **327**(4): p. 819-32.
22. Hatano, S.Y., M. Tada, H. Kimura, S. Yamaguchi, T. Kono, T. Nakano, H. Suemori, N. Nakatsuji, and T. Tada, *Pluripotential competence of cells associated with Nanog activity.* Mech Dev, 2005. **122**(1): p. 67-79.
23. Singh, A.M., T. Hamazaki, K.E. Hankowski, and N. Terada, *A heterogeneous expression pattern for Nanog in embryonic stem cells.* Stem Cells, 2007. **25**(10): p. 2534-42.
24. Pettersson, K., K. Svensson, R. Mattsson, B. Carlsson, R. Ohlsson, and A. Berkenstam, *Expression of a novel member of estrogen response element-binding nuclear receptors is restricted to the early stages of chorion formation during mouse embryogenesis.* Mech Dev, 1996. **54**(2): p. 211-23.
25. Lu, D., Y. Kiriyama, K.Y. Lee, and V. Giguere, *Transcriptional regulation of the estrogen-inducible pS2 breast cancer marker gene by the ERR family of orphan nuclear receptors.* Cancer Res, 2001. **61**(18): p. 6755-61.
26. Xie, W., H. Hong, N.N. Yang, R.J. Lin, C.M. Simon, M.R. Stallcup, and R.M. Evans, *Constitutive activation of transcription and binding of coactivator by estrogen-related receptors 1 and 2.* Mol Endocrinol, 1999. **13**(12): p. 2151-62.
27. Castet, A., A. Herledan, S. Bonnet, S. Jalaguier, J.M. Vanacker, and V. Cavailles, *Receptor-interacting protein 140 differentially regulates estrogen receptor-related receptor transactivation depending*

on target genes. Mol Endocrinol, 2006. **20**(5): p. 1035-47.
28. Ying, Q.L., M. Stavridis, D. Griffiths, M. Li, and A. Smith, *Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture.* Nat Biotechnol, 2003. **21**(2): p. 183-6.
29. Mullin, N.P., A. Yates, A.J. Rowe, B. Nijmeijer, D. Colby, P.N. Barlow, M.D. Walkinshaw, and I. Chambers, *The pluripotency rheostat Nanog functions as a dimer.* Biochem J, 2008. **411**(2): p. 227-31.
30. Dignam, J.D., R.M. Lebovitz, and R.G. Roeder, *Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei.* Nucleic Acids Res, 1983. **11**(5): p. 1475-89.
31. Wilm, M., A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis, and M. Mann, *Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry.* Nature, 1996. **379**(6564): p. 466-9.
32. Sanchez, C., I. Sanchez, J.A. Demmers, P. Rodriguez, J. Strouboulis, and M. Vidal, *Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbxl10/Jhdm1B histone demethylase and the Bcl6 interacting corepressor.* Mol Cell Proteomics, 2007. **6**(5): p. 820-34.
33. Gu, P., D. LeMenuet, A.C. Chung, M. Mancini, D.A. Wheeler, and A.J. Cooney, *Orphan nuclear receptor GCNF is required for the repression of pluripotency genes during retinoic acid-induced embryonic stem cell differentiation.* Mol Cell Biol, 2005. **25**(19): p. 8507-19.
34. Nakatake, Y., N. Fukui, Y. Iwamatsu, S. Masui, K. Takahashi, R. Yagi, K. Yagi, J. Miyazaki, R. Matoba, M.S. Ko, and H. Niwa, *Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells.* Mol Cell Biol, 2006. **26**(20): p. 7772-82.
35. Chew, J.L., Y.H. Loh, W. Zhang, X. Chen, W.L. Tam, L.S. Yeap, P. Li, Y.S. Ang, B. Lim, P. Robson, and H.H. Ng, *Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells.* Mol Cell Biol, 2005. **25**(14): p. 6031-46.

SUPPLEMENTAL FIGURE

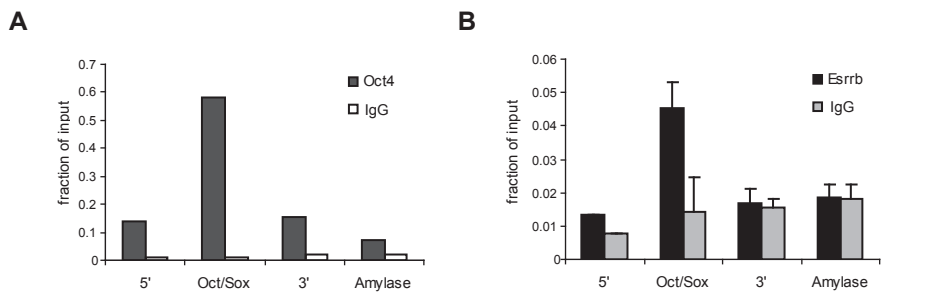


Figure S1. Oct4 and Esrrb bind to the *Nanog* promoter
(A) Chromatin immunoprecipitation with anti-Oct4 and goat IgG antibodies on dual cross-linked chromatin from 46C ES cells. Enrichments compared to input are shown. (B) Chromatin immunoprecipitation with anti-Esrrb and mouse IgG antibodies on dual cross-linked 46C ES cell chromatin. Precipitated material is shown as a fraction of input; error bars represent SEM of two independent experiments.

3

An Oct4-Centered Protein Interaction Network in Embryonic Stem Cells

Cell Stem Cell. 2010 Apr 2;6(4):369-81.

An Oct4-Centered Protein Interaction Network in Embryonic Stem Cells

Debbie L.C. van den Berg,[1] Tim Snoek,[1] Nick P. Mullin,[3] Adam Yates,[3] Karel Bezstarosti,[2] Jeroen Demmers,[2] Ian Chambers,[3] and Raymond A. Poot[1,*]

[1]Department of Cell Biology, [2]Proteomics Center, Erasmus MC, Dr.Molewaterplein 50, 3015GE Rotterdam, The Netherlands; [3]MRC Centre for Regenerative Medicine, Institute for Stem Cell Research, School of Biological Sciences, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JQ, Scotland

*Corresponding author

ABSTRACT

Transcription factors, such as Oct4, are critical for establishing and maintaining pluripotent cell identity. Whereas the genomic locations of several pluripotency transcription factors have been reported, the spectrum of their interaction partners is underexplored. Here, we use an improved affinity protocol to purify Oct4-interacting proteins from mouse embryonic stem (ES) cells. Subsequent purification of Oct4-partners Sall4, Tcfcp2l1, Dax1 and Esrrb resulted in an Oct4-interactome of 166 proteins, including transcription factors and chromatin-modifying complexes with documented roles in self-renewal, but also many factors novel to the ES cell network. We find Esrrb associated with the basal transcription machinery and also detect interactions between transcription factors and components of the TGFbeta, Notch and Wnt signaling pathways. Acute depletion of Oct4 reduced binding of Tcfcp2l1, Dax1 and Esrrb to several target genes. In conclusion, our purification protocol allowed us to bring greater definition to the circuitry controlling pluripotent cell identity.

INTRODUCTION

Embryonic stem (ES) cells are derived from the inner cell mass of mammalian embryos and have the unique ability to grow indefinitely in culture while retaining their pluripotency[1]. This self-renewal capacity is regulated by a set of transcription factors including Oct4, Nanog and Sox2[2]. ES cells are particularly sensitive to dosage alterations in Oct4; a 50% increase or decrease in the level of Oct4 causes differentiation into cells expressing markers of endoderm and mesoderm or trophectoderm, respectively[3]. Oct4 also plays a central role in the reprogramming of both human and mouse fibroblasts into induced pluripotent stem (iPS) cells[4-6]. Oct4 is one of a set of reprogramming factors that usually also includes Sox2, Klf4 and c-myc[7-8]. Sox2, Klf4 and c-myc can be replaced by family members such as Sox1, Sox3, Klf2, Klf5, L-Myc and N-Myc, but without Oct4 no reprogramming occurs[9].

Recently, genome-wide chromatin immunoprecipitation (ChIP) analyses in mouse ES cells have identified the genomic binding sites of Oct4 and a number of other ES cell transcription factors[10-12]. Oct4 clusters with a variable but overlapping set of transcription factors at many genomic locations, including promoters and enhancers (reviewed in[13]. Clusters with a relatively high number of different transcription factors appear to correlate with ES cell specific expression

of the near-by gene[10-11]. The mechanism for this molecular clustering may have similarities with the partnership of Oct4 with Sox2. Oct4 and Sox2 have low affinity for each other in solution[14-15], yet this affinity is critical for the co-operative binding of Oct and Sox proteins to adjacent sites on DNA[14, 16]. Therefore, identifying the interaction partners of transcription factors important for pluripotency could add novel components to the pluripotency transcriptional network and help to elucidate the assembly mechanism of transcription factor clusters. However, physical interactions between ES cell transcription factors remain underinvestigated. Low affinity interactions between transcription factors together with the generation of sufficient ES cell material for biochemical purification complicate an effective search for interaction partners. To address these drawbacks, we improved the FLAG-affinity based protein purification protocol. Using only small amounts of starting material, we initially purified FLAG-tagged Oct4 and its interacting proteins from mouse ES cells. Subsequently, we purified four of the identified Oct4-interacting ES cell transcription factors, Sall4, Esrrb, Dax1 and Tcfcp2l1. The resulting interaction network contains many transcriptional regulators and chromatin modifying complexes known to play roles in ES cell self-renewal, as well as transcriptional regulators not previously affiliated with pluripotency. We find associations between transcription factors and several signaling pathways and uncover the first physical connection of an ES cell transcription factor, Esrrb, with the basal transcription machinery. Thus, our methodology allowed for a much more detailed view of the physical interactions between factors that act in the ES cell pluripotency network.

RESULTS

Purification of Oct4-interacting proteins from ES cells

We have previously described a mouse ES cell line in which, under self-renewing conditions, all the Oct4 protein in the cell has an N-terminal triple FLAG-tag (F-Oct4)[17]. Both F-Oct4 and the parental ZHBTc4 cells have a normal ES cell morphology[3, 17] and express normal levels of ES cell markers Sox2, Sall4 (Figure S1A) and Klf4, Dax1, Zfp42 and Eras (Figure S1B). This indicates that the F-Oct4 protein present in the F-Oct4 cells maintains their ES cell identity. We prepared nuclear extracts from F-Oct4 cells and ZHBTc4 cells, which do not express F-Oct4 and serve as a control. FLAG-affinity purifications were performed from 1.5 ml of nuclear extract (equivalent to ~4 x 10^8 cells) using an improved protocol in which near-physiological salt conditions, low detergent concentrations and low-adherence tubes were employed (see Experimental Procedures for details). Benzonase nuclease was added to the extract to remove the remaining DNA (Figure S1C), thereby eliminating protein interactions mediated indirectly by DNA bridging. Virtually all F-Oct4 in the extract was bound to the FLAG-antibody beads and subsequently eluted by FLAG peptide competition (Figure S1D). An SDS polyacrylamide gel of the eluted fractions, stained with a sensitive Colloidal Coomassie protocol, showed Oct4 as the predominant band in the F-Oct4 sample (Figure 1A). The control sample showed only one prominent band, which was also present in the F-Oct4 sample but was otherwise devoid of major contaminants. This indicates that our

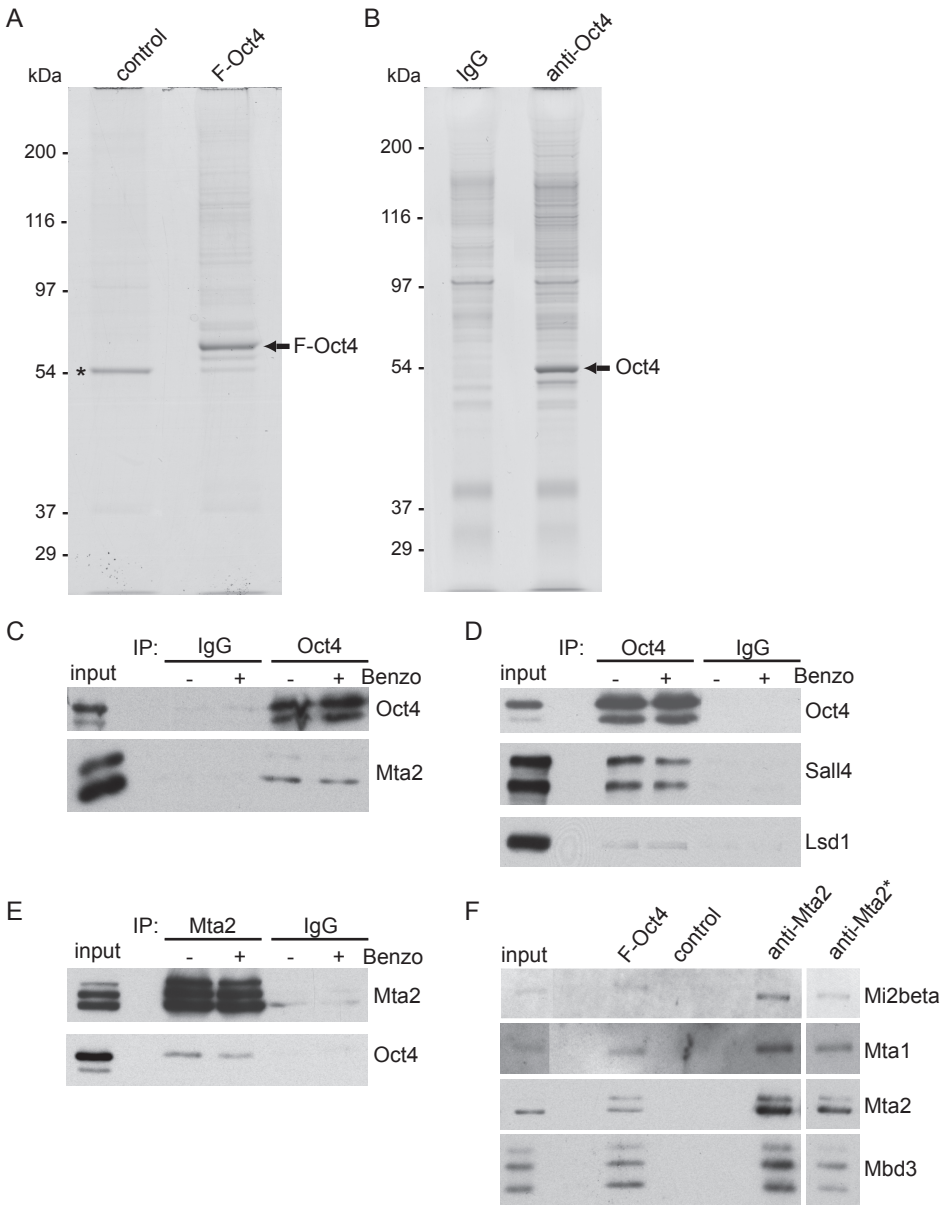


Figure 1. Purification of Oct4 and its interacting proteins
(A) Colloidal Coomassie-stained SDS-polyacrylamide gel of a FLAG-Oct4 (F-Oct4) and control purification. (*) indicates contaminating band. The F-Oct4 band is indicated. (B) Colloidal Coomassie-stained SDS-polyacrylamide gel of immunoprecipitated endogenous Oct4 and a control immunoprecipitation using IgG. The Oct4 band is indicated. (C, D) Oct4 immunoprecipitates analyzed by western blots with the indicated antibodies. Benzonase (Benzo) was added where indicated. (E) MTA2 immunoprecipitates analyzed by western blots with the indicated antibodies. (F) Subunit stoichiometry of F-Oct4-bound NuRD complex (F-Oct4) compared to anti-Mta2 co-immunoprecipitated NuRD complex (anti-Mta2) by western blot against the indicated NuRD subunits. (*) indicates a lighter exposure of the same experiment. See also Figure S1.

FLAG-mediated purification of Oct4 has a very good signal to background ratio. The presence of multiple bands of lower intensity in the F-Oct4 lane, suggests that Oct4 interacts with a variety of proteins at sub-stoichiometric levels. The majority of Oct4 runs at approximately its own molecular weight on a gel filtration column (Figure S1E), unlike a stable complex such as NuRD. Therefore most Oct4 interactions are likely to be weak and do not survive the 4 hour gel filtration procedure, in which dissociation causes an irreversible loss of the interaction. To independently verify candidate F-Oct4-interacting proteins, we also immunoprecipitated endogenous Oct4 from nuclear extracts of a different ES cell line, 46C[18], using an antibody that captured all Oct4 from the extract (Figure S1F). Although we used the same buffer conditions and low-adherence tubes, this procedure gives higher background compared to the FLAG-affinity purification (Figure 1B), as proteins that bind non-specifically to the antibody beads or the tubes cannot be excluded from the eluate by FLAG-peptide elution, as they can in the FLAG purification strategy.

We analyzed three independent F-Oct4 purifications and the endogenous Oct4 immunoprecipitation by mass spectrometry (Table 1). A representation of the identified proteins by a more quantitative measure, emPAI score[19] is shown in Table S1. Our list of over 50 putative Oct4-associated proteins (Table 1) contains 22 transcription factors of which half have a role in maintaining pluripotency (Table 2). These include Sall4, Klf5, Zfp143, Esrrb and Sox2, the best characterized Oct4 partner for which 3-D structures of the Oct4-Sox2-DNA ternary complex have been reported[16, 20]. We also identified a number of chromatin modifying complexes (CMCs). All of the subunits of the transcriptional repressor NuRD were specifically present, except for Rbbp4 (high background prevented inclusion of Rbbp4 in Table1). We detected subunits from the chromatin remodeling complexes SWI/SNF and Trrap/p400, the Lsd1 histone demethylase complex and components of the Polycomb Repression Complex 1 (PRC1).

Next we examined the presence of some of the identified interactors in Oct4 immunoprecipitates by immunoblotting. Indeed, we find that NuRD subunit Mta2 (Figure 1C), spalt-like protein Sall4 and histone-demethylase Lsd1 (Figure 1D), Sall1 and Wdr5 (Figure S1G) co-precipitate with Oct4, while immunoprecipitates of Mta2 (Figure 1E) and Wdr5 (Figure S1H) contain Oct4. Recently, it was suggested that a subset of the NuRD subunits (Mta1 and 2, Gatad2a and Gatad2b, Hdac1 and 2) forms an Oct4/ Nanog-associated complex called NODE (Nanog and Oct4 associated deacetylase)[21]. We found that Oct4 binds the classical NuRD complex, as it was originally defined[22], including catalytic subunit Mi2beta and Mbd3 and Rbbp7 (Table 1). Immunoblotting confirmed this; the proportionate amount of antigen detected for Mi2beta, Mbd3, Mta1 and Mta2 was the same in FLAG-Oct4 and Mta2 IP samples (Figure 1F). This suggests that Oct4-bound NuRD is similar or identical to classical NuRD in its composition and argues against the existence of Oct4-bound NuRD subcomplexes, such as NODE.

Oct4-interacting proteins correlate with gene regulation by Oct4 and ES cell self-renewal

Proteins that interact with Oct4 may be expected to be Oct4-cofactors in gene regulation and have DNA binding profiles that overlap with Oct4. Recently, two studies reported the genome-

Table 1. Oct4-Interacting Proteins as Identified by Mass Spectrometry Analysis of Purified Oct4 Samples

Protein	Accession	Flag#1		Flag#2		Flag#3		Oct4-IP		Average mascot
		Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	
Oct4 (Pou5f1)	gi|200118	532	7	987(161)	14(2)	990(233)	15(5)	909	11	854
NuRD complex										
Mi2beta (Chd4)	gi|39204553	2041(113)	33(4)	2576	42	2537(287)	50(7)	2959(143)	42(2)	2528
Mta1	gi|15077051	930	13	1433	22	1411	25	1235	17	1252
Gatad2a	gi|148696823	1030	14	1154	17	1575(358)	23(7)	926	12	1171
Mta2	gi|51491880	875	13	1294	19	1358	23	1003(58)	13(1)	1092
Gatad2b	gi|120577529	697	10	961	14	1346(181)	20(3)	623	7	906
Hdac1	gi|2347180	928(171)	13(3)	815(84)	12(1)	801(171)	15(4)	790(275)	9(5)	833
Mbd3	gi|7305261	563	7	525	7	793	14	1084(112)	21(1)	741
Mta3	gi|18381007	521	7	402	7	870	17	1010(69)	13(2)	700
Hdac2	gi|3023934	879(190)	11(3)	656(84)	11(1)	654	11	700(201)	9(3)	722
Rbbp7	gi|2494892	584(146)	9(4)	472	9	780(409)	14(7)	520	7	589
SWI/SNF complex										
Baf155(Smarcc1)	gi|30851572	616	13	558	9	872	15	420(328)	7(7)	616
Brg1 (Smarca4)	gi|76253779	384	7	618	11	444	10	525(362)	10(8)	492
PRC1 complex										
Phc1	gi|30923312	256	5	507	7	630	10	-	-	348
Ring1B (Rnf2)	gi|109157342	273	4	155	2	297(53)	6(2)	251(82)	4(1)	244
Rybp	gi|5381327	95(51)	1(1)	94	1	127	2	107	1	105
Trrap/p400 complex										
Trrap	gi|124486949	154	4	273	4	134	3	496	7	264
Ep400	gi|27348237	261	6	91	1	231	5	77	2	165
LSD1 complex										
Lsd1	gi|51315882	174	4	604	10	640	13	97	2	378
Zmym2	gi|28175571	189	2	533	8	296	5	-	-	254
Rcor2	gi|17298682	93	2	163	2	272	5	241	3	192
Transcription factors										
Sall4	gi|81913723	2622(709)	30(10)	2526(125)	31(2)	2574(594)	38(12)	2554(636)	29(8)	2569
Sall1	gi|14164331	1987	24	2371	30	2088	31	1822	26	2067
Zfp219	gi|30794418	297	4	620	10	430	7	505	6	463
Arid3b	gi|9790033	257	3	301	4	1030	17	113	2	426
Wdr5	gi|16554627	452(68)	5(2)	222	2	468(67)	7(2)	447	5	397
Zfp462	gi|114431238	-	-	64	2	256	7	985	16	326
Sox2	gi|127140986	214	4	-	-	344	5	444	4	250
Mga	gi|6692607	-	-	375	6	348	10	247	5	242
Ubp1	gi|7305605	290	6	131	2	236	5	430	5	242
Nac1	gi|31543309	-	-	315	5	287	5	269	5	217

Table 1. *Continued*

Protein	Accession	Flag#1		Flag#2		Flag#3		Oct4-IP		Average mascot
		Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	
Transcription Factors (continued)										
Hcfc1	gi|4098678	98	4	293	5	419(59)	10(1)	-	-	202
Hells	gi|12232371	-	-	316	6	287	6	53	2	164
Rbpj	gi|94400775	61	1	174	3	88	3	307	5	157
Tcfcp2l1	gi|90101766	227	4	125	2	61	2	213	4	156
Requiem	gi|6755314	304	4	150	2	-	-	157	3	153
Esrrb	gi|124375796	117	2	69	1	134	3	256	4	144
Pml	gi|9506979	136	2	66	2	333	7	-	-	134
Foxp4	gi|161016782	-	-	71	1	349	7	52	1	118
Ctbp2	gi|6753548	-	-	128	3	231	4	97	2	114
Dax1	gi|6671531	77	1	97	2	135	3	122	2	108
Zfp143	gi|22902397	186	3	-	-	118	3	84	1	97
Klf5	gi|31981873	-	-	70	1	132	2	111	1	78
Other										
Rif1	gi|47078460	2343	31	3370	40	2213	35	2421(1026)	31(12)	2587
L1td1	gi|148698953	271	3	311	5	497(337)	9(6)	196(58)	3(1)	319
Akap8	gi|31560394	92	1	298	4	358	6	311	4	264
Msh2	gi|30047836	298	5	142	4	522	11	54	1	254
Ogt	gi|13775066	148	2	149	4	671(160)	15(5)	-	-	242
Rbm14	gi|16307494	179	2	90	1	463(57)	9(2)	163	2	224
Frg1	gi|17376286	139	2	394	6	180	4	155	2	217
Smc1a	gi|123220915	74	3	433	10	243	8	-	-	187
Emsy	gi|124249084	144	3	104	1	429	7	-	-	170
0610010K14Rik	gi|81917220	103	2	175	2	222	3	-	-	125
2810474O19Rik	gi|148678819	69	3	69	1	213	5	-	-	88
Zcchc8	gi|148687677	97	2	84	1	106	2	-	-	72

Thresholds for inclusion of the identified proteins into Table 1 are in Experimental Procedures.

[a] Mascot score for the specified protein in the Oct4 sample, purified by FLAG-affinity or Oct4-immunoprecipitation (Oct4-IP), Mascot score for the specified protein in the corresponding control purification, if present, is between brackets.

[b] Number of identified unique, non-redundant peptides for the specified protein in the Oct4 sample. Number of identified unique peptides in the control purification is between brackets. See also Table S1.

wide binding sites of different sets of ES cell transcription factors[10-11]. Five of the Oct4-interacting transcription factors identified here (Sox2, Nac1, Tcfcp2l1, Esrrb, Dax1) were investigated in those studies and were found to co-localize frequently with Oct4 (Table 2), including at the promoters of important pluripotency genes such as *Nanog* and *Oct4*[10-11, 23].

Phenotypes are documented for ~60% of the identified Oct4-interacting proteins (Table 2). Of these, ~65% (21/32) of the tested factors (Table 2) affect the ability of ES cells to remain undifferentiated. This includes most of the aforementioned transcription factors and subunits of all the identified Oct4-associated chromatin modifying complexes (Table 2), except for the Lsd1 complex.

We then investigated whether genes encoding Oct4-interacting proteins are bound and regulated by Oct4. Gene expression profiling data from ZHBTc4 ES cells, which express Oct4 from a doxycycline-repressible transgene[24] was combined with two different sets of Oct4 ChIP data[10-11]. We find that 14 factors (26%) are encoded by genes bound by Oct4 that are down-regulated after 48 hours of doxycycline treatment (Table 2). This correlation of Oct4 binding and transcriptional regulation by Oct4 increases the interdependence of the associated proteins with Oct4, as previously observed[25].

Purification of interaction partners of Sall4, Esrrb, Dax1 and Tcfcp2l1

Having established that our FLAG-affinity purification protocol identifies novel interactions that are independently verifiable and biologically relevant, an expanded network of Oct4 interactions was sought. Sall4, Esrrb, Dax1 and Tcfcp2l1 were selected for purification because of their consistent presence in all Oct4 purifications (Table 1). The spalt-like transcription factor Sall4 is important for stabilizing ES cell self-renewal[26-27]. Orphan receptor Esrrb is important for ES cell self-renewal[28-29]. Esrrb positively regulates the expression of the key pluripotency gene Nanog[17] and overexpression of Esrrb allows short-term ESC maintenance without the addition of exogenous LIF[30]. Esrrb is also capable of replacing KLF4 in somatic cell reprogramming[31]. Dax1 is an orphan receptor that is important for ES cells self-renewal[32]. Tcfcp2l1 co-localizes with Oct4 on many ES cell promoters and may be important for optimal ES cell proliferation[11, 28]. FLAG-tagged cDNAs were stably introduced into ZHBTc4 ES cells and clones selected that express the encoded proteins at levels similar to the endogenous proteins (Figure S2A). These clones had comparable morphology and growth rate to the parental line (data not shown). Proteins were purified by our FLAG-affinity protocol and coomassie-stained gels of the purified fractions from F-Sall4, F-Esrrb and F-Tcfcp2l1 purifications showed prominent bands of the expected molecular weight (Figure 2A) that reacted with the FLAG-antibody (Figure S2B). The presence of additional bands in the transcription factor purifications suggests the efficient co-purification of associated proteins. F-Dax1 was not visible by coomassie blue staining (Figure 2A), although it was almost completely depleted from the nuclear extract by the purification (Figure S2B). Together with the weaker anti-FLAG western signals of F-Dax1 extracts and purified Dax1 fractions, compared to the other FLAG proteins (not shown), this suggests a relatively low expression level of F-Dax1 (and therefore of endogenous Dax1) in ES

Table 2. Transcriptional Network and Phenotype of Oct4-Interacting Proteins

Protein	Promoter Co-occup. with Oct4[a]	Gene Bound by Oct4[a]	Expression Change upon Oct4 Depletion[b]	ESC Depletion Phenotype[c]	Developmental Phenotype[d] (Emb. Day of Lethality)
NuRD complex					
Mi2beta	-	no	no	-	-
Mta1	-	no	no	differentiation	-
Gatad2a	-	no	no	not detected	~ E10.5
Mta2	-	yes	no	not detected	-
Gatad2b	-	no	no	-	-
Hdac1	-	no	no	reduced proliferation	before E10.5
Mbd3	-	yes	no	increased selfrenewal	~ E8.5
Mta3	-	yes	no	-	-
Hdac2	-	yes	no	not detected	viable
Rbbp7	-	no	down	-	-
SWI/SNF complex					
Baf155	-	yes	down	differentiation	before E5.5
Brg1	-	no	no	differentiation	before E6.5
PRC1/Mblr complex					
Phc1	-	yes	down	-	peri-natal
Ring1b	yes	no	no	differentiation	before E10.5
Rybp	-	yes	down	not detected	before E7.5
Trrap/p400 complex					
Ep400	-	no	no	differentiation	~ E9.5
Trrap	-	no	no	differentiation	~ E3.5
LSD1 complex					
Lsd1	-	no	no	reduced proliferation	before E7.5
Zmym2	-	no	no	-	-
Rcor2	-	yes	down	-	-
Transcription factors					
Sall4	-	yes	no	differentiation prone	before E5.5
Sall1	-	yes	no	not detected	peri-natal
Zfp219	-	yes	down	differentiation	-
Arid3b	-	no	no	-	before E11.5
Wdr5	-	no	down	differentiation	-
Zfp462	-	yes	down	-	-
Mga	-	yes	no	differentiation	-
Sox2	yes	yes	down	differentiation	before E7.5
Ubp1	-	no	no	-	~ E11.5
Nac1	yes	no	no	differentiation	viable
Hcfc1	-	no	no	differentiation	-
Hells	-	no	down	-	-
Rbpj	-	yes	no	not detected	before E10.5
Tcfcp2l1	yes	yes	down	reduced proliferation	-
Requiem	-	no	no	-	-
Esrrb	yes	yes	down	differentiation	~ E10.5
Pml	-	yes	down	-	viable
Foxp4	-	no	no	-	~ E12.5
Ctbp2	-	no	no	increased selfrenewal	~ E10.5
Dax1	yes	yes	down	differentiation	-
Zfp143	yes	yes	no	differentiation	-
Klf5	yes	yes	down	differentiation	before E8.5
Other					
Rif1	-	yes	down	differentiation	-
L1td1	-	no	no	-	-
Akap8	-	no	no	-	-
Msh2	-	yes	no	not detected	not detected
Smc1a	-	no	no	differentiation	-
Ogt	-	yes	no	lethality	~ E5
Rbm14	-	yes	down	-	-
Frg1	-	no	no	-	-
Emsy	-	no	no	-	-
0610010K14Rik	-	no	no	-	-
281047O19Rik	-	no	no	-	-
Zcchc8	-	no	no	-	-

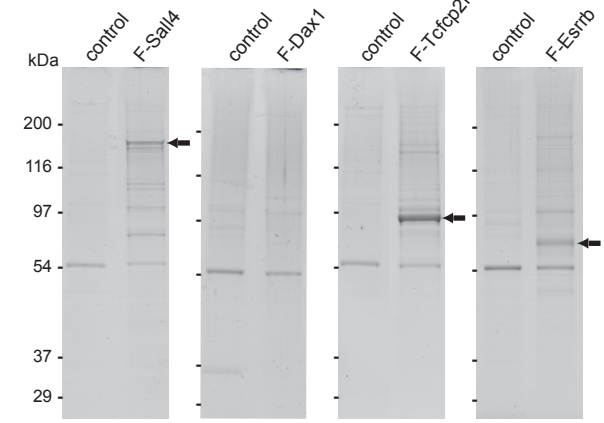
[a] Criteria and references for promoter co-occupancy with Oct4 and encoding gene bound by Oct4 are in the Experimental Procedures.

[b] Expression change upon Oct4 depletion in ZHBTc4 ES cells, criteria see Experimental Procedures.

[c] ES cell phenotype upon knock-out, or knock-down by RNA-interference, references in Supplemental Data,

[d] Developmental phenotype upon knock-out, references in Supplemental Data.

