Molecular Cell
Article

# Context-Dependent Wiring of Sox2 Regulatory Networks for Self-Renewal of Embryonic and Trophoblast Stem Cells

Kenjiro Adachi,[1,9,10,*] Itoshi Nikaido,[2,9,11] Hiroshi Ohta,[3,12] Satoshi Ohtsuka,[1] Hiroki Ura,[1] Mitsutaka Kadota,[5] Teruhiko Wakayama,[3,6] Hiroki R. Ueda,[2,4] and Hitoshi Niwa[1,7,8,*]
[1]Laboratory for Pluripotent Stem Cell Studies
[2]Functional Genomics Unit
[3]Laboratory for Genome Reprogramming
[4]Laboratory for Systems Biology
[5]Genome Resource and Analysis Unit
RIKEN Center for Developmental Biology, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe 6500047, Japan
[6]Faculty of Life and Environmental Sciences, University of Yamanashi, Yamanashi 4008510, Japan
[7]Laboratory for Development and Regenerative Medicine, Kobe University Graduate School of Medicine, 7-5-1 Kusunokicho, Chuo-ku, Kobe 6500017, Japan
[8]JST, CREST, Sanbancho, Chiyoda-ku, Tokyo, 1020075, Japan
[9]These authors contributed equally to this work
[10]Present address: Department of Cell and Developmental Biology, Max Planck Institute for Molecular Biomedicine, 48149 Münster, Germany
[11]Present address: Bioinformatics Research Unit, Advanced Center for Computing and Communication, RIKEN, Wako, Saitama 3510198, Japan
[12]Present address: Department of Anatomy and Cell Biology, Graduate School of Medicine, Kyoto University, Sakyo-ku, Kyoto 6068501, Japan
*Correspondence: kenjiro.adachi@mpi-muenster.mpg.de (K.A.), niwa@cdb.riken.jp (H.N.)
http://dx.doi.org/10.1016/j.molcel.2013.09.002

## SUMMARY

**Sox2 is a transcription factor required for the maintenance of pluripotency. It also plays an essential role in different types of multipotent stem cells, raising the possibility that Sox2 governs the common stemness phenotype. Here we show that Sox2 is a critical downstream target of fibroblast growth factor (FGF) signaling, which mediates self-renewal of trophoblast stem cells (TSCs). Sustained expression of *Sox2* together with *Esrrb* or *Tfap2c* can replace FGF dependency. By comparing genome-wide binding sites of Sox2 in embryonic stem cells (ESCs) and TSCs combined with inducible knockout systems, we found that, despite the common role in safeguarding the stem cell state, Sox2 regulates distinct sets of genes with unique functions in these two different yet developmentally related types of stem cells. Our findings provide insights into the functional versatility of transcription factors during embryogenesis, during which they can be recursively utilized in a variable manner within discrete network structures.**

## INTRODUCTION

The transcriptional output of a given cell type is controlled by unique combinations of transcription factors under the control of extrinsic signals that can modulate the expression and activity of transcription factors, forming a gene regulatory network that dictates a specific cellular phenotype. Tissue-specific transcription factors play deterministic roles in cell-type specification, which is manifested as lineage reprogramming by forced expression of such transcription factors (Graf and Enver, 2009; Zhou and Melton, 2008). Sox2 is one such transcription factor required for the maintenance of pluripotent stem cells in vivo (Avilion et al., 2003) and in vitro (Masui et al., 2007) and for the induction of pluripotency (Takahashi and Yamanaka, 2006). However, it is also preferentially expressed in neural, retinal, and trophoblast stem cells (TSCs) (Avilion et al., 2003; Pevny and Nicolis, 2010), suggesting a possible role for Sox2 in governing a common stemness phenotype.

In embryonic stem cells (ESCs), Sox2 forms a heterodimer with Oct3/4 (also known as Pou5f1) on DNA with the OCT-SOX composite motifs, and these factors cooperatively activate pluripotency-related target genes such as *Nanog*, *Fgf4*, *Utf1*, *Lefty1*, and *Fbxo15*, as well as their own expression (Nakatake et al., 2006, and references therein). *Oct3/4*-knockout ESCs are differentiated along the trophoblast lineage in a highly homogeneous manner (Niwa et al., 2000). In contrast, the loss of *Sox2* causes differentiation of ESCs accompanied by upregulation of markers for trophoblast and embryonic germ layers, although artificial maintenance of *Oct3/4* from the transgene can sustain self-renewal and pluripotency of *Sox2*-null ESCs (Masui et al., 2007), suggesting that the unique function of Sox2 may be to maintain *Oct3/4* expression. These two core transcription factors, along with *Nanog*, form an interconnected and hierarchical network downstream of the leukemia inhibitory factor (LIF)-Stat3 and LIF-phosphatidylinositol 3-kinase (PI3K) signaling

pathways, which are connected together via signal-responsive transcription factors, such as *Klf4* and *Tbx3* (Niwa et al., 2009). However, the expression of core transcription factors, including Sox2, is also maintained by shielding from ERK signaling even in the absence of LIF, so long as the GSK3 activity is simultaneously blocked (Ying et al., 2008), indicating that the ERK signaling, mainly activated by autocrine fibroblast growth factor (FGF) 4, negatively regulates the transcriptional network for pluripotency in ESCs.

The role of Sox2 as a crucial safeguard of stem cell state in ESCs could be phenotypically conserved in other types of stem cells. In neural stem cells and retinal progenitor cells, Sox2 can cooperate with other tissue-specific homeodomain transcription factors, such as Brn2 (also known as Pou3f2) and Pax6, respectively, to activate specific target genes (Inoue et al., 2007; Tanaka et al., 2004) and plays an essential role in stem cell maintenance (Favaro et al., 2009; Taranova et al., 2006). In TSCs, Sox2 is expressed predominantly in undifferentiated cells (Avilion et al., 2003). Although a previous study on *Sox2*-knockout mice revealed its critical role in the trophoblast lineage in vivo (Avilion et al., 2003), the mode of Sox2 function, as well as its possible partners, is unknown. TSCs are derived from embryos at a similar developmental stage as ESCs but originate from a distinct lineage and show completely different characteristics from ESCs. Self-renewal of TSCs is absolutely dependent on FGF-ERK signaling that supposedly activates the expression of TSC-specific transcription factors such as *Cdx2* and *Eomes* (Kunath et al., 2004; Tanaka et al., 1998), which is in contrast to the detrimental effect of ERK activation in ESCs. Therefore, it is intriguing to examine the extent to which Sox2 functions are conserved within discrete transcriptional networks in these well-characterized and developmentally similar stem cells.

Here, we performed a comprehensive analysis of Sox2 functions in ESCs and TSCs. First, we analyzed functions of transcription factors that are activated downstream of FGF signaling in TSCs and revealed the pivotal role of Sox2 in FGF-dependent self-renewal of TSCs. Then we compared the target genes regulated by Sox2 in ESCs and TSCs using transcriptome and chromatin immunoprecipitation analyses and found that Sox2 regulates quite different sets of genes in ESCs and TSCs. Finally, we identified Tfap2c as one of the TSC-specific Sox2 partners partly responsible for the recruitment of Sox2 to TSC-specific target genes. These data indicated that the function of Sox2 is highly flexible in the context of transcriptional networks controlled by different extrinsic signals.

## RESULTS

### FGF-ERK Signaling Converges on a Subset of Transcription Factors in TSCs

We have previously shown that trophoblast stem (TS)-like cells can be induced by forced repression of *Oct3/4* in ESCs, which results in gradual but homogenous downregulation of ESC markers, like Nanog, and concomitant upregulation of TSC markers, like Cdx2, when exogenous FGF4 is supplied (Figure 1A and see Figure S1A available online; Niwa et al., 2000, 2005). After subcloning, these TS-like cells (hereafter called ZHBTc4-TSCs) were stably maintained as TSCs with a global expression profile similar to that of embryo-derived EGFP-$TS_{3.5}$ TSCs (eTSCs) (Tanaka et al., 1998; Figure 3A). This in vitro system allowed us to compare transcriptional networks in two distinct types of stem cells with identical genotypes. We first investigated the effect of FGF4 on the expression of key transcription factors in TSCs, which are less well understood than the ESC transcription factors regulated by LIF. Increased expression was found for transcription factors previously implicated in TSC maintenance, including *Cdx2* (Chawengsaksophak et al., 1997; Strumpf et al., 2005), *Eomes* (Russ et al., 2000; Strumpf et al., 2005), *Elf5* (Donnison et al., 2005), *Esrrb* (Luo et al., 1997), and *Sox2* (Avilion et al., 2003) (Figure 1B). In contrast, the expression of differentiated trophoblast markers, such as *Ascl2* (also known as *Mash2*) and *Cdkn1c* (also known as *p57Kip2*), was repressed by FGF4 (Figure S1B). Other possible regulators of trophoblast development, such as *Ets2*, *Tfap2c*, and *Tead4*, did not appear to be activated by FGF4 (Figure S1B). The FGF responsiveness was confirmed by the rapid downregulation in ZHBTc4-TSCs and eTSCs after withdrawal of FGF4 or treatment with the FGF receptor inhibitor PD173074 (Figures 1C and S1C). Conversely, when TSCs were stimulated with FGF4 after starvation, the transcription factors were upregulated in 12 hr (Figure 1D). This effect was abrogated in the presence of the MEK inhibitor PD0325901, but not by the PI3K inhibitor LY294002 (Figure 1D), indicating that it was primarily mediated by the FGF-ERK signaling pathway. Consistent with this, these transcription factors showed largely overlapping expression with phosphorylated ERK in the extraembryonic ectoderm of postimplantation embryos (E5.5), although the expression of *Esrrb* and *Sox2* was even more highly restricted, and that of *Eomes* and *Elf5* was extended (Figure 1E). It should be noted that the response of *Esrrb* and *Sox2* to FGF is TSC specific, as the expression of these genes in ESCs did not depend on FGF signaling (Figure 1F). Instead, their expression was maintained by either activation of LIF signaling or intrinsic transcriptional circuitry without the counteracting FGF signal (Figure 1F; Niwa et al., 2009; Ying et al., 2008). Indeed, around the time of implantation (E3.5–E4.75), they are expressed exclusively in the epiblast without detectable phosphorylated ERK (Figure 1G; data not shown). These results also suggest that TSCs, with their FGF/ERK dependency, correspond to a population of trophoblasts in postimplantation, but not in preimplantation, embryos.

### Sox2 Is a Critical Downstream Target of FGF Signaling that Mediates Self-Renewal of TSCs

To evaluate the functional roles of these FGF-responsive transcription factors in self-renewal downstream of FGF-ERK signaling, we used a gain-of-function approach using eTSCs to test whether the requirement for FGF4 can be replaced by artificial maintenance of these transcription factors from transgene expression. Although sustained expression of a single factor alone did not confer independence from FGF4, combined expression of Sox2 and Esrrb enabled prolonged self-renewal in the absence of FGF4 (Figures S2A and S2B). These results were confirmed using ZHBTc4-derived TS-like cells, although

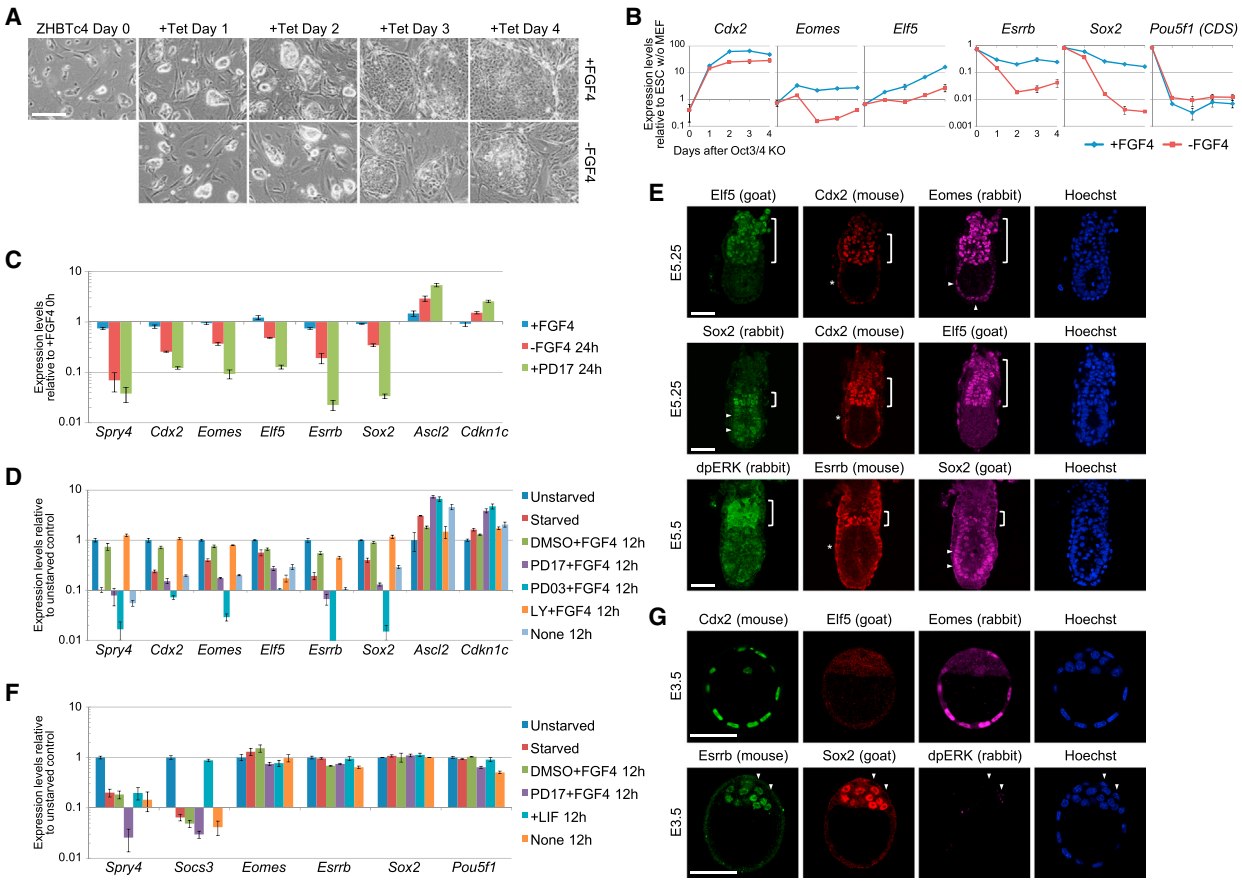



**Figure 1. Integration of FGF-ERK Signaling into the Transcription Factor Network in TSCs**

(A) Trophoblast differentiation of ZHBTc4 ESCs after treatment with Tet to repress *Oct3/4* expression. The cells formed TSC-like colonies on MEF feeders only in the presence of FGF4. Scale bar, 200 μm.