A



B

Protein	Average Mascot	Average peptides
Sall4	4254	47
NuRD complex (12)	1615	21
SWI/SNF complex (6)	225	3
Transcription factors		
Sall1	2536	32
Bend3	1962	27
Zfp219	704	9
Nac 1	703	9
Ruvbl2	640	9
Wdr5	506	7
Oct4	326	4
Grhl2	268	5
Sall 3	261	7
Cxxc5	210	3
Zfp143	194	3
Tcfcp2	180	3
Sall2	155	2
Klf5	142	2
Ctbp2	111	3
Zbtb2	109	2
Requiem	108	1
Ewsr1	106	2
Esrrb	101	2
Other		
Usp9x	634	11
7420416P09Rik	460	8
Set	365	4
L1td1	275	4

C

Protein	Average Mascot	Average peptides
Dax1 (Nr0b1)	845	14
Transcription factors		
Sall4	831	15
Esrrb	277	6
Snw1	206	4
Sall1	160	3
Wdr5	146	3
Oct4	136	2
Prmt1	56	1
Other		
Rif1	232	5
Ogt	201	4

Figure 2. Purification of F-Sall4, F-Dax1, F-Tcfcp2l1, F-Esrrb and their interacting proteins
(A) Colloidal Coomassie-stained SDS-polyacrylamide gels of representative purifications of the FLAG-tagged transcription factors and control purifications from the parental ES cell line. Arrows indicate the respective FLAG-tagged proteins.
(B-E) Summaries of the identified interacting proteins. The average Mascot score and number of identified unique peptides of two purifications without doxycycline addition are indicated for individual proteins or complexes. The number of identified subunits of a complex is between brackets. (F) F-Esrrb or control purifications analyzed by western blots with the indicated antibodies. Benzonase was added where indicated. (G) GST-Esrrb pull-downs analyzed by western blots with the indicated antibodies. Figure S3H (right panel) shows the purified GST proteins on a Coomassie-stained polyacrylamide gel. See also Figure S2 and Tables S2-S9.

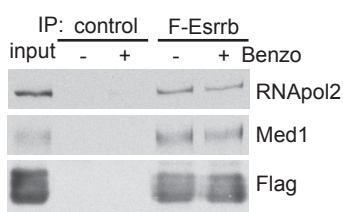
D

Protein	Average mascot	Average peptides
Tcfcp2l1	2391	27
NuRD complex (11)	998	14
Trrap/p400 complex (11)	669	10
SWI/SNF complex (9)	663	10
PRC1 complex (4)	143	3
Transcription factors		
Sall4	2770	35
Ubp1	1922	23
Tcfcp2	1027	12
Lsd1	882	11
Sall1	871	11
Esrrb	788	12
Smarca5	576	12
Wdr5	562	8
Peg10	549	7
Yeats2	509	8
Hcfc1	501	8
Zmym4	475	8
Hells	434	6
Mga	400	7
Zfp462	398	7
Oct4	348	5
Requiem	347	4
Pogz	293	4
Zfp143	287	5
Sin3a	245	4
Hnrnpab	236	3
Wiz	235	4
Adnp	234	4
Satb2	233	4
Klf5	221	3
Trim33	205	3
Mybl2	193	3
Bptf	196	3
Grhl2	169	3
L3Mbtl2	144	2
Zzz3	95	1
Ehmt1	68	1
Zfp828	63	2
Other		
Rif1	2084	27
Zcchc8	785	10
Lig3	481	9
Ogt	461	8
Xrcc6	423	6
Xrcc5	341	5
C130039O16Rik	327	5
Msh2	327	7
L1td1	297	5
Msh6	291	4
Rbm14	278	4
Xrcc1	243	4
Polb	230	4
EMSY	218	3
Rpa1	186	3
Prkdc	159	4
Pnkp	121	2
Asf1a	101	1
Cabin1	99	2
2310057J16Rik	67	1
Ubqln4	68	1

E

Protein	Average mascot	Average peptides
Esrrb	2391	27
SWI/SNF complex (5)	632	12
Trrap/p400 complex (7)	550	10
NuRD complex (11)	486	9
Mediator complex (23)	534	10
RNApol2 complex (4)	468	9
TFIID complex (5)	191	3
TRX/MLL complex (6)	218	5
Transcription factors		
Sall4	1147	17
Dax1	692	11
Tcfcp2l1	667	12
Fkbp15	629	9
Esrra	488	9
Ncoa3	421	8
Ubp1	408	8
Nrip1	408	9
Sall1	399	8
Zfp462	348	5
Cdc2a	295	6
Zbtb9	279	6
Snf2h	258	3
Wiz	186	4
Requiem	181	3
Jmjd1c	159	2
Tcfcp2	140	2
L3mbtl2	140	2
Oct4	133	2
Myst1	101	2
Ehmt1	88	2
Other		
Rif1	2084	27
Ogt	461	8

F



G

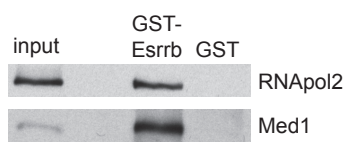


Figure 2. D-G

cells. Figures 2B-E provide summaries of the interacting proteins of Sall4, Dax1, Tcfcp2l1 and Esrrb (complete lists of identifications and information on Mascot scores, number of identified unique peptides and emPAI scores are shown in Tables S2-S5 and Tables S6-S9). To examine the Oct4-dependence of the interaction partner associations, we also performed the purifications 16 hours after doxycycline-mediated repression of Oct4, which removes essentially all Oct4 protein from ZHBTc4 derived cells[3, 17]. Purified fractions from two FLAG-purifications of cells with or without doxycycline addition were analyzed by mass spectrometry. Doxycycline addition had no consistent effect on the vast majority of the identified interactions (Tables S2-S9). Of the proteins affected by Oct4 modulation, only Esrrb was ever identified as an Oct4 interactor (Table 1). The interaction between Esrrb and Sall4 appears to be sensitive to removal of Oct4 in the F-Sall4 purifications (Tables S2 and S6). However, the mascot scores here are close to threshold, whereas in F-Esrrb purifications where Sall4 has a high Mascot and emPAI score, removal of Oct4 had no effect (Tables S5 and S9). Taken together, this suggests that the identified interactions are unlikely to be bridged by Oct4, although many of the identified proteins also interact with Oct4.

We independently verified a number of the putative interactors of F-Sall4, F-Dax1, F-Tcfcp2l1 and F-Esrrb. Immunoprecipitation of Sall4 co-precipitated Sall1 and MTA2 (Figure S3A), V5-tagged Zfp143 (Figure S3B) and F-Nac1 (Figure S3C), whereas Sall4 is present in immunoprecipitates of MTA2 (Figure S3D) and F-Nac1 (Figure S3E). GST-Dax1 pull downs precipitated Sall4, Sall1, Oct4, Wdr5 and Esrrb (Figure S3F). V5-Tcfcp2l1 immunoprecipitation brought down Esrrb and MTA2 (Figure S3G), whereas GST-Esrrb pull down co-precipitated MTA2, Sall4, Ep400 (Figure S3H), V5-Dax1 (Figure S3I) and F-Tcfcp2l1 (Figure S3J). MTA2 immunoprecipitation co-precipitated Esrrb (Figure S3K).

An Oct4-centered interaction network

We assembled the identified interactions of Oct4, Tcfcp2l1, Dax1, Sall4 and Esrrb into an interaction network containing 166 proteins (Figure 3). This allows the visualization of the interactions between the purified tagged transcription factors and their interaction with multiple chromatin modifying complexes (CMCs). The NuRD complex was associated with every tagged factor purified, except for Dax1 (Table 1, Figures 2B-E). The smaller set of interactors identified for Dax1 (Figure 2C), compared to the other purified proteins, may be due to the purification of relatively small amounts of F-Dax1 protein (Figure 2A). The Mascot and emPAI scores of NuRD are highest in the F-Sall4 purifications (Figure 2B, Tables S2, S6). Sall4 also interacts with Sall1, Sall2, Sall3 and associates with all the other tagged factors (Figures 2B-E). Binding of Sall4 to NuRD and Sall1 was previously observed[27]. Our data suggests that spalt proteins form a unit with NuRD, which then can associate with other transcription factors. Sall4 interactors Nac1 and Bend3 (Figure 2B) could also be part of this unit, as they were observed together in individual purifications of Tcfcp2l1 and Esrrb (data not shown). The SWI/SNF complex also associates with most tagged transcription factors (Table 1, Figures 2B, 2D and 2E). The Trrap/p400 complex is present with relatively high mascot and emPAI scores in Esrrb and Tcfcp2l1 purifications, with many subunits detected (Figures

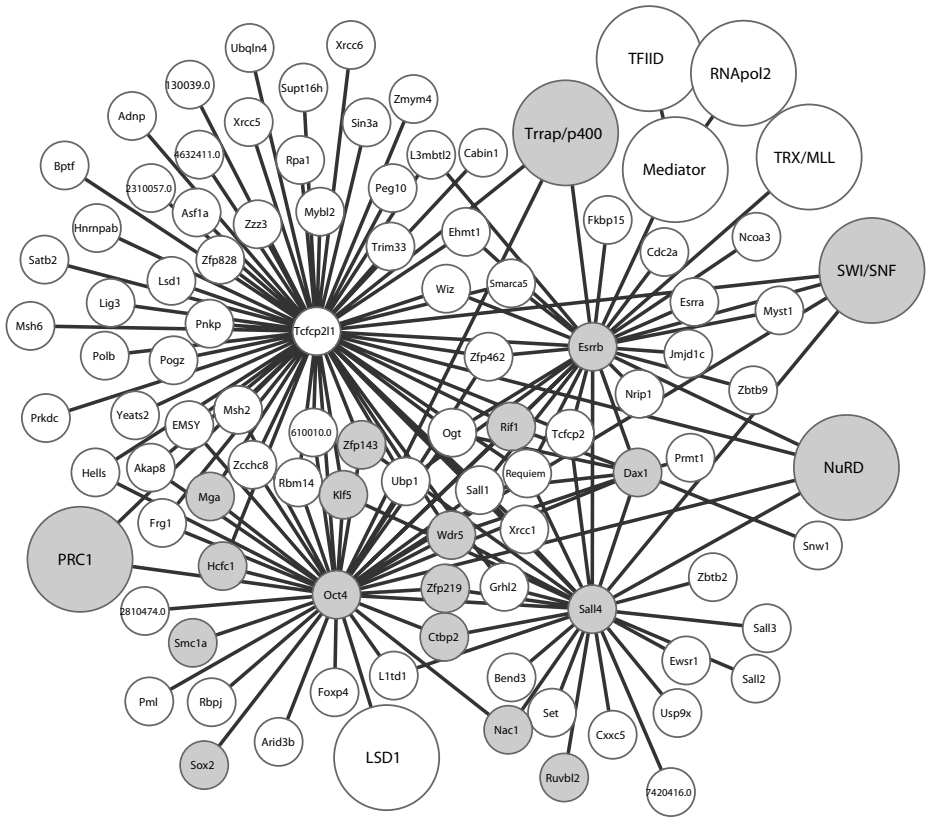


Figure 3. Protein interaction network of Oct4 and its associated proteins Sall4, Dax1, Tcfcp2l1 and Esrrb
The network represents the proteins present in both purifications (-Dox) of F-Sall4, F-Dax1, F-Tcfcp2l1 or F-Esrrb and/or present in F-Oct4 purifications as in Table 1 (complete lists of identifications and information on Mascot scores, number of identified unique peptides and emPAI scores are shown in Table 1, Tables S2-S5 and Tables S6-S9). Complexes are shown as larger circles. Grey shading indicates importance for ES cell self-renewal capacity (see Table 2). See also Figure S3.

2D and 2E, Tables S4, S5, S8 and S9). The PRC1/Mblr complex associates, besides with Oct4, also with Tcfcp2l1 (Figure 2D).

We find that the purified factors often bind efficiently to evolutionary related proteins. In addition to spalt proteins, we observed interactions between Tcfcp2l1, Tcfcp2, Ubp1 and Grhl2 (Figure 2D), all of which are related to the Drosophila Grainyhead transcription factors[33], whereas Esrrb binds the related protein Esrra (Figure 2E). This suggests that despite diversification, these proteins can still act together in transcription regulation.

Some of the purified factors harbor extensive sets of unique interacting proteins that may mediate their specific function in ES cells. For example, Tcfcp2l1 interacts with many proteins involved in DNA metabolic processes (Figure 2D) such as DNA replication (Polb, Asf1a, Rpa1) and DNA repair (Xrcc1, 5, 6, Msh2, 6, lig3, EMSY, Prkdc, pnkp) and related pathways such as cell-cycle

progression or cell proliferation (Hells, Msh2, Mybl2, EMSY).

Orphan receptor Esrrb, which is related to the estrogen receptor, was found to associate with Ncoa3 and Nrip and the TRX/Mll chromatin modifying complex (Figure 2E). Intriguingly, Esrrb also interacts with the Mediator complex, RNA polymerase II subunits (RNApol2) and TBP plus Tafs (TFIID complex, Figure 2E, Tables S5, S9), which are all components of the basal transcription machinery[34]. The association of Esrrb with Mediator and RNApol2 is DNA independent as it was not affected by benzonase treatment of the extract (Figure 2F). Moreover, recombinant GST-Esrrb also interacted efficiently with Mediator and RNA pol2 (Figure 2G).

The network provides links with protein modification and signaling pathways. For example, Oct4 associates with Rbpj, a transcription factor that acts as the nuclear effector of the Notch signaling pathway[35], suggesting a connection between Notch-regulated and Oct4-regulated gene-expression. Sall4 shows an interaction with Usp9x (Figure 2B), an essential component of the TGFbeta/BMP signaling pathway, which activates Smad4 by removing a mono-ubiquitin group[36]. Another Sall4 associated factor, Cxxc5 (Figure 2B), is regulated by TGFbeta signaling in neural stem cells, binds Wnt-signaling mediator Dvl and inhibits Wnt signaling[37]. By interacting with both Usp9x and Cxxc5, spalt proteins may provide a physical link between the TGFbeta and Wnt signaling pathways. Oct4, Esrrb, Tcfcp2l1 and Dax1 bind the glycosyl transferase Ogt (O-GlcNAc Transferase, Table 1, Figures 2B-2E), an enzyme that adds N-acetylglucosamine groups (O-GlcNAc) to proteins.

The network contains a number of transcription factors with a high level of inter-connectivity, characteristic of network hubs. Examples of such hubs are Zfp143 and Klf5. Zfp143 interacted with Oct4, Sall4, Tcfcp2l1 (Table1, Figures 2B and 2D) and was present in one Esrrb purification (not shown). Klf5 was present in Oct4, Sall4 and Tcfcp2l1 purifications (Table1, Figures 2B and 2D). The purified factors Esrrb, Tcfcp2l1, Dax1 and Sall4 were selected on their interaction with Oct4, but they also have an Oct4-independent interaction with one another. All these highly connected factors affect ES cell self-renewal when depleted (Table 2), suggesting that physical interaction may play a role in regulating this process. A possible rationale for this correlation, co-dependent recruitment to DNA, will be tested experimentally below.

Oct4-dependent recruitment of Dax1, Tcfcp2l1 and Esrrb

Our purifications showed the physical interaction of Oct4 with Dax1, Tcfcp2l1 and Esrrb. To investigate the relevance of these interactions for the ES cell transcriptional network, we tested the effect of acute Oct4 depletion, by 12 hours doxycycline treatment, on the recruitment of Dax1, Tcfcp2l1 and Esrrb to a number of genomic binding sites to which Oct4 also binds[10-11]. Indeed, depletion of Oct4 reduced recruitment of F-Dax1, F-Tcfcp2l1 and F-Esrrb to several of their targets (Figures 4A-C). For example, Dax1 recruitment to the *Rest* and *Nanog* promoters, which are both also occupied by many other ES cell transcription factors[10-11] is dependent on Oct4. Our data suggest that Oct4 can provide an anchor on the DNA for the recruitment of several of its associated factors.

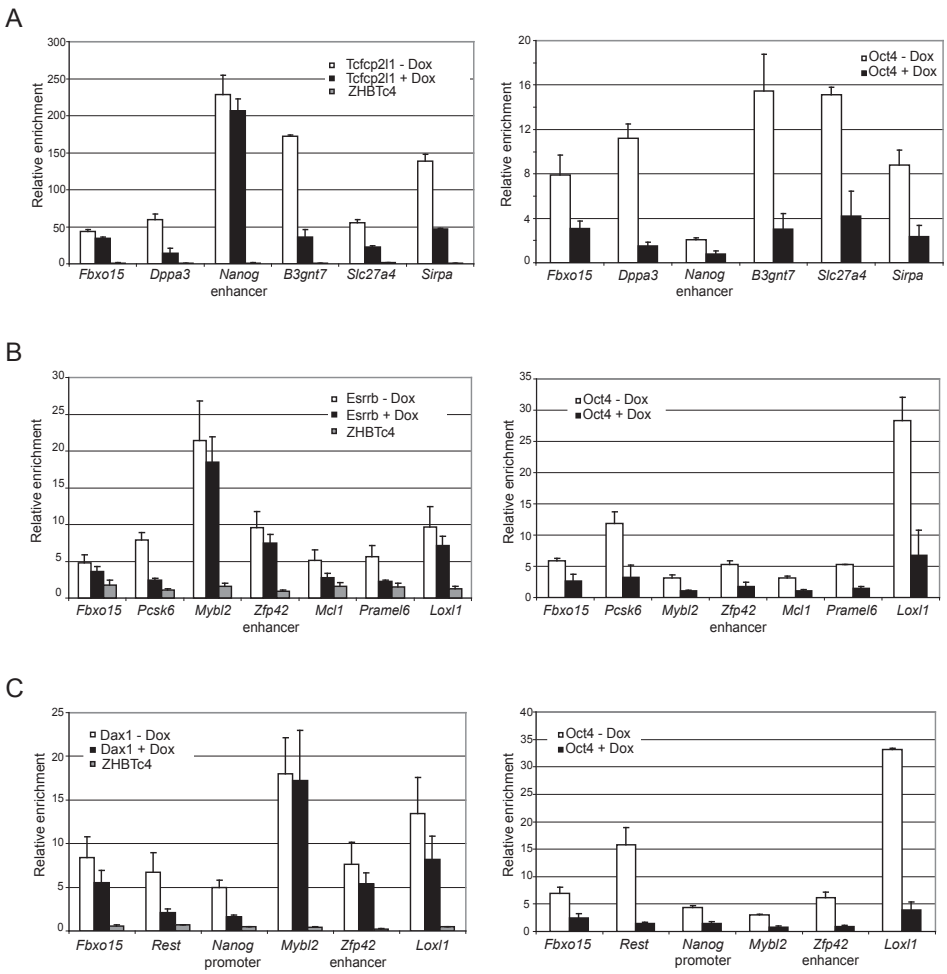


Figure 4. Oct4-dependent genome targeting by Dax1, Tcfcp2l1 and Esrrb
(A-C) Left panels indicate genome binding by F-Tcfcp2l1 (A), F-Esrrb (B) and V5-Dax1 (C) at the indicated genomic regions in the absence (-Dox) or presence (+Dox) of doxycycline, as assessed by ChIP against FLAG (F-Tcfcp2l1 and F-Esrrb) or V5 (V5-Dax1) in ZHBTc4 ES cells stably expressing these tagged proteins. The ZHBTc4 parental cell line functions as a specificity control (ZHBTc4). Right panels indicate Oct4 genome binding, as assessed by Oct4 antibody ChIP, on the same regions and in the same ES cells as the corresponding left panels. Note that the addition of doxycycline diminishes expression and thereby genome binding by Oct4. Graphs show the enrichment over a control region (*Amylase*). SEM is indicated by error bars.

DISCUSSION

Improved methodology to identify interaction networks in ES cells

We have improved the FLAG-affinity-based protein purification procedure by using near-physiological buffer conditions and very low detergent levels, which is possible due to our use of low-adherence plastic tubes. Previous approaches to identify interacting proteins of stem cell transcription factors used higher concentrations of detergent[21, 25] and salt[25], which can cause the

loss of bona-fide but weak protein-protein interactions. Non-specific elution from beads[21, 25] is likely to increase background, thereby reducing the detection sensitivity and further decreasing the number of identified specific interactors.

In support of the improved sensitivity and specificity of our procedure, we identified over 50 F-Oct4 interacting proteins by mass spectrometry (Table 1). Our increased sensitivity detected the efficient association of Oct4 with all components of NuRD. The previously claimed existence of a NuRD sub-complex with Oct4 may therefore have been the result of a limited detection efficiency[21]. We subsequently applied our protocol to purify four Oct4-interacting factors, Sall4, Tcfcp2l1, Dax1 and Esrrb and identify their associated proteins. The combined identified interactions of the five purified factors resulted in a dense interaction network that contains over 160 proteins. In a previous study, 35 proteins were identified in a Nanog-centered interaction network, resulting from six purified factors[25]. Proteins identified in the Nanog purifications included Oct4, Dax1, Zfp281 and Nac1, but in the reverse experiment, Nanog was not identified by mass spectrometry analyses of Oct4, Dax1, Zfp281 and Nac1 purifications[25]. We did not identify Nanog either in our purifications of Oct4 and Dax1. Nanog may be hard to detect by mass spectrometry, possibly due to a relative resistance to digestion into tryptic peptides.

The increased sensitivity of our procedure does not appear to come at the cost of a higher false positive rate. Three-quarter of the identified F-Oct4 interactors were also present in an endogenous Oct4 immunoprecipitation, providing a strong validation of our methodology. Further evidence of the reliability of our procedure is the reverse identification of Oct4 in all the samples of the purified transcription factors. Moreover, we indepently verified 23 interactions, of which several in two directions, by immunoprecipitations and GST-pulldowns combined with western blotting.

Multiple network connections with chromatin and protein modifying factors

Our interaction network shows the efficient association of the purified transcription factors with several chromatin remodeling complexes previously reported to be important for ES cell self renewal (Table 2). Genome-wide analyses of binding sites in mouse ES cells have been reported for SWI/SNF[38-39] and PRC1[40-41]. The SWI/SNF complex binds broadly to several kilobases around the start site of many genes expressed in ES cells, including Oct4 target genes[38-39]. PRC1 also covers several kilobases around promoters enriched for both H3K27me3 and H3K4me3 and shows overlapping binding with Oct4[40, 42]. ES cell transcription factors such as Oct4, Sox2, Nanog, Esrrb and Tcfcp2l1 often cluster more closely together[10-11]. This suggests that transcription factors may not be necessary for the continual targeting of these CMCs but recruitment may occur by initial local targeting followed by chromatin modification, thereby creating the appropriate binding surface that facilitates further spreading. CMCs often contain subunits with domains that recognize specific histone modifications[43] and are therefore well equipped to bind specific promoter chromatin environments. Dependence both on histone marks and transcription factors would allow for multiple mechanisms of fine-tuning CMC recruitment.

Oct4, Esrrb, Tcfcp2l1 and Dax1 all bind the glycosylating enzyme Ogt, which adds O-GlcNAc groups to proteins. Recently, human Oct4 was shown to be modified by O-GlcNAc[44]. O-GlcNAc modification can regulate the activity of many transcription factors[45]. Modification of Mll5 by Ogt was shown to be required for its histone H3K4 methylation activity and induction of granulocytic differentiation in HL60 cells[46]. The association of Ogt with multiple ES cell transcription factors suggests that the O-GlcNac modification may also regulate ES cell transcriptional networks.

Sall4, Tcfcp2l1 and Esrrb have unique sets of interacting proteins

Some of the purified factors have extensive sets of interacting proteins that were not observed in other purifications. For example, spalt protein Sall4 is linked to TGFbeta and Wnt signaling through association with Usp9x and Cxxc5, respectively. In Drosophila wings, *spalt* genes are regulated by TFGbeta signaling and disruption of TGFbeta signaling phenocopies the effect of *spalt* mutations on wing patterning[47]. The Sall4-Usp9x association shows that spalt proteins are also connected to the TGFbeta pathway by physical interaction. Tcfcp2l1 associates with several factors involved in DNA replication, DNA repair or cell-cycle regulation, suggests that Tcfcp2l1 may link these pathways in ES cells. Tcfcp2l1 knock-down affected cell growth but no effect on self-renewal was reported[28]. This may suggest that Tcfcp2l1 regulates cell-cycle progression in ES cells and senses input from DNA replication and repair processes. Consistent with a role of Tcfcp2l1 in cell cycle regulation, Tcfcp2l1 was shown to co-localize on many promoters with transcription factor E2f1[11], a cell cycle regulator that binds and regulates many DNA replication and DNA repair genes[48].

An intriguing interaction is that of Esrrb with basal transcription machinery complexes Mediator, TFIID and RNApol2, as well as with the TRX/Mll chromatin modifying complex and Ncoa3. Mediator, TRX/Mll and Ncoa3 also bind to the ligand-binding domain of the estrogen receptor, which is related to Esrrb, and are essential co-factors for estrogen receptor-dependent transcriptional activation in mammary cells[49-51]. To date it is unknown how ES cell transcription factor binding at promoters leads to the recruitment of the basal transcription machinery to activate transcription. By analogy to estrogen receptor in mammary cells, Esrrb may provide for such a function in ES cells.

Interactions between ES cell transcription factors

Our purifications identified a number of transcription factors as interaction hubs, as they interacted with many of the other transcription factors in the network. Examples of such hubs are Zfp143 and Klf5 but also the purified factors Oct4, Esrrb, Sall4, Dax1 and Tcfcp2l1 (Figure 3). Esrrb, Tcfcp2l1 and Dax1 were shown to cluster across the genome to distinct sets of Oct4 binding sites, suggesting the possibility of cooperativity. We indeed found that all three factors depend on Oct4 for efficient targeting of several of their shared binding sites with Oct4. This suggests that Oct4 DNA binding in some cases provides an anchor that, by physical interaction, facilitates the binding of other transcription factors. A paradigm for such a recruitment mechanism could be the proximal promoter of the *Nanog* gene, which contains an Oct-Sox motif 170 basepairs upstream from the transcription start site. Oct4 and Sox2 were shown to regulate *Nanog* expression by synergistic binding to this

motif[52-53]. Using ChIP and EMSA analysis, we have recently shown that the function of the *Nanog* proximal promoter depends on the co-operative interaction between Oct4 and Esrrb[17]. Here we show that Dax1 depends on Oct4 for its binding to the *Nanog* proximal promoter. Nac1 also binds to the *Nanog* proximal promoter [10], while binding of interaction hubs Klf5 and Zfp143 to sequences in the *Nanog* proximal promoter regulate its activity[54-55]. In summary, at least six Oct4-associated proteins (Sox2, Esrrb, Dax1, Nac1, Klf5, Zfp143) bind the *Nanog* proximal promoter, of which at least three do so in an Oct4-dependent manner (Sox2, Esrrb and Dax1). Such a strong correlation could be a coincidence, but may also reflect a scenario in which multiple transcription factors bind in close proximity, depending both on DNA sequence recognition and protein-protein interactions and together ensure the appropriate *Nanog* expression level. Interestingly, a predicted consensus motif for common target genes of two sets of ES cell transcription factors, including Oct4, Sox2, Dax1, Klf4, Nac1, Esrrb and Nanog, was found to be almost identical to the Oct4-Sox2 binding site[10-11]. This suggests that a recruitment mechanism dependent on DNA sequence and protein-protein interaction, as we propose here for the *Nanog* promoter, may have many equivalents in the ES cell genome.

EXPERIMENTAL PROCEDURES

Cell culture and DNA constructs

Mouse ES cell lines were grown on gelatin-coated dishes without feeders, as described previously[17]. The coding sequences of Sall4, Dax1, Tcfcp2l1 and Esrrb were amplified from mouse ES cell cDNA and inserted with an N-terminal double FLAG-tag (Sall4, Dax1, Esrrb), C-terminal double FLAG-tag (Tcfcp2l1) or N-terminal V5-tag (Dax1) into a pPyCAG-driven expression vector. ZHBTc4 ES cells[3] were transfected using Lipofectamine 2000 (Invitrogen), clones were selected by 1 µg/ml puromycin, and expression of the tagged proteins in selected clones was tested by western blot analysis with anti-FLAG (Sigma) and anti-V5 antibodies (Invitrogen). For transcription factor purifications from ES cells in the absence of Oct4, 1 µg/ml doxycycline (Sigma) was added for 16 hours before processing.

Protein purifications

FLAG-tagged transcription factor containing ZHBTc4 cells and control ZHBTc4 cells were expanded to five 14 cm diameter dishes, washed with PBS, scraped off, nuclear extracts prepared[56] and dialyzed to buffer C-100 (20 mM Hepes pH 7.6, 0.2 mM EDTA, 1.5 mM $MgCl_2$, 100 mM KCl, 20% glycerol). 60 µl of anti-FLAG M2 agarose beads (Sigma) equilibrated in buffer C-100 were added to 1.5ml of nuclear extract in No Stick microcentrifuge tubes (Alpha Laboratories) and incubated for 3 hours at 4ºC in the presence of 225 units of Benzonase (Novagen). Beads were washed five times for 5 minutes with buffer C-100 containing 0.02% NP-40 (C-100*) and bound proteins eluted four times for 15 min at 4ºC with buffer C-100* containing 0.2 mg/ml FLAG-tripeptide (Sigma). Elutions were pooled, TCA precipitated, proteins separated by polacrylamide gel electrophoresis stained

with the sensitive Colloidal Blue Staining Kit (Invitrogen) and analyzed by mass spectrometry (see supplementary experimental procedures). For immunoprecipitation of endogenous Oct4 complexes, 10 µg of anti-Oct3/4 antibody (sc-8628, Santa Cruz) or goat IgG (Santa Cruz) was cross-linked to 50 µl protein G Sepharose beads (Amersham). Antibody-beads, equilibrated in C-100* and blocked with 0.1 mg/ml insulin (Sigma), 0.2 mg/ml chicken egg albumin (Sigma), 1% fish skin gelatin (Sigma) were added to 1 ml of nuclear extracts made from 46C ES cells [18] containing Benzonase for 3 hours at 4ºC in No Stick microcentrifuge tubes, washed five times for 5 min with C-100* at 4ºC and boiled in SDS-loading dye. For smaller scale immunoprecipitations, 20 µl beads and 200 µl extract was used. The following antibodies were used: anti-Mi2beta, anti-Mbd3 (kind gifts from Paul Wade), anti-Mta2 (8106, Abcam), anti-Mta1 (sc-9445, Santa Cruz), anti-Sall4 (a gift of Matthias Treir)**,** anti-Lsd1 (ab17721, Abcam), anti-Med1 (sc-8998, Santa Cruz), anti-RNA polymerase II (largest subunit, sc-899, Santa Cruz).

GST pull down

The GST-fusion expression constructs were created by inserting mEsrrb, mDax1 or mTcfcp2l1 cDNA into pGEX-2TK. GST-fusions and GST were expressed in BL21 LysS bacteria (Invitrogen). Cells were lysed in bacterial lysis buffer (25mM Hepes pH 7.6, 5mM $MgCl_2$, 150mM NaCl, 10% glycerol, 0.1% NP-40, 50 µM $ZnCl_2$, protease inhibitors), sonicated and GST fusion proteins were bound to glutathione-sepharose beads (GE Healthcare), equilibrated in C-100* and incubated with 46C nuclear extract in No Stick tubes for 2 hrs at 4°C in the presence of Benzonase. Bound proteins were analyzed by Western blotting.

Chromatin immunoprecipitation

For ChIPs in the absence of Oct4, doxycycline was added to the cells for 12 hours before processing. $5*10^7$ ES cells were used per chromatin immunoprecipitation. Anti-Oct4 and anti-V5 ChIPs were performed on dual-crosslinked chromatin, as previously described [17]. For anti-FLAG ChIP chromatin was cross-linked for 10 min at RT with 0.4% formaldehyde. Cross-linking reactions were stopped by addition of 0.125M glycine. ChIPs were carried out according to the online Millipore protocol; anti-FLAG and anti-V5 beads (Sigma) were pre-blocked with 0.5 mg/ml BSA, 0.2 mg/ml salmon sperm DNA for 3 hours at 4°C. PCR-amplified genomic regions are in Supplementary Experimental Procedures.

Protein interaction network criteria and references

Criteria for inclusion as Oct4-interacting protein in Table 1 are present in 3 out of 4 experiments (3 F-Oct4 purifications and endogenous Oct4 immunoprecipitation) with a Mascot score higher than 50 and at least three fold higher than the corresponding control experiment. Criteria for inclusion in Tables S2-S9 are: Present in both tagged transcription factor purifications (-Dox) with a Mascot score higher than 50 and three fold higher than the corresponding control experiment. In case of protein identifications with mascot score values between 50 and 60 or protein identifications based on one peptide, individual peptide MS/MS spectra were checked manually and either interpreted

as valid identifications or discarded. Cytoskeletal and cytoplasmic proteins were removed from the data set. Transcription factor status and subunit composition of the complexes were assigned according to the Uniprot database. Correlation between transcription factor occupancy[10-11] was scored as positive when > 0.2. Promoter occupancy by Klf5 was assigned as overlapping with Klf4, as shown[57]. Genes bound by Oct4 were assigned according to the detection of Oct4 at their promoter[10] or ChIP sequencing data showing an association score > 0.3[11]. Microarray data on genes regulated by Oct4 (Table 2) are from Sharov *et al.*, Figure S4[24]. Genes were scored as regulated by Oct4 if they showed at least 1.5-fold up or down regulation within 48 hrs after shutdown of Oct4 transcription by addition of doxycycline to ES cell line ZHBTc4 and 2-fold difference within the time-course of the experiment (5 days).

ACKNOWLEDGEMENTS

We thank Erik Engelen for technical advice, Rooda Abdillahi Ibrahim for technical assistance, Nicola Festuccia for Dax1 cDNA, Drs. Paul Wade and Matthias Treir for antibodies, Dr. Austin Smith for 46C ES cells and Hitoshi Niwa for ZHBTc4 ES cells. This work was supported by an NWO Vidi grant to the R.P. lab and by the Wellcome Trust, the Juvenile Diabetes Research Foundation and the Medical Research Council of the UK.

REFERENCES

1. Smith, A.G., *Embryo-derived stem cells: of mice and men.* Annu Rev Cell Dev Biol, 2001. **17**: p. 435-62.
2. Niwa, H., *How is pluripotency determined and maintained?* Development, 2007. **134**(4): p. 635-46.
3. Niwa, H., J. Miyazaki, and A.G. Smith, *Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells.* Nat Genet, 2000. **24**(4): p. 372-6.
4. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.
5. Okita, K., T. Ichisaka, and S. Yamanaka, *Generation of germline-competent induced pluripotent stem cells.* Nature, 2007. **448**(7151): p. 313-7.
6. Wernig, M., A. Meissner, R. Foreman, T. Brambrink, M. Ku, K. Hochedlinger, B.E. Bernstein, and R. Jaenisch, *In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state.* Nature, 2007. **448**(7151): p. 318-24.
7. Yamanaka, S., *A fresh look at iPS cells.* Cell, 2009. **137**(1): p. 13-7.
8. Hochedlinger, K. and K. Plath, *Epigenetic reprogramming and induced pluripotency.* Development, 2009. **136**(4): p. 509-23.
9. Nakagawa, M., M. Koyanagi, K. Tanabe, K. Takahashi, T. Ichisaka, T. Aoi, K. Okita, Y. Mochiduki, N. Takizawa, and S. Yamanaka, *Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts.* Nat Biotechnol, 2008. **26**(1): p. 101-6.
10. Kim, J., J. Chu, X. Shen, J. Wang, and S.H. Orkin, *An extended transcriptional network for pluripotency of embryonic stem cells.* Cell, 2008. **132**(6): p. 1049-61.
11. Chen, X., H. Xu, P. Yuan, F. Fang, M. Huss, V.B. Vega, E. Wong, Y.L. Orlov, W. Zhang, J. Jiang, Y.H. Loh, H.C. Yeo, Z.X. Yeo, V. Narang, K.R. Govindarajan*, et al.*, *Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.* Cell, 2008. **133**(6): p. 1106-17.
12. Sridharan, R., J. Tchieu, M.J. Mason, R. Yachechko, E. Kuoy, S. Horvath, Q. Zhou, and K. Plath, *Role of the murine reprogramming factors in the induction of pluripotency.* Cell, 2009. **136**(2): p. 364-77.
13. Chambers, I. and S.R. Tomlinson, *The transcriptional foundation of pluripotency.* Development, 2009. **136**(14): p. 2311-22.
14. Ambrosetti, D.C., C. Basilico, and L. Dailey, *Synergistic activation of the fibroblast growth factor 4*

enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. Mol Cell Biol, 1997. **17**(11): p. 6321-9.
15. Wissmuller, S., T. Kosian, M. Wolf, M. Finzsch, and M. Wegner, *The high-mobility-group domain of Sox proteins interacts with DNA-binding domains of many transcription factors.* Nucleic Acids Res, 2006. **34**(6): p. 1735-44.
16. Remenyi, A., K. Lins, L.J. Nissen, R. Reinbold, H.R. Scholer, and M. Wilmanns, *Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers.* Genes Dev, 2003. **17**(16): p. 2048-59.
17. van den Berg, D.L., W. Zhang, A. Yates, E. Engelen, K. Takacs, K. Bezstarosti, J. Demmers, I. Chambers, and R.A. Poot, *Estrogen-related receptor beta interacts with Oct4 to positively regulate Nanog gene expression.* Mol Cell Biol, 2008. **28**(19): p. 5986-95.
18. Ying, Q.L., M. Stavridis, D. Griffiths, M. Li, and A. Smith, *Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture.* Nat Biotechnol, 2003. **21**(2): p. 183-6.
19. Ishihama, Y., Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, and M. Mann, *Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein.* Mol Cell Proteomics, 2005. **4**(9): p. 1265-72.
20. Williams, D.C., Jr., M. Cai, and G.M. Clore, *Molecular basis for synergistic transcriptional activation by Oct1 and Sox2 revealed from the solution structure of the 42-kDa Oct1.Sox2.Hoxb1-DNA ternary transcription factor complex.* J Biol Chem, 2004. **279**(2): p. 1449-57.
21. Liang, J., M. Wan, Y. Zhang, P. Gu, H. Xin, S.Y. Jung, J. Qin, J. Wong, A.J. Cooney, D. Liu, and Z. Songyang, *Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells.* Nat Cell Biol, 2008. **10**(6): p. 731-9.
22. Zhang, Y., H.H. Ng, H. Erdjument-Bromage, P. Tempst, A. Bird, and D. Reinberg, *Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation.* Genes Dev, 1999. **13**(15): p. 1924-35.
23. Levasseur, D.N., J. Wang, M.O. Dorschner, J.A. Stamatoyannopoulos, and S.H. Orkin, *Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells.* Genes Dev, 2008. **22**(5): p. 575-80.
24. Sharov, A.A., S. Masui, L.V. Sharova, Y. Piao, K. Aiba, R. Matoba, L. Xin, H. Niwa, and M.S. Ko, *Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data.* BMC Genomics, 2008. **9**: p. 269.
25. Wang, J., S. Rao, J. Chu, X. Shen, D.N. Levasseur, T.W. Theunissen, and S.H. Orkin, *A protein interaction network for pluripotency of embryonic stem cells.* Nature, 2006. **444**(7117): p. 364-8.
26. Zhang, J., W.L. Tam, G.Q. Tong, Q. Wu, H.Y. Chan, B.S. Soh, Y. Lou, J. Yang, Y. Ma, L. Chai, H.H. Ng, T. Lufkin, P. Robson, and B. Lim, *Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1.* Nat Cell Biol, 2006. **8**(10): p. 1114-23.
27. Yuri, S., S. Fujimura, K. Nimura, N. Takeda, Y. Toyooka, Y. Fujimura, H. Aburatani, K. Ura, H. Koseki, H. Niwa, and R. Nishinakamura, *Sall4 is essential for stabilization, but not for pluripotency, of embryonic stem cells by repressing aberrant trophectoderm gene expression.* Stem Cells, 2009. **27**(4): p. 796-805.
28. Ivanova, N., R. Dobrin, R. Lu, I. Kotenko, J. Levorse, C. DeCoste, X. Schafer, Y. Lun, and I.R. Lemischka, *Dissecting self-renewal in stem cells with RNA interference.* Nature, 2006. **442**(7102): p. 533-8.
29. Loh, Y.H., Q. Wu, J.L. Chew, V.B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K.Y. Wong, K.W. Sung, C.W. Lee, X.D. Zhao, K.P. Chiu*, et al.*, *The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells.* Nat Genet, 2006. **38**(4): p. 431-40.
30. Zhang, X., J. Zhang, T. Wang, M.A. Esteban, and D. Pei, *Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells.* J Biol Chem, 2008. **283**(51): p. 35825-33.
31. Feng, B., J. Jiang, P. Kraus, J.H. Ng, J.C. Heng, Y.S. Chan, L.P. Yaw, W. Zhang, Y.H. Loh, J. Han, V.B. Vega, V. Cacheux-Rataboul, B. Lim, T. Lufkin, and H.H. Ng, *Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb.* Nat Cell Biol, 2009. **11**(2): p. 197-203.
32. Niakan, K.K., E.C. Davis, R.C. Clipsham, M. Jiang, D.B. Dehart, K.K. Sulik, and E.R. McCabe, *Novel role for the orphan nuclear receptor Dax1 in embryogenesis, different from steroidogenesis.* Mol Genet Metab, 2006. **88**(3): p. 261-71.
33. Wilanowski, T., A. Tuckfield, L. Cerruti, S. O'Connell, R. Saint, V. Parekh, J. Tao, J.M. Cunningham, and S.M. Jane, *A highly conserved novel family of mammalian developmental transcription factors related to Drosophila grainyhead.* Mech Dev, 2002. **114**(1-2): p. 37-50.
34. Sikorski, T.W. and S. Buratowski, *The basal initiation machinery: beyond the general transcription factors.* Curr Opin Cell Biol, 2009. **21**(3): p. 344-51.
35. Bray, S.J., *Notch signalling: a simple pathway becomes complex.* Nat Rev Mol Cell Biol, 2006. **7**(9): p. 678-89.

36. Dupont, S., A. Mamidi, M. Cordenonsi, M. Montagner, L. Zacchigna, M. Adorno, G. Martello, M.J. Stinchfield, S. Soligo, L. Morsut, M. Inui, S. Moro, N. Modena, F. Argenton, S.J. Newfeld*, et al.*, *FAM/USP9x, a deubiquitinating enzyme essential for TGFbeta signaling, controls Smad4 monoubiquitination.* Cell, 2009. **136**(1): p. 123-35.
37. Andersson, T., E. Sodersten, J.K. Duckworth, A. Cascante, N. Fritz, P. Sacchetti, I. Cervenka, V. Bryja, and O. Hermanson, *CXXC5 is a novel BMP4-regulated modulator of Wnt signaling in neural stem cells.* J Biol Chem, 2009. **284**(6): p. 3672-81.
38. Kidder, B.L., S. Palmer, and J.G. Knott, *SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells.* Stem Cells, 2009. **27**(2): p. 317-28.
39. Ho, L., R. Jothi, J.L. Ronan, K. Cui, K. Zhao, and G.R. Crabtree, *An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network.* Proc Natl Acad Sci U S A, 2009. **106**(13): p. 5187-91.
40. Boyer, L.A., K. Plath, J. Zeitlinger, T. Brambrink, L.A. Medeiros, T.I. Lee, S.S. Levine, M. Wernig, A. Tajonar, M.K. Ray, G.W. Bell, A.P. Otte, M. Vidal, D.K. Gifford, R.A. Young*, et al.*, *Polycomb complexes repress developmental regulators in murine embryonic stem cells.* Nature, 2006. **441**(7091): p. 349-53.
41. Ku, M., R.P. Koche, E. Rheinbay, E.M. Mendenhall, M. Endoh, T.S. Mikkelsen, A. Presser, C. Nusbaum, X. Xie, A.S. Chi, M. Adli, S. Kasif, L.M. Ptaszek, C.A. Cowan, E.S. Lander*, et al.*, *Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains.* PLoS Genet, 2008. **4**(10): p. e1000242.
42. Endoh, M., T.A. Endo, T. Endoh, Y. Fujimura, O. Ohara, T. Toyoda, A.P. Otte, M. Okano, N. Brockdorff, M. Vidal, and H. Koseki, *Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity.* Development, 2008. **135**(8): p. 1513-24.
43. Taverna, S.D., H. Li, A.J. Ruthenburg, C.D. Allis, and D.J. Patel, *How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers.* Nat Struct Mol Biol, 2007. **14**(11): p. 1025-40.
44. Webster, D.M., C.F. Teo, Y. Sun, D. Wloga, S. Gay, K.D. Klonowski, L. Wells, and S.T. Dougan, *O-GlcNAc modifications regulate cell survival and epiboly during zebrafish development.* BMC Dev Biol, 2009. **9**(1): p. 28.
45. Issad, T. and M. Kuo, *O-GlcNAc modification of transcription factors, glucose sensing and glucotoxicity.* Trends Endocrinol Metab, 2008. **19**(10): p. 380-9.
46. Fujiki, R., T. Chikanishi, W. Hashiba, H. Ito, I. Takada, R.G. Roeder, H. Kitagawa, and S. Kato, *GlcNAcylation of a histone methyltransferase in retinoic-acid-induced granulopoiesis.* Nature, 2009. **459**(7245): p. 455-9.
47. de Celis, J.F., R. Barrio, and F.C. Kafatos, *A gene complex acting downstream of dpp in Drosophila wing morphogenesis.* Nature, 1996. **381**(6581): p. 421-4.
48. Ren, B., H. Cam, Y. Takahashi, T. Volkert, J. Terragni, R.A. Young, and B.D. Dynlacht, *E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints.* Genes Dev, 2002. **16**(2): p. 245-56.
49. Mo, R., S.M. Rao, and Y.J. Zhu, *Identification of the MLL2 complex as a coactivator for estrogen receptor alpha.* J Biol Chem, 2006. **281**(23): p. 15714-20.
50. Shang, Y., X. Hu, J. DiRenzo, M.A. Lazar, and M. Brown, *Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription.* Cell, 2000. **103**(6): p. 843-52.
51. Kang, Y.K., M. Guermah, C.X. Yuan, and R.G. Roeder, *The TRAP/Mediator coactivator complex interacts directly with estrogen receptors alpha and beta through the TRAP220 subunit and directly enhances estrogen receptor function in vitro.* Proc Natl Acad Sci U S A, 2002. **99**(5): p. 2642-7.
52. Rodda, D.J., J.L. Chew, L.H. Lim, Y.H. Loh, B. Wang, H.H. Ng, and P. Robson, *Transcriptional regulation of nanog by OCT4 and SOX2.* J Biol Chem, 2005. **280**(26): p. 24731-7.
53. Kuroda, T., M. Tada, H. Kubota, H. Kimura, S.Y. Hatano, H. Suemori, N. Nakatsuji, and T. Tada, *Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression.* Mol Cell Biol, 2005. **25**(6): p. 2475-85.
54. Chen, X., F. Fang, Y.C. Liou, and H.H. Ng, *Zfp143 regulates Nanog through modulation of Oct4 binding.* Stem Cells, 2008. **26**(11): p. 2759-67.
55. Parisi, S., F. Passaro, L. Aloia, I. Manabe, R. Nagai, L. Pastore, and T. Russo, *Klf5 is involved in self-renewal of mouse embryonic stem cells.* J Cell Sci, 2008. **121**(Pt 16): p. 2629-34.
56. Dignam, J.D., R.M. Lebovitz, and R.G. Roeder, *Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei.* Nucleic Acids Res., 1983. **11**(5): p. 1475-89.
57. Jiang, J., Y.S. Chan, Y.H. Loh, J. Cai, G.Q. Tong, C.A. Lim, P. Robson, S. Zhong, and H.H. Ng, *A core Klf circuitry regulates self-renewal of embryonic stem cells.* Nat Cell Biol, 2008. **10**(3): p. 353-60.

SUPPLEMENTAL FIGURES

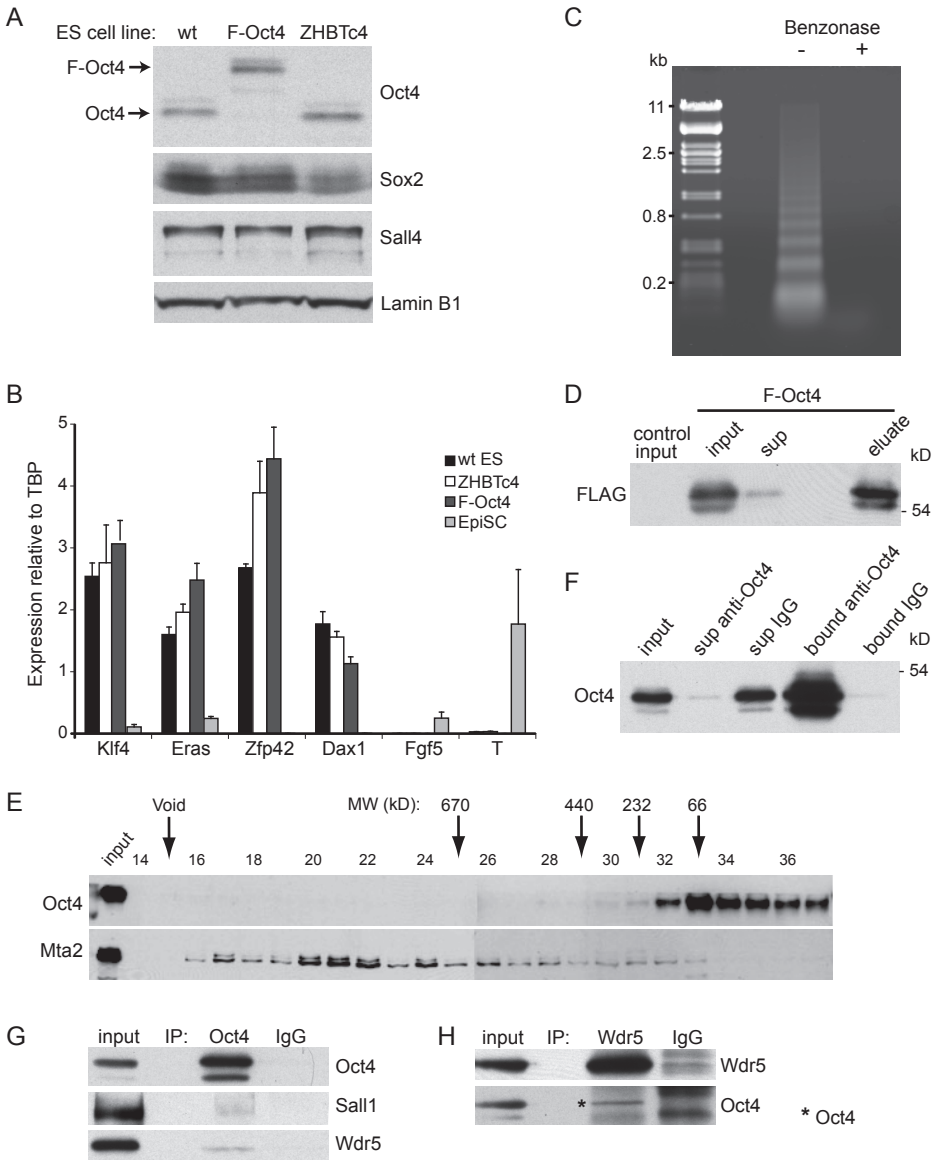


Figure S1. Additional data on Oct4 purifications

(A) Expression levels of ES cell markers in F-Oct4 and ZHBTc4 ES cells. Extracts from wild type (wt), F-Oct4 and ZHBTc4 ES cells were probed by western with the indicated antibodies. Lamin B1 served as a loading control. (B) Quantitative RT-PCR analysis of indicated transcript levels in wild type (wt), ZHBTc4, F-Oct4 ES cells and epiblast stem cells. (C) Treatment of ES cell nuclear extract with Benzonase removes DNA.DNA purified from 100 µl nuclear extract treated with 15 units Benzonase for 3 hrs at 4°C (+ Benzonase), or not treated (- Benzonase). Size markers are indicated. (D) F-Oct4 is depleted from nuclear extract by anti-FLAG purification, as shown by anti-FLAG western. Input, supernatant after purification (sup), eluate and control input from ZHBTc4 ES cells (not expressing F-Oct4) are indicated. (E) Gel filtration analysis of Oct4 and NuRD subunit Mta2. Mouse ES cell nuclear extract was size-fractionated on a Superose-6 gel filtration column. Fractions were resolved on an SDS-polyacrylamide gel and probed with the indicated antibodies. Molecular weights of gel filtration standards are indicated. (F) Oct4 is depleted from nuclear extract by anti-Oct4 immunoprecipitation. Input, supernatant (sup) and bound fraction are indicated. (G) Verification of Oct4 interactions. Oct4 immunoprecipitates analyzed by western blots with the indicated antibodies. (H) Wdr5 immunoprecipitates analyzed by western blots with the indicated antibodies, * indicates Oct4 band.

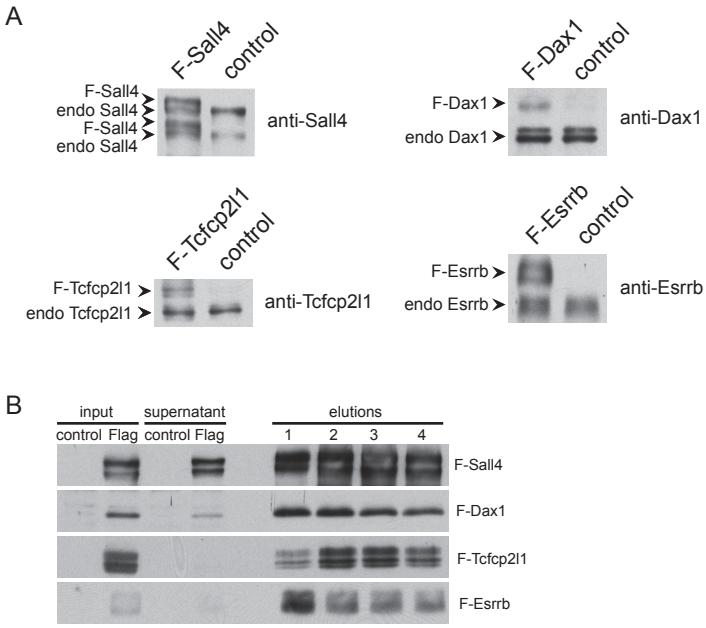


Figure S2. Additional data on the purifications of the FLAG-tagged transcription factors
(A) Expression levels of FLAG-tagged transcription factors and their endogenous counterparts in the used ES cell clones. Extracts from the ES cell clones used for FLAG-IP and FLAG-ChIP experiments were compared to control extracts from the parental ZHBTc4 ES cell line by western with the indicated antibodies. FLAG-tagged transcription factors and endogenous (endo) transcription factors are indicated by arrows. (B) anti-FLAG western blot analysis of the purifications of FLAG-tagged Sall4 (F-Sall4), Dax1 (F-Dax1), Tcfcp2l1 (F-Tcfcp2l1) and Esrrb (F-Esrrb) from ES cell clones expressing these tagged proteins. Nuclear extracts before purification (input) and after purification (supernatant) are shown, as well as the elution fractions from anti-FLAG beads. Control extract is from the parental ZHBTc4 ES cells.

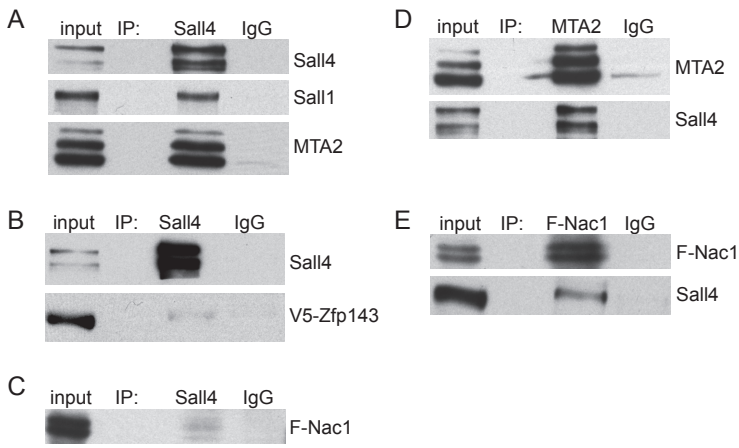


Figure S3. Verification of interactions of Sall4, Dax1, Tcfcp2l1 and Esrrb by GST-pull down and immunoprecipitation

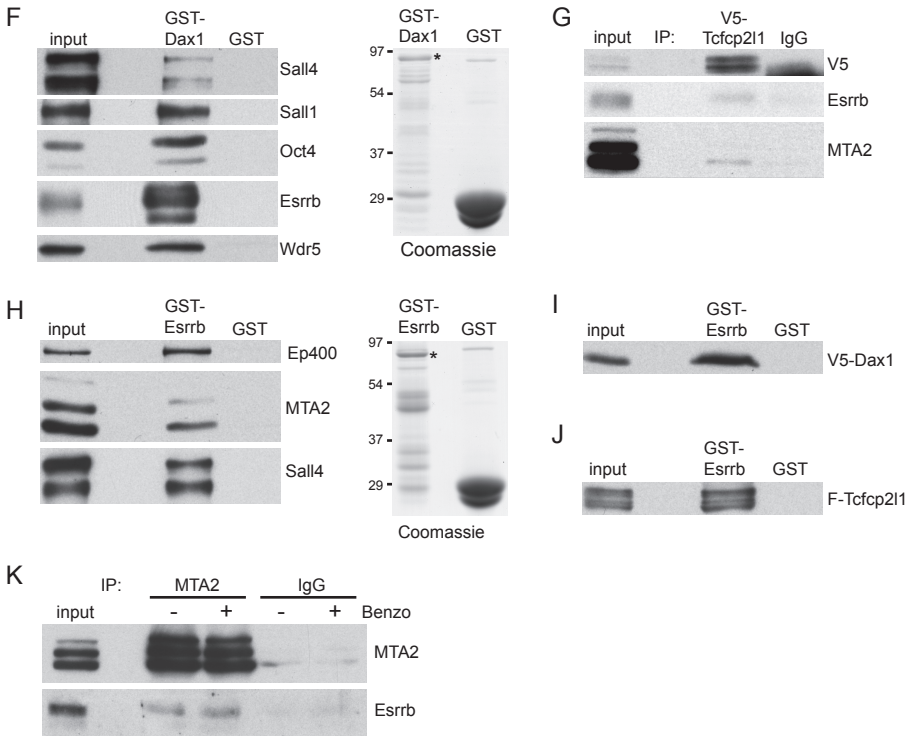


Figure S3. Verification of interactions of Sall4, Dax1, Tcfcp2l1 and Esrrb by GST-pull down and immunoprecipitation
(A) Sall4 immunoprecipitates analyzed by western blots with the indicated antibodies. (B) Sall4 immunoprecipitates from extracts of V5-Zfp143 transfected ES cells analyzed by western blots with Sall4 and V5 antibodies. (C) Sall4 immunoprecipitates from extracts of F-Nac1 transfected ES cells analyzed by western blot with Flag antibody. (D) MTA2 immunoprecipitates analyzed by western blots with the indicated antibodies. (E) F-Nac1 immunoprecipitates from extracts of F-Nac1 transfected ES cells analyzed by western blots with Flag and Sall4 antibodies. (F, left panel) GST-Dax1 precipitates analyzed by western blots with the indicated antibodies. (F, right panel) GST-Dax1 (*) and GST bound to beads, as used in the GST pull down, analyzed by Coomassie stained PAA gel. (G) V5-Tcfcp2l1 immunoprecipitates analyzed by western blots with the indicated antibodies. (H, left panel) GST-Esrrb precipitates analyzed by western blots with the indicated antibodies. (H, right panel) GST-Esrrb (*) and GST bound to beads, as used in the GST pull down, analyzed by Coomassie stained PAA gel. (I) GST-Esrrb precipitates analyzed by western blots with V5 antibody. (J) GST-Esrrb precipitates analyzed by western blots with Flag antibody. (K) MTA2 immunoprecipitates analyzed by western blots with the indicated antibodies. Benzonase was added where indicated.

SUPPLEMENTAL TABLES

Table S1. Oct4-interacting proteins: emPAI scores

Protein	Accession	Flag#1 emPAI [a]	Flag#2 emPAI [a]	Flag#3 emPAI [a]	Oct4-IP emPAI [a]	average emPAI [a]
Oct4 (Pou5f1)	gi|200118	1.36	7.03(0.22)	12.38(0.32)	2.63	5.85
NuRD complex						
Mi2beta (Chd4)	gi|39204553	0.68(0.02)	1.38	1.57(0.12)	1.22(0.06)	1.21
Mta1	gi|15077051	0.72	1.97	2.27	0.98	1.49
Gatad2a	gi|148696823	0.9	1.48	3.47(0.45)	0.71	1.64
Mta2	gi|51491880	0.69	1.48	2.33	0.86(0.05)	1.34
Gatad2b	gi|120577529	0.73	1.16	4.16(0.18)	0.55	1.65
Hdac1	gi|2347180	2.66(0.14)	2.43(0.07)	2.36(0.21)	1.82(0.21)	2.32
Mbd3	gi|7305261	1.39	1.41	8.24	6.26(0.12)	4.33
Mta3	gi|18381007	0.44	0.45	1.96	1.21(0.05)	1.02
Hdac2	gi|3023934	2.2(0.21)	1.47(0.07)	2.55	1.32(0.21)	1.89
Rbbp7	gi|2494892	0.82(0.18)	1.09	1.66(0.82)	1.09	1.17
SWI/SNF complex						
Baf155(Smarcc1)	gi|30851572	0.34	0.39	0.78	0.16(0.16)	0.42
Brg1 (Smarca4)	gi|76253779	0.08	0.22	0.18	0.17(0.11)	0.16
PRC1 complex						
Phc1	gi|30923312	0.13	0.19	0.51	-	0.21
Ring1B (Rnf2)	gi|109157342	1.51	0.21	1.02(0.1)	0.45(0.1)	0.80
Rybp	gi|5381327	0.15(0.15)	0.15	0.3	0.15	0.19
Trrap/p400 complex						
Trrap	gi|124486949	0.01	0.04	0.02	0.05	0.03
Ep400	gi|27348237	0.03	0.01	0.05	0.02	0.03
LSD1 complex						
Lsd1	gi|51315882	0.12	0.35	0.54	0.08	0.27
Zmym2	gi|28175571	0.03	0.2	0.11	-	0.09
Rcor2	gi|17298682	0.07	0.14	0.34	0.21	0.19
Transcription factors						
Sall4	gi|81913723	2.78(0.29)	3.03(0.09)	6.23(0.42)	2.44(0.43)	3.62
Sall1	gi|14164331	1.05	1.58	1.8	1	1.36
Zfp219	gi|30794418	0.18	0.44	0.29	0.32	0.31
Arid3b	gi|9790033	0.19	0.27	1.7	0.13	0.57
Wdr5	gi|16554627	0.62(0.17)	0.21	1.05(0.1)	0.62	0.63
Zfp462	gi|114431238	-	0.04	0.07	0.23	0.09
Mga	gi|6692607	-	0.07	0.1	0.04	0.05
Sox2	gi|127140986	0.23	-	3.29	1.07	1.15
Ubp1	gi|7305605	0.38	0.14	0.32	0.29	0.28
Nac1	gi|31543309	-	0.37	0.34	0.37	0.27
Hcfc1	gi|4098678	0.02	0.09	0.14(0.02)	-	0.06
Hells	gi|12232371	-	0.26	0.19	0.04	0.12
Rbpj	gi|94400775	1.04	0.23	0.12	0.32	0.43
Tcfcp2l1	gi|90101766	0.3	0.14	0.06	0.14	0.16
Requiem	gi|6755314	0.49	0.17	-	0.09	0.19
Esrrb	gi|124375796	0.16	0.08	0.22	0.34	0.20
Pml	gi|9506979	0.08	0.08	0.27	-	0.11
Foxp4	gi|161016782	-	0.04	0.37	0.04	0.11
Ctbp2	gi|6753548	-	0.26	0.23	0.08	0.14
Dax1	gi|6671531	0.07	0.14	0.21	0.07	0.12
Zfp143	gi|22902397	0.05	-	0.1	0.06	0.05
Klf5	gi|31981873	-	0.07	0.14	0.07	0.07
Other						
Rif1	gi|47078460	0.43	0.82	0.63	0.43(0.16)	0.58
L1td1	gi|148698953	0.12	0.23	0.36(0.28)	0.13(0.04)	0.21
Akap8	gi|31560394	0.05	0.21	0.3	0.21	0.19
Msh2	gi|30047836	0.15	0.11	0.29	0.04	0.15
Smc1a	gi|123220915	0.04	0.26	0.1	-	0.10
Ogt	gi|13775066	0.06	0.1	0.58(0.06)	-	0.19
Rbm14	gi|16307494	0.12	0.05	0.55(0.05)	0.11	0.21
Frg1	gi|17376286	0.27	1.07	0.57	0.27	0.55
Emsy	gi|124249084	0.07	0.05	0.19	-	0.08
0610010K14Rik	gi|81917220	0.53	0.53	1.48	-	0.64
2810474O19Rik	gi|148678819	0.02	0.09	0.08	-	0.05
Zcchc8	gi|148687677	0.11	0.05	0.14	-	0.08

[a]emPAI score for the specified protein in the Oct4 sample. emPAI score for the specified protein in the corresponding control purification, if present, is between brackets.