(B) Transcription factors positively regulated by FGF4 during the trophoblast differentiation. Expression levels relative to ESCs cultured without MEFs are shown in logarithmic scale.

(C) Downregulation of the TSC transcription factors by the inhibition of FGF signaling. ZHBTc4-TSCs were deprived of FGF4 or treated with the FGF receptor inhibitor PD173074 (PD17) for 24 hr before harvest. *Spry4*, a direct target of the FGF signaling, served as a control.

(D) Induction of the TSC transcription factors by the FGF-ERK signaling pathway. ZHBTc4-TSCs were starved in 1% FBS for 20 hr and restimulated with FGF4 for 12 hr. The FGF receptor inhibitor PD173074 (PD17), the MEK inhibitor PD0325901 (PD03), or the PI3K inhibitor LY294002 (LY) was added 30 min prior to re-stimulation. Expression levels relative to unstarved control are shown.

(E) Expression of FGF-responsive transcription factors and phosphorylated ERK in postimplantation embryos. Embryos were immunostained for TSC transcription factors or dually phosphorylated ERK (dpERK) and analyzed with a confocal laser scanning microscope. Two different antibodies against Sox2 (rabbit and goat) yielded similar results. Nonspecific binding of anti-mouse IgG to endogenous maternal immunoglobulins concentrated in the cytoplasm of vacuolated visceral endoderm was observed (asterisks). Note that, besides their expression in the trophoblast lineage (brackets), Eomes and Sox2 were also detected in the visceral endoderm and epiblast, respectively (arrowheads). Scale bars, 50 μm.

(F) Response to FGF or LIF in ESCs. EB5 ESCs were starved in 1% FBS for 20 hr and restimulated with FGF4 or LIF for 12 hr. The FGF receptor inhibitor PD173074 (PD17) was added 30 min prior to restimulation. Note that autocrine FGFs are enough to mask the effect of exogenously added FGF4 on, for instance, *Spry4* expression. Socs3, a direct target of the LIF signaling, served as a control.

(G) Expression of FGF-responsive transcription factors and phosphorylated ERK in preimplantation embryos. Note that Esrrb and Sox2 were not detected in the polar trophectoderm (arrowheads). The expression of Elf5, Esrrb, Sox2, and dpERK in the trophoblast lineage was observed only after E4.0, E4.75, E5.25, and E5.0, respectively (data not shown). Scale bars, 50 μm.

(B–D and F) Error bars indicate standard deviation (SD) of three replicates.

See also Figure S1.

Sox2 alone slightly promoted self-renewal (Figures S2C and S2D). We next assessed the multipotency of the transgenic TSCs by chimera formation assay. The stable transfection of an expression vector for *Sox2* and *Esrrb*, linked by a self-cleaving 2A peptide sequence and flanked by *loxP* sites, enabled the self-renewal of eTSCs at a clonal density in the absence of FGF4 (Figure 2A). These cells maintained *Cdx2* and *Eomes* expression even in the presence of PD173074, confirming their independence from FGF signaling (Figure 2B). After the culture without FGF4 for 4 weeks, the cells were transiently transfected

with *Cre* recombinase to excise the floxed transgene in the presence of FGF4, followed by negative selection against thymidine kinase with gancyclovir. The resulting transgene-deleted clones restored FGF dependency and readily differentiated after withdrawal of FGF4 (Figures 2C and S2E). When the reverted TSCs were injected into blastocysts, they contributed to all of the placental cell lineages in chimeras at the same efficiency as the parental line maintained in the presence of FGF4 (Figures 2D, 2E, and S2F), demonstrating that the expression of *Sox2* and *Esrrb* is minimally sufficient to maintain the multipotency of TSCs in the absence of FGF4.

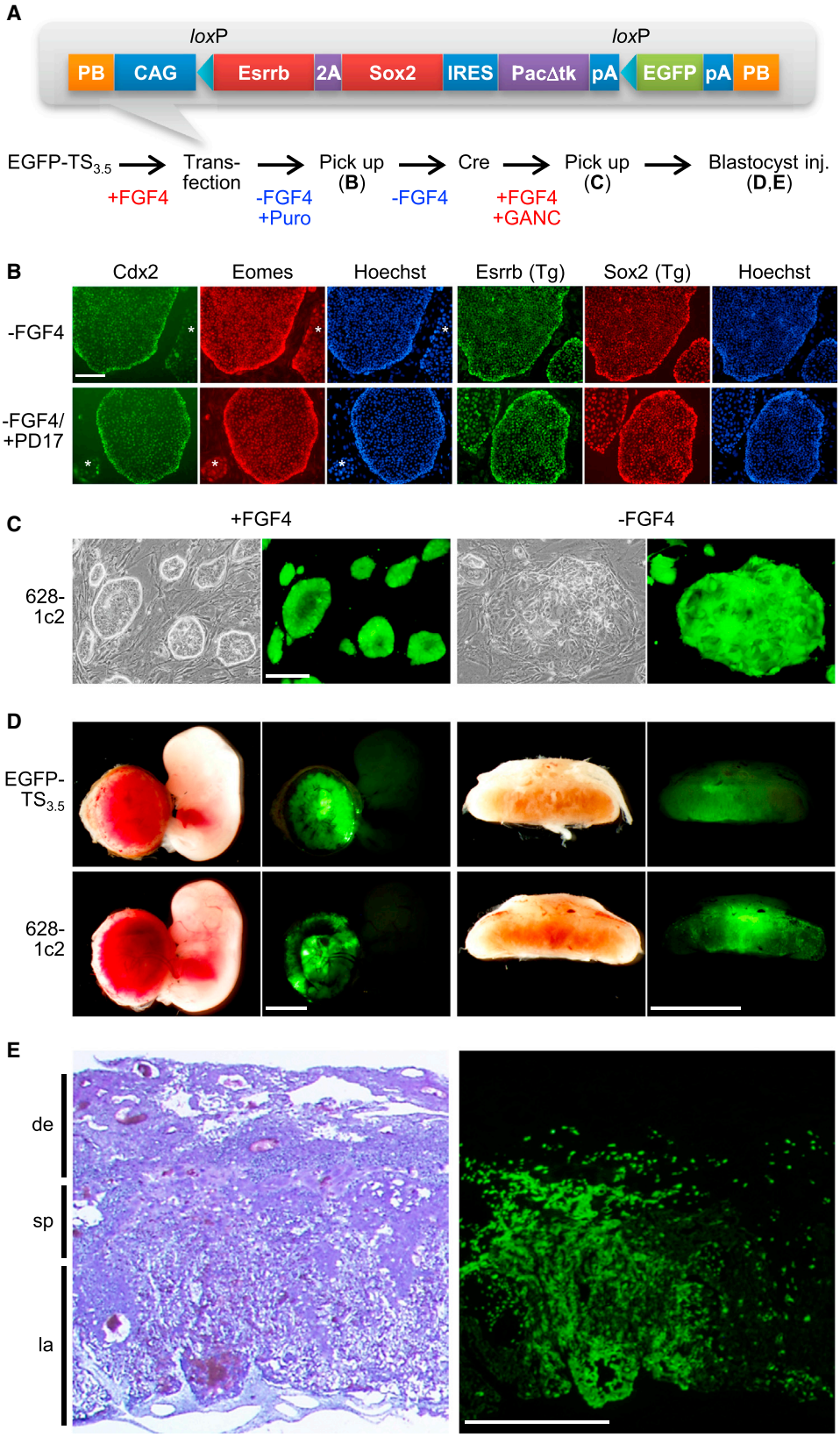
## Induced Knockout of Sox2 in ESCs and TSCs Causes Differentiation along Their Respective Developmental Pathways

The results described above indicate that *Sox2* plays a central role in regulating self-renewal of TSCs downstream of FGF signaling. The function of Sox2 in TSCs stands in contrast to that in ESCs, in which Sox2 acts as a core transcription factor downstream of LIF signaling to govern pluripotency (Niwa et al., 2009). In ESCs, Sox2 prevents trophoblast differentiation mainly by maintaining *Oct3/4* expression or repressing *Eomes* expression and restricts the specification of the primary germ layers as well (Ivanova et al., 2006; Masui et al., 2007).

To compare the functional role of Sox2 in ESCs and TSCs in more detail, we performed whole-genome expression analysis after induced knockout of *Sox2* in each cell type. For this, we first established a stable TSC line by transient induction of Cdx2 (Niwa et al., 2005) in the Tet-inducible *Sox2*-knockout ESC line, 2TS22C (Masui et al., 2007). Global gene expression profiles of ZHBTc4- and 2TS22C-derived TSCs, together with embryo-derived EGFP-$TS_{3.5}$, revealed that these TSCs are clustered together and separated from the ESCs along the primary principal component (PC1) axis (Figure 3A). Upon knockout of *Sox2* in TSCs, expression of TSC markers, especially Esrrb, gradually decreased even in the presence of FGF4, concomitant with upregulation of differentiated trophoblast markers (Figures 3B and S3A). These TSCs gradually lost their self-renewal capacity, manifesting flat morphological changes (Figure 3C). Global expression analysis identified 3,440 and 1,066 genes that were affected by *Sox2* knockout in ESCs and TSCs, respectively (Figures 3D, 3E, and S3B–S3F; Table S1). The smaller number of affected genes in TSCs, especially upregulated genes, could represent the restricted developmental capacity of TSCs compared to ESCs. Only a few genes were commonly downregulated in ESCs and TSCs, while larger numbers of genes, especially genes involved in placental development, were upregulated in both cell types; this makes sense, considering Sox2's role in preventing trophoblast commitment in ESCs (Masui et al., 2007). Downregulation of pluripotency markers and concomitant upregulation of mesodermal and endodermal markers, but not neuroectodermal markers, was observed exclusively in ESCs (Figures 3F and 3G). In contrast, genes involved in cell-cycle regulation, especially in mitotic checkpoint control, were downregulated only in TSCs (Figure 3G); this likely reflects the endoreduplication cycle, characteristic of differentiating trophoblast giant cells (Ullah et al., 2008). Gene ontology (GO) analysis using DAVID (Huang da et al., 2009) showed that GO terms related to placental function, like blood vessel development and negative regulation of immune system processes, are overrepresented in genes upregulated in TSCs and, to a lesser extent, in genes upregulated in ESCs (Figures 3G, S3G, and S3H; Table S2). Gene set enrichment analysis (GSEA; Subramanian et al., 2005) using published microarray data (Ralston et al., 2010; Schulz et al., 2009) further revealed that genes affected by *Sox2* knockout are preferentially enriched in genes whose expression is similarly regulated during normal differentiation of the respective cell types (Figures 3H, 3I, and S3I). These results indicate that Sox2 regulates unique sets of genes to maintain stemness and prevent precocious differentiation along the possible developmental pathways in each stem cell type.

## Sox2 Occupies Distinct Sets of Genes Implicated in Later Differentiation in ESCs and TSCs

To better understand the mechanisms underlying the versatility in transcriptional regulation by Sox2, we next analyzed genome-wide binding sites of Sox2 during the differentiation of ZHBTc4 ESCs into TS-like cells by chromatin immunoprecipitation, followed by high-throughput sequencing (ChIP-seq) (Figures 4A–4C and S4A–S4F; Table S3). During the differentiation, Sox2 binding sites changed dynamically. For instance, a conserved region 4 within a distal enhancer of *Pou5f1* was preferentially occupied in ESCs (Figure 4A), while the *Cdx2* locus was specifically occupied in differentiating TSCs (Figure 4B). Persistent binding was less frequently observed, as exemplified at the *Col4a1* locus (Figure 4C). As a whole, the numbers of peaks (FDR < 0.05) that overlapped with those of other samples were quite limited (Figure S4G). K-means clustering of normalized ChIP-seq signals around the merged peak regions over the course of differentiation revealed that the vast majority of Sox2 binding sites are cell-type specific (Figure 4D). About half of the Sox2 binding sites are ESC specific (cluster 2 [c2] and c5), while the other half are predominant in differentiating TSCs (c1, c3, and c4), which are further classified as either increased (c1), transiently increased (c3), or gained (c4) binding. The Sox2 biding sites in any clusters are predominantly found at transcription start site (TSS)-distal intronic or intergenic regions (Figure 4E). To gain insights into the underlying mechanisms for such a context-dependent Sox2 binding, we performed de novo motif analysis using MEME (Bailey et al., 2009) and found that different noncanonical motifs, besides the canonical SOX motif, were enriched in the peak centers of each cluster (Figure 4F). We then mapped these motifs to the peak regions (Figure 4G). As expected, overrepresentation of the OCT-SOX composite motif was observed only in ESC-specific clusters (c2 and c5), confirming the pivotal role of Oct3/4 as a recruiter of Sox2 to the ESC-specific target sites. Interestingly, the AP-2 binding motif was overrepresented only in the TSC-predominant clusters (c1, c3, and c4) and found together with the canonical SOX motif within a subset of peaks (Figure 4H), suggesting co-binding of Sox2 and AP-2 transcription factors. It also should be noted that although the canonical SOX motif was enriched in all clusters, it was more overrepresented in c1 and c3, where Sox2 binding is relatively persistent. Although the ESRRB motif was also overrepresented in Sox2 peak clusters, we could not




(legend on next page)

find a significant difference of the motif enrichment in each peak cluster (data not shown). We also mapped these motifs to the peak summits of original samples, which were ranked by the $-\log_{10}$ (q-value) for peak call (Figure 4I). The more confident peaks in ESCs were highly enriched for the OCT-SOX motif compared to the less confident peaks. In contrast, the more confident peaks in differentiating TSCs were enriched for the canonical SOX motif, suggesting a stronger dependence of Sox2 binding on its own recognition motif. These results suggest that sequence features could, in part, explain the dynamic behavior of Sox2 binding.

To explore the functional significance of differential Sox2 binding, GO enrichment analysis was performed using GREAT (McLean et al., 2010), considering the predominance of TSS-distal binding sites (Figure 4J; Table S4). Genes involved in stem cell maintenance were enriched for the peaks of all clusters, but especially c5, while genes for early placental development were enriched for c3, c4, and c5. Enrichment of genes associated with placental functions, such as artery development, regulation of lymphocyte differentiation, and regulation of T cell activation, are observed in c1, c2, or c3. These results suggest that contextual Sox2 binding is related to specific developmental processes.

Next we investigated the functional consequence of Sox2 binding by comparing the ChIP-seq data with the microarray data. The peaks of each cluster were assigned to RefSeq genes whose TSSs are located within a 10 kb distance (Table S5). Enrichment of these genes in genes affected by *Sox2* knockout in ESC or TSC was analyzed by GSEA (Figures 4K, 4L, and 4N). We found that genes associated with the ESC-specific clusters (c2 and c5) are significantly enriched in genes downregulated in *Sox2*-knockout ESCs, while genes near the TSC-predominant clusters (c1, c3, and c4) are enriched in genes downregulated in *Sox2*-knockout TSCs. GSEA using the microarray data of ZHBTc4 ESCs differentiating into TSCs showed that genes associated with the ESC-specific clusters (c2 and c5) and with the cluster showing gains of Sox2 binding in TSCs (c4) are significantly enriched in genes downregulated and upregulated during the trophoblast differentiation, respectively (Figures 4M and 4N). It should be noted that when more distal binding sites were assigned to genes, such a positive effect of Sox2 binding on expression of nearby genes was less evident (data not shown). These results suggest that Sox2 primarily contributes to activation of neighboring genes in both ESCs and TSCs. Nevertheless, this does not exclude a role of distal Sox2 binding in transcriptional regulation and a role of Sox2 binding in priming gene expression without significant consequences (see Discussion). Taken together, self-renewal of ESCs and TSCs is regulated by a context-dependent role of Sox2.

We also performed ChIP-seq analysis of Esrrb using the same chromatin samples (Figures S4A–S4F). In stark contrast to Sox2, Esrrb bindings were less dynamically regulated during the trophoblast differentiation (Figure S4G). K-means clustering of normalized ChIP-seq signals around the merged peak regions identified ESC-specific binding peaks (c5) as a major cluster (Figure S4I). The remaining peaks were classified as relatively persistent clusters, with a TSC-predominant cluster not apparent. The Esrrb biding sites, especially of c2 and c3, were frequently found around the TSS (Figure S4J), consistent with the suggested role of Esrrb in collaboration with basal transcription machinery (van den Berg et al., 2010). The peaks of all clusters were enriched for the canonical ESRRB recognition motif (a motif for the nuclear receptor superfamily), and strong binding sites were associated with the occurrence of the ESRRB motifs (Figures S4L and S4M). These results suggest a common mechanism for Esrrb in regulating stemness in ESCs and TSCs. Although a minor fraction of Sox2 peak regions was also bound by Esrrb, we found only a slight enrichment of genes related to specific biological processes near these regions (Figures S4N–S4P), suggesting that Sox2 and Esrrb might play complementary roles via different sets of target genes.

## Tfap2c Is a Cooperative Factor of Sox2 in TSCs

The motif analysis described above suggests a possible role of AP-2 transcription factors in cooperating with Sox2 in TSCs. Among the five members of the AP-2 family of transcription factors, *Tfap2c* (also known as *Tcfap2c* or AP-2γ) is most abundantly expressed in TSCs (Figure 5A), and its expression is restricted to the trophoblast lineages in vitro and in vivo (Figures 5B, 5C, and S1B). The forced expression of *Tfap2c* rapidly suppressed *Oct3/4* expression and induced trophoblast differentiation in ESCs (Figures S5A and S5B), confirming previous reports (Berg et al., 2011; Kuckenberg et al., 2010). When *Tfap2c* expression was induced in ESCs, the enhanced recruitment of

**Figure 2. FGF-Independent Self-Renewal of TSCs by Sustained Expression of *Sox2* and *Esrrb***

(A) Scheme for assessment of developmental capacity of FGF-independent TSCs with sustained expression of *Sox2* and *Esrrb*. A *piggyBac* vector carrying a floxed transgene cassette for bicistronic expression of *Sox2* and *Esrrb* was transfected into the eTSC line EGFP-$TS_{3.5}$ (Tanaka et al., 1998). PB, terminal repeats of the *piggyBac* transposon; 2A, a 2A peptide from porcine teschovirus-1; *pacΔtk*, a fusion gene composed of a puromycin N-acetyltransferase and a truncated version of HSV-1 thymidine kinase; GANC, gancyclovir. The letters in brackets refer to the panels in this figure corresponding to each step.

(B) Immunostaining of the Sox2/Esrrb transgenic TSCs (clone 628-1) grown in the absence of FGF4 with or without PD173074. Without FG4, transgenic cells were positive for the TSC markers Cdx2 and Eomes, although some spontaneously differentiated cells (asterisks) were negative or only weakly positive. Note that the parental EGFP-$TS_{3.5}$ line constitutively expresses *EGFP*, but the fluorescence intensity was low and almost negligible compared to the immunofluorescence signals. Scale bar, 200 μm.

(C) Restoration of FGF dependency in reverted TSCs. The self-renewal of TSCs after Cre-mediated excision of the transgenes (clone 628-1c2) became FGF dependent. Note that the reverted clone showed higher *EGFP* expression than the parental line. Scale bar, 200 μm.

(D) E13.5 chimeric embryos and placentas obtained by blastocyst injection of the parental (EGFP-$TS_{3.5}$, upper) or reverted (clone 628-1c2, lower) TSCs maintained in the presence of FGF4. Scale bars, 4 mm.

(E) Transverse section of placentas from the chimeras (clone 628-1c2). The differentiated progeny of the injected TSCs were observed in the spongiotrophoblast (sp) and labyrinth (la) layers, as well as in the interstitial spaces of the maternal decidua (de). Scale bar, 1 mm.

See also Figure S2.

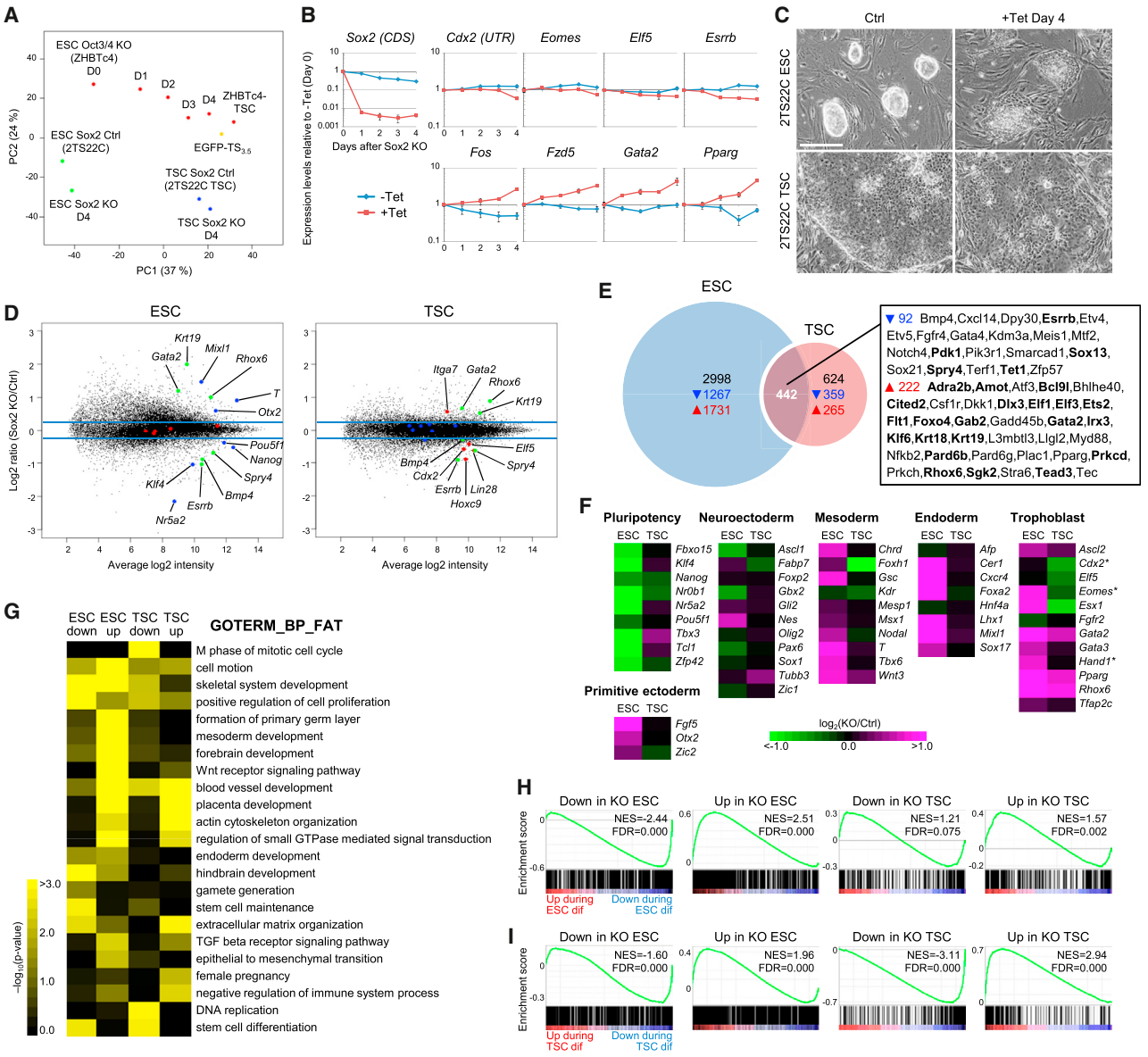



**Figure 3. Microarray Analysis of Genes Affected by *Sox2* Knockout in ESCs and TSCs**

(A) Principal component analysis (PCA) of global gene expression profiles. The first two PCs explain 37% and 24%, respectively, of the total variance.

(B) Changes in gene expression after Tet-induced *Sox2* knockout in TSCs. Error bars indicate SD of three replicates.

(C) Photomicrographs of Tet-inducible *Sox2*-knockout ESC line 2TS22C and its derivative TSC line cultured in the absence or presence of Tet for 4 days. Scale bar, 200 μm.

(D) MA plots showing the $\log_2$ ratio and average $\log_2$ intensity of control and *Sox2*-knockout ESCs or TSCs. Whole-genome expression analysis was performed 4 days after *Sox2* knockout. Blue lines indicate $\log_2$ ratio cutoff of 0.25. Blue, red, or green dots indicate selected genes affected only in ESCs, in TSCs, or in both, respectively.

(E) The number of genes affected by *Sox2* knockout in ESCs and TSCs. Cutoffs for q-value and $\log_2$ ratio were set to 0.1 and 0.25, respectively. The numbers of genes upregulated or downregulated are shown in red or blue, respectively. The remaining genes in the intersection were affected in an opposite manner in ESCs and TSCs. Selected genes downregulated or upregulated in common are shown. Genes in bold were also considered significant when more stringent cutoff values (q-value < 0.05) were used (Figure S3F).

(F) Changes in expression levels of lineage markers. Note that some markers for trophoblast lineage (asterisks) are also expressed in embryonic tissues.

(G) GO analysis of genes affected by *Sox2* knockout using DAVID. The $-\log_{10}$ (p value) for selected GO terms in the Biological Process category (GOTERM_BP_FAT) is shown in a heatmap.

(H and I) GSEA of changes in expression of genes affected by *Sox2* knockout during the differentiation of ESCs (H) and TSCs (I). Genes are ordered according to their expression levels in differentiating ESCs (embryoid bodies) relative to undifferentiated ESCs (H) or levels in differentiating TSCs relative to undifferentiated TSCs (I). Normalized enrichment scores (NES) and FDR are shown.

See also Figure S3.

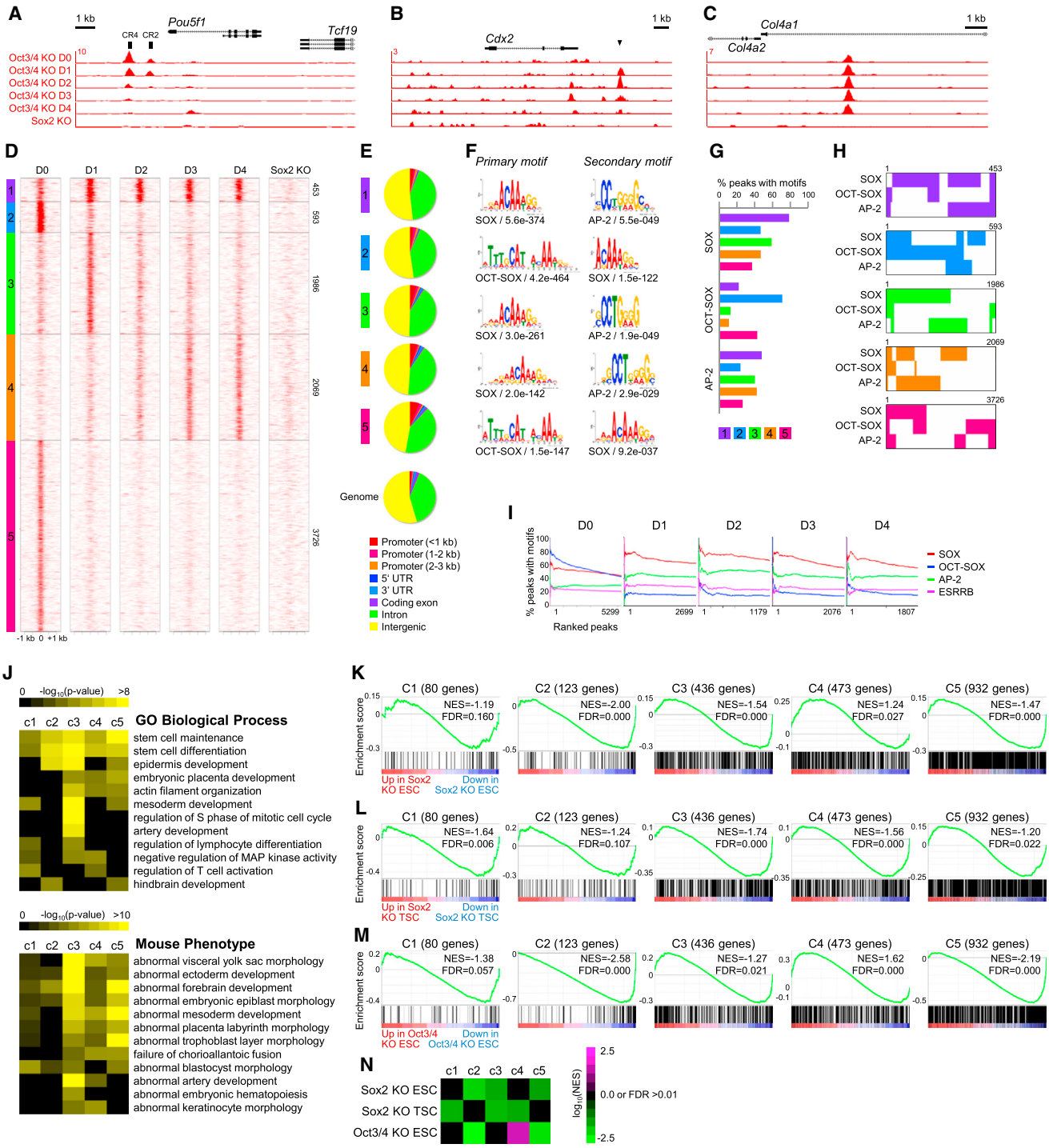



**Figure 4. ChIP-Seq Analysis of Sox2 Occupancy during Differentiation of ESCs into TSCs**

(A–C) Normalized Sox2 ChIP-seq signals at the *Pou5f1* (A), *Cdx2* (B), and *Col4a1* (C) loci during the differentiation of ZHBTc4 ESCs into TS-like cells. ChIP-seq signals were normalized as levels per million mapped reads. The chromatin of Sox2 KO ESCs was used as a negative control. The *Cdx2* region analyzed by ChIP-PCR in Figure 5 is shown (an arrowhead).