Table S2. Sall4-interacting proteins as identified by mass spectrometry analysis of purified F-Sall4 samples

		Flag#1		Flag#2		Flag#3		Flag#4			
		-Dox		+Dox		-Dox		+Dox		-Dox	+Dox
Protein	Accession	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Average mascot	
Sall4	gi|117553631	3978(70)	44(1)	3991(330)	43(4)	4530(321)	50(4)	4378(295)	48(4)	4254	4185
NuRD complex											
Chd4	gi|39204553	3716	54	4645	68	6021	80	5230	74	4867	4938
Gatad2a	gi|148696823	1896	21	1967	24	2055	23	2107	25	1976	2037
Mta2	gi|51491880	1717	22	1803	24	1930	26	1945	28	1824	1874
Mta1	gi|86577662	1423	20	1575	22	1727	25	1610	23	1575	1593
Gatad2b	gi|21314854	1412	16	1503	18	1598	18	1531	17	1505	1517
Mta3	gi|18381007	1336	16	1321	16	1479	19	1421	17	1408	1371
Mbd3	gi|7305261	1147	16	915	12	1112	13	1044	12	1130	980
Hdac1	gi|2347180	1144	14	1316	16	1586(200)	20(3)	1425(161)	18(2)	1365	1371
Hdac2	gi|87162464	1002	13	1055	14	1336	18	1109	15	1169	1082
Rbbp4	gi|5032027	888(120)	11(3)	934(62)	12(1)	1222(295)	15(5)	1311(241)	16(4)	1055	1123
Rbbp7	gi|157909799	709	11	967	13	1222	16	1156	15	966	1062
Mbd2	gi|7706609	477	7	371	5	601	8	386	5	539	379
SWI/SNF complex											
Actl6a	gi|4001805	427(128)	6(2)	606(86)	8(1)	653(211)	8(3)	628(88)	8(1)	540	617
Smarcc1	gi|30851572	239(43)	4(1)	509	10	409(81)	8(3)	304(55)	6(1)	324	407
Smarcd1	gi|576884	211	5	326	7	165	3	81	2	188	204
Arid1a	gi|124249109	158	2	129	2	92	1	-	-	125	65
Smarcb1	gi|6755578	68	1	76	1	174	2	60	1	121	68
Smarca4	gi|76253779	55	2	254	5	50	2	116	2	53	185
Transcription factors											
Sall1	gi|14164331	1978	26	1986	27	3094	38	2535	34	2536	2261
Bend3	gi|39841055	1585	23	1386	20	2339	30	1966	27	1962	1676
Zfp219	gi|47940209	730	9	543	6	677	9	294	4	704	419
Nac1	gi|81886163	417	6	423	6	988	12	556	8	703	490
Ruvbl2	gi|6755382	330(76)	6(1)	362(78)	5(1)	950(139)	12(2)	557(355)	8(6)	640	460
Wdr5	gi|16554627	329	4	230	4	683	10	645	11	506	437
Oct4	gi|53501	334	4	-	-	318	4	-	-	326	-
Grhl2	gi|46810275	109	2	105	2	427	7	293	6	268	199
Sall3	gi|49257163	521	7	-	-	580	7	-	-	551	-
Cxxc5	gi|19526854	132	2	230	3	287	4	449	5	210	340
Zfp143	gi|81908410	185	4	164	3	202	4	140	3	194	152
Tcfcp2	gi|15628025	82	2	-	-	277	3	163	2	180	82
Sall2	gi|49087134	145	2	157	2	164	2	172	3	155	165
Klf5	gi|31981873	90	1	105	1	194	2	-	-	142	53
Ctbp2	gi|2909779	123	3	62	1	99	2	98	2	111	80

Table S2. ***Continued***

		Flag#1		Flag#2		Flag#3		Flag#4			
		-Dox		+Dox		-Dox		+Dox		-Dox	+Dox
Protein	**Accession**	**Mascot[a]**	**Pept.[b]**	**Mascot[a]**	**Pept.[b]**	**Mascot[a]**	**Pept.[b]**	**Mascot[a]**	**Pept.[b]**	**Average mascot**	
Zbtb2	gi|85701993	76	2	-	-	141	2	54	1	109	27
Requiem	gi|6755314	106	1	208	2	109	1	-	-	108	104
Ewsr1	gi|88853581	129	2	132	2	82	1	-	-	106	66
Esrrb	gi|6166153	60	1	-	-	141	2	-	-	101	-
Other											
Usp9x	gi|115511018	203	2	158	2	1064	19	-	-	634	79
7420416P09Rik	gi|76253890	260	5	-	-	660	11	413	6	460	207
Set	gi|123295280	431	6	-	-	298	4	161	2	365	81
L1td1	gi|148698953	93	1	105	1	456	6	76	1	275	91

[a] Mascot score for the specified protein in the F-Sall4 sample. Mascot score for the specified protein in the corresponding control purification, if present, is between brackets.
[b] Number of identified unique, non-redundant peptides for the specified protein in the F-Sall4 sample. Number of identified unique peptides in the control purification is between brackets.

Table S3. Dax1-interacting proteins as identified by mass spectrometry analysis of purified F-Dax1 samples

		Flag#1		Flag#2		Flag#3		Flag#4			
		-Dox		+Dox		-Dox		+Dox		-Dox	+Dox
Protein	**Accession**	**Mascot[a]**	**Pept.[b]**	**Mascot[a]**	**Pept.[b]**	**Mascot[a]**	**Pept.[b]**	**Mascot[a]**	**Pept.[b]**	**Average mascot**	
Dax1 (Nr0b1)	gi|6671531	928	15	860	10	736	12	711	12	832	786
Transcription factors											
Sall4	gi|117553631	932(245)	15(5)	646(124)	9(2)	729(204)	12(4)	894(227)	16(4)	831	770
Esrrb	gi|6166153	332	5	163	5	195	6	55	2	264	109
Snw1	gi|146149191	161	3	-	-	251	5	292	5	206	146
Sall1	gi|14164331	162	2	-	-	158	4	309	7	160	155
Wdr5	gi|16554627	130	2	90	1	161	3	65	2	146	78
Oct4	gi|200118	148	2	-	-	123	2	-	-	136	-
Prmt1	gi|30185908	57	1	-	-	55	1	80	2	56	40
Other											
Rif1	gi|47078460	105	2	-	-	358	7	581	11	232	291
Ogt	gi|13775066	200	3	164	3	201	4	211	4	201	188

[a, b] Equivalent to Table S2.

Table S4. Tcfcp2l1-interacting proteins as identified by mass spectrometry analysis of purified F-Tcfcp2l1 samples

		Flag#1		Flag#2		Flag#3		Flag#4			
		-Dox		+Dox		-Dox		+Dox		-Dox	+Dox
Protein	Accession	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Average mascot	
Tcfcp2l1	gi|134053939	2187(85)	25(1)	2133(217)	25(2)	2595	28	2234	26	2391	2184
NuRD complex											
Chd4	gi|39204553	2587	37	2235	35	3469	55	2487	41	3028	2361
Gatad2a	gi|30795222	1283	17	1026	14	1302	18	900	12	1293	963
Mta1	gi|15077051	909	14	937	14	1292	19	303	5	1101	620
Mta2	gi|51491880	922	13	825	12	1155	17	225	5	1039	525
Gatad2b	gi|21314854	868	10	821	9	970	11	280	3	919	551
Hdac1	gi|2347180	918(200)	13(3)	1025(161)	14(2)	728	10	492	9	823	759
Rbbp7	gi|157909799	811	10	568	9	682	12	660	12	747	614
Hdac2	gi|3023934	868	13	742	11	554	9	391	7	711	567
Mta3	gi|18381007	850	13	470	8	480	8	114	2	665	292
Mbd3	gi|7305261	498	6	585	8	641	8	447	6	570	516
Mbd2	gi|7706609	97	1	-	-	64	1	-	-	81	-
Trrap/p400 complex											
Trrap	gi|124486949	1741	28	1952	31	2474	44	1779	30	2108	1866
Ruvbl2	gi|6755382	1062(139)	13(2)	1196(355)	15(6)	1568(76)	19(1)	1513(78)	17(1)	1315	1355
Ep400	gi|27348237	1367	17	1158	14	1051	14	883	12	1209	1021
Ruvbl1	gi|9790083	738(128)	11(2)	830(202)	12(3)	1123(257)	15(4)	732(206)	10(3)	931	781
Actl6a	gi|4001805	679(221)	9(3)	651(88)	8(1)	479(128)	6(2)	630(86)	8(1)	579	641
Yeats4	gi|13386064	317	5	328	5	392	7	357	6	355	343
Vps72	gi|83921607	387	5	85	2	134	3	169	3	261	127
Dmap1	gi|12963557	239	4	437	6	241	3	253	4	240	345
Brd8	gi|19344050	124	2	99	2	167	3	-	-	146	50
Ing3*	gi|17390799	87	1	-	-	73	1	78	1	80	39
1600027Rik	gi|110625965	68	1	-	-	192	2	177	2	130	89
SWI/SNF complex											
Smarcc1	gi|30851572	1748(81)	24(3)	1523(55)	23(1)	1786(43)	26(1)	632	11	1767	1078
Smarca4	gi|76253779	749	13	753	12	1508	23	724	11	1129	739
Arid1a	gi|124249109	794	12	791	13	989	15	720	12	892	756
Actl6a	gi|4001805	679(221)	9(3)	651(88)	8(1)	479(128)	6(2)	630(86)	8(1)	579	641
Smarcc2	gi|38565930	504	8	-	-	647	9	-	-	576	-
Smarce1	gi|10181166	322(76)	5(1)	310	5	361	7	297	4	342	304
Smarcd1	gi|1549249	250	5	369	8	434	7	314	4	342	342
Smarcb1	gi|6755578	236	3	263	3	379	5	237	5	308	250
Smarcd2	gi|57013098	68	1	70	2	140	3	69	2	104	70
PRC1/mblr complex											
Rnf2	gi|33563274	247	4	150	3	157	3	260	4	202	205
Phc1	gi|1490546	176	3	105	3	193	2	93	1	185	99
Mblr (Pcgf6)	gi|28076973	106	2	96	2	87	2	87	2	97	92
Rybp	gi|5381327	85	1	216	3	92	1	-	-	89	108

Table S4. ***Continued***

		Flag#1		Flag#2		Flag#3		Flag#4			
		-Dox		+Dox		-Dox		+Dox		-Dox	+Dox
Protein	Accession	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Average mascot	
Transcription factors											
Sall4	gi|81913723	2689(321)	35(4)	2252(295)	30(4)	2851(70)	34(1)	1564(330)	22(4)	2770	1908
Ubp1	gi|134032032	1775	22	2081	23	2068	24	1759	22	1922	1920
Tcfcp2	gi|15628025	868	11	1113	13	1185	13	941	12	1027	1027
Lsd1	gi|51315882	728	9	667	11	1035	15	109	3	882	388
Sall1	gi|14164331	994	12	967	10	748	10	421	7	871	694
Esrrb	gi|28277057	966	14	1451	18	610	9	868	12	788	1160
Smarca5	gi|14028669	670	12	461	8	482	10	56	1	576	259
Wdr5	gi|16554627	660	9	740	10	464	6	567	7	562	653
Peg10	gi|98985814	631(71)	8(1)	688	9	466	6	397	5	549	543
Yeats2	gi|84794613	616	9	234	4	402	6	-	-	509	117
Hcfc1	gi|34328130	323	5	379	6	679	10	105	2	501	242
Zmym4	gi|167555112	350	5	465	7	600	10	156	3	475	311
Hells	gi|12232371	345	4	229	3	522	7	150	2	434	190
Mga	gi|120444914	359	7	258	7	440	7	59	2	400	159
Zfp462	gi|85740499	711	12	692	13	85	2	56	2	398	374
Oct4	gi|125490392	375	5	-	-	321	4	-	-	348	-
Requiem	gi|6648956	332	4	295	3	362	4	197	2	347	246
Pogz	gi|111598687	496	6	452	6	90	1	-	-	293	226
Zfp143	gi|121247390	320	4	198	4	254	5	138	3	287	168
Sin3a	gi|726286	334	4	421	5	156	4	-	-	245	211
Hnrnpab	gi|6754222	300	4	251	3	171	2	401	6	236	326
Wiz	gi|46909565	208	3	91	2	262	5	64	1	235	78
Adnp	gi|55930867	268	4	198	3	199	3	51	1	234	125
Satb2	gi|20982839	220	5	100	2	246	5	162	3	233	131
Klf5	gi|31981873	209	3	151	2	232	3	84	1	221	118
Trim33	gi|119637828	98	1	344	5	311	4	97	1	205	221
Mybl2	gi|6678974	210	3	158	3	175	3	126	2	193	142
Bptf	gi|123241372	217	3	136	2	174	3	-	-	196	68
Grhl2	gi|46810275	218	4	-	-	120	2	-	-	169	-
L3mbtl2	gi|27734414	209	3	67	2	78	1	-	-	144	34
Zzz3*	gi|47847456	78	1	59	1	112	1	-	-	95	30
Ehmt1	gi|34784556	54	1	-	-	81	1	-	-	68	-
Zfp828	gi|32469497	51	2	-	-	75	2	-	-	63	-

Table S4. *Continued*

		Flag#1		Flag#2		Flag#3		Flag#4			
		-Dox		+Dox		-Dox		+Dox		-Dox	+Dox
Protein	Accession	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Average mascot	
Other											
Rif1	gi|47078460	1930	25	1833	26	2238	29	1403	24	2084	1618
Zcchc8	gi|169808385	822	10	780	9	747	9	-	-	785	390
Supt16h	gi|15637171	465	7	430	7	577	9	242	5	521	336
Lig3	gi|3913496	426	8	423	9	535	9	247	4	481	335
Ogt	gi|27499606	424	7	292	7	497	8	-	-	461	146
Xrcc6	gi|145587104	416	5	411	5	430	7	-	-	423	206
Xrcc5	gi|22137748	265	4	506	8	417	6	120	2	341	313
C130039O16Rik	gi|148670819	430	6	398	4	224	3	171	2	327	285
Msh2	gi|726086	260	5	251	5	393	8	-	-	327	126
L1td1	gi|124487095	378	6	520	8	216	3	74	2	297	297
Msh6	gi|6754744	448	5	101	1	134	3	-	-	291	51
Rbm14	gi|86262142	199	3	188	4	356	4	-	-	278	94
Xrcc1	gi|55391482	212	3	65	2	273	5	64	2	243	65
Polb	gi|21729749	294	5	259	4	166	3	154	3	230	207
EMSY	gi|124249084	280	3	286	4	155	3	296	4	218	291
Rpa1	gi|18390321	146	3	-	-	226	3	-	-	186	-
Prkdc	gi|124517706	119	4	-	-	198	4	-	-	159	-
4632411B12Rik	gi|37360322	176	2	-	-	82	1	-	-	129	-
Akap8	gi|5931618	149	2	178	2	107	1	-	-	128	89
Pnkp	gi|7108591	143	2	-	-	99	2	66	1	121	33
Asf1a*	gi|13384964	101	1	101	1	101	1	101	1	101	101
Cabin1	gi|70995287	135	3	-	-	62	1	-	-	99	-
2310057J16Rik	gi|61213696	78	1	104	3	55	1	-	-	67	52
Ubqln4*	gi|15805016	72	1	75	1	63	1	-	-	68	38

[a, b] Equivalent to Table S2. * Identification by one unique peptide.

Table S5. Esrrb-interacting proteins as identified by mass spectrometry analysis of purified F-Esrrb samples

		Flag#1		Flag#2		Flag#3		Flag#4			
		-Dox		+Dox		-Dox		+Dox		-Dox	+Dox
Protein	Accession	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Average mascot	
Esrrb	gi|6166153	2156	25	2452	28	2005	27	1945	24	2081	2199
SWI/SNF complex											
Smarcc1	gi|30851572	1212(95)	20(2)	992(107)	16(3)	1048	22	568	12	1130	780
Arid1a	gi|124249109	258(51)	4	418	7	593	14	144	4	426	281
Smarca4	gi|76253779	952	16	623	12	743	15	508	10	848	566
Smarcd1	gi|1549249	573	10	310	5	451	9	362	8	512	336
Pbrm1	gi|116284015	93	1	122	2	392	9	144	3	243	133
Trrap/p400 complex											
Trrap	gi|124486949	1452	24	1241	22	2176	49	-	-	1814	621
Ep400	gi|27348237	828	14	778	13	1159	21	62	2	994	420
Yeats4	gi|13386064	308	6	411	8	109	3	-	-	209	206
Dmap1	gi|12963557	452	5	314	4	339	7	55	1	396	185
Brd8	gi|19344050	323	5	-	-	201	5	66	2	262	33
Ing3	gi|17390799	127	2	110	2	100	3	-	-	114	55
1600027Rik	gi|110625965	81	1	71	1	69	1	-	-	75	36
NuRD complex											
Chd4	gi|39204553	1332	23	714	13	653	17	327	9	993	521
Gatad2a	gi|148696823	585	8	538	7	638	11	556	10	611	547
Mta2	gi|51491880	493	9	573	9	515	10	298	6	504	436
Mta1	gi|86577662	461	6	258	4	325	7	130	3	393	194
Gatad2b	gi|21314854	464	6	538	7	416	7	241	4	440	390
Mta3	gi|18381007	128	2	312	5	172	3	-	-	150	156
Mbd3	gi|7305261	252	4	402	5	210	4	-	-	231	201
Hdac1	gi|2347180	722(198)	13(3)	791(95)	14(2)	514(98)	11(2)	488	8	618	640
Hdac2	gi|87162464	391	8	402	8	505(98)	10(2)	333	6	448	368
Rbbp4	gi|5032027	464(145)	7(3)	510(204)	7(3)	473(152)	6(2)	317(251)	5(4)	469	414
Mediator complex											
Med14	gi|115270972	1332	24	2113	36	1518	29	1294	24	1425	1704
Med12	gi|123226656	1092	19	1577	24	1538	28	995	20	1315	1286
Med23	gi|61651678	853	14	1239	26	1252	25	980	19	1052	1110
Med24	gi|119220579	1081	16	1024	16	867	15	675	11	974	850
Med17	gi|21450345	1040	18	1290	17	845	15	801	15	942	1046
Med1	gi|14193713	708	12	976	18	1150	23	663	15	929	820
Med16	gi|148699683	867	13	809	13	692	13	673	13	779	741
Med15	gi|32451779	798	12	805	13	615	13	465	10	706	635
Med13	gi|124286862	389	7	465	7	668	12	224	5	528	345
Med27	gi|16741439	785	12	739	13	376	7	343	5	580	541
Med25	gi|47940179	466	6	479	8	322	5	329	5	394	404
Med13l	gi|49257394	301	5	279	5	395	10	185	6	348	232

Table S5. *Continued*

		Flag#1		Flag#2		Flag#3		Flag#4			
		-Dox		+Dox		-Dox		+Dox		-Dox	+Dox
Protein	Accession	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Average mascot	
Mediator complex											
Med4	gi|13385626	285	4	733	10	405	7	326	6	345	530
Med26	gi|28466971	334	7	518	9	321	8	69	2	327	294
Med6	gi|27754027	384	6	429	6	260	6	346	8	322	388
Med8	gi|29366816	290	5	370	6	191	4	164	4	240	267
Cncc	gi|38382739	181	4	312	6	270	4	133	4	226	223
Med30	gi|19882231	274	5	209	3	114	2	78	1	194	144
Med7	gi|157266302	244	4	216	7	101	2	67	1	172	142
Med29	gi|27754101	203	4	185	3	135	2	137	3	169	161
Cdk8	gi|31652272	104	2	99	2	108	2	155	4	106	127
Med18	gi|21313064	195	4	257	5	54	1	149	3	124	203
Med19	gi|28277157	88	2	348	7	57	1	-	-	72	174
RNApol2 complex											
Polr2a	gi|2145091	797	13	775	13	813	18	550	11	805	663
Polr2b	gi|24418911	498	9	548	9	792	16	459	10	645	504
Polr2c	gi|29336059	334	5	454	6	214	3	91	2	274	273
Polr2g	gi|4505947	109	1	307	5	191	3	86	2	150	197
TFIID complex											
Taf9	gi|28175808	162	3	410	8	352	8	68	2	257	239
Taf6	gi|6678215	162	4	140	3	270	5	60	1	216	100
Taf4a	gi|123288532	109	2	71	1	270	5	146	2	190	109
Taf10*	gi|46518499	147	1	129	1	113	1	-	-	130	65
Tbp*	gi|29477183	57	1	47	1	66	1	-	-	62	24
TRX/MLL complex											
Wdr5	gi|16554627	414	6	475	7	275	6	244	5	345	360
Hcfc1	gi|34328130	263	5	-	-	414	7	152	3	339	76
Mll2	gi|149266757	117	2	116	2	338	8	-	-	228	58
Mll3	gi|37999865	73	2	96	2	272	7	-	-	173	48
Ashl2	gi|4009338	89	2	130	2	184	5	-	-	137	65
Rbbp5	gi|34784634	50	1	68	1	126	3	-	-	88	34
Transcription factors											
Sall4	gi|117553631	1329(204)	19(4)	1104(227)	17(4)	965(245)	15(5)	861(124)	14(2)	1147	983
Dax1	gi|6671531	882	12	644	9	502	9	283	6	692	464
Tcfcp2l1	gi|90101766	768	12	524	8	566	11	390	7	667	457
Fkbp15	gi|38614309	602	9	774	11	656	10	235	4	629	505
Esrra	gi|112293262	436	7	874	12	540	10	265	6	488	570
Ncoa3	gi|118026946	294	5	122	2	549	11	-	-	421	61
Ubp1	gi|134032032	480	8	587	9	336	7	46	1	408	317
Nrip1	gi|27734110	385	7	888	15	430	10	528	11	408	708
Sall1	gi|14164331	360	5	267	4	438	11	272	6	399	270

Table S5. ***Continued***

		Flag#1		Flag#2		Flag#3		Flag#4			
		-Dox		+Dox		-Dox		+Dox		-Dox	+Dox
Protein	Accession	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Mascot[a]	Pept.[b]	Average mascot	
Transcription factors											
Zfp462	gi|148670321	189	5	82	3	507	12	-	-	348	41
Cdc2a	gi|13542826	278	6	360	7	313	6	71	1	295	216
Zbtb9	gi|54400753	205	5	210	4	353	6	101	3	279	156
Smarca5	gi|14028669	144	3	100	2	373	9	296	6	258	198
Wiz	gi|46909565	106	1	60	1	266	6	98	3	186	79
Requiem	gi|6648956	178	2	261	3	184	4	69	1	181	165
Jmjd1c	gi|149260924	115	2	49	1	204	4	-	-	159	25
Tcfcp2	gi|15628025	154	2	176	2	125	3	-	-	140	88
L3mbtl2	gi|27734414	63	1	-	-	217	3	73	1	140	37
Oct4	gi|125490392	107	1	-	-	158	3	-	-	133	-
Myst1	gi|21312790	64	1	71	1	138	3	-	-	101	36
Ehmt1	gi|34784556	85	1	67	1	90	3	-	-	88	34
Other											
Rif1	gi|47078460	1930	25	1370	20	2238	29	1656	27	2084	1513
Ogt	gi|27499606	424	7	292	7	497	8	214	5	461	253

[a, b] Equivalent to Table S2. * Identification by one unique peptide.

Table S6. Sall4-interacting proteins: emPAI scores

		Flag#1	Flag#2	Flag#3	Flag#4		
		-Dox	+Dox	-Dox	+Dox	-Dox	+Dox
Protein	Accession	emPAI [a]	emPAI [a]	emPAI [a]	emPAI [a]	Average emPAI	
Sall4	gi|117553631	20.62(0.04)	17.72(0.18)	32.85(0.18)	20.65(0.13)	26.74	19.18
NuRD complex							
Chd4	gi|39204553	1.78	3.24	7.52	4.6	4.65	3.92
Gatad2a	gi|148696823	5.02	4.74	5.03	6	5.03	5.37
Mta2	gi|51491880	2.99	3.18	3.37	3.57	3.18	3.38
Mta1	gi|86577662	1.64	2.28	2.89	2.57	2.27	2.43
Gatad2b	gi|21314854	2.43	2.62	3.68	3	3.06	2.81
Mta3	gi|18381007	2.51	2.04	3.6	3.34	3.06	2.69
Mbd3	gi|7305261	9.72	4.22	8.69	6.1	9.21	5.16
Hdac1	gi|2347180	5.54	7.88	18.54(0.2)	8.42(0.2)	12.04	8.15
Hdac2	gi|87162464	3.52	3.81	8.93	5.12	6.23	4.47
Rbbp4	gi|5032027	1.63(0.21)	1.64(0.07)	2.2(0.29)	3.42(0.3)	1.92	2.53
Rbbp7	gi|157909799	1.66	1.86	3.67	4.37	2.67	3.12
Mbd2	gi|7706609	0.58	0.36	1.15	0.47	0.87	0.42
SWI/SNF complex							
Actl6a	gi|4001805	0.42(0.14)	0.75(0.07)	1.02(0.23)	0.88(0.07)	0.72	0.82
Smarcc1	gi|30851572	0.09(0.03)	0.29	0.18(0.04)	0.15(0.03)	0.14	0.22
Smarcd1	gi|576884	0.29	0.47	0.21	0.14	0.25	0.31
Arid1a	gi|124249109	0.03	0.03	0.02	0.02	0.03	0.03
Smarcb1	gi|6755578	0.08	0.08	0.16	0.08	0.12	0.08
Smarca4	gi|76253779	0.02	0.08	0.02	0.04	0.02	0.06
Transcription factors							
Sall1	gi|14164331	1.05	1.15	2.2	1.6	1.62	1.38
Bend3	gi|39841055	1.18	1.11	3.3	2.23	2.24	1.67
Zfp219	gi|47940209	0.6	0.37	0.52	0.23	0.56	0.3
Nac1	gi|81886163	0.34	0.42	1.03	0.6	0.69	0.51
Ruvbl2	gi|6755382	0.48(0.07)	0.39(0.07)	1.2(0.14)	0.69	0.84	0.54
Wdr5	gi|16554627	0.57	0.57	1.46	1.46	1.02	1.02
Oct4	gi|53501	0.45	-	0.45	-	0.45	-
Grhl2	gi|46810275	0.1	0.1	0.33	0.37	0.22	0.24
Sall3	gi|49257163	0.26	-	0.23	-	0.25	-
Cxxc5	gi|19526854	0.22	0.35	0.65	0.85	0.44	0.6
Zfp143	gi|81908410	0.16	0.16	0.32	0.22	0.24	0.19
Tcfcp2	gi|15628025	0.06	-	0.19	0.12	0.13	0.06
Sall2	gi|49087134	0.13	0.1	0.1	0.1	0.12	0.1
Klf5	gi|31981873	0.07	0.07	0.14	-	0.11	0.04
Ctbp2	gi|2909779	0.16	0.08	0.16	0.16	0.16	0.12
Zbtb2	gi|85701993	0.12	-	0.12	0.06	0.12	0.03
Requiem	gi|6755314	0.08	0.17	0.08	-	0.08	0.09
Ewsr1	gi|88853581	0.1	0.1	0.05	-	0.08	0.05
Esrrb	gi|6166153	0.07	-	0.15	-	0.11	-
Other							
Usp9x	gi|115511018	0.02	0.02	0.23	-	0.13	0.01
7420416P09Rik	gi|76253890	0.13	-	0.38	0.27	0.26	0.14
Set	gi|123295280	0.65	-	0.37	0.23	0.51	0.12
L1td1	gi|148698953	0.04	0.04	0.2	0.04	0.12	0.04

[a] emPAI score for the specified protein in the F-Sall4 sample. emPAI score for the specified protein in the control sample, if present, is between brackets.

Table S7. Dax1-interacting proteins: emPAI scores

		Flag#1	Flag#2	Flag#3	Flag#4		
		-Dox	+Dox	-Dox	+Dox	-Dox	+Dox
Protein	Accession	emPAI [a]	emPAI [a]	emPAI [a]	emPAI [a]	Average emPAI	
Dax1 (Nr0b1)	gi|6671531	1.12	1.26	1.56	1.57	1.34	1.42
Transcription factors							
Sall4	gi|117553631	0.56(0.18)	0.52(0.09)	0.43(0.13)	0.61(0.13)	0.5	0.57
Esrrb	gi|6166153	0.51	0.15	0.51	0.07	0.51	0.11
Snw1	gi|146149191	0.12	-	0.25	0.25	0.19	0.13
Sall1	gi|14164331	0.05	-	0.07	0.18	0.06	0.09
Wdr5	gi|16554627	0.34	0.16	0.56	0.35	0.45	0.26
Oct4	gi|200118	0.2	-	0.2	-	0.2	-
Prmt1	gi|30185908	0.08	-	0.08	0.08	0.08	0.04
Rif1	gi|47078460	0.03	-	0.08	0.14	0.06	0.07
Ogt	gi|13775066	0.09	0.09	0.12	0.09	0.11	0.09

[a] equivalent to Table S6.

Table S8. Tcfcp2l1-interacting proteins: emPAI scores

		Flag#1	Flag#2	Flag#3	Flag#4		
		-Dox	+Dox	-Dox	+Dox	-Dox	+Dox
Protein	Accession	emPAI [a]	emPAI [a]	emPAI [a]	emPAI [a]	Average emPAI	
Tcfcp2l1	gi|134053939	13.57(0.06)	14.48(0.2)	19.04	14.53	16.31	14.51
NuRD complex							
Chd4	gi|39204553	0.89	0.95	1.7	0.95	1.30	0.95
Gatad2a	gi|30795222	1.71	1.11	1.58	0.91	1.65	1.01
Mta1	gi|15077051	0.89	0.89	1.24	0.24	1.07	0.57
Mta2	gi|51491880	0.63	0.63	1.14	0.14	0.89	0.39
Gatad2b	gi|21314854	1.16	0.76	1.06	0.17	1.11	0.47
Hdac1	gi|2347180	2.56(0.2)	2.79(0.2)	1.49	0.83	2.03	1.81
Rbbp7	gi|157909799	1.31	1.15	1.71	1.32	1.51	1.24
Hdac2	gi|3023934	2.14	1.33	1.19	0.72	1.67	1.03
Mta3	gi|18381007	0.87	0.35	0.42	0.1	0.65	0.23
Mbd3	gi|7305261	1.28	1.28	1.81	0.86	1.55	1.07
Mbd2	gi|7706609	0.14	-	0.08	-	0.11	-
Trrap/p400 complex							
Trrap	gi|124486949	0.22	0.23	0.33	0.24	0.28	0.24
Ruvbl2	gi|6755382	1.86(0.14)	2.05(0.39)	3.54(0.07)	3.24(0.07)	2.70	2.65
Ep400	gi|27348237	0.19	0.15	0.15	0.12	0.17	0.14
Ruvbl1	gi|9790083	1.38(0.14)	2.31(0.22)	2.33(0.22)	0.82(0.22)	1.86	1.57
Actl6a	gi|4001805	1.16(0.23)	1.32(0.07)	0.64(0.14)	1.16(0.07)	0.90	1.24
Yeats4	gi|13386064	1.23	1.61	1.13	0.87	1.18	1.24
Vps72	gi|83921607	0.51	0.18	0.28	0.28	0.40	0.23
Dmap1	gi|12963557	0.29	0.37	0.21	0.29	0.25	0.33
Brd8	gi|19344050	0.07	0.03	0.07	-	0.07	0.02
Ing3	gi|17390799	0.08	-	0.08	0.08	0.08	0.04
1600027Rik	gi|110625965	0.15	-	0.34	0.34	0.25	0.17
SWI/SNF complex							
Smarcc1	gi|30851572	1.39(0.04)	1.32(0.03)	1.46(0.03)	0.31	1.43	0.82
Smarca4	gi|76253779	0.23	0.23	0.51	0.23	0.37	0.23
Arid1a	gi|124249109	0.18	0.17	0.23	0.17	0.21	0.17
Actl6a	gi|4001805	1.16(0.23)	1.32(0.07)	0.64(0.14)	1.16(0.07)	0.90	1.24
Smarcc2	gi|38565930	0.28	-	0.4	-	0.34	-
Smarce1	gi|10181166	0.54(0.07)	0.54	0.44	0.43	0.49	0.49
Smarcd1	gi|1549249	0.29	0.54	0.67	0.36	0.48	0.45
Smarcb1	gi|6755578	0.35	0.25	0.47	0.35	0.41	0.30
Smarcd2	gi|57013098	0.07	0.14	0.21	0.07	0.14	0.11
PRC1/mblr complex							
Rnf2	gi|33563274	0.42	0.19	0.3	0.42	0.36	0.31
Phc1	gi|1490546	0.1	0.1	0.07	0.03	0.09	0.07
Mblr (Pcgf6)	gi|28076973	0.18	0.18	0.18	0.09	0.18	0.14
Rybp	gi|5381327	0.14	0.5	0.14	-	0.14	0.25

Table S8. ***Continued***

		Flag#1	Flag#2	Flag#3	Flag#4		
		-Dox	+Dox	-Dox	+Dox	-Dox	+Dox
Protein	Accession	emPAI [a]	emPAI [a]	emPAI [a]	emPAI [a]	Average emPAI	
Transcription factors							
Sall4	gi|81913723	3.25(0.18)	2.35(0.18)	3.95(0.04)	1.22(0.18)	3.60	1.79
Ubp1	gi|134032032	6	8.77	7.32	4.31	6.66	6.54
Tcfcp2	gi|15628025	1.14	1.71	1.72	1.41	1.43	1.56
Lsd1	gi|51315882	0.37	0.53	0.58	0.11	0.48	0.32
Sall1	gi|14164331	0.33	0.3	0.24	0.15	0.29	0.23
Esrrb	gi|28277057	1.61	5.36	1.13	1.61	1.37	3.49
Smarca5	gi|14028669	0.28	0.25	0.25	0.03	0.27	0.14
Wdr5	gi|16554627	0.72	0.88	1.25	1.46	0.99	1.17
Peg10	gi|98985814	0.38(0.08)	0.38	0.26	0.19	0.32	0.29
Yeats2	gi|84794613	0.21	0.09	0.13	-	0.17	0.05
Hcfc1	gi|34328130	0.1	0.1	0.17	0.03	0.14	0.07
Zmym4	gi|167555112	0.08	0.1	0.19	0.08	0.14	0.09
Hells	gi|12232371	0.15	0.11	0.24	0.07	0.20	0.09
Mga	gi|120444914	0.06	0.04	0.1	0.01	0.08	0.03
Zfp462	gi|85740499	0.15	0.17	0.02	0.01	0.09	0.09
Oct4	gi|125490392	0.45	-	0.45	-	0.45	-
Requiem	gi|6648956	0.36	0.26	0.37	0.17	0.37	0.22
Pogz	gi|111598687	0.14	0.14	0.03	-	0.09	0.07
Zfp143	gi|121247390	0.28	0.16	0.28	0.17	0.28	0.17
Sin3a	gi|726286	0.1	0.12	0.1	-	0.10	0.06
Hnrnpab	gi|6754222	0.54	0.38	0.24	0.91	0.39	0.65
Wiz	gi|46909565	0.07	0.02	0.17	0.02	0.13	0.02
Adnp	gi|55930867	0.12	0.12	0.12	0.04	0.12	0.08
Satb2	gi|20982839	0.23	0.08	0.18	0.13	0.21	0.11
Klf5	gi|31981873	0.22	0.14	0.22	0.07	0.22	0.11
Trim33	gi|119637828	0.03	0.12	0.11	0.06	0.07	0.09
Mybl2	gi|6678974	0.14	0.14	0.09	0.09	0.12	0.12
Bptf	gi|123241372	0.02	0.01	0.03	-	0.03	0.01
Grhl2	gi|46810275	0.3	0.07	0.14	-	0.22	0.04
L3mbtl2	gi|27734414	0.13	0.04	0.04	-	0.09	0.02
Zzz3	gi|47847456	0.03	0.03	0.03	-	0.03	0.02
Ehmt1	gi|34784556	0.12	-	0.03	-	0.08	-
Zfp828	gi|32469497	0.05	-	0.08	-	0.07	-
Other							
Rif1	gi|47078460	0.32	0.34	0.48	0.29	0.40	0.32
Zcchc8	gi|169808385	0.6	0.75	0.83	-	0.72	0.38
Supt16h	gi|15637171	0.18	0.18	0.25	0.12	0.22	0.15
Lig3	gi|3913496	0.27	0.27	0.23	0.13	0.25	0.20
Ogt	gi|27499606	0.22	0.15	0.22	-	0.22	0.08
Xrcc6	gi|145587104	0.27	0.27	0.42	0.06	0.35	0.17
Xrcc5	gi|22137748	0.18	0.33	0.27	0.08	0.23	0.21
C130039O16Rik	gi|148670819	0.16	0.13	0.10	0.10	0.13	0.12
Msh2	gi|726086	0.17	0.17	0.25	-	0.21	0.09
L1td1	gi|124487095	0.21	0.3	0.12	0.08	0.17	0.19
Msh6	gi|6754744	0.12	0.02	0.07	-	0.10	0.01
Rbm14	gi|86262142	0.1	0.16	0.21	-	0.16	0.08
Xrcc1	gi|55391482	0.16	0.05	0.22	0.1	0.19	0.08
Polb	gi|21729749	0.55	0.42	0.3	0.25	0.43	0.34
EMSY	gi|124249084	0.08	0.11	0.08	0.11	0.08	0.11
Rpa1	gi|18390321	0.1	-	0.16	-	0.13	-
Prkdc	gi|124517706	0.01	-	0.01	-	0.01	-
4632411B12Rik	gi|37360322	0.07	0.04	0.04	-	0.06	0.02
Akap8	gi|5931618	0.09	0.09	0.05	-	0.07	0.05
Pnkp	gi|7108591	0.13	-	0.12	0.06	0.13	0.03
Asf1a	gi|13384964	0.15	0.15	0.15	0.15	0.15	0.15
Cabin1	gi|70995287	0.04	-	0.07	-	0.06	-
2310057J16Rik	gi|61213696	0.03	0.1	0.03	-	0.03	0.05
Ubqln4	gi|15805016	0.05	0.05	0.05	-	0.05	0.03

[a] equivalent to Table S6.