(D) A heatmap of the ChIP-seq signals around the merged peak regions. ChIP-seq peaks of each sample (q-value < 0.05) were merged, and normalized ChIP-seq signals over ± 1 kb regions around the merged peak centers were clustered with the k-means algorithm into five clusters.

(E) Annotation of ChIP-seq clusters with genomic features. The merged peak centers of each cluster were annotated with genomic features using CEAS (Shin et al., 2009). Annotation of the whole genome is shown as a control.

*(legend continued on next page)*

Sox2 was observed at the Tfap2c binding sites, including the *Cdx2* locus, as early as one day after induction, even with sustained expression of *Oct3/4* by the transgene (Figures 5D and S5C). Conversely, knockdown of *Tfap2c* significantly decreased TSC-specific recruitment of Sox2 (Figures 5E and S5E). We confirmed that overexpression or knockdown of *Tfap2c* did not increase or decrease Sox2 protein levels, respectively (Figure S5F), which excludes the possibility that the observed changes in Sox2 binding are simply due to the modulation of Sox2 protein levels. To elucidate the mechanistic basis of cooperative binding, we performed coimmunoprecipitation experiments of endogenous Sox2 and Tfap2c proteins in TSCs. Tfap2c was coimmunoprecipitated with Sox2 (Figure 5F, left) and, reciprocally, Sox2 was coimmunoprecipitated with Tfap2c (Figure 5F, right), indicating that these proteins physically interact with each other. Importantly, overexpression of Tfap2c together with Sox2 also enabled the self-renewal of TSCs without FGF4 (Figures 5G and 5H), whereas knockdown of *Tfap2c* prevented the formation of undifferentiated TSC colonies (Figure S5D). Taken together, these results suggest that Tfap2c is a crucial cooperative factor for Sox2 to maintain self-renewal of TSCs.

To gain more global insights into the cooperativity between Sox2 and Tfap2c, we performed ChIP-seq analysis of these factors, along with Cdx2, in ZHBTc4-TSCs. We also analyzed the Sox2 binding sites in eTSCs, confirming a large overlap with those in ZHBTc4-TSCs (Figure S6I). As expected, a number of previously identified TSC-predominant Sox2 binding sites, including the *Esrrb*, *Eomes*, and *Atrx/Magt1* loci, were co-occupied by Sox2 and Tfap2c in ZHBTc4-TSCs (Figures 6A–6C and S6A–S6H). A significant number of Sox2 peaks overlapped with Tfap2c peaks, while Cdx2 peaks that overlapped with Sox2 or Tfap2c peaks were limited, irrespective of the relatively large numbers of peaks (Figure 6D). Interestingly, Tfap2c peaks were significantly enriched at proximal promoters of known genes, while Cdx2 binding sites were found exclusively at TSS-distal regions (Figure 6E). De novo motif discovery identified the canonical recognition motifs for each factor as primary motifs (Figure 6F). The motifs for AP-1 and AP-2, and TEAD and AP-1 transcription factors were also enriched in Sox2 and Tfap2c peaks, respectively. Motif mapping into the ranked peaks suggests that Tfap2c and Cdx2 bindings primarily depend on their recognition sequences, while the presence of canonical SOX motif is less predictive for Sox2 bindings in vivo, further emphasizing the contextual binding of Sox2 (Figure 6G). Finally, we found TSC-predominant Sox2 peak clusters (c1, c3, and c4), but not ESC-specific clusters (c2 and c5), shown in Figure 4D were significantly overlapped with the Sox2 and Tfap2c peaks in TSCs (Figure 6H). Taken together, these results suggest a global cooperativity of Sox2 and Tfap2c in regulating transcriptional network unique to TSCs.

## DISCUSSION

Tissue-specific transcription factors are primarily responsible for directing transcriptional programs that confer unique cellular phenotypes. However, it is frequently found that the same transcription factors are recurrently expressed in different cell types at different stages during development. Sox2 is one such factor whose expression is dynamically regulated: inner cell mass at blastocyst stage; primitive ectoderm and extraembryonic ectoderm at egg-cylinder stage; neuroectoderm around gastrulation; and central nervous system, sensory placodes, and foregut endoderm thereafter (Collignon et al., 1996; Wood and Episkopou, 1999). In addition, Sox2 is also preferentially expressed in putative stem cell populations of various adult tissues (Arnold et al., 2011). Studies of germline or conditional knockout mice revealed that Sox2 is definitely required for the development and maintenance of these tissues (Arnold et al., 2011; Avilion et al., 2003; Favaro et al., 2009; Que et al., 2009; Taranova et al., 2006). Is such a pleiotropic role of Sox2 based on its common functions in different cell types or different modes of action in different contexts? Here, we have shown the different modes of Sox2 action in ESCs and TSCs underpinning the recursive deployment of Sox2 in different cell types during embryonic development.

First we found that Sox2 is both necessary and sufficient, if combined with Esrrb or Tfap2c, to mediate the effect of FGF on self-renewal of TSCs. Although some transcription factors are considered to be involved in this process, no factors have been identified that can replace FGF dependency. Although *Sox2*-deficient mice die shortly after implantation due to defective epiblast development (Avilion et al., 2003), chimeric analysis using *Sox2*-null blastocyst injected with wild-type ESCs suggests a critical role of Sox2 in the trophoblast lineage around E7.5, when *Erk2* (also known as *Mapk1*)- or *Frs2*-deficient mice

(F) Motifs found by MEME to be overrepresented in 600 randomly selected ± 100 bp regions around the merged peak centers of each cluster with E-value scores (from c1 to c5, top to bottom).
(G) Occurrence of the motifs in peaks of each cluster. The motifs were mapped to the ± 200 bp regions around the merged peak centers of each cluster, and the percentages of peaks containing at least one sequence for a given motif are shown.
(H) Co-occurrence of the motifs in peak centers of each cluster (from c1 to c5, top to bottom).
(I) Occurrence of the motifs in the ranked Sox2 peaks during differentiation of ESCs into TSCs. The peak summits of each sample were ranked by the $-\log_{10}$ (q-value) for peak call. The cumulative percentages of peaks (summits ± 200 bp) containing at least one sequence for a given motif are plotted against ranks of peaks. The ESRRB motif is a motif discovered in the Esrrb ChIP-seq data (Figure S4K).
(J) GREAT ontology enrichments for the peaks of each cluster. The binomial raw p values for selected GO Biological Process and Mouse Phenotype are shown in a heatmap.
(K–N) GSEA of changes in expression of genes associated with the peaks of each cluster after *Sox2* knockout in ESCs (K), after *Sox2* knockout in TSCs (L), and during the trophoblast differentiation of ESCs induced by *Oct3/4* knockout (M). The ChIP-seq peaks of each cluster were assigned to RefSeq genes whose TSSs are located within 10 kb of the peak centers. The NES for gene sets with a FDR < 0.01 is shown in a heatmap (N). The core enrichment genes that contributed most to the enrichment results are listed in Table S5.
See also Figure S4.

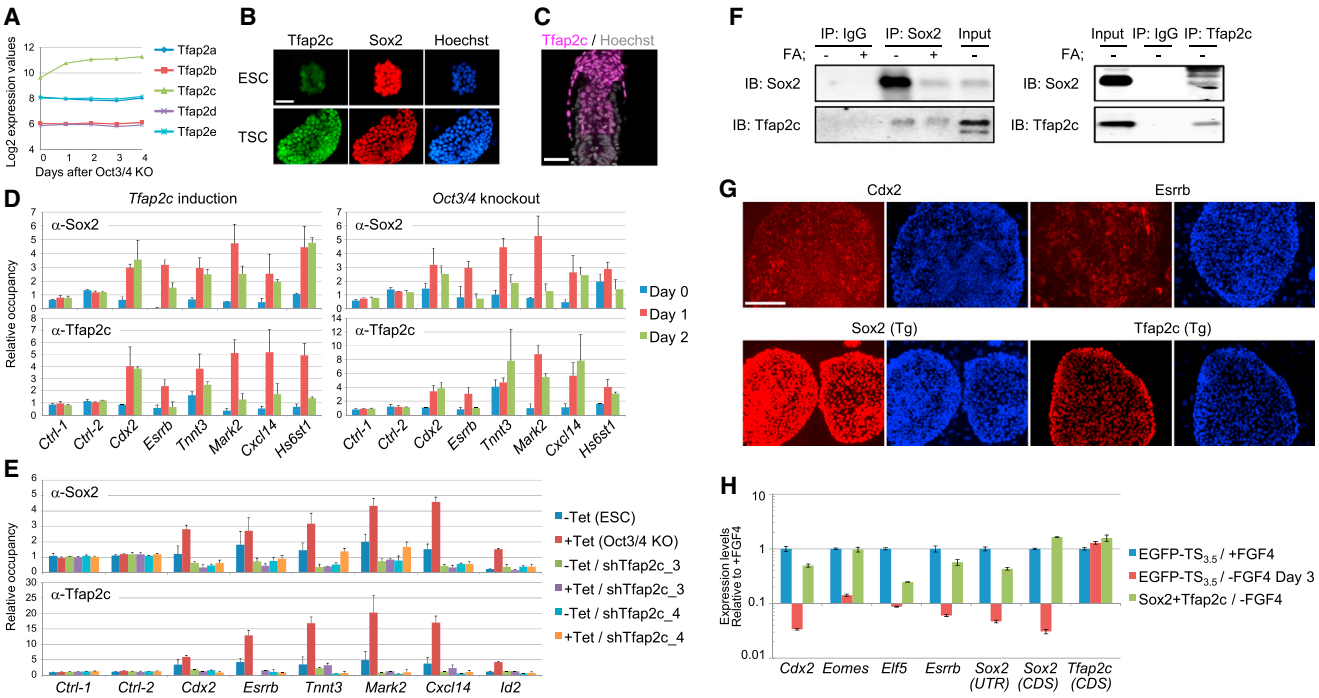



**Figure 5. Tfap2c-Dependent Recruitment of Sox2 in TSCs**

(A) Expression of AP-2 family members during the differentiation of ZHBTc4 ESCs into TS-like cells. $Log_2$-transformed expression values from the Affymetrix Exon Array data are shown. Note that AP-2 transcription factors, except Tfap2c, are expressed at negligible levels.

(B) Immunostaining for Tfap2c in ZHBTc4 ESCs and TSCs. Scale bar, 100 μm.

(C) Immunostaining for Tfap2c in postimplantation embryos (E5.5). Scale bar, 50 μm.

(D) Recruitment of Sox2 after induced expression of *Tfap2c* or knockout of *Oct3/4*. *Oct3/4* expression was regulated by the Tet-off transgene, and expression of *Tfap2c* was induced using an ecdysone receptor (EcR)-based system in the same cell line. Relative occupancies to control regions (Ctrl-1 and Ctrl-2) are shown.

(E) Abrogation of Sox2 recruitment by knockdown of *Tfap2c*. ESCs stably expressing miR-155-based shRNA against *Tfap2c* were induced to differentiate into TS-like cells for 2 days by Tet-induced repression of *Oct3/4*.

(F) Coimmunoprecipitation of Tfap2c with Sox2 in eTSCs. The cell lysates were prepared from EGFP-$TS_{3.5}$ treated with or without formaldehyde (FA) for 10 min and immunoprecipitated with anti-Sox2 (left) or anti-Tfap2c (right) antibodies. The whole-cell lysate (input) and immunoprecipitates (IP) were analyzed by immunoblotting (IB).

(G) Immunostaining of the *Sox2*/*Tfap2c* transgenic TSCs cultured in the absence of FGF4 for more than five passages. Scale bar, 200 μm.

(H) Maintained expression of TSC markers in *Sox2*/*Tfap2c*-transfected TSCs grown in the absence of FGF4. The parental EGFP-$TS_{3.5}$ line was cultured in the presence or absence of FGF4 for 3 days.

(D–E and H) Error bars indicate SD of three replicates.

See also Figure S5.

also result in lethality (Gotoh et al., 2005; Hadari et al., 2001; Saba-El-Leil et al., 2003). The function of Esrrb is also essential but auxiliary, since, unlike Sox2, Esrrb alone did not affect self-renewal of TSCs in the absence of FGF4 (Figure S2C). It should be noted that, as exemplified by the FGF-ERK dependency, extraembryonic ectoderm in postimplantation embryos is phenotypically and molecularly distinct from the trophectoderm in preimplantation embryos, in which Tead4 and Cdx2 play central roles. Thus, Sox2 is deployed for FGF-dependent self-renewal of TSCs after implantation (Figures 1E and 1G).

In two types of stem cells, Sox2 can connect different extrinsic signals, LIF-STAT and FGF-ERK, to different sets of transcription factors, enabling cell-type-specific responses to environmental stimuli (Figure 6I). We found that one of the mechanisms for such context-dependent functions of Sox2, with intrinsic low-affinity DNA-binding activity (Kamachi et al., 2000), is attributable to the cell-type-specific transcription factors that assist in its recruitment or binding to the target sites: Oct3/4 in ESCs and Tfap2c in TSCs. Mechanistically, Sox2 physically interacts with Tfap2c in TSCs, which may facilitate the efficient binding of Sox2 to the targets co-occupied with Tfap2c. The cooperative role of Sox2 and Tfap2c in vivo is supported by the evidence that *Tfap2c*-knockout embryos die between E7.5 and E8.5 due to a defect in trophoblast-derived tissues (Auman et al., 2002; Werling and Schorle, 2002), as is the case in *Sox2*-null embryos rescued by wild-type ESCs. Interestingly, such cooperative action of Sox2 and Tfap2c in TSCs was found on *Cdx2*, which is the most important regulator of trophoblast development. It should be noted that although Sox2 activates *Cdx2* in TSCs, it is expressed in a highly reciprocal manner to Cdx2 from E3.0 to E5.5, in which Sox2 begins to be expressed in trophoblast lineages as well (Figures 1E and 1G; data not shown). Although epigenetic constraints imposed on the chromatin might be involved in different accessibility of target genes in different

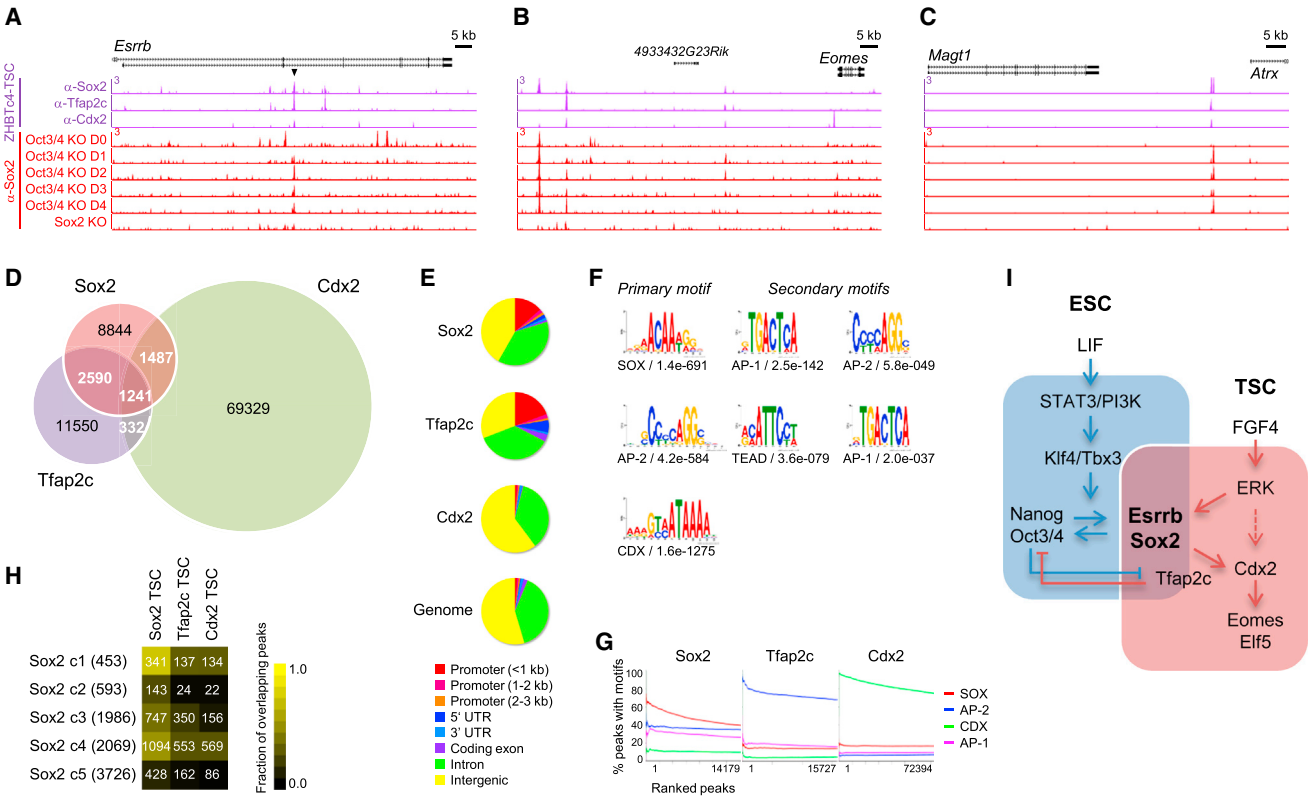



**Figure 6. Global Cooperativity of Sox2 and Tfap2c in TSCs**

(A–C) Normalized Sox2, Tfap2c, and Cdx2 ChIP-seq signals at the *Esrrb* (A), *Eomes* (B), and *Atrx/Magt1* (C) loci in ZHBTc4-TSCs. The *Esrrb* region analyzed by ChIP-PCR in Figure 5 is shown (an arrowhead).

(D) A Venn diagram showing the number of overlapping peaks in each data set. Peak regions were defined as ± 100 bp regions around peak summits.

(E) Annotation of ChIP-seq peaks with genomic features using CEAS.

(F) Motifs found by MEME to be overrepresented in the top 1,000 ranked peaks (summits ± 100 bp).

(G) Occurrence of the motifs in the ranked peaks. The cumulative percentage of peaks (summits ± 200 bp) containing at least one sequence for a given motif are plotted against ranks of peaks.

(H) A heatmap representing the fractions of overlapping peaks of Sox2 ChIP-seq clusters found in Figure 4D with Sox2, Tcfap2c, and Cdx2 peaks in TSCs. The numbers of overlapping peaks are shown in the grid.

(I) Self-renewal networks regulated by transcription factors and extrinsic signals in ESCs (Niwa et al., 2009) and TSCs. Context-dependent role of Sox2 is in part mediated by the lineage-specific and mutually antagonistic transcription factors Oct3/4 and Tfap2c.

cellular contexts, its contribution would be minimal, since the epigenetic state of ESCs is considered as a tabula rasa (Murakami et al., 2011; Niwa, 2007); this was confirmed by the dynamic change of the Sox2 binding sites immediately after elimination of *Oct3/4*.

However, even in a different mode of action with different partners, Sox2 plays a conceptually common role in ESCs and TSCs: the maintenance of self-renewal. This role is achieved by the activation of genes involved in maintaining the undifferentiated states and, albeit less commonly, restricts the expression of genes promoting differentiation. Furthermore, although a large fraction of the Sox2-bound genes whose functions are related to later developmental processes were not significantly affected by *Sox2* knockout, Sox2 may prime these genes for subsequent activation (Wegner, 2011). A flexible mode of regulation by Sox2 fits well with the idea that Sox2 is one of the putative stemness genes that is predominantly expressed and serves as a molecular hub to respond to diverse environmental cues for stem cell maintenance in different types of stem cells. This is in sharp contrast to the case of Esrrb, as shown here, and Myc, which regulates a common set of genes involved in cell proliferation and metabolism in wide variety of stem or progenitor cells (Kim et al., 2010). The versatile functions of Sox2 provide a clear example of how tissue-specific transcription factors have maximum flexibility to play divergent roles in different networks. Considering that the sequence specificities of individual transcription factors are generally limited, and only a few of the potential recognition sites are bound in vivo, the context-dependent redeployment of these transcription factors to drive tissue-specific transcriptional programs, as shown here, may be prevailing. Such mechanisms should contribute to maximizing the complexity and diversity of cellular functions with a minimal set of transcription factors. Reorganization of transcription factor networks might occur extensively during

embryogenesis, in which networks transit from one state to another along the developmental trajectory. This strategy also should be beneficial in evolution, since it allows establishment of novel networks, such as those for placental formation, only from existing transcription factors in the genome. This idea is supported by the evidence that very few transcription factors are expressed only in the sole cell type, and their combinatorial usage is quite flexible. We expect that the concept found in this study will emphasize the significance of the entire transcription factor network for the understanding of refined strategies in evolution and development.

## EXPERIMENTAL PROCEDURES

### Cell Culture and Cell Lines

ESCs were maintained in GMEM supplemented with 10% FBS, 1 × nonessential amino acids, 1 mM sodium pyruvate, 0.1 mM 2-mercaptoethanol, and 1,000 U/ml LIF on gelatin-coated dishes. TSCs were induced and maintained in GMEM/10% FBS supplemented with 50 ng/ml human FGF4, 2 μg/ml heparin, 10 ng/ml human Activin A, and 2 mM L-glutamine on mitomycin C treated mouse embryonic fibroblasts (MEFs). Tet-inducible *Oct3/4*-knockout ESC line ZHBTc4 and its derivative (4S2IP14) carrying an *IRES-pac* cassette knocked in downstream of the *Sox2* coding region for selection of *Sox2*-positive cells were maintained as described previously (Niwa et al., 2000, 2005). Tet-inducible *Sox2*-knockout ESC line 2TS22C and its TSC derivative (22CEROH2) carrying a 4-hydroxytamoxifen (4-OHT)-inducible *Cdx2-ER*$^{TM}$ transgene and a hygromycin-thymidine kinase fusion gene targeted into the *Pou5f1* locus were maintained as described previously (Masui et al., 2007; Niwa et al., 2005). ESC-derived TSC lines were established from ZHBTc4 in TSC medium supplemented with Tet on MEF feeders followed by subcloning, or from 22CEROH2 treated with 4-OHT followed by culture without 4-OHT and with GANC to eliminate undifferentiated ESCs. Inducers and inhibitors were used at the following concentrations: Tet, 1 μg/ml; doxycycline (Dox), 1 μg/ml; 4-OHT, 200 ng/ml; GANC, 1 μM; PD173074, 100 nM; PD0325901, 1 μM; LY294002, 20 μM.

### ChIP

ChIP assays were performed using the ChIP-IT Express Enzymatic Kit (Active Motif) following the manufacturer's instructions. Briefly, cells were crosslinked with 1% formaldehyde for 10 min at room temperature, and chromatin was enzymatically sheared into fragments with an average size of 300–500 bp. The chromatin fragments were immunoprecipitated with the antibodies against Sox2 (raised by us), Esrrb (H6705, Perseus Proteomics), Tfap2c (sc-8977, Santa Cruz Biotechnology), and Cdx2 (CDX2-88, BioGenex). Precipitated DNA was analyzed by real-time PCR. For each experiment, the percentage of input was determined and normalized to the value obtained at the 28S gene and intergenic spacer of ribosomal DNA, which gave reliable amplification due to their multiple copies in a genome. All primers are listed in Table S6.

### Mice and Embryos

All animal experiments were performed according to the guidelines for animal experiments of RIKEN Center for Developmental Biology and approved by the Animal Experiment Committee of the RIKEN Kobe Institute.

## ACCESSION NUMBERS

Microarray and ChIP-seq data have been deposited at the Gene Expression Omnibus (GEO) database under accession number GSE28455.

## SUPPLEMENTAL INFORMATION

Supplemental Information for this article includes six figures, six tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2013.09.002.

## AUTHOR CONTRIBUTIONS

K.A. and H.N. designed the study; K.A. performed experiments and analyzed the data; K.A., I.N., and H.R.U. performed bioinformatic analysis; H.O. and T.W. performed blastocyst injection; S.O. performed coimmunoprecipitation experiments; K.A., H.U., and M.K. performed ChIP-seq experiments; and K.A. and H.N. interpreted the data and wrote the paper.

## ACKNOWLEDGMENTS

We thank A. Smith and G. Guo (University of Cambridge, Cambridge) for *piggyBac* transposon vectors, S. Tanaka (University of Tokyo, Tokyo) and S. Kuraku (Genome Resource and Analysis Unit, RIKEN, Kobe) for the EGFP-$TS_{3.5}$ line, the RIKEN Omics Science Center (Yokohama) for high-throughput sequencing analysis, K. Uno (Functional Genomics Unit, CDB, RIKEN, Kobe) for microarray analysis, H. Schöler (Max Planck Institute for Molecular Biomedicine, Münster) for generous support, and Y. Nakai-Futatsugi and D. Sipp (CDB, RIKEN, Kobe) for their critical reading of the manuscript. This work was supported by a RIKEN research grant and by the Cell Innovation Program from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to H.N. and H.R.U. It was also supported by the CREST Program, from the Japan Science and Technology Agency (JST) to H.N.

Received: March 1, 2013
Revised: July 8, 2013
Accepted: August 29, 2013
Published: October 10, 2013

## REFERENCES

Arnold, K., Sarkar, A., Yram, M.A., Polo, J.M., Bronson, R., Sengupta, S., Seandel, M., Geijsen, N., and Hochedlinger, K. (2011). Sox2(+) adult stem and progenitor cells are important for tissue regeneration and survival of mice. Cell Stem Cell *9*, 317–329.

Auman, H.J., Nottoli, T., Lakiza, O., Winger, Q., Donaldson, S., and Williams, T. (2002). Transcription factor AP-2gamma is essential in the extra-embryonic lineages for early postimplantation development. Development *129*, 2733–2747.

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

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*(Web Server issue), W202–W208.

Berg, D.K., Smith, C.S., Pearton, D.J., Wells, D.N., Broadhurst, R., Donnison, M., and Pfeffer, P.L. (2011). Trophectoderm lineage determination in cattle. Dev. Cell *20*, 244–255.

Chawengsaksophak, K., James, R., Hammond, V.E., Köntgen, F., and Beck, F. (1997). Homeosis and intestinal tumours in Cdx2 mutant mice. Nature *386*, 84–87.

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

Donnison, M., Beaton, A., Davey, H.W., Broadhurst, R., L'Huillier, P., and Pfeffer, P.L. (2005). Loss of the extraembryonic ectoderm in Elf5 mutants leads to defects in embryonic patterning. Development *132*, 2299–2308.

Favaro, R., Valotta, M., Ferri, A.L., Latorre, E., Mariani, J., Giachino, C., Lancini, C., Tosetti, V., Ottolenghi, S., Taylor, V., and Nicolis, S.K. (2009). Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. Nat. Neurosci. *12*, 1248–1256.