Table S9. Esrrb-interacting proteins: emPAI scores

		Flag#1	Flag#2	Flag#3	Flag#4		
		-Dox	+Dox	-Dox	+Dox	-Dox	+Dox
Protein	Accession	emPAI[a]	emPAI[a]	emPAI[a]	emPAI[a]	Average emPAI	
Esrrb	gi|6166153	29.49	55.48	29.03	26.16	29.26	40.82
SWI/SNF complex							
Smarcc1	gi|30851572	0.86(0.03)	0.76(0.06)	1.01	0.48	0.94	0.62
Arid1a	gi|124249109	0.06(0.01)	0.09	0.22	0.05	0.14	0.07
Smarca4	gi|76253779	0.35	0.23	0.34	0.2	0.35	0.22
Smarcd1	gi|1549249	0.85	0.47	0.77	0.69	0.81	0.58
Pbrm1	gi|116284015	0.07	0.03	0.13	0.07	0.10	0.05
Trrap/p400 complex							
Trrap	gi|124486949	0.19	0.16	0.45	-	0.32	0.08
Ep400	gi|27348237	0.14	0.13	0.25	0.02	0.20	0.08
Yeats4	gi|13386064	1.41	1.41	0.47	-	0.94	0.71
Dmap1	gi|12963557	0.46	0.29	0.39	0.07	0.43	0.18
Brd8	gi|19344050	0.2	-	0.15	0.04	0.18	0.02
Ing3	gi|17390799	0.16	0.16	0.25	-	0.21	0.08
1600027Rik	gi|110625965	0.15	0.15	0.16	-	0.16	0.08
NuRD complex							
Chd4	gi|39204553	0.39	0.22	0.25	0.1	0.32	0.16
Gatad2a	gi|148696823	0.42	1.12	0.67	0.71	0.55	0.92
Mta2	gi|51491880	0.43	0.5	0.52	0.27	0.48	0.39
Mta1	gi|86577662	0.29	0.18	0.3	0.15	0.30	0.17
Gatad2b	gi|21314854	0.36	0.59	0.45	0.18	0.41	0.39
Mta3	gi|18381007	0.1	0.28	0.17	-	0.14	0.14
Mbd3	gi|7305261	0.68	0.86	0.7	0.12	0.69	0.49
Hdac1	gi|2347180	1.64(0.2)	2.17(0.13)	1.26(0.06)	1.04	1.45	1.61
Hdac2	gi|87162464	0.62	0.72	1.4(0.06)	0.68	1.01	0.70
Rbbp4	gi|5032027	0.79(0.14)	0.68(0.21)	0.49(0.14)	0.41(0.3)	0.64	0.55
Mediator complex							
Med14	gi|115270972	0.77	1.17	1.06	0.75	0.92	0.96
Med12	gi|123226656	0.32	0.4	0.48	0.31	0.40	0.36
Med23	gi|61651678	0.38	0.75	0.81	0.47	0.60	0.61
Med24	gi|119220579	0.73	0.73	0.74	0.45	0.74	0.59
Med17	gi|21450345	1.41	2.34	1.05	1.2	1.23	1.77
Med1	gi|14193713	0.25	0.42	0.56	0.36	0.41	0.39
Med16	gi|148699683	0.63	0.69	0.54	0.51	0.59	0.60
Med15	gi|32451779	0.86	0.79	0.75	0.58	0.81	0.69
Med13	gi|124286862	0.13	0.12	0.18	0.06	0.16	0.09
Med27	gi|16741439	3.94	2.74	1.16	0.98	2.55	1.86
Med25	gi|47940179	0.41	0.41	0.31	0.38	0.36	0.40
Med13l	gi|49257394	0.1	0.12	0.14	0.08	0.12	0.10
Med4	gi|13385626	0.75	4.36	2.17	1.6	1.46	2.98
Med26	gi|28466971	0.37	0.6	0.46	0.09	0.42	0.35
Med6	gi|27754027	1.79	1.41	1.13	2.48	1.46	1.95
Med8	gi|29366816	0.94	1.42	0.77	0.6	0.86	1.01
Cncc	gi|38382739	0.43	0.88	0.45	0.1	0.44	0.49
Med30	gi|19882231	1.24	1.24	0.39	0.19	0.82	0.72
Med7	gi|157266302	0.84	0.84	0.46	0.14	0.65	0.49
Med29	gi|27754101	1.17	0.86	0.38	0.39	0.78	0.63
Cdk8	gi|31652272	0.13	0.07	0.14	0.14	0.14	0.11
Med18	gi|21313064	0.75	1.01	0.16	0.81	0.46	0.91
Med19	gi|28277157	0.29	1.43	0.14	0.14	0.22	0.79
RNApol2 complex							
Polr2a	gi|2145091	0.22	0.21	0.27	0.2	0.25	0.21
Polr2b	gi|24418911	0.26	0.22	0.52	0.31	0.39	0.27
Polr2c	gi|29336059	0.69	0.69	0.39	0.25	0.54	0.47
Polr2g	gi|4505947	0.19	1.34	0.69	0.2	0.44	0.77
TFIID complex							
Taf9	gi|28175808	0.41	1.23	1.58	0.13	1.00	0.68
Taf6	gi|6678215	0.1	0.1	0.1	0.05	0.10	0.08
Taf4a	gi|123288532	0.1	0.05	0.28	0.11	0.19	0.08
Taf10	gi|46518499	0.16	0.16	0.17	-	0.17	0.08
Tbp	gi|29477183	0.1	0.1	0.1	-	0.10	0.05

Table S9. ***Continued***

		Flag#1	Flag#2	Flag#3	Flag#4		
		-Dox	+Dox	-Dox	+Dox	-Dox	+Dox
Protein	Accession	emPAI[a]	emPAI[a]	emPAI[a]	emPAI[a]	Average emPAI	
TRX/MLL complex							
Wdr5	gi|16554627	0.72	0.72	0.59	0.47	0.66	0.60
Hcfc1	gi|34328130	0.08	-	0.12	0.05	0.10	0.03
Mll2	gi|149266757	0.01	0.01	0.05	-	0.03	0.01
Mll3	gi|37999865	0.01	0.01	0.04	-	0.03	0.01
Ashl2	gi|4009338	0.1	0.05	0.29	-	0.20	0.03
Rbbp5	gi|34784634	0.06	0.06	0.19	-	0.13	0.03
Transcription factors							
Sall4	gi|117553631	0.92(0.13)	0.75(0.13)	0.58(0.18)	0.56(0.09)	0.75	0.66
Dax1	gi|6671531	1.41	0.87	1.04	0.5	1.23	0.69
Tcfcp2l1	gi|90101766	0.96	0.63	1	0.48	0.98	0.56
Fkbp15	gi|38614309	0.25	0.32	0.3	0.14	0.28	0.23
Esrra	gi|112293262	0.67	1.59	1.29	0.48	0.98	1.04
Ncoa3	gi|118026946	0.12	0.05	0.23	-	0.18	0.03
Ubp1	gi|134032032	0.48	0.71	0.5	0.07	0.49	0.39
Nrip1	gi|27734110	0.17	0.45	0.25	0.37	0.21	0.41
Sall1	gi|14164331	0.13	0.1	0.25	0.11	0.19	0.11
Zfp462	gi|148670321	0.02	0.02	0.14	-	0.08	0.01
Cdc2a	gi|13542826	0.81	0.81	1.03	0.11	0.92	0.46
Zbtb9	gi|54400753	0.4	0.31	0.62	0.15	0.51	0.23
Smarca5	gi|14028669	0.09	0.06	0.19	0.23	0.14	0.15
Wiz	gi|46909565	0.03	0.02	0.18	0.11	0.11	0.07
Requiem	gi|6648956	0.17	0.26	0.36	0.09	0.27	0.18
Jmjd1c	gi|149260924	0.03	0.01	0.04	-	0.04	0.01
Tcfcp2	gi|15628025	0.12	0.19	0.2	-	0.16	0.10
L3mbtl2	gi|27734414	0.07	-	0.16	0.05	0.12	0.03
Oct4	gi|125490392	0.1	-	0.33	-	0.22	-
Myst1	gi|21312790	0.07	0.07	0.14	-	0.11	0.04
Ehmt1	gi|34784556	0.03	0.03	0.1	-	0.07	0.02
Other							
Rif1	gi|47078460	0.39	0.26	0.41	0.06	0.40	0.16
Ogt	gi|27499606	0.3	0.15	0.39	0.1	0.35	0.13

[a] equivalent to Table S6.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Antibodies used in Figures S1, S2, S3

anti-Sox2 (Y17, Santa Cruz), anti-Sall4 (PP-PPZ 0601-00, Perseus Proteomics), anti-Oct4 (C10, Santa Cruz), anti-Dax1 (sc-841, Santa Cruz), anti-Tcfcp2l1 (ARP32606, Aviva), anti-Esrrb (PP-H6707-00, R&D systems), anti-Flag (M2, Sigma), anti-Sall1 (PP-K9814-00, R&D), anti-Wdr5 (07-706, Upstate), Ep400 (A300-541a, Bethyl Laboratories), anti-Lamin B1 (C20, Santa Cruz).

Western and RT-PCR analysis of ES cell lines

Whole cell extracts of F-Oct4 ES cells, ZHBTc4 ES cells and CGR8 ES cells were analysed by western blot with the indicated antibodies. For RT-PCR analysis, RNA was purified from $5x10^6$ cells using the RNeasy protocol (Qiagen). cDNA was synthesised using 2.5µg RNA primed with random hexamers according to the Superscript First Strand Synthesis System (Invitrogen). PCRs were performed on a Roche Lightcycler by an initial denaturation at 95° for 5 mins, followed by 45 cycles of denaturation (95°, 5s), annealing (58°, 10 s) and elongation (72°, 20 s).

Primers used for RT-PCR analysis of ES cell lines

TBP FW	ggggagctgtgatgtgaagt
TBP RV	ccaggaaataattctggctca
Klf4 FW	cgggaagggagaagact
Klf RV	gacttcctcacgccaacg
Zfp42 FW	cagctcctgcacacagaaga
Zfp42 RV	actgatccgcaaacacctg
Eras FW	gcccctcatcagactgctac
Eras RV	gcagctcaaggaagaggtgt
Dax1 FW	accgtgctctttaacccaga
Dax1 RV	ccggatgtgctcagtaagg
Fgf5 FW	gtttccagtggagcccttc
Fgf5RV	gagacacagcaaatatttccaaaa
T (Brachyury) FW	cagcccacctactggctcta
T (Brachyury) RV	gagcctggggtgatggta

Superose 6 gelfiltration of ES cell nuclear extracts

200 µl ES cell nuclear extract was separated on a 25 ml Superose 6 gel filtration column (GE Healthcare) with a flow rate of 0.1 ml per minute in C-100 buffer (20 mM Hepes pH 7.6, 0.2 mM EDTA, 1.5 mM $MgCl_2$, 100 mM KCl, 20% glycerol). 0.5 ml elution fractions were TCA precipitated, separated on SDS polyacrylamide and western blots probed with anti-Oct3/4 antibody (sc-8628, Santa Cruz) or anti-Mta2 (8106, Abcam). The Superose 6 column was calibrated with gel filtration calibration standards (GE Healthcare).

Mass spectrometric analysis

1D SDS-PAGE gel lanes were cut into 2-mm slices using an automatic gel slicer and subjected to in-gel reduction with dithiothreitol, alkylation with iodoacetamide and digestion with trypsin (Promega, sequencing grade), essentially as described by [1]. Nanoflow LC-MS/MS was performed on an 1100 series capillary LC system (Agilent Technologies) coupled to an LTQ-Orbitrap mass spectrometer (Thermo) operating in positive mode and equipped with a nanospray source. Peptide mixtures were trapped on a ReproSil C18 reversed phase column (Dr Maisch GmbH; column dimensions 1.5 cm × 100 µm, packed in-house) at a flow rate of 8 µl/min. Peptide separation was performed on ReproSil C18 reversed phase column (Dr Maisch GmbH; column dimensions 15 cm × 50 µm, packed in-house) using a linear gradient from 0 to 80% B (A = 0.1 % formic acid; B = 80% (v/v) acetonitrile, 0.1 % formic acid) in 70 min and at a constant flow rate of 200 nl/min using a splitter. The column eluent was directly sprayed into the ESI source of the mass spectrometer. Mass spectra were acquired in continuum mode; fragmentation of the peptides was performed in data-dependent mode. Peak lists were automatically created from raw data files using the Mascot Distiller software (version 2.1; MatrixScience). The Mascot search algorithm (version 2.2, MatrixScience) was used for searching against the NCBInr database (release NCBInr_20090222; taxonomy: Mus musculus) or the IPI_mouse_database (release 20090924). The peptide tolerance was typically set to 10 ppm and the fragment ion tolerance to 0.8 Da. A maximum number of 2 missed cleavages by trypsin were allowed and carbamidomethylated cysteine and oxidized methionine were set as fixed and variable modifications, respectively. The Mascot score cut-off value for a positive protein hit was set to 60, based on at least two peptides. In case of protein identifications with Mascot scores between 50 and 60, or that were based on only one peptide, individual peptide MS/MS spectra were checked manually and either interpreted as valid identifications or discarded. We also show a more quantitative measure of our identified proteins, emPAI score [2]. emPAI score incorporates the

number of peptides identified per protein (spectral counts) normalized by the theoretical number of peptides. This is a superior method over just counting the number of identified peptides, because it takes account of the fact that, for the same number of molecules, larger proteins and proteins with many peptides in the preferred mass range for mass spectrometry will generate more observed peptides.

Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996). Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. Nature *379*, 466-469.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics *4*, 1265-1272.

Immunoprecipitation with tagged proteins

Coding sequences were amplified from mouse ES cell cDNA and inserted with an C-terminal V5-tag (Tcfcp2l1 and Zfp143) or FLAG-tag (Nac1) into a pPyCAG-driven expression vector. 46C ES cells[3] were transfected with the constructs using Lipofectamine 2000 (Invitrogen), nuclear extracts were made 24 hrs after transfection and immunoprecipitations were done, as described in the Experimental Procedures.

Primers used for amplification ChIP targets

Gene	FW primer	RV primer
Fbxo15 (-0.6kb)	TCCCCCTGTAAATTCACTCA	TAGCTAGCTGGTTGGTCCAC
Rest (-3.1kb)	CTCCCCTGGACAATAGCTTC	CGTCCTTCATTTCCTCAGTG
Nanog enhancer (-5kb)	GTCCCCGCTCCTTTTCAGCACTAACCATAC	CGGTTTGAATAGGGAGGAGGGCGTCT
Nanog promoter (-0.2kb)	AAGATGAATAAAGTGAAATGAGGTAAAGC	ACTGGGAGGGAGGGAAAGC
Dppa3 (-1.7kb)	GATCCAGCTGGTCTGAGCTA	GTGCAGGGATCATAGGAGTG
Zfp42 enhancer (-14kb)	GTGTGGTGTTGAGCAGGTGT	TGACACAAAGCTTCACTACGG
Mybl2 (-2.4kb)	GACCACTCCCAGGTTTGACT	AGGAATCTGGTGACCTCCAC
Pcsk6 (+11.5kb)	ACTTGGGATCCTCCCTTCTT	ATCCTAGGCAGTGCTGGTCT
Mcl1 (+4.7kb)	CTCCCCTTGGAAGTTAACCA	GATGGCTGACTGGAGTCTCA
Pramel6 (+8.5kb)	CTCAGGGAACGCCAGTTTAT	ATGGGATCCCCACATAGAAA
Loxl1 (+33.6kb)	TGCATTGTCAAGAAACAAAGG	TTTCTGGATAGCCCATCTCC
B3gnt7 (+7.6kb)	ATCCCACTGTTACCCAGAGC	CCCTACTCCCCGGTACTACA
Slc27a4 (+17.9kb)	TAGTCTTTGGCGGCAGTTTA	CTTCCTCCTCCCATTCTTGT
Sirpa (-5.0kb)	CTGGACTCATTGTGGATTGG	TCTGGGGATCTGGTTCTACC

References for Oct4 interaction network

References for (genome-wide) chromatin immunoprecipitation data; Oct4[4-6], Sox2[4-5], Esrrb[4, 7], Klf5[5, 8-9], Dax1, Nac1[5], Tcfcp2l1[4], PRC1 complex[6, 10-11]. References for ES cell and developmental phenotypes (Table 2); NuRD subunits[12-17], SWI/SNF subunits[18-21], PRC1 subunits[6, 22-26], Trrap subunits[21, 27-28], Lsd1 complex[29-31], Sall4[32-34], Sall1[35], Zfp219[26], Wdr5[36], Ubp1[37], Mga[26], Arid3b[38], Sox2[39-40], Nac1[41-42], Tcfcp2l1[43], Rbpj[44], Esrrb[43, 45-46], hcfc1[47], Dax1[48], Zfp143[49], Pml[50], Foxp4[51], Ctbp2[29, 52], Klf5[8, 53], Rif1[46], Smc1a[26], Msh2[54], Ogt[55-56].

SUPPLEMENTAL REFERENCES

1. Wilm, M., A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis, and M. Mann, *Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry.* Nature, 1996. **379**(6564): p. 466-9.
2. Ishihama, Y., Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, and M. Mann, *Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein.* Mol Cell Proteomics, 2005. **4**(9): p. 1265-72.
3. Ying, Q.L., M. Stavridis, D. Griffiths, M. Li, and A. Smith, *Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture.* Nat Biotechnol, 2003. **21**(2): p. 183-6.
4. Chen, X., H. Xu, P. Yuan, F. Fang, M. Huss, V.B. Vega, E. Wong, Y.L. Orlov, W. Zhang, J. Jiang, Y.H. Loh, H.C. Yeo, Z.X. Yeo, V. Narang, K.R. Govindarajan*, et al.*, *Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.* Cell, 2008. **133**(6): p. 1106-17.
5. Kim, J., J. Chu, X. Shen, J. Wang, and S.H. Orkin, *An extended transcriptional network for pluripotency of embryonic stem cells.* Cell, 2008. **132**(6): p. 1049-61.
6. Endoh, M., T.A. Endo, T. Endoh, Y. Fujimura, O. Ohara, T. Toyoda, A.P. Otte, M. Okano, N. Brockdorff, M. Vidal, and H. Koseki, *Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity.* Development, 2008. **135**(8): p. 1513-24.
7. van den Berg, D.L., W. Zhang, A. Yates, E. Engelen, K. Takacs, K. Bezstarosti, J. Demmers, I. Chambers, and R.A. Poot, *Estrogen-related receptor beta interacts with Oct4 to positively regulate Nanog gene expression.* Mol Cell Biol, 2008. **28**(19): p. 5986-95.
8. Parisi, S., F. Passaro, L. Aloia, I. Manabe, R. Nagai, L. Pastore, and T. Russo, *Klf5 is involved in self-renewal of mouse embryonic stem cells.* J Cell Sci, 2008. **121**(Pt 16): p. 2629-34.
9. Jiang, J., Y.S. Chan, Y.H. Loh, J. Cai, G.Q. Tong, C.A. Lim, P. Robson, S. Zhong, and H.H. Ng, *A core Klf circuitry regulates self-renewal of embryonic stem cells.* Nat Cell Biol, 2008. **10**(3): p. 353-60.
10. Boyer, L.A., K. Plath, J. Zeitlinger, T. Brambrink, L.A. Medeiros, T.I. Lee, S.S. Levine, M. Wernig, A. Tajonar, M.K. Ray, G.W. Bell, A.P. Otte, M. Vidal, D.K. Gifford, R.A. Young*, et al.*, *Polycomb complexes repress developmental regulators in murine embryonic stem cells.* Nature, 2006. **441**(7091): p. 349-53.
11. Ku, M., R.P. Koche, E. Rheinbay, E.M. Mendenhall, M. Endoh, T.S. Mikkelsen, A. Presser, C. Nusbaum, X. Xie, A.S. Chi, M. Adli, S. Kasif, L.M. Ptaszek, C.A. Cowan, E.S. Lander*, et al.*, *Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains.* PLoS Genet, 2008. **4**(10): p. e1000242.
12. Liang, J., M. Wan, Y. Zhang, P. Gu, H. Xin, S.Y. Jung, J. Qin, J. Wong, A.J. Cooney, D. Liu, and Z. Songyang, *Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells.* Nat Cell Biol, 2008. **10**(6): p. 731-9.
13. Marino, S. and R. Nusse, *Mutants in the mouse NuRD/Mi2 component P66alpha are embryonic lethal.* PLoS ONE, 2007. **2**(6): p. e519.
14. Hendrich, B., J. Guy, B. Ramsahoye, V.A. Wilson, and A. Bird, *Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development.* Genes Dev, 2001. **15**(6): p. 710-23.
15. Kaji, K., I.M. Caballero, R. MacLeod, J. Nichols, V.A. Wilson, and B. Hendrich, *The NuRD component Mbd3 is required for pluripotency of embryonic stem cells.* Nat Cell Biol, 2006. **8**(3): p. 285-92.
16. Lagger, G., D. O'Carroll, M. Rembold, H. Khier, J. Tischler, G. Weitzer, B. Schuettengruber, C. Hauser, R. Brunmeir, T. Jenuwein, and C. Seiser, *Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression.* Embo J, 2002. **21**(11): p. 2672-81.
17. Zimmermann, S., F. Kiefer, M. Prudenziati, C. Spiller, J. Hansen, T. Floss, W. Wurst, S. Minucci, and M. Gottlicher, *Reduced body size and decreased intestinal tumor rates in HDAC2-mutant mice.* Cancer Res, 2007. **67**(19): p. 9047-54.
18. Bultman, S., T. Gebuhr, D. Yee, C. La Mantia, J. Nicholson, A. Gilliam, F. Randazzo, D. Metzger, P. Chambon, G. Crabtree, and T. Magnuson, *A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes.* Mol Cell, 2000. **6**(6): p. 1287-95.
19. Kim, J.K., S.O. Huh, H. Choi, K.S. Lee, D. Shin, C. Lee, J.S. Nam, H. Kim, H. Chung, H.W. Lee, S.D. Park, and R.H. Seong, *Srg3, a mouse homolog of yeast SWI3, is essential for early embryogenesis and*

involved in brain development. Mol Cell Biol, 2001. **21**(22): p. 7787-95.
20. Klochendler-Yeivin, A., L. Fiette, J. Barra, C. Muchardt, C. Babinet, and M. Yaniv, *The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression.* EMBO Rep, 2000. **1**(6): p. 500-6.
21. Fazzio, T.G., J.T. Huff, and B. Panning, *An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity.* Cell, 2008. **134**(1): p. 162-74.
22. Takihara, Y., D. Tomotsune, M. Shirai, Y. Katoh-Fukui, K. Nishii, M.A. Motaleb, M. Nomura, R. Tsuchiya, Y. Fujita, Y. Shibata, T. Higashinakagawa, and K. Shimada, *Targeted disruption of the mouse homologue of the Drosophila polyhomeotic gene leads to altered anteroposterior patterning and neural crest defects.* Development, 1997. **124**(19): p. 3673-82.
23. van der Stoop, P., E.A. Boutsma, D. Hulsman, S. Noback, M. Heimerikx, R.M. Kerkhoven, J.W. Voncken, L.F. Wessels, and M. van Lohuizen, *Ubiquitin E3 ligase Ring1b/Rnf2 of polycomb repressive complex 1 contributes to stable maintenance of mouse embryonic stem cells.* PLoS ONE, 2008. **3**(5): p. e2235.
24. Voncken, J.W., B.A. Roelen, M. Roefs, S. de Vries, E. Verhoeven, S. Marino, J. Deschamps, and M. van Lohuizen, *Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition.* Proc Natl Acad Sci U S A, 2003. **100**(5): p. 2468-73.
25. Pirity, M.K., J. Locker, and N. Schreiber-Agus, *Rybp/DEDAF is required for early postimplantation and for central nervous system development.* Mol Cell Biol, 2005. **25**(16): p. 7193-202.
26. Hu, G., J. Kim, Q. Xu, Y. Leng, S.H. Orkin, and S.J. Elledge, *A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal.* Genes Dev, 2009. **23**(7): p. 837-48.
27. Ueda, T., R. Watanabe-Fukunaga, H. Ogawa, H. Fukuyama, Y. Higashi, S. Nagata, and R. Fukunaga, *Critical role of the p400/mDomino chromatin-remodeling ATPase in embryonic hematopoiesis.* Genes Cells, 2007. **12**(5): p. 581-92.
28. Herceg, Z., W. Hulla, D. Gell, C. Cuenin, M. Lleonart, S. Jackson, and Z.Q. Wang, *Disruption of Trrap causes early embryonic lethality and defects in cell cycle progression.* Nat Genet, 2001. **29**(2): p. 206-11.
29. Hildebrand, J.D. and P. Soriano, *Overlapping and unique roles for C-terminal binding protein 1 (CtBP1) and CtBP2 during mouse development.* Mol Cell Biol, 2002. **22**(15): p. 5296-307.
30. Wang, J., K. Scully, X. Zhu, L. Cai, J. Zhang, G.G. Prefontaine, A. Krones, K.A. Ohgi, P. Zhu, I. Garcia-Bassets, F. Liu, H. Taylor, J. Lozach, F.L. Jayes, K.S. Korach*, et al.*, *Opposing LSD1 complexes function in developmental gene activation and repression programmes.* Nature, 2007. **446**(7138): p. 882-7.
31. Wang, J., S. Hevi, J.K. Kurash, H. Lei, F. Gay, J. Bajko, H. Su, W. Sun, H. Chang, G. Xu, F. Gaudet, E. Li, and T. Chen, *The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation.* Nat Genet, 2009. **41**(1): p. 125-9.
32. Zhang, J., W.L. Tam, G.Q. Tong, Q. Wu, H.Y. Chan, B.S. Soh, Y. Lou, J. Yang, Y. Ma, L. Chai, H.H. Ng, T. Lufkin, P. Robson, and B. Lim, *Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1.* Nat Cell Biol, 2006. **8**(10): p. 1114-23.
33. Elling, U., C. Klasen, T. Eisenberger, K. Anlag, and M. Treier, *Murine inner cell mass-derived lineages depend on Sall4 function.* Proc Natl Acad Sci U S A, 2006. **103**(44): p. 16319-24.
34. Yuri, S., S. Fujimura, K. Nimura, N. Takeda, Y. Toyooka, Y. Fujimura, H. Aburatani, K. Ura, H. Koseki, H. Niwa, and R. Nishinakamura, *Sall4 is essential for stabilization, but not for pluripotency, of embryonic stem cells by repressing aberrant trophectoderm gene expression.* Stem Cells, 2009. **27**(4): p. 796-805.
35. Nishinakamura, R., Y. Matsumoto, K. Nakao, K. Nakamura, A. Sato, N.G. Copeland, D.J. Gilbert, N.A. Jenkins, S. Scully, D.L. Lacey, M. Katsuki, M. Asashima, and T. Yokota, *Murine homolog of SALL1 is essential for ureteric bud invasion in kidney development.* Development, 2001. **128**(16): p. 3105-15.
36. Ding, L., M. Paszkowski-Rogacz, A. Nitzsche, M.M. Slabicki, A.K. Heninger, I. de Vries, R. Kittler, M. Junqueira, A. Shevchenko, H. Schulz, N. Hubner, M.X. Doss, A. Sachinidis, J. Hescheler, R. Iacone*, et al.*, *A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity.* Cell Stem Cell, 2009. **4**(5): p. 403-15.
37. Parekh, V., A. McEwen, V. Barbour, Y. Takahashi, J.E. Rehg, S.M. Jane, and J.M. Cunningham, *Defective extraembryonic angiogenesis in mice lacking LBP-1a, a member of the grainyhead family of transcription factors.* Mol Cell Biol, 2004. **24**(16): p. 7113-29.
38. Takebe, A., T. Era, M. Okada, L. Martin Jakt, Y. Kuroda, and S. Nishikawa, *Microarray analysis of PDGFR alpha+ populations in ES cell differentiation culture identifies genes involved in differentiation*

of mesoderm and mesenchyme including ARID3b that is essential for development of embryonic mesenchymal cells. Dev Biol, 2006. **293**(1): p. 25-37.

39. Masui, S., Y. Nakatake, Y. Toyooka, D. Shimosato, R. Yagi, K. Takahashi, H. Okochi, A. Okuda, R. Matoba, A.A. Sharov, M.S. Ko, and H. Niwa, *Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells.* Nat Cell Biol, 2007. **9**(6): p. 625-35.
40. Avilion, A.A., S.K. Nicolis, L.H. Pevny, L. Perez, N. Vivian, and R. Lovell-Badge, *Multipotent cell lineages in early mouse development depend on SOX2 function.* Genes Dev, 2003. **17**(1): p. 126-40.
41. Wang, J., S. Rao, J. Chu, X. Shen, D.N. Levasseur, T.W. Theunissen, and S.H. Orkin, *A protein interaction network for pluripotency of embryonic stem cells.* Nature, 2006. **444**(7117): p. 364-8.
42. Mackler, S., A. Pacchioni, R. Degnan, Y. Homan, A.C. Conti, P. Kalivas, and J.A. Blendy, *Requirement for the POZ/BTB protein NAC1 in acute but not chronic psychomotor stimulant response.* Behav Brain Res, 2008. **187**(1): p. 48-55.
43. Ivanova, N., R. Dobrin, R. Lu, I. Kotenko, J. Levorse, C. DeCoste, X. Schafer, Y. Lun, and I.R. Lemischka, *Dissecting self-renewal in stem cells with RNA interference.* Nature, 2006. **442**(7102): p. 533-8.
44. Oka, C., T. Nakano, A. Wakeham, J.L. de la Pompa, C. Mori, T. Sakai, S. Okazaki, M. Kawaichi, K. Shiota, T.W. Mak, and T. Honjo, *Disruption of the mouse RBP-J kappa gene results in early embryonic death.* Development, 1995. **121**(10): p. 3291-301.
45. Luo, J., R. Sladek, J.A. Bader, A. Matthyssen, J. Rossant, and V. Giguere, *Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-beta.* Nature, 1997. **388**(6644): p. 778-82.
46. Loh, Y.H., Q. Wu, J.L. Chew, V.B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K.Y. Wong, K.W. Sung, C.W. Lee, X.D. Zhao, K.P. Chiu*, et al.*, *The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells.* Nat Genet, 2006. **38**(4): p. 431-40.
47. Dejosez, M., J.S. Krumenacker, L.J. Zitur, M. Passeri, L.F. Chu, Z. Songyang, J.A. Thomson, and T.P. Zwaka, *Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells.* Cell, 2008. **133**(7): p. 1162-74.
48. Niakan, K.K., E.C. Davis, R.C. Clipsham, M. Jiang, D.B. Dehart, K.K. Sulik, and E.R. McCabe, *Novel role for the orphan nuclear receptor Dax1 in embryogenesis, different from steroidogenesis.* Mol Genet Metab, 2006. **88**(3): p. 261-71.
49. Chen, X., F. Fang, Y.C. Liou, and H.H. Ng, *Zfp143 regulates Nanog through modulation of Oct4 binding.* Stem Cells, 2008. **26**(11): p. 2759-67.
50. Wang, Z.G., L. Delva, M. Gaboli, R. Rivi, M. Giorgio, C. Cordon-Cardo, F. Grosveld, and P.P. Pandolfi, *Role of PML in cell growth and the retinoic acid pathway.* Science, 1998. **279**(5356): p. 1547-51.
51. Li, S., D. Zhou, M.M. Lu, and E.E. Morrisey, *Advanced cardiac morphogenesis does not require heart tube fusion.* Science, 2004. **305**(5690): p. 1619-22.
52. Tarleton, H.P. and I.R. Lemischka, *Delayed differentiation in embryonic stem cells and mesodermal progenitors in the absence of CtBP2.* Mech Dev, 2010. **127**(1-2): p. 107-119.
53. Shindo, T., I. Manabe, Y. Fukushima, K. Tobe, K. Aizawa, S. Miyamoto, K. Kawai-Kowase, N. Moriyama, Y. Imai, H. Kawakami, H. Nishimatsu, T. Ishikawa, T. Suzuki, H. Morita, K. Maemura*, et al.*, *Kruppel-like zinc-finger transcription factor KLF5/BTEB2 is a target for angiotensin II signaling and an essential regulator of cardiovascular remodeling.* Nat Med, 2002. **8**(8): p. 856-63.
54. de Wind, N., M. Dekker, A. Berns, M. Radman, and H. te Riele, *Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer.* Cell, 1995. **82**(2): p. 321-30.
55. O'Donnell, N., N.E. Zachara, G.W. Hart, and J.D. Marth, *Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability.* Mol Cell Biol, 2004. **24**(4): p. 1680-90.
56. Shafi, R., S.P. Iyer, L.G. Ellies, N. O'Donnell, K.W. Marek, D. Chui, G.W. Hart, and J.D. Marth, *The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny.* Proc Natl Acad Sci U S A, 2000. **97**(11): p. 5735-9.

4

Characterization of Mediator Complexes in Mouse Embryonic Stem Cells

Work in progress

Characterization of Mediator Complexes in Mouse Embryonic Stem Cells

Debbie L.C. van den Berg,[1] Karel Bezstarosti,[2] Jeroen Demmers,[2] and Raymond A. Poot[1]

[1]Department of Cell Biology, [2]Proteomics Center, Erasmus MC, Dr.Molewaterplein 50, 3015GE Rotterdam, The Netherlands

ABSTRACT

Transcription factors Oct4 and Esrrb are important for establishing and maintaining ES cell pluripotency. Genomic binding sites and regulated genes have been described for both factors. We recently established physical interactions between Oct4 and Esrrb and between Esrrb and components of the basal transcription machinery, including Mediator complex, RNA pol2 and TFIID. Here we purified the ES cell Mediator complex and show that only two transcription factors, Esrrb and Obox4, reproducibly associate with core Mediator. Purification of the Med12/Cdk8-kinase submodule that optionally associates with core Mediator revealed a number of unique interactors not identified in core Mediator preparations, including the PCAF/SAGA complex and ubiquitin E3 ligases SCF and APC/C. These interactions suggest a role for the Med12/Cdk8 module as a substrate priming kinase for SCF and APC/C activity that could limit unlicensed transcriptional noise at tissue-specific enhancer elements.

INTRODUCTION

Embryonic stem (ES) cells are derived from the inner cell mass of blastocyst stage embryos and can be cultured *in vitro* indefinitely while retaining the ability to differentiate into derivatives of any of the three germ layers. ES cell maintenance additionally relies on a core transcriptional circuitry, set around the key pluripotency factors Oct4, Sox2 and Nanog[1]. An RNAi screen aimed at identifying novel regulators of ES cell maintenance further expanded the core transcription factor network and demonstrated an essential role for Esrrb in preservation of pluripotency[2]. Esrrb is an orphan nuclear receptor related to the estrogen receptor that physically interacts with Oct4 and Nanog[3-4]. In murine ES cell colonies it is expressed in a mosaic fashion and expression levels correlate with those of several other transcription factors that are presumed to mark the pluripotent ground state, such as Nanog, Zfp42 and Tbx3[3, 5-6]. Overexpression of Esrrb sustains ES cell self-renewal in the absence of exogenously added LIF[7]. In addition, Esrrb can substitute for Klf4 in somatic cell reprogramming[8]. Despite identification of increasing numbers of transcription factors that contribute to maintenance of the pluripotent state, relatively little is known about how these factors communicate to the basal transcription machinery. We have recently identified an association between Esrrb and the Mediator complex, RNA pol2 and TFIID[9]. Mammalian Mediator constitutes a 26 subunit, 1.2 MDa complex that associates with RNA pol2 to stimulate basal and activator-dependent transcription by enhancing RNA pol2 recruitment and stimulating phosphorylation of its C-terminal domain, a prerequisite for productive transcription elongation[10]. A kinase submodule comprised of Med12, Med13, Cdk8 and Cyclin C optionally associates with

core Mediator and exerts both positive and negative effects on its coactivator function[11-12].

We employed a FLAG-affinity based purification protocol to characterize Mediator complexes in ES cells. FLAG-tagged Med15 copurified all known Mediator subunits, including all four components of the kinase submodule. We found merely two transcription factors, Esrrb and Obox4, consistently associated with ES cell Mediator. In addition we purified Mediator via Med12 and find a number of cofactors, including the PCAF/SAGA complex and ubiquitin E3 ligases SCF and APC/C, to uniquely copurify with kinase submodule-containing Mediator. We propose a role for the Med12/Cdk8 submodule as a substrate priming kinase for ubiquitin E3 ligase activity.

RESULTS

Purification of core Mediator from ES cells

To identify Mediator complexes and associated proteins we introduced an expression construct containing FLAG-tagged Med15 into mouse ZHBTc4 ES cells[13]. Med15 is part of the core Mediator tail domain and by purifying Mediator via this subunit we expect to capture all complexes present in ES cells. FLAG-Med15 expressing cells were expanded, nuclear extracts were prepared and FLAG-Med15 complexes were purified using FLAG affinity (Figure 1A). Western blot analysis of FLAG-affinity resin bound complexes with Med12 antibody suggested incorporation of FLAG-Med15 into Mediator complexes (Figure 1B). Purified fractions were separated on a polyacrylamide gel and stained with Colloidal Coomassie (Figure 1C). A prominent band was detected that corresponds to the size of FLAG-Med15 and reacts with FLAG antibody (Figure 1A). Despite the fact that Med15 constitutes a stoichiometric subunit of Mediator, no other bands of similar intensity were detected, suggesting our FLAG-Med15 is overexpressed compared to the low abundant endogenous Mediator complex. Our control purification on the parental ZHBTc4 ES cell line, apart from one common background band, is otherwise devoid of any major contaminants. To identify interacting factors, two independent FLAG-Med15 purifications were analyzed by mass spectrometry. Proteins present in both purifications and not in the controls were considered to be bona fide interactors and are summarized in Table 1. Except for Med9 and Med12l, which were detected in only one FLAG-purification, we find all Mediator core subunits and all components of the kinase submodule to specifically co-purify with FLAG-Med15. In addition components of the basal transcription machinery, i.e. RNA pol2 (7 subunits), TFIIF (2 subunits) and TAF9, are found to be associated with Mediator. We further identified proteins involved in transcription regulation (e.g. SWI/SNF and NuRD subunits) and RNA processing (e.g. Prpf4, Raly). Notably only two sequence specific transcription factors were consistently present in the FLAG-Med15 purification, namely Esrrb and Obox4.

Purification of endogenous kinase module-Mediator complexes from ES cells

To independently verify the observed interactors of ES cell Mediator, we conducted an immunoprecipitation of endogenous Mediator complexes from a different ES cell line, 46C[14], using

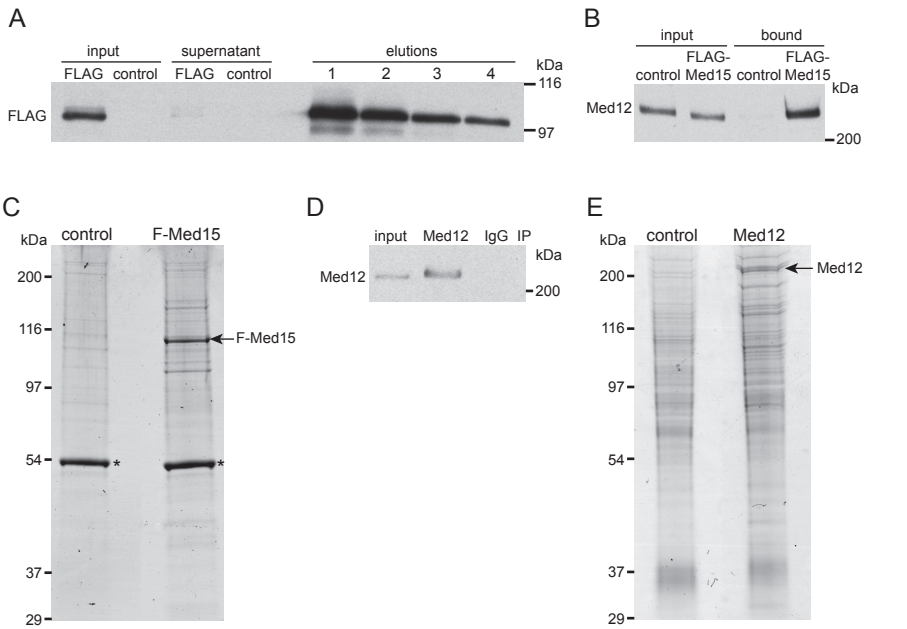


Figure 1. Purification of Mediator complexes
(A) FLAG-Med15 is depleted from nuclear extract of FLAG-Med15 expressing ES cells and eluted by addition of 3xFLAG-peptide. (B) FLAG-Med15 purification analyzed by western blot with Med12 antibody. (C) Colloidal Coomassie stained polyacrylamide gel of FLAG-Med15 (F-Med15) and control purification. FLAG-Med15 band is indicated, * designates contaminating band. Note additional bands in FLAG-Med15 sample (D) Med12 and control IgG immunoprecipitates analyzed by western blotting with Med12 antibody. (E) Colloidal Coomassie stained polyacrylamide gel of Med12 and control IgG immunoprecipitates. Band corresponding to Med12 is indicated.

an antibody directed against the kinase module subunit Med12 that efficiently precipitates Med12 protein (Figure 1D). As is evident from the Colloidal Coomassie stained polyacrylamide gel this approach results in a higher background, as bound complexes cannot be peptide-eluted from the FLAG-affinity beads (Figure 1E). The band that most likely represents Med12 is indicated by an arrow. Immunoprecipitated complexes were analyzed by mass spectrometry. Except for Med25 all Mediator subunits that were identified in the FLAG-Med15 purification were also found to specifically co-purify with Med12 (Table 1). Inherent to this approach several subunits, including Med12 itself, were present in the control purification, albeit at lower levels. In addition, high background levels hampered independent verification of several interactions identified based on FLAG-affnity purification (designated NS in Table 1). However, association of both Esrrb and Obox4 with Mediator could be confirmed.

Interaction with kinase module alters the stochiometry within Mediator

It has been suggested that stoichiometry of individual subunits within Mediator alters as a consequence of kinase submodule association and in particular that association of Med26 and the

Table 1. FLAG-Med15 and Med12 interacting proteins as identified by Mass Spectrometry analysis

Protein	Accession	FLAG-Med15 #1		FLAG-Med15 #2		Avg. score	Med12 IP	
		Mascot score[a]	Unique pept.[b]	Mascot score[a]	Unique pept.[b]		Mascot score[a]	Unique pept.[b]
Mediator complex								
Med12	IPI00620781	2832	57	2200	46	2516	3531(1502)	57(35)
Med14	IPI00417070	2375	41	2331	41	2353	2489(186)	43(7)
Med23	IPI00606087	1970	39	1628	30	1799	1790(300)	36(7)
Med15	IPI00462457	1648	27	1690	25	1669	937	18
Med1	IPI00648562	1487	29	1600	30	1544	2038	39
Med24	IPI00165717	1457	27	1525	25	1491	1462	26
Med16	IPI00352130	1333	26	1176	22	1255	1054	19
Med17	IPI00459787	1146	19	1304	22	1225	1282	21
Med13	IPI00118296	1407	28	1039	23	1223	1824(368)	31(9)
Med13l	IPI00420457	1066	24	1010	24	1038	2217(381)	40(10)
Med27	IPI00119036	709	13	835	14	772	809	13
Med26	IPI00354323	741	16	620	12	681	150	4
Med4	IPI00132328	574	9	560	9	567	776	10
Med6	IPI00177199	594	12	516	11	555	557	12
Med25	IPI00139172	537	9	529	8	533	416	7
Cdk8	IPI00405475	558	13	471	11	515	807	18
Med20	IPI00128183	485	9	460	8	473	447	8
Ccnc	IPI00228171	406	8	359	6	383	703	11
Med8	IPI00119169	377	8	276	6	327	180	4
Med30	IPI00132580	347	9	287	5	317	247	4
Med18	IPI00112230	311	6	242	4	277	181	4
Med29	IPI00119185	270	6	272	5	271	252	5
Med28	IPI00471067	229	5	288	5	259	240	5
Med31	IPI00321251	164	4	339	6	252	190	4
Med22	IPI00123503	315	7	179	4	247	158	3
Med7	IPI00331242	219	5	183	3	201	328	6
Med10	IPI00377931	181	5	209	3	195	200	3
Med19	IPI00224399	187	4	165	4	176	444	9
Med11	IPI00111353	111	3	155	2	133	109	2
Med21	IPI00131967	109	2	54	1	82	ND	ND
RNA polymerase II								
Polr2a	IPI00136207	2111	52	1784	40	1948	NS	NS
Polr2b	IPI00320034	1438	33	1306	26	1372	233(62)	8(2)
Polr2g	IPI00263106	254	6	407	8	331	ND	ND
Polr2c	IPI00129026	274	5	381	7	328	NS	NS
Polr2e	IPI00337955	326	7	204	3	265	181	4
Polr2h	IPI00124284	193	4	128	3	161	ND	ND
Polr2l	IPI00850336	80	2	102	2	91	ND	ND
General transcription factors								
Gtf2f2	IPI00135812	293	7	510	11	402	ND	ND
Gtf2f1	IPI00153986	109	2	202	4	156	ND	ND
Taf9	IPI00128702	142	4	154	5	148	NS	NS
SWI/SNF complex								
Smarca4	IPI00460668	485	11	486(88)	13(3)	486	NS	NS
Smarcb1	IPI00129145	180	5	116	2	148	ND	ND
Dpf2	IPI00117727	71	2	94	2	83	NS	NS
NuRD complex								
Mta1	IPI00330304	151	5	229	6	190	NS	NS
Mbd3	IPI00131067	73	3	62	1	68	NS	NS

Table 1. ***Continued***

Protein	Accession	FLAG-Med15 #1		FLAG-Med15 #2		Avg. score	Med12 IP	
		Mascot score[a]	Unique pept.[b]	Mascot score[a]	Unique pept.[b]		Mascot score[a]	Unique pept.[b]
Transcription factors								
Esrrb	IPI00752694	133	3	79	3	106	238(64)	5(2)
Obox4	IPI00128293	69	3	91	3	80	92	3
Transcription related factors								
Tcea3	IPI00137142	202	4	85	2	144	NS	NS
Hic2	IPI00267297	155	4	99	2	127	NS	NS
Fubp3	IPI00379513	92	3	108	3	100	NS	NS
RNA processing								
Prpf4	IPI00458908	227	6	221(43)	5(1)	224	NS	NS
Ints1	IPI00877221	257	7	169	5	213	NS	NS
Raly	IPI00130147	229	6	145	4	187	NS	NS
Lrpprc	IPI00420706	109	4	57	2	83	ND	ND
Other								
Trim11	IPI00480525	526	9	732	12	629	ND	ND
Ppp1ca	IPI00130185	218	6	170(43)	4(1)	194	NS	NS
L1td1	IPI00378700	157	5	180	4	169	NS	NS
Ogt	IPI00420870	214(48)	7(1)	114	3	164	NS	NS
Rfc4	IPI00319874	98	3	90	3	94	NS	NS
4933424B01Rik	IPI00154065	67	2	76	3	72	56	2

[a] Mascot score for the specified protein in the FLAG-Med15 or Med12 IP sample. Mascot score for corresponding protein in control purification between brackets.
[b] Number of unique, non-redundant peptides for the specified protein in the FLAG-Med15 or Med12 IP sample. Number of unique, non-redundant peptides in the control purification between brackets.
NS, not specific; ND, not detected.

kinase unit are mutually exclusive[15-16]. To address whether similar compositional changes occur in ES cell Mediator we calculated emPAI values of all subunits identified in the three independent purifications. emPAI scores reflect the number of identified peptides normalized to the theoretical number of peptides and therefore constitute a quantitative measure far superior to spectral counts[17]. emPAI score and relative contribution of each individual subunit are summarized in Table 2. Overall subunit composition is similar between the independently isolated Mediator complexes. Some differences between the individual FLAG-purifications exist, for example unique presence of Med9 and an increase in relative amounts Med31 and Med28 in the second FLAG-Med15 isolation. As expected, we detect an elevated proportionate contribution of Cdk8 and Cyclin C in Mediator complexes purified via its Med12 subunit. This is accompanied by a relative decrease in Med26 incorporation and strongly elevated levels of Med19.

Kinase submodule-specific interactions

In addition to the Med12-interacting proteins summarized in Table 1, a large number of factors copurified with Med12 that were not detected in any of the Med15 purifications. Represented in Table 3, these entail additional general transcription factors and a number of proteins involved in transcription regulation, such as several subunits of the Trrap/p400 and PCAF/STAGA chromatin

Table 2. Quantification of Mediator subunits in FLAG-Med15 and Med12 immunopurifications

Mediator subunit	FLAG-Med15 #1		FLAG-Med15 #2		Med12 IP	
	emPAI[a]	Contribution[b]	emPAI[a]	Contribution[b]	emPAI[a]	Contribution[b]
Med1	1.28	1.86	1.35	2.04	1.58	2.75
Med4	7.06	10.26	4.69	7.10	6.18	10.74
Med6	2.94	4.27	3.47	5.25	3.62	6.29
Med7	0.91	1.32	0.68	1.03	1.66	2.89
Med8	3.44	5.00	1.58	2.39	1.45	2.52
Med9	ND	-	3.69	5.59	ND	-
Med10	0.55	0.80	1.42	2.15	1.31	2.28
Med11	1.84	2.67	1.19	1.80	1.10	1.91
Med12	1.59	2.31	1.27	1.92	1.78	3.09
Med12l	0.11	0.16	ND	-	ND	-
Med13	0.55	0.80	0.41	0.62	0.71	1.23
Med13l	0.50	0.73	0.50	0.76	0.85	1.48
Med14	2.59	3.76	2.93	4.44	2.62	4.55
Med15	7.13	10.36	7.13	10.80	1.55	2.69
Med16	1.74	2.53	1.19	1.80	1.17	2.03
Med17	2.29	3.33	2.63	3.98	2.07	3.60
Med18	2.29	3.33	0.81	1.23	0.76	1.32
Med19	0.96	1.39	1.25	1.89	4.69	8.15
Med20	5.02	7.29	3.46	5.24	2.57	4.47
Med21	0.25	0.36	0.25	0.38	ND	-
Med22	5.75	8.35	1.89	2.86	0.56	0.97
Med23	2.02	2.93	1.63	2.47	1.54	2.68
Med24	2.03	2.95	1.74	2.63	1.75	3.04
Med25	0.91	1.32	0.82	1.24	0.61	1.06
Med26	1.57	2.28	1.18	1.79	0.23	0.40
Med27	3.10	4.50	4.01	6.07	5.08	8.83
Med28	1.92	2.79	4.00	6.06	1.76	3.06
Med29	2.73	3.97	2.73	4.13	1.97	3.42
Med30	1.36	1.98	0.98	1.48	0.91	1.58
Med31	1.39	2.02	4.71	7.13	1.81	3.15
Cdk8	1.40	2.03	1.10	1.67	3.31	5.75
Ccnc	1.61	2.34	1.35	2.04	2.33	4.05

[a] emPAI score for specified Mediator subunit.
[b] Relative contribution of specified subunit to Mediator complex.
ND, not detected

modifying complexes and known nuclear receptor co-regulators (e.g. NCoR complex, Ncoa5, Jmjd1c). Apart from Esrrb and Obox4, we detect additional DNA-binding transcription factors in the kinase module-enriched Mediator fraction, such as Sox2, Mga and Foxp1. We further find factors involved in nuclear organization (i.e. cohesins and nuclear pore complex components), RNA processing, DNA repair and cell cycle regulation to be present in the Med12 purification. Intriguingly, two major RING domain ubiquitin E3 ligase complexes that control degradation of cell cycle regulators interact with Mediator's kinase submodule. Skp1a and an F-box protein assemble onto a cullin family scaffold protein to form the substrate-recognition module of the SCF (Skp1/Cul1/F-box) ubiquitin ligase complex. The mouse genome contains 74 F-box proteins[18],

Table 3. Med12-interacting proteins as Identified by Mass Spectrometry Analysis.

Protein	Accession	Med12 IP	
		Mascot score[a]	Unique pept.[b]
General transcription factors			
Gtf2i	IPI00229538	1906(565)	35(14)
Taf1	IPI00330385	142	5
Gtf2h2	IPI00279701	79	2
Tbp	IPI00120505	70	2
Taf8	IPI00459354	64	2
Trrap/p400 complex			
Trrap	IPI00330902	1463(219)	37(7)
Ep400	IPI00229659	661(105)	16(3)
Morf4l1	IPI00553815	312	7
Yeats4	IPI00132946	139	4
PCAF/SAGA complex			
Tada1	IPI00117034	269	7
Crebbp	IPI00751842	195	7
Tada3	IPI00153220	192	5
Taf5l	IPI00128308	180	4
Ccdc101	IPI00320317	132	3
Taf6l	IPI00153596	127	4
Supt7l	IPI00113175	100	3
Kat2b	IPI00471164	66	2
NCoR complex			
NCoR1	IPI00274795	279	7
Tbl1xr1	IPI00308283	275	6
NCoR2	IPI00123871	246	6
Tbl1x	IPI00223056	233	5
Transcription factors			
Dido1	IPI00227469	182	6
Zfp592	IPI00308391	159	5
Sox2	IPI00830717	152	4
Zfp687	IPI00134718	150	4
Mga	IPI00135072	130	4
Zfp655	IPI00112984	113	3
Ubp1	IPI00330019	108	2
Zfp462	IPI00467729	99	4
Zfp532	IPI00411014	88	2
Foxp1	IPI00230542	85	3
Transcription regulation			
Jmjd1c	IPI00755780	386(104)	13(3)
Mta3	IPI00221805	260	5
Safb	IPI00944159	232	6
Ep300	IPI00461822	173	6
Jarid2	IPI00124575	164(48)	5(1)
Dnmt1	IPI00469323	163(44)	3(1)
Cecr2	IPI00344641	152	4
Mbtd1	IPI00420364	148	3
Prmt5	IPI00229845	126	3
Ssbp3	IPI00341944	81	2
Chd8	IPI00407590	70	2

Table 3. *Continued*

Protein	Accession	Med12 IP	
		Mascot score[a]	Unique pept.[b]
Transcription regulation (*continued*)			
Setd1a	IPI00323238	67	2
Ncoa5	IPI00313525	65	2
Epc2	IPI00223821	61	2
Paxip1	IPI00421197	60	2
Anaphase Promoting Complex/Cyclosome			
Anapc1	IPI00930890	1566(79)	32(3)
Anapc7	IPI00331074	1269	20
Cdc23	IPI00221793	1204	24
Anapc5	IPI00380355	1202	21
Anapc2	IPI00227654	1069	22
Anapc4	IPI00127493	1014	20
Cdc27	IPI00461309	1004	17
Cdc16	IPI00221621	940	18
Ube2s	IPI00121891	279	5
Anapc10	IPI00108367	260	6
Fzr1	IPI00128734	242	6
Cdc26	IPI00115426	105	2
Fbxo5	IPI00132013	185	6
Cdc20	IPI00320406	94	2
SCF complex			
Fbxl19	IPI00420711	445	9
Fbxw7	IPI00162935	168	5
Skp1a	IPI00331163	108	2
Cohesin complex			
Nipbl	IPI00357096	257	7
Pds5a	IPI00669709	205(53)	6(1)
Cdca5	IPI00132175	106	2
Pds5b	IPI00317401	82	3
Nuclear pore complex			
Nup214	IPI00229722	90	4
Sec13	IPI00133920	81	2
Nup93	IPI00222307	80	3
Nup88	IPI00229021	67	2
Nup107	IPI00221767	62	2
DNA repair			
Atrx	IPI00322707	184	4
Apex1	IPI00224152	137	4
Xrcc1	IPI00118139	61	2
Cell cycle			
Cdk1	IPI00114491	313(49)	8(2)
Rcc2	IPI00222509	158	4
Orc5l	IPI00125261	111	3
Bub1b	IPI00353560	107	3
RNA processing			
Ints6	IPI00461472	252(74)	6(1)
Ints9	IPI00223422	168	4
Wbp11	IPI00123333	116	3

Table 3. ***Continued***

Protein	Accession	Med12 IP	
		Mascot score[a]	Unique pept.[b]
RNA processing (*continued*)			
Pum2	IPI00112203	96	3
Cpsf3l	IPI00467084	95	3
Upf1	IPI00420949	96	2
Sf3a2	IPI00380281	67	2
Ints5	IPI00229728	62	2
Other			
Kif4	IPI00881456	1421	28
C130039O1	IPI00225777	672(132)	12(3)
Cdk19	IPI00226248	431	9
Ppp1cb	IPI00311873	321	7
Zmynd8	IPI00108978	228	5
Hspa1a	IPI00798482	177	4
Ranbp2	IPI00337844	176	5
LOC627816	IPI00752131	150	4
Rprd1b	IPI00338515	122	2
Gltscr1	IPI00340401	111	4
2210018M11	IPI00416315	93	3
Smarcad1	IPI00223926	93	3
2310033P09	IPI00117270	88	2
Dock6	IPI00222462	81	3
Ppp2r1b	IPI00222306	80	2
5730528L13	IPI00112237	77	3
Zc3h4	IPI00625723	75	3
Mett10d	IPI00387384	74	2
Ubap2	IPI00457533	72	2
2810046L04	IPI00227274	71	2
2610301G19	IPI00123624	64	2
Csnk1d	IPI00138790	62	2

[a] Mascot score for the specified protein in the Med12 IP sample. Mascot score for corresponding protein in control purification, if present, between brackets.
[b] Number of unique, non-redundant peptides for the specified protein in the Med12 IP sample. Number of unique, non-redundant peptides in the control purification, if present, between brackets.

two of which (i.e. Fbxl19 and Fbxw7) interact with Med12. Furthermore, we detect 10 out of 12 core APC/C (Anaphase Promoting Complex/Cyclosome) E3 ligase subunits and its two alternative substrate recognizing cofactors (i.e. Cdc20 and Fzr1) in the Med12 immunoprecipitation. We additionally find Fbxo5 (also known as Emi1), an inhibitor of APC/C activity, associated with Med12-bound APC/C.

DISCUSSION

Transcription factors Esrrb and Obox4 interact with ES cell Mediator

We had previously identified an interaction between nuclear receptor Esrrb and the Mediator complex[9]. The reverse approach, purification of Mediator complexes from ES cell nuclear extracts, showed that Esrrb is one of just two transcription factors that reproducibly co-purified with

Mediator. Analogous to the estrogen receptor in mammary cells[19], Esrrb may activate target gene transcription by recruitment of Mediator and its associated TFIIF-RNA pol2 complex in ES cells. Mediator-recruitment by nuclear receptors has been shown to depend on coactivators, such as CCAR1 and PGC-1α[20-21]. Interaction between Esrrb and coactivators NCoA2 and NCoA3 has been reported[9, 22]. However, despite consistent Esrrb presence in FLAG-Med15 purifications we failed to detect stable association of nuclear receptor coregulators. Although this may be a consequence of the detection method (e.g. incapacity to generate a sufficient amount of tryptic peptides), it is also possible that Mediator binding to Esrrb per se does not require the presence of a coactivator.

Obox4 is the second transcription factor consistently associated with ES cell Mediator complexes. It belongs to the oocyte-specific-homeobox gene family and is preferentially expressed in ovaries, testis and oocytes[23]. Obox4 and Obox6 transcripts are detected in mESCs and microarray analysis of Obox4-overexpressing mESCs has suggested a role in transcription regulation of histone genes[24], although overexpression also induces ES cell differentiation.

Chromatin modifying complexes associate with Med12-kinase module

Med12 is part of the Cdk8-kinase submodule that can be found as a free entity or associated with core Mediator[25]. We have identified a large number of factors in the Med12-enriched fraction that were not present in the FLAG-Med15 purifications, which may reflect enrichment for specific Mediator subcomplexes or represent interactions of free kinase submodules. On the other hand, the use of a polyclonal antibody for complex purification demands careful interpretation of detected interactions due to possible cross-reactivity. However, several of the Med12-associated proteins identified here have independently been demonstrated to co-purify with Cdk8-Mediator complexes, which strengthens the validity of our results. For example biochemical purification of Mediator complexes exposed Trrap and the histone acetyltransferase Gcn5 as stable partners of Cdk8-containing Mediator[26]. This so called T/G-Mediator complex can phosphoacetylate histone H3 in chromatin templates, but does not support activator-induced transcription activation[26]. Here we identify Trrap as Med12-associated factor and in addition demonstrate association of other subunits of the Trrap-p400 complex. This complex catalyzes deposition of histone variant H2A.Z and in ESCs is found at over half of all promoter regions[27]. Targeting to these regions is dependent on H3K4me3, a hallmark of active promoters. Protein-protein interaction between subunits of Trrap-p400 and Mediator complexes may further stabilize binding at promoters actively engaged in transcription. Trrap was recently found specifically associated with activator-bound Mediator and absent from core Mediator preparations[28]. Indeed in our mild purification conditions we find several potential transcription activators associated with Med12-Mediator (e.g. Sox2, Mga, Ubp1).

Apart from Trrap-p400 we have shown that the histone acetyltransferase PCAF/SAGA complex interacts with Med12. In yeast binding of Cdk8-Mediator to the *GAL1* upstream activating sequence requires SAGA-subunit Spt3, but is independent of SAGA-incorporated HAT-activity[29]. Mediator recruitment to Myc target genes in mammalian cells is likewise dependent on the SAGA/STAGA subunit Supt7l (also known as STAF65γ) and downregulation of Supt7l affects gene

expression post-recruitment of the basal transcription machinery[30]. Association between these two complexes is evolutionary conserved, indicating functional importance of this interaction in transcription regulation.