Gotoh, N., Manova, K., Tanaka, S., Murohashi, M., Hadari, Y., Lee, A., Hamada, Y., Hiroe, T., Ito, M., Kurihara, T., et al. (2005). The docking protein FRS2alpha is an essential component of multiple fibroblast growth factor responses during early mouse development. Mol. Cell. Biol. *25*, 4105–4116.

Graf, T., and Enver, T. (2009). Forcing cells to change lineages. Nature *462*, 587–594.

Hadari, Y.R., Gotoh, N., Kouhara, H., Lax, I., and Schlessinger, J. (2001). Critical role for the docking-protein FRS2 alpha in FGF receptor-mediated signal transduction pathways. Proc. Natl. Acad. Sci. USA *98*, 8578–8583.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Inoue, M., Kamachi, Y., Matsunami, H., Imada, K., Uchikawa, M., and Kondoh, H. (2007). PAX6 and SOX2-dependent regulation of the Sox2 enhancer N-3 involved in embryonic visual system development. Genes Cells *12*, 1049–1061.

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y., and Lemischka, I.R. (2006). Dissecting self-renewal in stem cells with RNA interference. Nature *442*, 533–538.

Kamachi, Y., Uchikawa, M., and Kondoh, H. (2000). Pairing SOX off: with partners in the regulation of embryonic development. Trends Genet. *16*, 182–187.

Kim, J., Woo, A.J., Chu, J., Snow, J.W., Fujiwara, Y., Kim, C.G., Cantor, A.B., and Orkin, S.H. (2010). A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. Cell *143*, 313–324.

Kuckenberg, P., Buhl, S., Woynecki, T., van Fürden, B., Tolkunova, E., Seiffe, F., Moser, M., Tomilin, A., Winterhager, E., and Schorle, H. (2010). The transcription factor TCFAP2C/AP-2gamma cooperates with CDX2 to maintain trophectoderm formation. Mol. Cell. Biol. *30*, 3310–3320.

Kunath, T., Strumpf, D., and Rossant, J. (2004). Early trophoblast determination and stem cell maintenance in the mouse–a review. Placenta *25*(*Suppl A*), S32–S38.

Luo, J., Sladek, R., Bader, J.A., Matthyssen, A., Rossant, J., and Giguère, V. (1997). Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-beta. Nature *388*, 778–782.

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

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Murakami, K., Araki, K., Ohtsuka, S., Wakayama, T., and Niwa, H. (2011). Choice of random rather than imprinted X inactivation in female embryonic stem cell-derived extra-embryonic cells. Development *138*, 197–202.

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

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

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

Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell *123*, 917–929.

Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature *460*, 118–122.

Pevny, L.H., and Nicolis, S.K. (2010). Sox2 roles in neural stem cells. Int. J. Biochem. Cell Biol. *42*, 421–424.

Que, J., Luo, X., Schwartz, R.J., and Hogan, B.L. (2009). Multiple roles for Sox2 in the developing and adult mouse trachea. Development *136*, 1899–1907.

Ralston, A., Cox, B.J., Nishioka, N., Sasaki, H., Chea, E., Rugg-Gunn, P., Guo, G., Robson, P., Draper, J.S., and Rossant, J. (2010). Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. Development *137*, 395–403.

Russ, A.P., Wattler, S., Colledge, W.H., Aparicio, S.A., Carlton, M.B., Pearce, J.J., Barton, S.C., Surani, M.A., Ryan, K., Nehls, M.C., et al. (2000). Eomesodermin is required for mouse trophoblast development and mesoderm formation. Nature *404*, 95–99.

Saba-El-Leil, M.K., Vella, F.D., Vernay, B., Voisin, L., Chen, L., Labrecque, N., Ang, S.L., and Meloche, S. (2003). An essential function of the mitogen-activated protein kinase Erk2 in mouse trophoblast development. EMBO Rep. *4*, 964–968.

Schulz, H., Kolde, R., Adler, P., Aksoy, I., Anastassiadis, K., Bader, M., Billon, N., Boeuf, H., Bourillot, P.Y., Buchholz, F., et al.; Functional Genomics in Embryonic Stem Cells Consortium. (2009). The FunGenES database: a genomics resource for mouse embryonic stem cell differentiation. PLoS ONE *4*, e6804.

Shin, H., Liu, T., Manrai, A.K., and Liu, X.S. (2009). CEAS: cis-regulatory element annotation system. Bioinformatics *25*, 2605–2606.

Strumpf, D., Mao, C.A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F., and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development *132*, 2093–2102.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

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

Tanaka, S., Kunath, T., Hadjantonakis, A.K., Nagy, A., and Rossant, J. (1998). Promotion of trophoblast stem cell proliferation by FGF4. Science *282*, 2072–2075.

Tanaka, S., Kamachi, Y., Tanouchi, A., Hamada, H., Jing, N., and Kondoh, H. (2004). Interplay of SOX and POU factors in regulation of the Nestin gene in neural primordial cells. Mol. Cell. Biol. *24*, 8834–8846.

Taranova, O.V., Magness, S.T., Fagan, B.M., Wu, Y., Surzenko, N., Hutton, S.R., and Pevny, L.H. (2006). SOX2 is a dose-dependent regulator of retinal neural progenitor competence. Genes Dev. *20*, 1187–1202.

Ullah, Z., Kohn, M.J., Yagi, R., Vassilev, L.T., and DePamphilis, M.L. (2008). Differentiation of trophoblast stem cells into giant cells is triggered by p57/Kip2 inhibition of CDK1 activity. Genes Dev. *22*, 3024–3036.

van den Berg, D.L., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2010). An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell *6*, 369–381.

Wegner, M. (2011). SOX after SOX: SOXession regulates neurogenesis. Genes Dev. *25*, 2423–2428.

Werling, U., and Schorle, H. (2002). Transcription factor gene AP-2 gamma essential for early murine development. Mol. Cell. Biol. *22*, 3149–3156.

Wood, H.B., and Episkopou, V. (1999). Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. Mech. Dev. *86*, 197–201.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

Zhou, Q., and Melton, D.A. (2008). Extreme makeover: converting one cell into another. Cell Stem Cell *3*, 382–388.

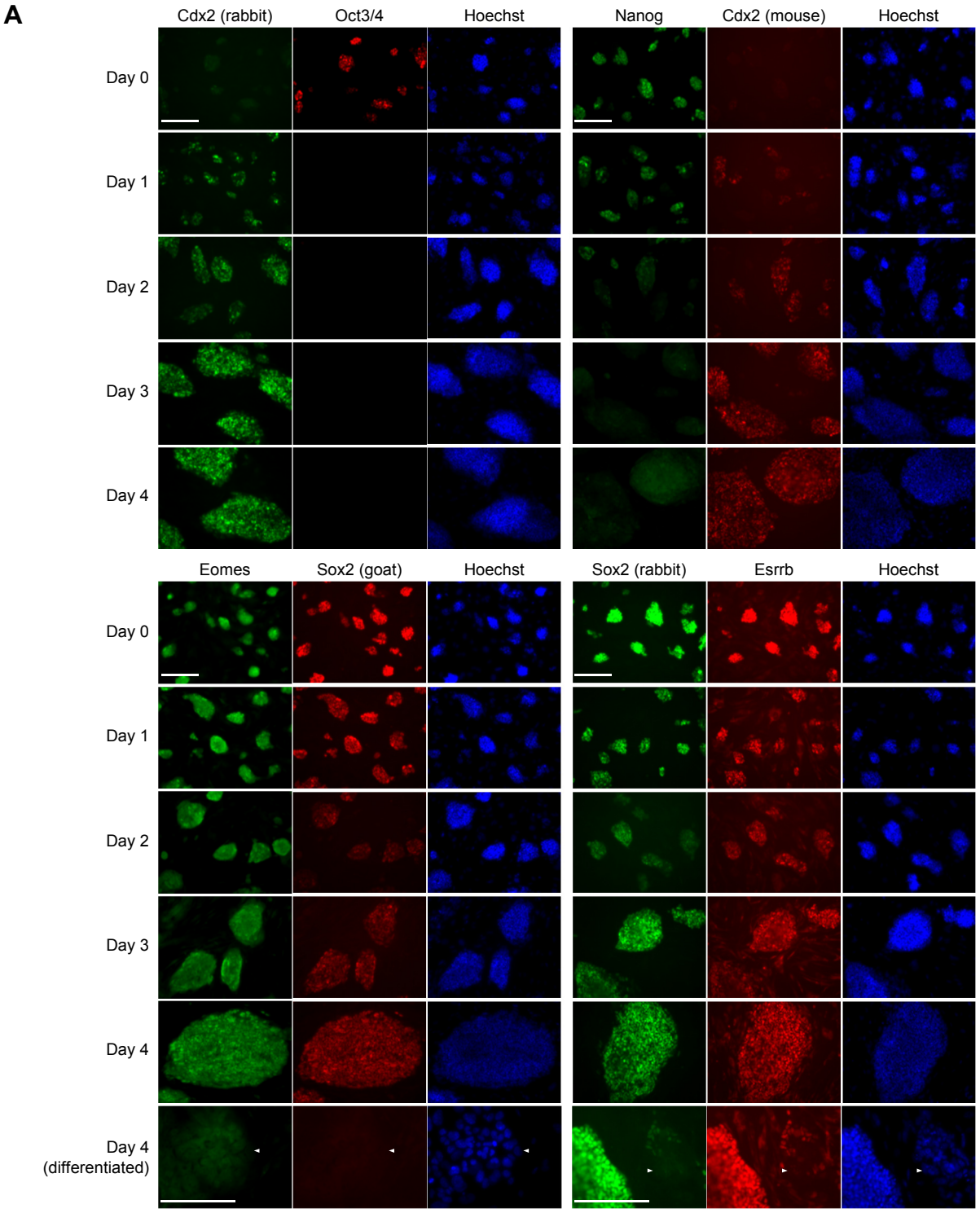
**Molecular Cell, Volume *52***

## Supplemental Information

# Context-Dependent Wiring of Sox2 Regulatory Networks for Self-Renewal of Embryonic and Trophoblast Stem Cells

**Kenjiro Adachi, Itoshi Nikaido, Hiroshi Ohta, Satoshi Ohtsuka, Hiroki Ura, Mitsutaka Kadota, Teruhiko Wakayama, Hiroki R. Ueda, and Hitoshi Niwa**

# Figure S1


A
Cdx2 (rabbit)
Oct3/4
Hoechst
Nanog
Cdx2 (mouse)
Hoechst
Day 0
Day 1
Day 2
Day 3
Day 4
Eomes
Sox2 (goat)
Hoechst
Sox2 (rabbit)
Esrrb
Hoechst
Day 0
Day 1
Day 2
Day 3
Day 4
Day 4
(differentiated)

# Figure S1

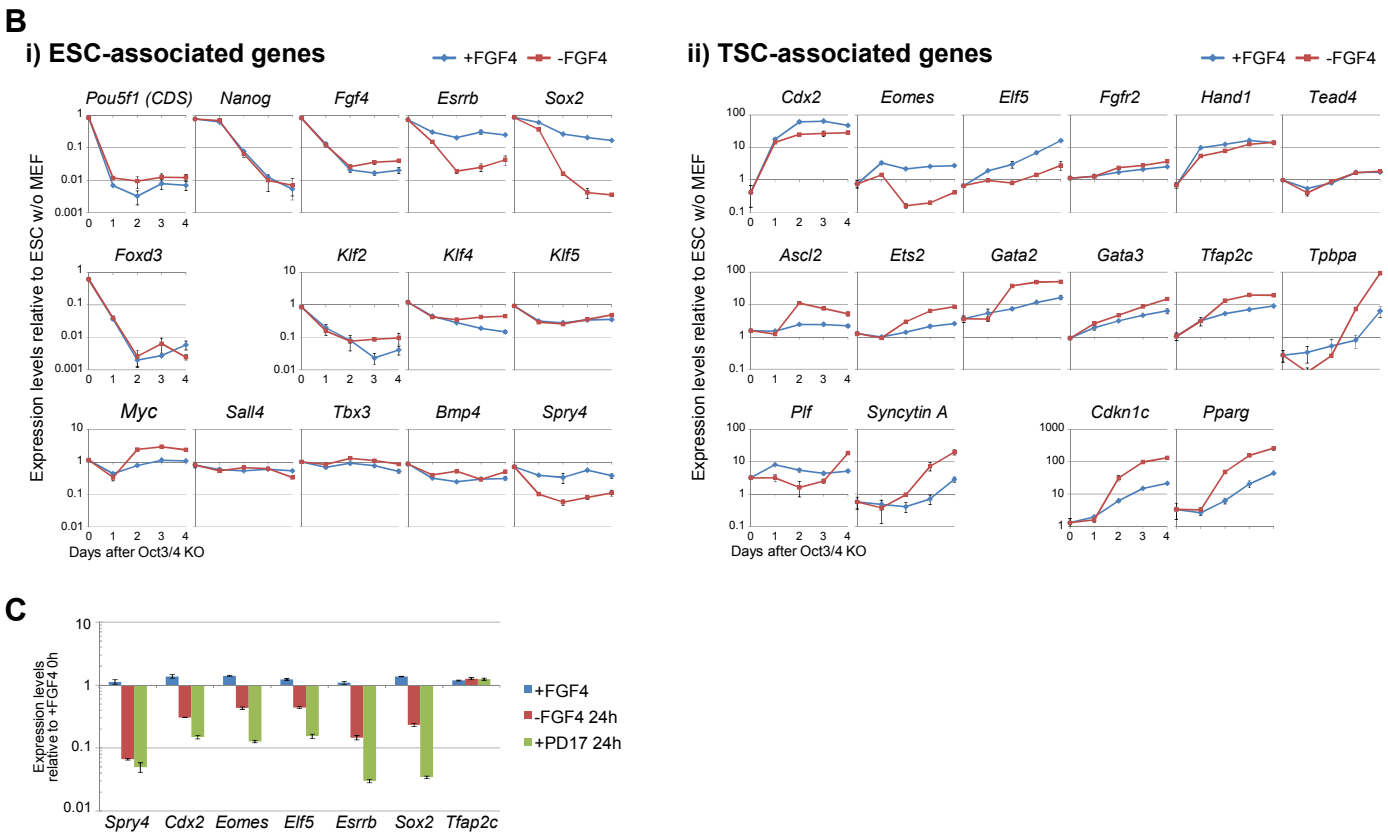



**Figure S1. Screening for FGF4-Responsive Transcription Factors Expressed in Trophoblasts, Related to Figure 1**

(A) Immunostaining for ESC and TSC markers after treatment of ZHBTc4 with Tet to repress Oct3/4 expression. The cells were cultured in the presence of FGF4 on MEF feeders. Oct3/4 expression from the Tet-off transgene was rapidly repressed upon Tet treatment, while Nanog expression was gradually decreased. Note that expression of Sox2 and Esrrb was maintained only in the undifferentiated TS-like cells, but not in the spontaneously differentiated trophoblast-like cells (arrowheads). Eomes is expressed, albeit at a low level, in ZHBTc4 ESCs and upregulated during the differentiation into TS-like cells (see also Figure S1B). Scale bars, 200 μm.