Med12 interacts with nuclear receptor corepressor complex

In addition to factors involved in transcription activation, we also detected multiple subunits of the nuclear receptor corepressor (NCoR/SMRT) complexes among the Med12-interacting proteins. In several instances Cdk8-Mediator, in contrast to core Mediator, has been implicated in transcription repression. Association of the Med12-Med13-Cdk8-Cyclin C submodule with Mediator can induce a structural shift that impedes interaction with RNA pol2 and negatively affects both PIC-assembly and transcription re-initiation[12]. In addition, phosphorylation of the Cdk7/Cyclin H kinase module of TFIIH by Cdk8 interferes with the ability of Cdk7 to phosphorylate and thereby activate RNA pol2[31]. NCoR complexes associate with unliganded nuclear receptors to prevent association of coactivators, but also actively participate in removal of co-repressors in a process known as derepression[32]. The latter involves Tbl1x/Tbl1xr1-mediated recruitment of the ubiquitin conjugating enzyme E2 D1 and the 19S proteasome, resulting in co-repressor removal[33]. Association of Mediator, possibly in an inactive form, with transcription factor-bound NCoR complexes could ensure rapid transcription activation upon elimination of repressor proteins.

Ubiquitin E3 ligase complexes associate with Med12-Cdk8 kinase module

We identified an interaction between Med12-complexes and two ubiquitin E3 ligases, the SKP1/CUL1/F-box (SCF) complex and its related Anaphase Promoting complex/Cyclosome (APC/C). These factors function at the final substrate recognition step of the ubiquitination cascade carried out by consecutive action of E1, E2 and E3 enzymes. Both complexes form multisubunit RING E3 ligases, which do not actively participate in ubiquitination, but rather facilitate ubiquitin transfer by bringing ubiquitin conjugating E2 enzymes in close proximity to their substrates. Notably, we identified Skp1a and two substrate specific F-box proteins among the Med12-interactors, but failed to detect the CUL1 scaffold subunit of SCF. CUL1 requires covalent attachment of the ubiquitin-like protein NEDD8 to form an active complex that can bind the Skp1a/F-box subcomplex[34]. The NEDD8 moiety can be removed by the COP9 signalosome, which is likely to occur during our nuclear extract preparation and protein purification procedure and would explain why we were unable to detect CUL1. Of the two detected substrate-recognizing subunits Fbxw7 is the best studied. Its targets include cyclin E, c-Myc, c-Jun and Notch and genetic ablation of *Fbxw7* results in aberrant cell cycle entry of hematopoietic stem cells as well as an incapacity to exit the cell cycle in immature T cells[35].

The role of RING E3 ligase APC/C is also closely related to cell cycle regulation. APC/C alternates between two adapter proteins for substrate recognition (i.e. Fzr1/Cdh1 and Cdc20), which are both found to interact with Med12. Cdc20 associates with APC/C during mitosis and initiates anaphase by targeting negative regulators of separase for degradation, allowing cohesion cleavage and sister chromatid separation[36]. Dephosphorylated Cdh1 subsequently binds APC/C in anaphase, induces mitotic exit and maintains G1 phase by targeting degradation of cyclins. APC/C additionally has

been attributed a role independent of cell cycle regulation, as it was demonstrated to suppress axonal growth in postmitotic neurons through degradation of transcription corepressor SnoN and inhibitor of differentiation (Id) proteins[37-38].

Association of ubiquitin E3 ligases with Mediator (sub)complexes has not been described before. Their identification in our purifications may be due to our relatively mild conditions or could reflect a unique feature of ES cells. The potential functional significance of such an interaction can be sought in the fact that ubiquitination often requires phosphorylation of specific consensus motifs in the substrate. For example c-Myc and Klf5, two transcription factors that together with Oct4 and Sox2 can drive somatic cell reprogramming, are bound and ubiquitinated by Fbxw7 upon phosphorylation of their respective phospho degrons[39-40]. Although glycogen synthase kinase (GSK) 3β acts as priming kinase for most Fbxw7 substrates[35], Cdk8 could also fulfill such a role. In fact the yeast Cdk8 ortholog, Srb10, has been shown to phosphorylate transcription factor Gcn4 and thereby prime it for degradation by SCF ubiquitin ligases[41]. Mammalian Cdk8 phosphorylates Smad1 and Smad3, enabling Smad-dependent transcription activation but simultaneously targeting them for proteasomal degradation[42]. Cdk8-Cyclin C furthermore has been demonstrated to phosphorylate Notch receptor intracellular domain, which induces Fbxw7-dependent ubiquitin-mediated degradation[43]. Direct interaction between the substrate priming kinase containing Med12-submodule and ubiquitin E3 ligases would enhance efficiency of regulated proteasomal degradation.

An interesting case of ubiquitin-targeted protein degradation in transcription regulation is the tissue-specific *λ5-VpreB1* locus[44]. This locus contains an intergenic enhancer harboring an early transcription competence mark (ETCM), which in ES cells is characterized by H3Ac and H3K4me2 modifications and binding of components of the basal transcription machinery (i.e. Taf10, RNA pol2) and sequence specific transcription factors (i.e. Sox2 and FoxD3)[45-46]. Sox2 and FoxD3 are thought to prime the enhancer for activation in later pro- and pre-B cell stages, in part by Sox2-dependent deposition of H3K4me2 marks. In ES cells intergenic transcripts are produced from the enhancer region, which normally is a characteristic of active enhancers. Addition of proteasomal inhibitors or knockdown of proteasome subunits resulted in increased binding of transcription activators (TBP, RNA pol2, Trrap) and elevated intergenic transcript levels, suggesting that transcription activity at the intergenic enhancer is limited by the proteasome[44]. Chromatin immunoprecipitation indeed demonstrated the localization of several proteasome subunits to the *λ5-VpreB1* enhancer[44]. Based on our Med12-interactome, in which we have identified ubiquitin E3 ligases and significant amounts of Sox2, we speculate that Sox2 may play a role in recruitment of combined Med12/Cdk8-kinase and ubiquitin E3 ligase activity to intergenic enhancer elements. Successive action of Cdk8 and E3 ligase would target transcription activators for proteasomal degradation, thereby constraining enhancer activity in ES cells, while simultaneously maintaining a primed state for future activation in differentiated cell types.

EXPERIMENTAL PROCEDURES

ES cell culture and DNA construct

Mouse ZHBTc4[13] or 46C[14] embryonic stem cells were cultured on gelatin coated dishes without feeders, as described[3]. The coding sequence of Med15 was amplified from full length cDNA clone IRAVp968H08129D (imaGenes) and inserted with a C-terminal double FLAG-tag in a pPyCAG-driven expression vector. ZHBTc4 ES cells were transfected with Lipofectamine 2000 (Invitrogen), clones were selected by 1 µg/ml puromycin and expression of the tagged protein was assessed by Western blotting with FLAG antibody (Sigma).

Protein purification

FLAG-Med15 expressing and control ZHBTc4 cells were expanded to five 14 cm dishes, washed with PBS, scraped off and collected in ice cold PBS with complete, EDTA-free protease inhibitors (Roche). Nuclear extracts were prepared as described[47] and dialyzed to C-100 (20 mM Hepes [pH7.6], 0.2 mM EDTA, 1.5 mM $MgCl_2$, 100 mM KCl, 20% glycerol). Details on protein purification have been described[9]. Briefly, 1.5 ml of nuclear extract was incubated with 60 µl M2 anti-FLAG agarose beads (Sigma) for 3 hrs at 4°C in the presence of 225 units of Benzonase (Novagen). Beads were washed five times for 5 min with C-100* (C-100 containing 0.02% NP-40) and bound proteins were eluted with 0.2 mg/ml 3xFLAG-peptide. Elutions were pooled, TCA precipitated, separated by polyacrylamide gel electrophoresis, stained with Colloidal Blue Staining kit (Invitrogen) and analyzed by mass spectrometry as previously described[9]. For Med12 immunoprecipitation 10 µg Med12 antibody (A300-774A, Bethyl laboratories) or rabbit IgG (sc-2027, Santa Cruz) was cross-linked to 60 µl protein A Sepharose beads (Amersham). Beads were blocked with 0.1 mg/ml insulin (Sigma), 0.2 mg/ml chicken egg albumin (Sigma) and 1% fish skin gelatin (Sigma), incubated with 1.5 ml 46C ES cell nuclear extract in the presence of Benzonase for 3 hrs at 4°C, washed five times for 5 min with C-100* and bound proteins were eluted in SDS loading dye, separated by polyacrylamide gel electrophoresis and analyzed by mass spectrometry.

Protein interaction criteria

To be included in Table 1, proteins had to be present in both independent FLAG-Med15 purifications with a minimum mascot score of 60 and at least three fold higher than the score in the corresponding control purification. Med12-interacting proteins were included in Table 3 based on a minimum mascot score of 60 and at least three fold enrichment over the corresponding control purification.

REFERENCES

1. Rao, S. and S.H. Orkin, *Unraveling the transcriptional network controlling ES cell pluripotency.* Genome Biol, 2006. **7**(8): p. 230.
2. Ivanova, N., R. Dobrin, R. Lu, I. Kotenko, J. Levorse, C. DeCoste, X. Schafer, Y. Lun, and I.R. Lemischka, *Dissecting self-renewal in stem cells with RNA interference.* Nature, 2006. **442**(7102): p. 533-8.
3. van den Berg, D.L., W. Zhang, A. Yates, E. Engelen, K. Takacs, K. Bezstarosti, J. Demmers, I. Chambers,

and R.A. Poot, *Estrogen-related receptor beta interacts with Oct4 to positively regulate Nanog gene expression.* Mol Cell Biol, 2008. **28**(19): p. 5986-95.
4. Wang, J., S. Rao, J. Chu, X. Shen, D.N. Levasseur, T.W. Theunissen, and S.H. Orkin, *A protein interaction network for pluripotency of embryonic stem cells.* Nature, 2006. **444**(7117): p. 364-8.
5. Nichols, J. and A. Smith, *Naive and primed pluripotent states.* Cell Stem Cell, 2009. **4**(6): p. 487-92.
6. Toyooka, Y., D. Shimosato, K. Murakami, K. Takahashi, and H. Niwa, *Identification and characterization of subpopulations in undifferentiated ES cell culture.* Development, 2008. **135**(5): p. 909-18.
7. Zhang, X., J. Zhang, T. Wang, M.A. Esteban, and D. Pei, *Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells.* J Biol Chem, 2008. **283**(51): p. 35825-33.
8. Feng, B., J. Jiang, P. Kraus, J.H. Ng, J.C. Heng, Y.S. Chan, L.P. Yaw, W. Zhang, Y.H. Loh, J. Han, V.B. Vega, V. Cacheux-Rataboul, B. Lim, T. Lufkin, and H.H. Ng, *Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb.* Nat Cell Biol, 2009. **11**(2): p. 197-203.
9. van den Berg, D.L., T. Snoek, N.P. Mullin, A. Yates, K. Bezstarosti, J. Demmers, I. Chambers, and R.A. Poot, *An Oct4-centered protein interaction network in embryonic stem cells.* Cell Stem Cell, 2010. **6**(4): p. 369-81.
10. Malik, S. and R.G. Roeder, *Dynamic regulation of pol II transcription by the mammalian Mediator complex.* Trends Biochem Sci, 2005. **30**(5): p. 256-63.
11. Belakavadi, M. and J.D. Fondell, *Cyclin-dependent kinase 8 positively cooperates with Mediator to promote thyroid hormone receptor-dependent transcriptional activation.* Mol Cell Biol, 2010. **30**(10): p. 2437-48.
12. Knuesel, M.T., K.D. Meyer, C. Bernecky, and D.J. Taatjes, *The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function.* Genes Dev, 2009. **23**(4): p. 439-51.
13. Niwa, H., J. Miyazaki, and A.G. Smith, *Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells.* Nat Genet, 2000. **24**(4): p. 372-6.
14. Ying, Q.L., M. Stavridis, D. Griffiths, M. Li, and A. Smith, *Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture.* Nat Biotechnol, 2003. **21**(2): p. 183-6.
15. Paoletti, A.C., T.J. Parmely, C. Tomomori-Sato, S. Sato, D. Zhu, R.C. Conaway, J.W. Conaway, L. Florens, and M.P. Washburn, *Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors.* Proc Natl Acad Sci U S A, 2006. **103**(50): p. 18928-33.
16. Sato, S., C. Tomomori-Sato, T.J. Parmely, L. Florens, B. Zybailov, S.K. Swanson, C.A. Banks, J. Jin, Y. Cai, M.P. Washburn, J.W. Conaway, and R.C. Conaway, *A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology.* Mol Cell, 2004. **14**(5): p. 685-91.
17. Ishihama, Y., Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, and M. Mann, *Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein.* Mol Cell Proteomics, 2005. **4**(9): p. 1265-72.
18. Jin, J., T. Cardozo, R.C. Lovering, S.J. Elledge, M. Pagano, and J.W. Harper, *Systematic analysis and nomenclature of mammalian F-box proteins.* Genes Dev, 2004. **18**(21): p. 2573-80.
19. Zhang, X., A. Krutchinsky, A. Fukuda, W. Chen, S. Yamamura, B.T. Chait, and R.G. Roeder, *MED1/TRAP220 exists predominantly in a TRAP/ Mediator subpopulation enriched in RNA polymerase II and is required for ER-mediated transcription.* Mol Cell, 2005. **19**(1): p. 89-100.
20. Chen, W., Q. Yang, and R.G. Roeder, *Dynamic interactions and cooperative functions of PGC-1alpha and MED1 in TRalpha-mediated activation of the brown-fat-specific UCP-1 gene.* Mol Cell, 2009. **35**(6): p. 755-68.
21. Kim, J.H., C.K. Yang, K. Heo, R.G. Roeder, W. An, and M.R. Stallcup, *CCAR1, a key regulator of mediator complex recruitment to nuclear receptor transcription complexes.* Mol Cell, 2008. **31**(4): p. 510-9.
22. Xie, W., H. Hong, N.N. Yang, R.J. Lin, C.M. Simon, M.R. Stallcup, and R.M. Evans, *Constitutive activation of transcription and binding of coactivator by estrogen-related receptors 1 and 2.* Mol Endocrinol, 1999. **13**(12): p. 2151-62.
23. Rajkovic, A., C. Yan, W. Yan, M. Klysik, and M.M. Matzuk, *Obox, a family of homeobox genes preferentially expressed in germ cells.* Genomics, 2002. **79**(5): p. 711-7.
24. Kim, H.M., H.J. Ahn, H.S. Lee, K.A. Lee, S.M. Lee, H.H. Kim, K.S. Kim, and K.S. Park, *Obox4 regulates the expression of histone family genes and promotes differentiation of mouse embryonic stem cells.* FEBS Lett, 2010. **584**(3): p. 605-11.
25. Knuesel, M.T., K.D. Meyer, A.J. Donner, J.M. Espinosa, and D.J. Taatjes, *The human CDK8 subcomplex is a histone kinase that requires Med12 for activity and can function independently of mediator.* Mol Cell Biol, 2009. **29**(3): p. 650-61.
26. Meyer, K.D., A.J. Donner, M.T. Knuesel, A.G. York, J.M. Espinosa, and D.J. Taatjes, *Cooperative activity of cdk8 and GCN5L within Mediator directs tandem phosphoacetylation of histone H3.* EMBO J, 2008. **27**(10): p. 1447-57.

27. Fazzio, T.G., J.T. Huff, and B. Panning, *An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity.* Cell, 2008. **134**(1): p. 162-74.
28. Ebmeier, C.C. and D.J. Taatjes, *Activator-Mediator binding regulates Mediator-cofactor interactions.* Proc Natl Acad Sci U S A, 2010. **107**(25): p. 11283-8.
29. Larschan, E. and F. Winston, *The Saccharomyces cerevisiae Srb8-Srb11 complex functions with the SAGA complex during Gal4-activated transcription.* Mol Cell Biol, 2005. **25**(1): p. 114-23.
30. Liu, X., M. Vorontchikhina, Y.L. Wang, F. Faiola, and E. Martinez, *STAGA recruits Mediator to the MYC oncoprotein to stimulate transcription and cell proliferation.* Mol Cell Biol, 2008. **28**(1): p. 108-21.
31. Akoulitchev, S., S. Chuikov, and D. Reinberg, *TFIIH is negatively regulated by cdk8-containing mediator complexes.* Nature, 2000. **407**(6800): p. 102-6.
32. Perissi, V., K. Jepsen, C.K. Glass, and M.G. Rosenfeld, *Deconstructing repression: evolving models of co-repressor action.* Nat Rev Genet, 2010. **11**(2): p. 109-23.
33. Perissi, V., A. Aggarwal, C.K. Glass, D.W. Rose, and M.G. Rosenfeld, *A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors.* Cell, 2004. **116**(4): p. 511-26.
34. Skaar, J.R. and M. Pagano, *Control of cell growth by the SCF and APC/C ubiquitin ligases.* Curr Opin Cell Biol, 2009. **21**(6): p. 816-24.
35. Crusio, K.M., B. King, L.B. Reavie, and I. Aifantis, *The ubiquitous nature of cancer: the role of the SCF(Fbw7) complex in development and transformation.* Oncogene, 2010.
36. Manchado, E., M. Eguren, and M. Malumbres, *The anaphase-promoting complex/cyclosome (APC/C): cell-cycle-dependent and -independent functions.* Biochem Soc Trans, 2010. **38**(Pt 1): p. 65-71.
37. Lasorella, A., J. Stegmuller, D. Guardavaccaro, G. Liu, M.S. Carro, G. Rothschild, L. de la Torre-Ubieta, M. Pagano, A. Bonni, and A. Iavarone, *Degradation of Id2 by the anaphase-promoting complex couples cell cycle exit and axonal growth.* Nature, 2006. **442**(7101): p. 471-4.
38. Stegmuller, J., Y. Konishi, M.A. Huynh, Z. Yuan, S. Dibacco, and A. Bonni, *Cell-intrinsic regulation of axonal morphogenesis by the Cdh1-APC target SnoN.* Neuron, 2006. **50**(3): p. 389-400.
39. Liu, N., H. Li, S. Li, M. Shen, N. Xiao, Y. Chen, Y. Wang, W. Wang, R. Wang, Q. Wang, J. Sun, and P. Wang, *The Fbw7/human CDC4 tumor suppressor targets proproliferative factor KLF5 for ubiquitination and degradation through multiple phosphodegron motifs.* J Biol Chem, 2010. **285**(24): p. 18858-67.
40. Welcker, M., A. Orian, J. Jin, J.E. Grim, J.W. Harper, R.N. Eisenman, and B.E. Clurman, *The Fbw7 tumor suppressor regulates glycogen synthase kinase 3 phosphorylation-dependent c-Myc protein degradation.* Proc Natl Acad Sci U S A, 2004. **101**(24): p. 9085-90.
41. Chi, Y., M.J. Huddleston, X. Zhang, R.A. Young, R.S. Annan, S.A. Carr, and R.J. Deshaies, *Negative regulation of Gcn4 and Msn2 transcription factors by Srb10 cyclin-dependent kinase.* Genes Dev, 2001. **15**(9): p. 1078-92.
42. Alarcon, C., A.I. Zaromytidou, Q. Xi, S. Gao, J. Yu, S. Fujisawa, A. Barlas, A.N. Miller, K. Manova-Todorova, M.J. Macias, G. Sapkota, D. Pan, and J. Massague, *Nuclear CDKs drive Smad transcriptional activation and turnover in BMP and TGF-beta pathways.* Cell, 2009. **139**(4): p. 757-69.
43. Fryer, C.J., J.B. White, and K.A. Jones, *Mastermind recruits CycC:CDK8 to phosphorylate the Notch ICD and coordinate activation with turnover.* Mol Cell, 2004. **16**(4): p. 509-20.
44. Szutorisz, H., A. Georgiou, L. Tora, and N. Dillon, *The proteasome restricts permissive transcription at tissue-specific gene loci in embryonic stem cells.* Cell, 2006. **127**(7): p. 1375-88.
45. Liber, D., R. Domaschenz, P.-H. Holmqvist, L. Mazzarella, A. Georgiou, M. Leleu, A.G. Fisher, P.A. Labosky, and N. Dillon, *Epigenetic Priming of a Pre-B Cell-Specific Enhancer through Binding of Sox2 and Foxd3 at the ESC Stage.* Cell Stem Cell, 2010. **7**(1): p. 114-126.
46. Szutorisz, H., C. Canzonetta, A. Georgiou, C.M. Chow, L. Tora, and N. Dillon, *Formation of an active tissue-specific chromatin domain initiated by epigenetic marking at the embryonic stem cell stage.* Mol Cell Biol, 2005. **25**(5): p. 1804-20.
47. Dignam, J.D., R.M. Lebovitz, and R.G. Roeder, *Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei.* Nucleic Acids Res, 1983. **11**(5): p. 1475-89.

Discussion

Discussion

Embryonic stem cells have been extensively studied over the past two decades, as they constitute an attractive model system to address aspects of early embryonic development and have potential use in cell replacement therapy. Self-renewal and pluripotency, two key properties of ES cells, are governed by cytokine signaling cascades and a core transcriptional circuit. In fact, ES cell identity can be maintained in the absence of exogenous signals if differentiation inducing pathways are inhibited and the transcription network is wired correctly[1]. The core of this network is formed by Oct4, Sox2 and Nanog. Genomics approaches have identified binding sites and putative target genes of these key transcription factors. The molecular mechanisms by which they regulate gene expression however have remained under-investigated, possibly due to difficulty in generating sufficient material for biochemical studies.

We developed an improved FLAG affinity based purification strategy that requires relatively small amounts of starting material, yields high numbers of identified interacting proteins and has a low background. We subsequently employed this methodology to purify Oct4, Sall4, Dax1, Tcfcp2l1 and Esrrb complexes from ES cells (**Chapter 3**). The functional relevance in gene regulation of the newly identified interaction between Oct4 and Esrrb was addressed in more detail (**Chapter 2**). Esrrb was one of only two transcription factors identified in purifications of Mediator complexes from ES cells (**Chapter 4**), confirming our previously identified link between Esrrb and the basal transcription machinery.

Protein complex isolation and characterization

Recent advances in mass spectrometry technology, such as increased sensitivity and the ability to handle more complex samples, inspired us to try to improve existing protein purification strategies based on peptide tag affinity so they would be applicable to limited amounts of stem cell nuclear extract. In the past, identification and characterization of novel protein complexes was usually performed by purification steps over several analytical columns to gain a significantly enriched and purified sample for mass spectrometric analysis. We improved methodology to such an extent that we could identify a wide range of specific interactions following a single round of protein purification from ES cell extracts. This particularly proved advantageous for the identification of transcription factor partner proteins, as these interactions in general are relatively weak and likely to be irreversibly lost during conventional column purification. For example, a modest increase in salt concentration from 100 to 150 mM results in a significant reduction in amounts of TF co-precipitating partner protein (our unpublished data, [2]). Therefore, to identify functionally relevant TF-interactors it is important to carry out protein purification at physiological conditions (for reference, nuclear ion concentrations are ± 100 mM K^+/Na^+ with K^+>Na^+)[3].

A disadvantage of using relatively mild conditions is the concomitant increase in a-specific background binding. To circumvent this problem we have made use of a FLAG-affinity based

purification strategy, in which tagged proteins can be specifically eluted from FLAG affinity beads by a FLAG peptide. We additionally observed a major reduction in a-specific background when low-adherence tubes were employed. Combined these improvements allowed us to reproducibly identify over 50 putative Oct4 interaction partners following a single round of purification, most of which could be verified by an independent approach (**Chapter 3**). That these interactions can be functionally relevant, even if they are relatively weak, was shown by the role of the Oct4-Esrrb interaction in activating the *Nanog* gene (**Chapter 2**).

Transcription factor networks and combinatorial gene regulation

Transcription factors provide specificity to gene regulatory networks by their ability to bind particular sequences. They can form an essential component among multiple factors controlling cell identity (e.g. Esrrb, Klf5 in ESCs[4-5]) or act as master regulator to drive transcriptional programs of terminally differentiated cell types (e.g. MyoD in skeletal muscle formation[6]). Regulated genes in metazoans commonly receive input from multiple transcription factors and several recent studies have focused on trying to discern the general features of combinatorial gene regulation. Whereas initial attention was given to co-occurrence of *cis*-regulatory elements[7], focus has now also been put on TF co-expression patterns and TF-TF protein interactions. One report addressing dynamic TF expression changes during *Drosophila* embryonic development showed that expression of most TFs is not restricted to a single tissue, suggesting specificity in gene expression has to arise from combinatorial regulation[8]. Physical interactions between TFs are likely to form an important constituent of such combinatorial gene regulatory networks, as they likely contribute to co-recruitment to or co-occupancy of target genes. In a systematic mammalian two-hybrid screen, the FANTOM consortium has mapped 762 human TF-TF and 877 mouse TF-TF interactions[9]. The majority of factors reportedly interact with several other TFs and in general are broadly expressed across different tissues. Based on observed physical interactions, TF subnetworks were constructed that are conserved between mouse and human and to some extent can predict tissue specificity. For each of the ES cell transcription factors we purified (**Chapter 3**), we also detected interactions with multiple additional transcription factors. Although our strategy requires significant amounts of a sufficiently homogeneous cell population, which in the case of certain tissue-specific transcription factors may be problematic, we feel it may be more sensitive than a two-hybrid screen. For example, we found 33 Tcfcp2l1- and 7 Dax1-interacting transcription factors compared to respectively 1 and 0 interactors identified by two-hybrid screening. We would therefore advocate integration of our TF-interaction data in the FANTOM Atlas of combinatorial gene regulation.

What would be the implications of TF-TF interactions on the mode of operation in combinatorial gene regulation? In **Chapter 2** we functionally characterized a newly identified interaction between two ES cell TFs, Oct4 and Esrrb. We demonstrated cooperative binding of Oct4 and Esrrb, in conjunction with Sox2, to the *Nanog* proximal promoter. The cooperative partnership between octamer binding proteins and Sox factors has been well described and is assisted by

direct protein-protein interactions between their respective DNA binding domains[10]. Sox2 locks the otherwise optionally free POU_S domain of Oct factors onto its DNA recognition sequence[11]. We similarly mapped the interaction of Esrrb to the POU_H domain of Oct4 (W. van Aert, unpublished data), which is in concordance with published data on the direct interaction of Oct1 and -2 POU_H domains with the glucocorticoid receptor (GR)[12] and the association of estrogen receptor α (ER) with Oct1 POU_H domain[13].

Co-occurrence of binding site motifs in *cis*-regulatory elements provides additional clues on combinatorial regulatory modules. Indeed, in the case of Oct4 and Sox2 their respective motifs are frequently found together in a composite element, whose relative orientation and spacing is constrained by structural demands of the heterodimer. Esrrb on the other hand binds palindromic EREs or an extended half-site (a so called ERRE), neither of which resembles an octamer binding site. However, the octamer binding sequence (ATGCAAAT) was found to be significantly enriched among ER-bound sites analyzed on a genome-wide scale[14]. Analysis of Oct motif position relative to the ERE revealed a multimodal distribution, in which Oct motifs identified in close proximity of EREs showed a slight preference for positioning upstream of the ER-bound site, but otherwise displayed a bimodal distribution with relative clustering at 200 bp up- and downstream of the ERE. It therefore seems likely that co-recruitment of Oct4 and Esrrb, as we have described in **Chapter 2**, has been preserved in more distantly related family members.

In **Chapters 2 and 3** we provide further evidence for a role of TF-TF interactions in co-recruitment to binding sites by demonstrating that, for several Oct4-interacting TFs, rapid downregulation of Oct4 expression causes a reduction in target site occupancy. These data again suggest that protein interactions make up an important component of combinatorial gene regulation. Genome-wide binding studies have shown extensive clustering of ES cell transcription factors, including Oct4 and its interactors Sox2, Esrrb, Dax1, Sall4, Klf5 and Tcfcp2l1[15-18]. Sequential chromatin immunoprecipitations (ChIPs) however have not been conducted, leaving the possibility that this genome wide clustering represents mixed cell populations in which the actual overlap is far less extensive than a simple comparison of binding site profiles may suggest. By establishing that target site binding of at least a subset of targets is Oct4-dependent we have provided evidence for physical co-occupancy. Since an Oct/Sox motif was identified as consensus sequence for binding sites of multiple ES cell transcription factors[17], many more examples of Oct4-dependent TF-targeting may exist across the genome. It therefore would be of great interest to study interdependent TF binding on a genome wide scale, as this information, combined with gene expression data, could provide a far more detailed analysis on combinatorial modes of gene regulation than currently available.

What could be the benefit of multiple TF targeting to certain sites? It may yield possibilities to increase specificity in regulation of target gene expression, but may also provide some robustness to the gene regulatory network. The *Nanog* proximal promoter is a paradigm example of a multiple transcription factor binding locus (MTL)[17]. We have demonstrated that binding of at least three factors (i.e. Esrrb, Dax1 and Sox2) to this site is strongly dependent on Oct4 (**Chapters**

2, 3 and our unpublished data). Downregulation of Oct4, however, does not immediately affect Nanog expression[19], suggesting that additional TFs that bind independently of Oct4 act as fail-safe mechanism.

Nanog is expressed in a mosaic fashion in the ICM of preimplantation blastocysts as well as in ES cells[20-21] and low levels of Nanog predispose cells to differentiation[20]. In **Chapter 2** we have shown that Esrrb levels co-fluctuate with Nanog, wich could have suggested that Esrrb is the rate-limiting regulator of Nanog expression. However, forced overexpression of Esrrb did not change Nanog heterogeneity in ES cell colonies (our unpublished data), suggesting the involvement of additional factors. Expression profiling of Nanog^{high} versus Nanog^{low} cells has revealed overrepresentation of several TFs (e.g. Dax1, Zfp42, Sox2, Tcl1 and Klf4) in the Nanog^{high} population[22]. Of these Dax1, Sox2 and Klf4 have been shown to bind the *Nanog* promoter, further hinting at possible roles for these TFs in the regulation of Nanog expression[17].

To activate or to repress?

In **Chapter 3** we describe interactions of Oct4 with chromatin modifying complexes that have been assigned primarily activating (i.e. SWI/SNF) or repressive (i.e. NuRD, PRC1) roles. However, time-course analysis of expression changes following rapid downregulation of Oct4 has indicated a predominantly activating role for Oct4[19]. How can these observations be reconciled? One possible explanation could involve dynamic target gene regulation by Oct4, where one or more rounds of active transcription are followed by a refractory period in which target gene transcription is shut down. Induction of silencing at this stage could involve recruitment of repressive complexes such as NuRD, analogous to what has been described to occur during ERα-induced transcriptional cycles[23]. Alternatively, NuRD complex-recruitment could play a direct role in target gene activation, similar to its recently attributed function in transcription activation of adult-type globin genes[24].

However, in contrast to its assigned role as predominant activator, an in depth statistical analysis of genome wide binding profiles of 12 transcription factors combined with gene expression data has suggested that Oct4 forms part of a group of TFs that on, a fraction of target genes, can act as repressor[25]. Other members of this group are Nanog, Sox2, Esrrb, Tcfcp2l1, Smad1 and Stat3, all of which have binding profiles that significantly overlap with Oct4[15]. We have determined interaction partners for three members of this TF group (Oct4, Esrrb and Tcfcp2l1) and found associations with NuRD (Oct4, Esrrb, Tcfcp2l1) and PRC1 (Oct4, Tcfcp2l1) complexes (**Chapter 3**), whose chromatin remodeling activities could account for TF-targeted gene repression.

Genomic binding profiles of PRC1 (Ring1b, Phc1) and PRC2 (Eed, Suz12) complexes in ES cells denote widespread targeting to silenced genes encoding developmental regulators[26]. Comparison of binding profiles in human ES cells reveals that a significant percentage (~25%) of Oct4 bound sites overlaps with Suz12 (PRC2) bound regions[27]. Indeed, a role for Oct4 in PRC1- and PRC2-target gene engagement has been demonstrated and was proposed to depend on a physical association between Oct4 and PRC1[28]. PRC complexes, however, normally occupy large genomic regions, whereas transcription factors localize to distinct sites, making it unlikely that Oct4 is solely

responsible for the targeting of individual PRC complexes. We therefore have previously proposed a role for Oct4 as nucleator of PRC binding. From a mechanistic point of view such a role may involve stimulating production of short non-coding RNAs transcribed from promoter regions of Polycomb target genes[29]. These short RNAs can interact with PRC2 and supposedly are involved in PRC2-targeting to genomic regions. PRC2-catalyzed deposition of histone H3K27 methylation marks would subsequently allow for Oct4-mediated PRC1 recruitment, docking and spreading. In this model transcription activation by Oct4 would ultimately result in gene silencing.

Touching the basal transcription machinery

We described association of Mediator complex and RNA pol2 with orphan nuclear receptor Esrrb (**Chapter 3**). Detectable presence of Esrrb in purified Mediator fractions suggests it may constitute an important targeting factor for the basal transcription machinery in ES cells (**Chapter 4**). The related estrogen receptor-α interacts with the Med1 subunit and recruits Mediator complex to ER-target genes in a ligand-dependent manner[30]. No endogenous ligands for Esrrb have been described, although the synthetic estrogen diethylstilbestrol (DES) has been shown to function as ERR-antagonist and can induce an *Esrrb-/-* phenotype upon administration to pregnant mice. Addition of DES to ES cell cultures however failed to reproduce the effects of Esrrb knock-down (our unpublished data), suggesting fundamental differences between Esrrb function in early embryonic development and ES cell maintenance. Ligand-dependency is furthermore unlikely to play a major role in the interaction between Esrrb and Mediator, as association could be reproduced using recombinant Esrrb and dialyzed nuclear extracts.

An interesting question that remains unsolved is whether Esrrb is indeed involved in targeting of Mediator and/or RNA pol2 complexes to transcription start sites. Rapid depletion of Oct4 was shown to reduce RNA pol2 occupancy at the *Nanog* proximal promoter and several additional Oct4 target genes[31]. As we have failed to detect physical association of the basal transcription machinery with Oct4 (**Chapter 3**), but have shown that Esrrb occupancy of the *Nanog* proximal promoter is Oct4-dependent (**Chapter 2**), we would like to speculate that Esrrb in this instance is responsible for RNA pol2 binding at the promoter. Future work will be directed at finding concrete evidence for the role of Esrrb in recruitment of the basal transcription machinery.

It is important to note that a large number of Esrrb binding sites in ES cells are found distal (>5 kb) to transcription start sites[15]. These possibly have an enhancer function and, indeed, a correlation was detected between Esrrb-only bound sites and association of p300[15], a known mark of active enhancers[32]. A similar promoter-distal binding pattern was observed for ERα in hormone-treated MCF-7 cells[14]. Chromosomal interactions of these ERα binding sites have been mapped on a genome wide scale[33]. A significant fraction of highly enriched ERα-binding sites is engaged in duplex (i.e. two interacting loci) intrachromosomal interactions spanning less than 100 kb, or in complex (i.e. multiple connecting duplex interactions) interactions typically spanning between 100 kb and 1 Mb. These interacting ERα binding sites are enriched for RNA pol2 and H3K4me3 marks and are reminiscent of enhancer-promoter communication. It would be interesting to

discern whether a physical interaction between Esrrb and Mediator/RNA pol2 can induce or sustain similar chromatin loops in ES cells. Structural organization of the *Nanog* gene locus has been examined in the presence and absence of Oct4, revealing loss of interaction between the proximal promoter and an upstream (-44 kb) hypersensitive site upon Oct4 depletion[34]. Esrrb, but not Oct4, detectably binds at the -44 kb site and hence may be involved in structural maintenance of an active *Nanog* locus.

REFERENCES

1. Ying, Q.L., J. Wray, J. Nichols, L. Batlle-Morera, B. Doble, J. Woodgett, P. Cohen, and A. Smith, *The ground state of embryonic stem cell self-renewal.* Nature, 2008. **453**(7194): p. 519-23.
2. Wissmuller, S., T. Kosian, M. Wolf, M. Finzsch, and M. Wegner, *The high-mobility-group domain of Sox proteins interacts with DNA-binding domains of many transcription factors.* Nucleic Acids Res, 2006. **34**(6): p. 1735-44.
3. Rippe, K., *Dynamic organization of the cell nucleus.* Curr Opin Genet Dev, 2007. **17**(5): p. 373-80.
4. Ivanova, N., R. Dobrin, R. Lu, I. Kotenko, J. Levorse, C. DeCoste, X. Schafer, Y. Lun, and I.R. Lemischka, *Dissecting self-renewal in stem cells with RNA interference.* Nature, 2006. **442**(7102): p. 533-8.
5. Parisi, S., F. Passaro, L. Aloia, I. Manabe, R. Nagai, L. Pastore, and T. Russo, *Klf5 is involved in self-renewal of mouse embryonic stem cells.* J Cell Sci, 2008. **121**(Pt 16): p. 2629-34.
6. Tapscott, S.J., *The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription.* Development, 2005. **132**(12): p. 2685-95.
7. Ochoa-Espinosa, A. and S. Small, *Developmental mechanisms and cis-regulatory codes.* Curr Opin Genet Dev, 2006. **16**(2): p. 165-70.
8. Adryan, B. and S.A. Teichmann, *The developmental expression dynamics of Drosophila melanogaster transcription factors.* Genome Biol, 2010. **11**(4): p. R40.
9. Ravasi, T., H. Suzuki, C.V. Cannistraci, S. Katayama, V.B. Bajic, K. Tan, A. Akalin, S. Schmeier, M. Kanamori-Katayama, N. Bertin, P. Carninci, C.O. Daub, A.R. Forrest, J. Gough, S. Grimmond*, et al.*, *An atlas of combinatorial transcriptional regulation in mouse and man.* Cell, 2010. **140**(5): p. 744-52.
10. Ambrosetti, D.C., C. Basilico, and L. Dailey, *Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites.* Mol Cell Biol, 1997. **17**(11): p. 6321-9.
11. Williams, D.C., Jr., M. Cai, and G.M. Clore, *Molecular basis for synergistic transcriptional activation by Oct1 and Sox2 revealed from the solution structure of the 42-kDa Oct1.Sox2.Hoxb1-DNA ternary transcription factor complex.* J Biol Chem, 2004. **279**(2): p. 1449-57.
12. Prefontaine, G.G., M.E. Lemieux, W. Giffin, C. Schild-Poulter, L. Pope, E. LaCasse, P. Walker, and R.J. Hache, *Recruitment of octamer transcription factors to DNA by glucocorticoid receptor.* Mol Cell Biol, 1998. **18**(6): p. 3416-30.
13. Prefontaine, G.G., R. Walther, W. Giffin, M.E. Lemieux, L. Pope, and R.J. Hache, *Selective binding of steroid hormone receptors to octamer transcription factors determines transcriptional synergism at the mouse mammary tumor virus promoter.* J Biol Chem, 1999. **274**(38): p. 26713-9.
14. Carroll, J.S., C.A. Meyer, J. Song, W. Li, T.R. Geistlinger, J. Eeckhoute, A.S. Brodsky, E.K. Keeton, K.C. Fertuck, G.F. Hall, Q. Wang, S. Bekiranov, V. Sementchenko, E.A. Fox, P.A. Silver*, et al.*, *Genome-wide analysis of estrogen receptor binding sites.* Nat Genet, 2006. **38**(11): p. 1289-97.
15. Chen, X., H. Xu, P. Yuan, F. Fang, M. Huss, V.B. Vega, E. Wong, Y.L. Orlov, W. Zhang, J. Jiang, Y.H. Loh, H.C. Yeo, Z.X. Yeo, V. Narang, K.R. Govindarajan*, et al.*, *Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.* Cell, 2008. **133**(6): p. 1106-17.
16. Jiang, J., Y.S. Chan, Y.H. Loh, J. Cai, G.Q. Tong, C.A. Lim, P. Robson, S. Zhong, and H.H. Ng, *A core Klf circuitry regulates self-renewal of embryonic stem cells.* Nat Cell Biol, 2008. **10**(3): p. 353-60.
17. Kim, J., J. Chu, X. Shen, J. Wang, and S.H. Orkin, *An extended transcriptional network for pluripotency of embryonic stem cells.* Cell, 2008. **132**(6): p. 1049-61.
18. Lim, C.Y., W.L. Tam, J. Zhang, H.S. Ang, H. Jia, L. Lipovich, H.H. Ng, C.L. Wei, W.K. Sung, P. Robson, H.

Yang, and B. Lim, *Sall4 regulates distinct transcription circuitries in different blastocyst-derived stem cell lineages.* Cell Stem Cell, 2008. **3**(5): p. 543-54.

19. Sharov, A.A., S. Masui, L.V. Sharova, Y. Piao, K. Aiba, R. Matoba, L. Xin, H. Niwa, and M.S. Ko, *Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data.* BMC Genomics, 2008. **9**: p. 269.
20. Chambers, I., J. Silva, D. Colby, J. Nichols, B. Nijmeijer, M. Robertson, J. Vrana, K. Jones, L. Grotewold, and A. Smith, *Nanog safeguards pluripotency and mediates germline development.* Nature, 2007. **450**(7173): p. 1230-4.
21. Chazaud, C., Y. Yamanaka, T. Pawson, and J. Rossant, *Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway.* Developmental Cell, 2006. **10**(5): p. 615-624.
22. Singh, A.M., T. Hamazaki, K.E. Hankowski, and N. Terada, *A heterogeneous expression pattern for Nanog in embryonic stem cells.* Stem Cells, 2007. **25**(10): p. 2534-42.
23. Metivier, R., G. Penot, M.R. Hubner, G. Reid, H. Brand, M. Kos, and F. Gannon, *Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter.* Cell, 2003. **115**(6): p. 751-63.
24. Miccio, A. and G.A. Blobel, *Role of the GATA-1/FOG-1/NuRD pathway in the expression of human beta-like globin genes.* Mol Cell Biol, 2010. **30**(14): p. 3460-70.
25. Ouyang, Z., Q. Zhou, and W.H. Wong, *ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells.* Proc Natl Acad Sci U S A, 2009. **106**(51): p. 21521-6.
26. Boyer, L.A., K. Plath, J. Zeitlinger, T. Brambrink, L.A. Medeiros, T.I. Lee, S.S. Levine, M. Wernig, A. Tajonar, M.K. Ray, G.W. Bell, A.P. Otte, M. Vidal, D.K. Gifford, R.A. Young*, et al.*, *Polycomb complexes repress developmental regulators in murine embryonic stem cells.* Nature, 2006. **441**(7091): p. 349-53.
27. Lee, T.I., R.G. Jenner, L.A. Boyer, M.G. Guenther, S.S. Levine, R.M. Kumar, B. Chevalier, S.E. Johnstone, M.F. Cole, K. Isono, H. Koseki, T. Fuchikami, K. Abe, H.L. Murray, J.P. Zucker*, et al.*, *Control of developmental regulators by Polycomb in human embryonic stem cells.* Cell, 2006. **125**(2): p. 301-13.
28. Endoh, M., T.A. Endo, T. Endoh, Y. Fujimura, O. Ohara, T. Toyoda, A.P. Otte, M. Okano, N. Brockdorff, M. Vidal, and H. Koseki, *Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity.* Development, 2008. **135**(8): p. 1513-24.
29. Kanhere, A., K. Viiri, C.C. Araujo, J. Rasaiyaah, R.D. Bouwman, W.A. Whyte, C.F. Pereira, E. Brookes, K. Walker, G.W. Bell, A. Pombo, A.G. Fisher, R.A. Young, and R.G. Jenner, *Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2.* Mol Cell, 2010. **38**(5): p. 675-88.
30. Zhang, X., A. Krutchinsky, A. Fukuda, W. Chen, S. Yamamura, B.T. Chait, and R.G. Roeder, *MED1/TRAP220 exists predominantly in a TRAP/ Mediator subpopulation enriched in RNA polymerase II and is required for ER-mediated transcription.* Mol Cell, 2005. **19**(1): p. 89-100.
31. Rahl, P.B., C.Y. Lin, A.C. Seila, R.A. Flynn, S. McCuine, C.B. Burge, P.A. Sharp, and R.A. Young, *c-Myc regulates transcriptional pause release.* Cell, 2010. **141**(3): p. 432-45.
32. Heintzman, N.D., R.K. Stuart, G. Hon, Y. Fu, C.W. Ching, R.D. Hawkins, L.O. Barrera, S. Van Calcar, C. Qu, K.A. Ching, W. Wang, Z. Weng, R.D. Green, G.E. Crawford, and B. Ren, *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.* Nat Genet, 2007. **39**(3): p. 311-8.
33. Fullwood, M.J., M.H. Liu, Y.F. Pan, J. Liu, H. Xu, Y.B. Mohamed, Y.L. Orlov, S. Velkov, A. Ho, P.H. Mei, E.G. Chew, P.Y. Huang, W.J. Welboren, Y. Han, H.S. Ooi*, et al.*, *An oestrogen-receptor-alpha-bound human chromatin interactome.* Nature, 2009. **462**(7269): p. 58-64.
34. Levasseur, D.N., J. Wang, M.O. Dorschner, J.A. Stamatoyannopoulos, and S.H. Orkin, *Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells.* Genes Dev, 2008. **22**(5): p. 575-80.

S

Summary

Samenvatting

Summary

Embryonic stem cell maintenance is regulated by a transcriptional circuitry. Over the past few years the number of transcription factors reported to play role in control of the pluripotent state has greatly increased. However, although genomic binding sites and regulated genes have been described for most factors, the scope of their interacting partners has remained underexplored. Dissecting the interactome of ES cell transcription factors can generate a better understanding of the molecular mechanisms involved in target gene regulation and potentially leads to identification of novel factors involved in control of ES cell identity.

In Chapter 3 we describe an improved FLAG affinity based methodology to purify transcription factor complexes from relatively small amounts of starting material. The use of relatively mild purification conditions and low-adherence tubes, in combination with peptide elution, enabled the generation of significant amounts of purified transcription factor and associated proteins. We initially applied this strategy to purify ES cell transcription factor Oct4 and, by mass spectrometry analysis, identified over 50 putative interaction partners, including many transcription factors involved in maintenance of pluripotency, whose encoding genes are often regulated by Oct4. Subsequent purification of Oct4-interacting transcription factors Sall4, Tcfcp2l1, Dax1 and Esrrb resulted in an interaction network of 166 proteins, including several chromatin modifying complexes (e.g. NuRD, SWI/SNF, PRC1 and Trrap/p400) and components of TGFβ, Notch and Wnt signaling pathways. In addition, physical association of Esrrb with the basal transcription machinery (i.e. Mediator, RNA pol2 and TFIID) was detected. Acute depletion of Oct4 reduced the binding of Oct4-interacting transcription factors to several common binding sites, demonstrating the functional significance of physical associations in genomic binding site occupancy.

In Chapter 2 the newly identified interaction between ES cell transcription factors Oct4 and Esrrb is investigated in more detail. We show that Esrrb binds the *Nanog* proximal promoter *in vivo* in a manner that is dependent both on Oct4 and on a promoter sequence element that resembles a degenerate ERRE. Esrrb positively regulates *Nanog* promoter activity and, thereby, Nanog expression. We further demonstrate that Esrrb protein levels co-fluctuate with Nanog in ES cell colonies, suggesting that, similar to Nanog, Esrrb expression levels positively correlate with ESC self-renewal efficiency.

In Chapter 4 we describe the isolation of Mediator complexes from ES cells and show that Esrrb is one of only two transcription factors reproducibly found associated with Mediator, suggesting it may constitute an important targeting factor. Purification of the Med12/Cdk8-kinase submodule, which optionally associates with the core Mediator complex, furthermore identified a range of interacting proteins not found present in core Mediator preparations. These included additional transcription factors and several members of the PCAF/SAGA and NCoR complexes. We also detected co-purification of two ubiquitin E3 ligases, APC/C and SCF, suggesting that the Med12/Cdk8 module may act as a substrate priming kinase for these two factors.

In brief, our improved peptide-affinity based purification methodology allowed us to bring greater definition to the transcriptional network controlling pluripotency. We demonstrated the functional importance of several newly identified physical associations in binding site occupancy and target gene regulation. Our strategy is widely applicable and, as it requires relatively small amounts of starting material, may benefit biochemical studies in cell or tissue types that are currently impeded by an inadequate supply of input material.

Samenvatting

Het behoud van embryonale stamcellen (ES cellen) wordt gereguleerd door een transcriptienetwerk. In de afgelopen jaren is het aantal gerapporteerde transcriptiefactoren dat een controlerende rol speelt in pluripotentie sterk toegenomen. Hoewel de genomische bindingsplaatsen en gereguleerde genen van de meeste transcriptiefactoren beschreven zijn, is er relatief weinig bekend over het spectrum van hun bindingspartners. Ontcijfering van het interactoom van ES cel transcriptiefactoren kan een beter inzicht geven in de moleculaire mechanismen waarmee genen gereguleerd worden en leidt mogelijk tot identificatie van nieuwe factoren die betrokken zijn bij het behoud van ES cel identiteit.

In Hoofdstuk 3 beschrijven we een verbeterde, op FLAG affiniteit gebaseerde methodologie om complexen van transcriptiefactoren op te zuiveren uit relatief kleine hoeveelheden startmateriaal. Het gebruik van relatief milde zuiveringscondities en lage-adherentie buizen, in combinatie met peptide eluties, maakte het mogelijk om significante hoeveelheden transcriptiefactor en geassocieerde eiwitten op te zuiveren. We pasten deze strategie toe om de ES cel transcriptiefactor Oct4 op te zuiveren en identificeerden, met behulp van massaspectrometrische analyse, meer dan 50 mogelijke interactiepartners. Hieronder bevonden zich vele transcriptiefactoren die betrokken zijn bij het behoud van pluripotentie en wiens coderende genen vaak gereguleerd worden door Oct4. Opzuivering van de Oct4-interacterende transcriptiefactoren Sall4, Tcfcp2l1, Dax1 en Esrrb resulteerde in een interactie-netwerk bestaande uit 166 eiwitten, waaronder verschillende chromatine modificerende complexen (b.v. NuRD, SWI/SNF, PRC1 en Trrap/p400) en componenten van de TGFβ, Notch en Wnt signaaltransductie cascades. Bovendien werd er associatie van Esrrb met het basale transcriptie-apparaat (nl. Mediator, RNA pol2 en TFIID) gedetecteerd. Acute Oct4-depletie verlaagde de bezetting van een aantal gemeenschappelijke bindingsplaatsen door Oct4-interacterende transcriptiefactoren, hetgeen duidt op een functioneel belang van fysieke interactie in het binden van genomische bindingsplaatsen.

In Hoofdstuk 2 wordt de geïdentificeerde interactie tussen Oct4 en Esrrb gedetailleerder onderzocht. We laten zien dat Esrrb *in vivo* aan de *Nanog* promoter bindt op een wijze die afhankelijk is van zowel Oct4, als van een korte sequentie in de promoter, welke lijkt op een gedegenereerd ERRE. Esrrb reguleert *Nanog* promoter activiteit, en daarmee Nanog expressie, op een positieve manier. Verder laten we zien dat Esrrb eiwit niveaus co-fluctueren met Nanog in ES cel kolonies, hetgeen doet vermoeden dat Esrrb expressie niveaus, net als die van Nanog, positief correleren met de efficiëntie van ESC zelfvernieuwing.

In Hoofdstuk 4 beschrijven we de isolatie van Mediator complexen uit ES cellen, waarbij we aantonen dat Esrrb één van slechts twee transcriptiefactoren is die herhaaldelijk wordt gedetecteerd in Mediator fracties. Opzuivering van de Med12/Cdk8-kinase submodule, die optioneel bindt aan het kern Mediator complex, resulteerde in de identificatie van een reeks interacterende eiwitten die niet eerder gevonden werden in preparaties van het kern Mediator

complex. Hieronder bevonden zich extra transcriptiefactoren en enkele eiwitten die onderdeel uitmaken van de PCAF/SAGA en NCoR complexen. Ook detecteerden we co-purificatie van twee ubiquitine E3 ligasen, APC/C en SCF, hetgeen suggereert dat de Med12/Cdk8 submodule een rol kan spelen in het markeren van substraten voor deze twee ubiquitinerende factoren.

In het kort heeft onze verbeterde, op peptide-affiniteit gebaseerde zuiveringsmethodologie een gedetailleerdere beschrijving van het pluripotentie-regulerende transcriptiefactor netwerk opgeleverd. We hebben laten zien dat een aantal van de hiermee geïdentificeerde fysieke associaties van functioneel belang zijn voor het bezetten van bindingsplaatsen en het reguleren van genen. Onze strategie is breed toepasbaar en kan, aangezien zij slechts relatief kleine hoeveelheden startmateriaal behoeft, biochemische studies bevorderen in cel- of weefseltypen die momenteel bemoeilijkt worden door ontoereikende hoeveelheden startmateriaal.

Curriculum Vitae

Personal details

Name:	Deborah Louisa Carolina van den Berg
Birthdate:	05-03-1980
Birthplace:	Breda, The Netherlands

Education

1998-2003	MSc Biomedical Sciences, *cum laude* Utrecht University, Utrecht, The Netherlands
1992-1998	VWO Gymnasium β Mencia de Mendoza Lyceum, Breda, The Netherlands

Research

2005-2010:	PhD research Department of Cell Biology, Erasmus MC, Rotterdam, The Netherlands (Prof.dr. F.G. Grosveld and Dr. R.A. Poot)
2003-2005:	PhD research Department of Biochemistry, Erasmus MC, Rotterdam, The Netherlands (Prof.dr. C.P. Verrijzer)
2003:	Internship Marie Curie Research Institute, Oxted, United Kingdom (Drs. P.D. Varga-Weisz and R.A. Poot)
2002:	MSc Research project Department of Immunology, University Medical Center Utrecht, The Netherlands (Prof.dr. J.G. van de Winkel and Dr. J.H. Leusen)
2001:	MSc Research project Department of Physiological Chemistry, University Medical Center Utrecht, The Netherlands (Prof.dr. H.T.H.M. Timmers and Ir. C.G. Zwartjes)

List of publications

van den Berg DL, Snoek T, Mullin NP, Yates A, Bezstarosti K, Demmers J, Chambers I, Poot RA. An Oct4-centered protein interaction network in embryonic stem cells.
Cell Stem Cell. 2010 Apr 2;6(4):369-81.

van den Berg DL, Zhang W, Yates A, Engelen E, Takacs K, Bezstarosti K, Demmers J, Chambers I, Poot RA. Estrogen-related receptor beta interacts with Oct4 to positively regulate *Nanog* gene expression.
Mol Cell Biol. 2008 Oct;28(19):5986-95.

Beekman JM, van der Poel CE, van der Linden JA, **van den Berg DL**, van den Berghe PV, van de Winkel JG, Leusen JH. Filamin A stabilizes Fc gamma RI surface expression and prevents its lysosomal routing.
J Immunol. 2008 Mar 15;180(6):3938-45.

Poot RA, Bozhenok L, **van den Berg DL**, Hawkes N, Varga-Weisz PD. Chromatin remodeling by WSTF-ISWI at the replication site: opening a window of opportunity for epigenetic inheritance?
Cell Cycle. 2005 Apr;4(4):543-6.

Poot RA*, Bozhenok L*, **van den Berg DL***, Steffensen S, Ferreira F, Grimaldi M, Gilbert N, Ferreira J, Varga-Weisz PD. The Williams syndrome transcription factor interacts with PCNA to target chromatin remodelling by ISWI to replication foci.
Nat Cell Biol. 2004 Dec;6(12):1236-44.

Zwartjes CG, Jayne S, **van den Berg DL**, Timmers HT. Repression of promoter activity by CNOT2, a subunit of the transcription regulatory Ccr4-not complex.
J Biol Chem. 2004 Mar 19;279(12):10848-54.

(* equal author contribution)

Erasmus MC
Universitair Medisch Centrum Rotterdam

PhD Portfolio

Department of Cell Biology	October 2005 - October 2010
Promotor: Prof.dr. F.G. Grosveld	Supervisor: Dr. R.A. Poot

1. PhD training	
General courses	Year
- From Development to Disease	2005
- Experimental Approach to Molecular and Cell Biology	2006
- EuTRACC Proteomics Course	2010
Seminars and Workshops	
- MGC PhD student workshop poster presentation oral presentation	2006-2008 2006-2007 2008
- Winterschool Transcriptional Control of Developmental Processes oral presentation	2006-2010
National Conferences	
- Dutch Chromatin Meeting	2005-2006
- NWO Nucleic Acids Meeting oral presentation	2007
- Dutch Stem Cell Meeting oral presentation	2006-2009 2009
International Conferences	
- UK Chromatin Meeting, Manchester, UK	2006
- EuroSTELLS Workshop, Montpellier, France Exploring Chromatin in Stem Cells poster presentation	2007
- International Stem Cell Symposium, Amsterdam poster presentation	2007-2009 2009
- MC-GARD Meeting, Edinburgh, Scotland Higher Order Genome Architecture poster presentation	2009
- EMBO Meeting, Amsterdam	2009
- Keystone Symposium Stem Cell Differentiation and Dedifferentiation, Keystone, Colorado, USA scholarship award, poster presentation	2010
- EuroSystem Consortium Meeting, Schiermonnikoog poster presentation	2010
2. Teaching	
- Master's theses	2007, 2009

Dankwoord

"*Things won are done; joy's soul lies in the doing*"
William Shakespeare

Het heeft zo moeten zijn: exact 5 jaar en 5,5 dag geleden begon ik aan mijn promotieonderzoek en ik ben blij en trots dat ik het nu kan afronden met dit proefschrift. Wat voor mij een nieuw begin in Londen had moeten worden, werd uiteindelijk een herstart op de zevende verdieping. Gelukkig kan ik terugkijken op een fijne periode, waarin ik veel heb geleerd en gelachen en waarin ik mijn plezier in de wetenschap heb hervonden.

Eerst en bovenal wil ik Raymond bedanken. Toen je in Engeland mijn stagebegeleider was vormden we al een goed team; ik heb de (waarschijnlijk) permanente schade aan mijn circadian clock maar op de koop toegenomen. En Londen? Ach, Rotterdam is ook best leuk... Dankzij jouw ideeën, projectkeuzes en het altijd één stap vooruit denken zijn we in relatief korte tijd van niets tot een paar mooie papers gekomen. Ik kon altijd mijn ideeën (en m'n ei) bij jou kwijt en gelukkig kon je hopeloze projecten ook snel naar de prullenbak verwijzen. Bedankt voor je enorme enthousiasme, humor en leermomentjes; voor het delen van mijn 'highs' en me uit de put praten als dingen eens niet lukten. Ik had me geen betere begeleider kunnen wensen, kortom (om de cursus Brabants voort te zetten): dache bedaankt zijt da witte!

Natuurlijk ben ik ook Frank, mijn promotor, dankbaar voor het feit dat ik mijn promotieonderzoek binnen de afdeling Celbiologie heb kunnen verrichten. Hoewel de Monday-Morning-Frank-Meetings er de afgelopen jaren bij in zijn geschoten, werkt jouw invloed door in mijn proefschrift, dankzij de door jou gecreëerde omgeving waarin, in een prettige sfeer, goede wetenschap bedreven kan worden; daar heb ik zeker van geprofiteerd én genoten.

Leden van mijn kleine commissie, Sjaak, Joost en Dies, dank jullie wel voor het lezen en becommentariëren van mijn proefschrift. Sjaak tevens bedankt voor de organisatie van de KWT winterschool meetings, s-Keystone is er niks bij! Dies, sorry dat ik voor één keer je Oct-6 dag heb ingepikt; ik had het zelf ook liever anders gezien, maar helaas, de maandag behoorde niet tot de mogelijkheden. Overige commissieleden Elaine, Ian en Jeroen, fijn dat jullie de moeite willen nemen om in de oppositie plaats te nemen.

Ian, I feel honoured that you are willing to come all the way from Scotland to join my doctoral committee. I've really enjoyed our collaboration, and it's been a fruitful one indeed! Thanks for all the constructs and cell lines you've sent across the pond over the past few years, for improving the manuscripts and for my visit to Edinburgh.

Jeroen, aan jou zou ik een pagina moeten wijden, hoewel je hem dan natuurlijk wel met Karel zou moeten delen... Ik durf niet eens te tellen hoeveel gels ik de afgelopen jaren bij jullie heb afgeleverd, maar ik vrees dat het een substantieel deel is van de 400-en-nog-wat mass specs die jullie in die tijd hebben gerund (of zijn jullie misschien stiekem opnieuw begonnen met

nummeren?). Hartstikke bedankt voor al dat werk dat jullie hebben verricht, gelukkig heeft het iets moois opgeleverd. En, dit was écht de laatste...uhum uhum.

Erik, vanaf dag één klikte het gewoon tussen ons en gelukkig maar, want alleen was ook maar alleen. Dank voor al die keren dat je m'n cellen in leven hebt gehouden, voor je organisatietalent, voor het aanhoren van mijn therapeutische klaagzangen ('ik haat ChIP'), voor het verzinnen van nieuwe gezegden, voor gedeelde interesses (Star Trek, Formule 1 en whiskey, om er maar een paar op te noemen), voor wetenschappelijke input en onzin output; die laatste vooral op vrijdagmiddag. Ik ben heel blij dat jij mijn paranimf wilt zijn en wens je natuurlijk heel veel succes met jouw resterende aio-jaren.

Eugin, beter een goede buur dan een verre vriend; jammer genoeg zullen binnen afzienbare tijd al deze kwalificaties voor jou opgaan. Ik ga onze heerlijk ongedwongen celkweekhok gesprekken missen; bedankt voor alle steun en medeleven tijdens deze laatste fase en voor het als eerste uitproberen van een Arminius-promotie! Fijn dat je mijn paranimf wilt zijn.

Alle 706 (oftewel Seven.O.Six) collega's hebben er voor een belangrijk deel aan bijgedragen dat ik de afgelopen 5 jaar geen enkele dag met tegenzin naar het lab ben gegaan. Ernie, alleswetende lab-oudste (wie mocht er ook alweer als eerste beginnen?), bedankt voor alle gezellige lunches, voor je openheid en eerlijkheid en natuurlijk voor die legendarische gamesnight! Umut, thanks for brightening up the office atmosphere when it's needed most. Alireza, I enjoyed many of our conversations about science, politics, life or combinations thereof and wish you all the best with the final bit of your PhD. Maureen, als je ooit nog eens assertiviteitstrainingen gaat geven, je weet me te vinden! Willem, Tim en Rooda, bedankt voor jullie inzet tijdens jullie stages en veel succes in de toekomst. Chantal, misschien tot in London 2012? Maaike, succes met de volgende stap. To all ex-706 lab members that contributed to a nice atmosphere: Miyata, Katy, Marianne, Vesna, Javi, Patrick - thanks, it's been fun to work in such an international environment. Marike en Jasperina, bedankt voor de hulp bij het regelen van al die vervelende regeldingetjes.

Sociale contacten bleven natuurlijk niet beperkt tot 706 en ik wil graag eenieder bedanken die er mede voor heeft gezorgd dat de afgelopen 5 jaren voorbij zijn gevlogen, in het bijzonder: Robert-Jan, Eric, Charlotte, Ralph (voor allerhande wetenschappelijke kletspraat), Dubi, Mariëtte (voor alle Illustrator tips), Farzin, Petros, Laura, Athina, Kris, Sanja, Iris, Akiko, Eskeatnaf, Filippo, Ekim, Rick, thanks for being such a cool bunch of people. De 'de Laatjes': Daan, Petra, Erik, Marieke, Sjoerd en Wouter, het is helaas wat stil aan de overkant sinds jullie verhuisd zijn! Thamar, bedankt voor de hulp met de lentivirussen; Nynke en Ton voor al jullie fiets-tips. Op de 10^{e} natuurlijk niet te vergeten Frank en Jeffrey, waar is de tijd van BBQs en Macaco Blanco gebleven? Zijn we nu echt oud?

Anna, thanks for your friendship and understanding, for brightening up my cloudy days and for

discovering Rotterdam night life with me! I wish I could write something in Greek, but I only know swear words.... Hartstikke bedankt that you're coming over from Greece for my defense!

Audrey, the seventh floor hasn't been the same since you left. Your lively company, our serious and less serious conversations definitely made life more interesting. Good luck on the 1st of October and all the best in La Ville-Lumière.

Wendy, Harald en Chris, jullie waren altijd van de partij (en meestal van de organiserende hand) bij last moment beers of pokernights; bedankt voor het goede gezelschap.

Noorie, with your perseverance, finishing a PhD will be a walk in the park, nothing compared to the Dodentocht! Good luck! Prashanth, my benchmate in the early days, thanks for all those (mostly) scientific discussions on Nieuwe Binnenweg, they will be dearly missed now that you've completely moved to Amsterdam. Hope we'll meet again, somewhere!

Dan is er gelukkig toch ook nog een wereld buiten het Erasmus MC, welke af en toe voor de broodnodige afleiding kan zorgen. Margriet, waar is de Mencia tijd gebleven? We hadden toen niet kunnen bedenken dat we zoveel jaren later, elk via onze eigen omweg, nog bijna tegelijk zouden promoveren. Meneer de Bruin zal trots op ons zijn... En nou heb ik eindelijk weer tijd om af te spreken!

Ellen en Debby, nu is het officieel: het laatste schaap is over de dam! Bedankt dat het met jullie altijd zo vanzelfsprekend gezellig is.

Geert en Maaike, onze vaste prik vrijdagavond tennis-date: het is altijd heerlijk om het wetenschapswereldje even achter ons te laten en het te hebben over kinderen, Friese zuipketen en latex pakjes. Dank jullie wel voor het meeleven met onze experimenten en schrijfperikelen!

Sarah, thanks for being my good friend and for being invaluable company at a scientific meeting or two. Never a dull moment - hope we'll soon be able to meet more frequently in the UK.

Marjorie, I'm sorry for not writing to you for so long – finishing my thesis left me without energy to write anything else, but I promise I'll make up for it. Good luck with your PhD in DC!

Para a minha família portuguesa, muito obrigada por todas as férias em Portugal; eu espero que vocês gostem de visitar a Holanda.

Als laatste wil ik mijn familie down (Z)under(t) bedanken. Erik, Olaf en Reg, ondanks alle keren dat ik jullie in jurkjes heb gehesen, ben ik toch maar mooi blij met m'n drie broers. Ik ben supertrots op jullie! Judith, je bent een fijne schoonzus. Pa & ma, zonder jullie bezorgdheid en onvoorwaardelijke steun, ook bij het nemen van soms moeilijke beslissingen, was ik hier waarschijnlijk niet aanbeland. Dank jullie wel.

Tiago, we made it!

Debbie