(B) FGF4-responsiveness of ESC- and TSC-associated genes during the differentiation of ZHBTc4 ESCs into TS-like cells. Kinetics of gene expression was analyzed by real-time PCR during the trophoblast differentiation in the presence or absence of FGF4. The expression of *Sox2* and *Esrrb* was rapidly downregulated without exogenous FGF4. The persistent activation of TSC markers, *Cdx2*, *Eomes* and *Elf5*, was also dependent on FGF4. In contrast, markers of differentiated trophoblast such as *Cdkn1c* (trophoblast giant cells), *Ascl2* (spongiotrophoblasts) and *syncytin A* (syncytiotrophoblasts) were reduced by the addition of FGF4. Only genes positively regulated by FGF4 are shown in Figure 1B.

(C) Downregulation of the TSC transcription factors by the inhibition of FGF signaling in the eTSC line EGFP-TS$_{3.5}$. The cells were deprived of FGF4 or treated with the FGF receptor inhibitor PD173074 (PD17) for 24 h before harvest.

(B and C) Error bars indicate standard deviation (SD) of three replicates.

# Figure S2

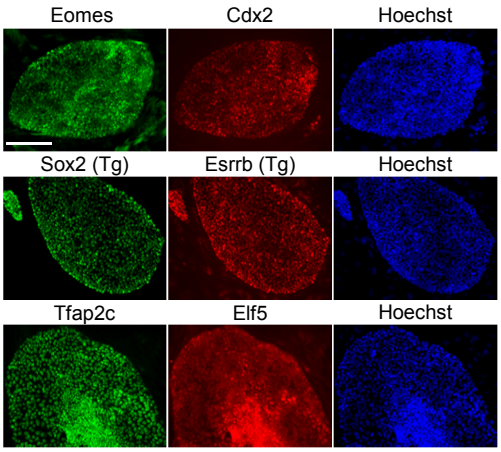
**A**

| Transgene | Without FGF4 |
|---|---|
| Cdx2 | ± |
| Eomes | − |
| Elf5 | − |
| Esrrb | − |
| Sox2 | − |
| Tfap2c | − |
| Cdx2+Esrrb | ± |
| Cdx2+Sox2 | ± |
| Esrrb+Sox2 | + |

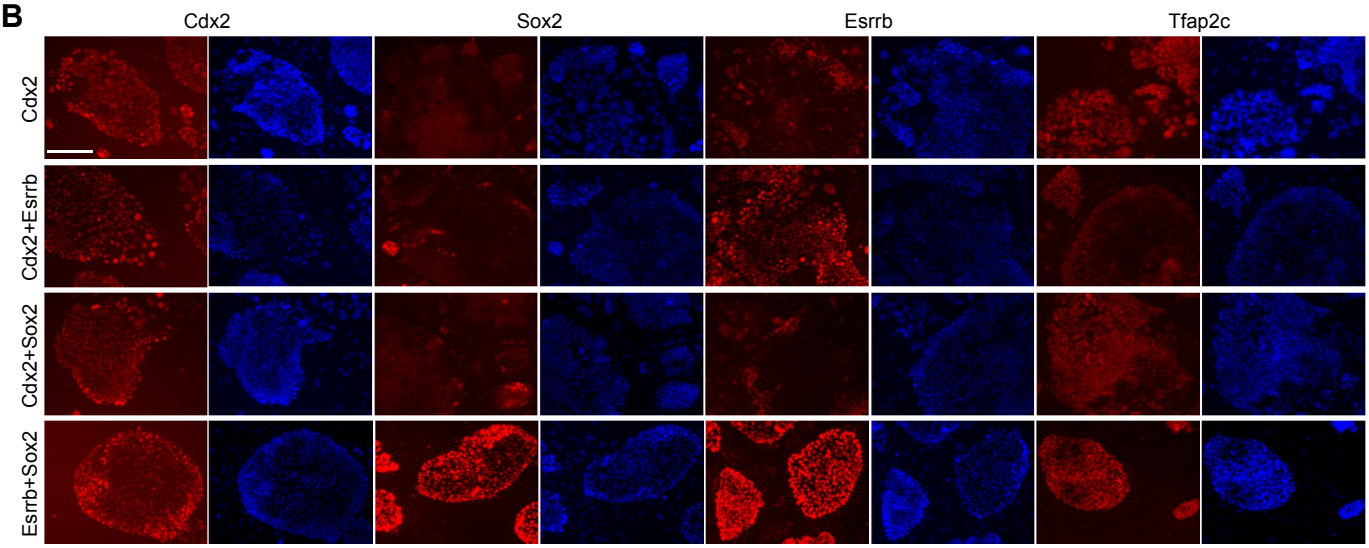
**C**

| Transgene | Without FGF4 |
|---|---|
| Cdx2-ER$^{TM}$ | − |
| Eomes-ER$^{TM}$ | − |
| Elf5-ER$^{TM}$ | − |
| Esrrb | − |
| Sox2 | ± |
| Sox2+Cdx2ER$^{TM}$ | ± |
| Sox2+Esrrb | + |

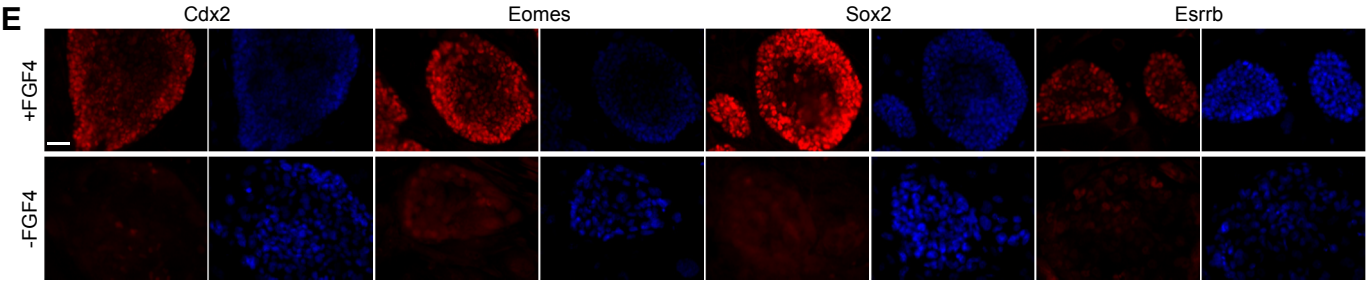
**D**




**B**




**E**




**F**

| Cell line | Recovered/Transferred | Chimeric (% recovered) | Highly contributed (% recovered) |
|---|---|---|---|
| EGFP-TS$_{3.5}$ | 22/50 | 6 (27.3) | 3 (13.6) |
| 628-1c1 | 28/50 | 7 (25.0) | 4 (14.3) |
| 628-1c2 | 32/53 | 6 (18.8) | 6 (18.8) |

# Figure S2

**Figure S2. FGF-Independent Self-Renewal of Transgenic TSCs, Related to Figure 2**

(A) Gain-of-function screen for transcription factors that support FGF-independent self-renewal of eTSCs. The constitutive expression vectors were stably transfected into embryo-derived EGFP-$TS_{3.5}$ line using *piggyBac* transposon system. After selection, the cells were cultured in the absence of FGF4. The self-renewal capacity was examined by serial passaging. Combined expression of *Sox2* and *Esrrb* enabled the self-renewal of TSCs. Although *Cdx2* also promoted proliferation, the transgenic cells tended to lose undifferentiated morphology and expression of other TSC transcription factors (see Figure S2B).

(B) Immunostaining of the transgenic TSCs cultured without FGF4 for more than five passages. Scale bar, 200 μm.

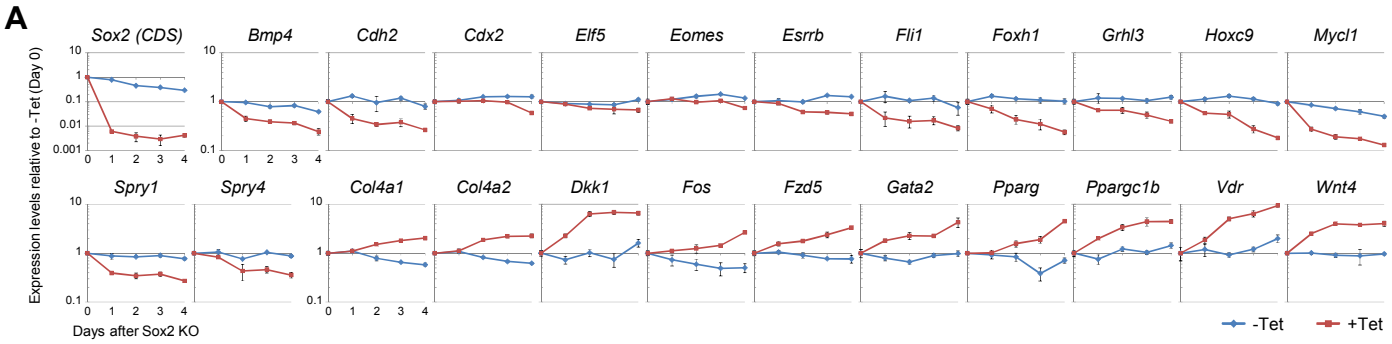
(C) Gain-of-function screen for transcription factors that support FGF-independent self-renewal of ZHBTc4-derived TS-like cells. The constitutive or 4-OHT-inducible expression vectors for these factors were stably transfected into ZHBTc4 ESCs by electroporation. The resulting clones were then induced to differentiate into trophoblast by the addition of Tet (and 4-OHT) in the absence of FGF4. Note that *Sox2* alone promoted short-term proliferation (less than two passages).

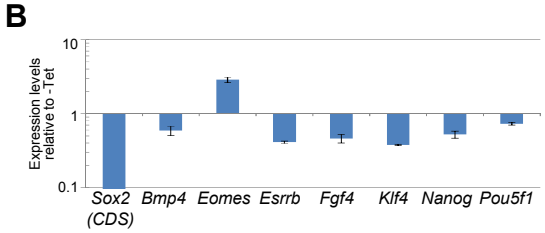
(D) Immunostaining of the ZHBTc4-derived transgenic TSCs grown in the absence of FGF4. Scale bar, 200 μm.

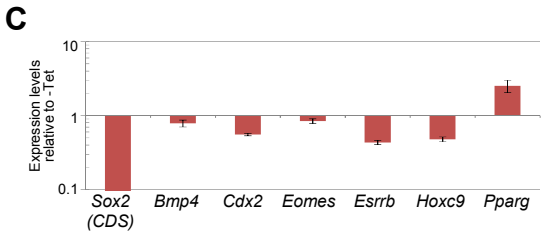
(E) Immunostaining of the transgene-deleted eTSCs (clone 628-1c2) cultured in the presence or absence of FGF4 for four days. Scale bar, 200 μm.

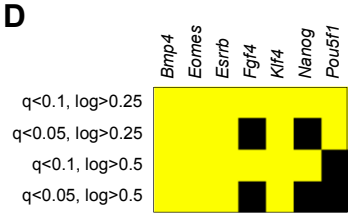
(F) Contribution of the transgene-deleted eTSCs to the chimeric placentas. The two transgene-excised TSC clones (628-1c1 and 628-1c2) and the parental EGFP-$TS_{3.5}$ line were injected into host blastocysts. The embryos and placentas were recovered at E13.5 and the contribution of TSCs was assessed by EGFP fluorescence.

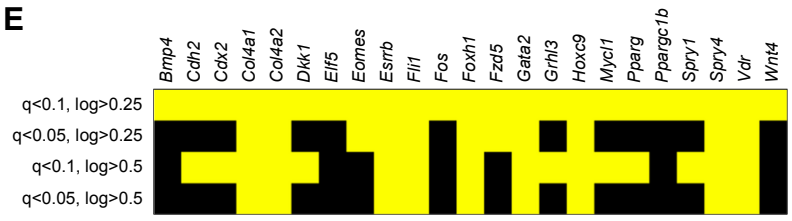
# Figure S3

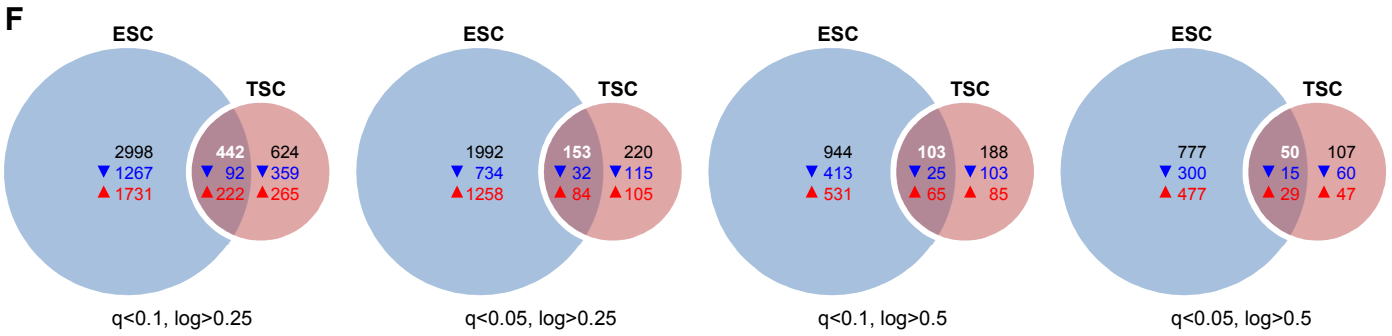

A
Sox2 (CDS)
Bmp4
Cdh2
Cdx2
Elf5
Eomes
Esrrb
Fli1
Foxh1
Grhl3
Hoxc9
Mycl1
Spry1
Spry4
Col4a1
Col4a2
Dkk1
Fos
Fzd5
Gata2
Pparg
Ppargc1b
Vdr
Wnt4
Expression levels relative to -Tet (Day 0)
Days after Sox2 KO
-Tet
+Tet

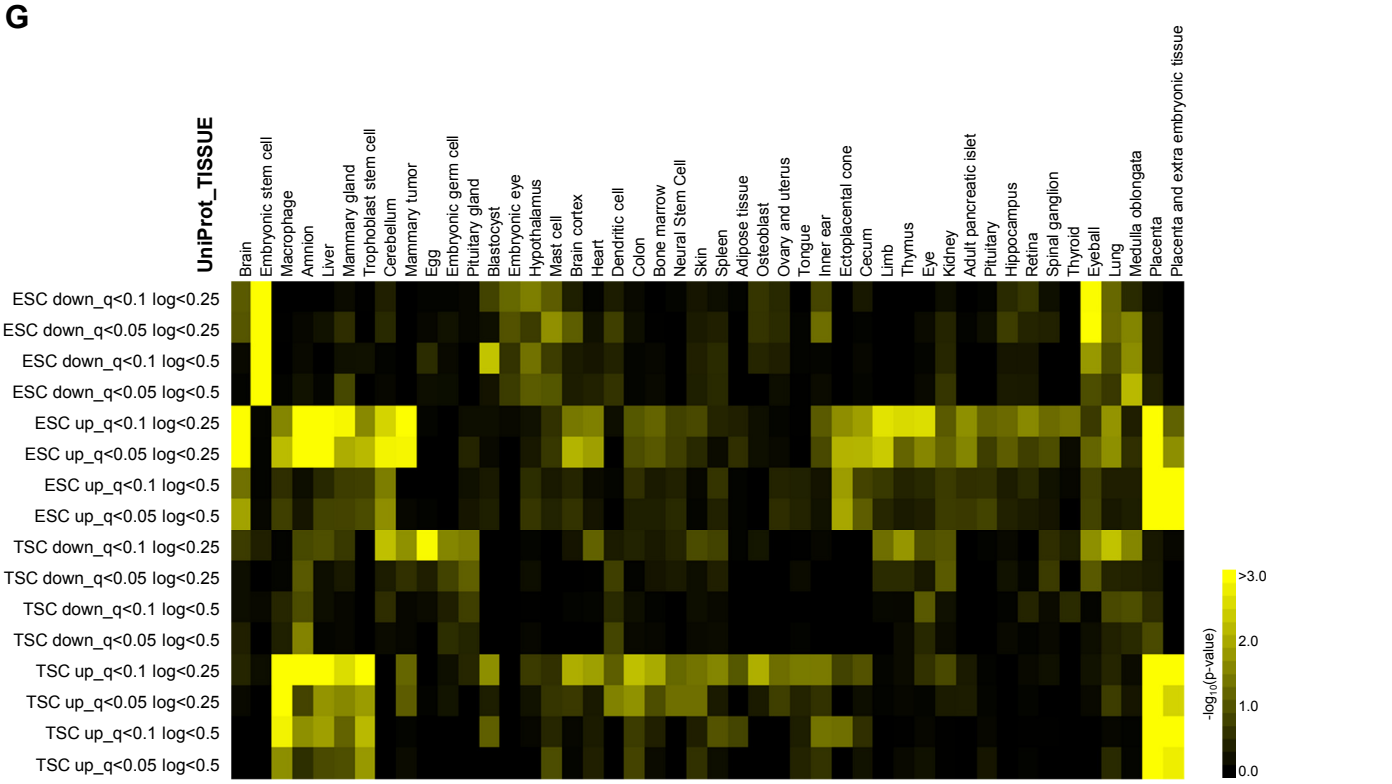

B
Expression levels relative to -Tet
Sox2 (CDS)
Bmp4
Eomes
Esrrb
Fgf4
Klf4
Nanog
Pou5f1


C
Expression levels relative to -Tet
Sox2 (CDS)
Bmp4
Cdx2
Eomes
Esrrb
Hoxc9
Pparg


D
Bmp4
Eomes
Esrrb
Fgf4
Klf4
Nanog
Pou5f1
q<0.1, log>0.25
q<0.05, log>0.25
q<0.1, log>0.5
q<0.05, log>0.5


E
Bmp4
Cdh2
Cdx2
Col4a1
Col4a2
Dkk1
Elf5
Eomes
Esrrb
Fli1
Fos
Foxh1
Fzd5
Gata2
Grhl3
Hoxc9
Mycl1
Pparg
Ppargc1b
Spry1
Spry4
Vdr
Wnt4
q<0.1, log>0.25
q<0.05, log>0.25
q<0.1, log>0.5
q<0.05, log>0.5


F
ESC
TSC
2998
▼1267
▲1731
442
▼92
▲222
624
▼359
▲265
q<0.1, log>0.25
1992
▼734
▲1258
153
▼32
▲84
220
▼115
▲105
q<0.05, log>0.25
944
▼413
▲531
103
▼25
▲65
188
▼103
▲85
q<0.1, log>0.5
777
▼300
▲477
50
▼15
▲29
107
▼60
▲47
q<0.05, log>0.5


G
UniProt_TISSUE
Brain
Embryonic stem cell
Macrophage
Amnion
Liver
Mammary gland
Trophoblast stem cell
Cerebellum
Mammary tumor
Egg
Embryonic germ cell
Pituitary gland
Blastocyst
Embryonic eye
Hypothalamus
Mast cell
Brain cortex
Heart
Dendritic cell
Colon
Bone marrow
Neural Stem Cell
Skin
Spleen
Adipose tissue
Osteoblast
Ovary and uterus
Tongue
Inner ear
Ectoplacental cone
Cecum
Limb
Thymus
Eye
Kidney
Adult pancreatic islet
Pituitary
Hippocampus
Retina
Spinal ganglion
Thyroid
Eyeball
Lung
Medulla oblongata
Placenta
Placenta and extra embryonic tissue
ESC down_q<0.1 log<0.25
ESC down_q<0.05 log<0.25
ESC down_q<0.1 log<0.5
ESC down_q<0.05 log<0.5
ESC up_q<0.1 log<0.25
ESC up_q<0.05 log<0.25
ESC up_q<0.1 log<0.5
ESC up_q<0.05 log<0.5
TSC down_q<0.1 log<0.25
TSC down_q<0.05 log<0.25
TSC down_q<0.1 log<0.5
TSC down_q<0.05 log<0.5
TSC up_q<0.1 log<0.25
TSC up_q<0.05 log<0.25
TSC up_q<0.1 log<0.5
TSC up_q<0.05 log<0.5
-log10(p-value)
>3.0
2.0
1.0
0.0

# Figure S3

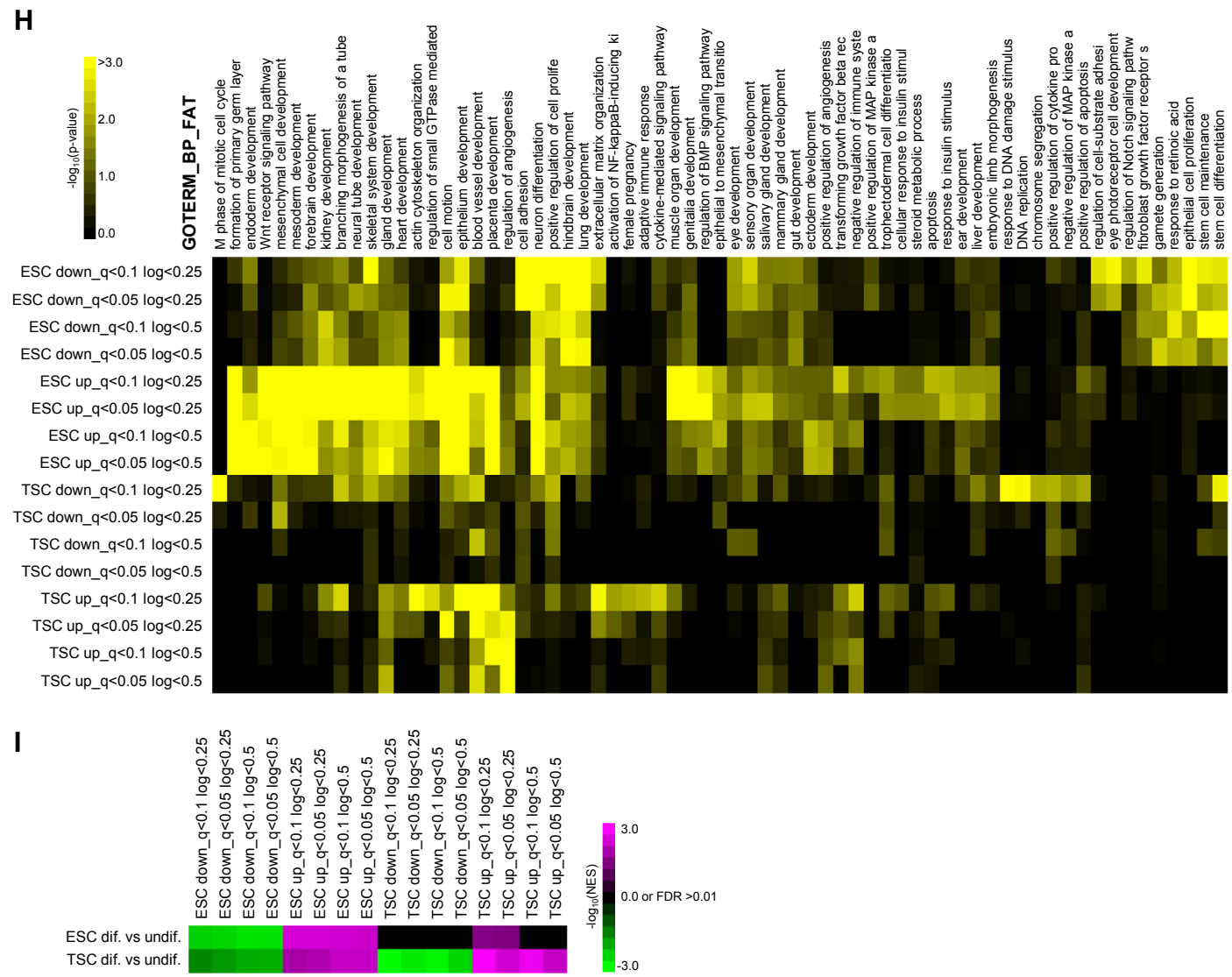



**Figure S3. Microarray Analysis of *Sox2* Knockout ESCs and TSCs, Related to Figure 3**

(A) Changes in gene expression after Tet-induced *Sox2* knockout in TSCs.

(B and C) Changes in gene expression four days after *Sox2* knockout in 2TS22C-ESCs (B) and -TSCs (C) analyzed by RT-PCR.

(D and E) Genes affected by *Sox2* knockout in ESCs (D) and TSCs (E) identified by microarray with different q-value and log2 ratio cutoffs (colored in yellow). Note that the sensitivity to identify genes found to be affected by RT-PCR was maximized with the less stringent cutoff.

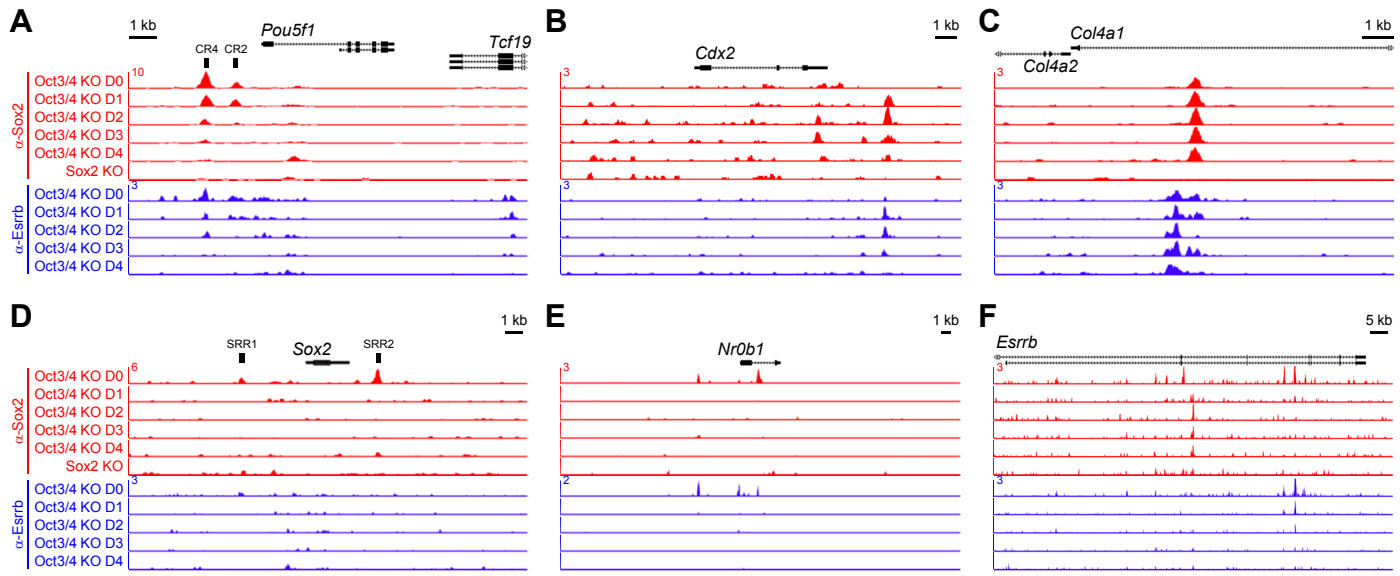
(F) The number of genes affected by *Sox2* knockout in ESCs and TSCs. Use of the relaxed and stringent cutoffs resulted in identification of similar proportions of affected genes in each group.

(G and H) Tissue Expression (G) and GO (H) enrichment analysis of genes affected by *Sox2* knockout. The $-\log_{10}$(p-value) for UniProt report for expressed tissues (UniProt_TISSUE) and GO terms in the Biological Process category (GOTERM_BP_FAT) are shown in a heatmap. Note that use of the less stringent cutoffs showed better sensitivity with reasonable specificity.

(I) GSEA analysis of changes in expression of genes affected by *Sox2* knockout during the differentiation of ESCs and TSCs. The $-\log_{10}$(NES) for gene sets with a FDR <0.01 are shown in a heatmap. The core enrichment genes that contributed most to the enrichment results are listed in the Table S2.

(A-C) Error bars indicate standard deviation (SD) of three replicates.

# Figure S4

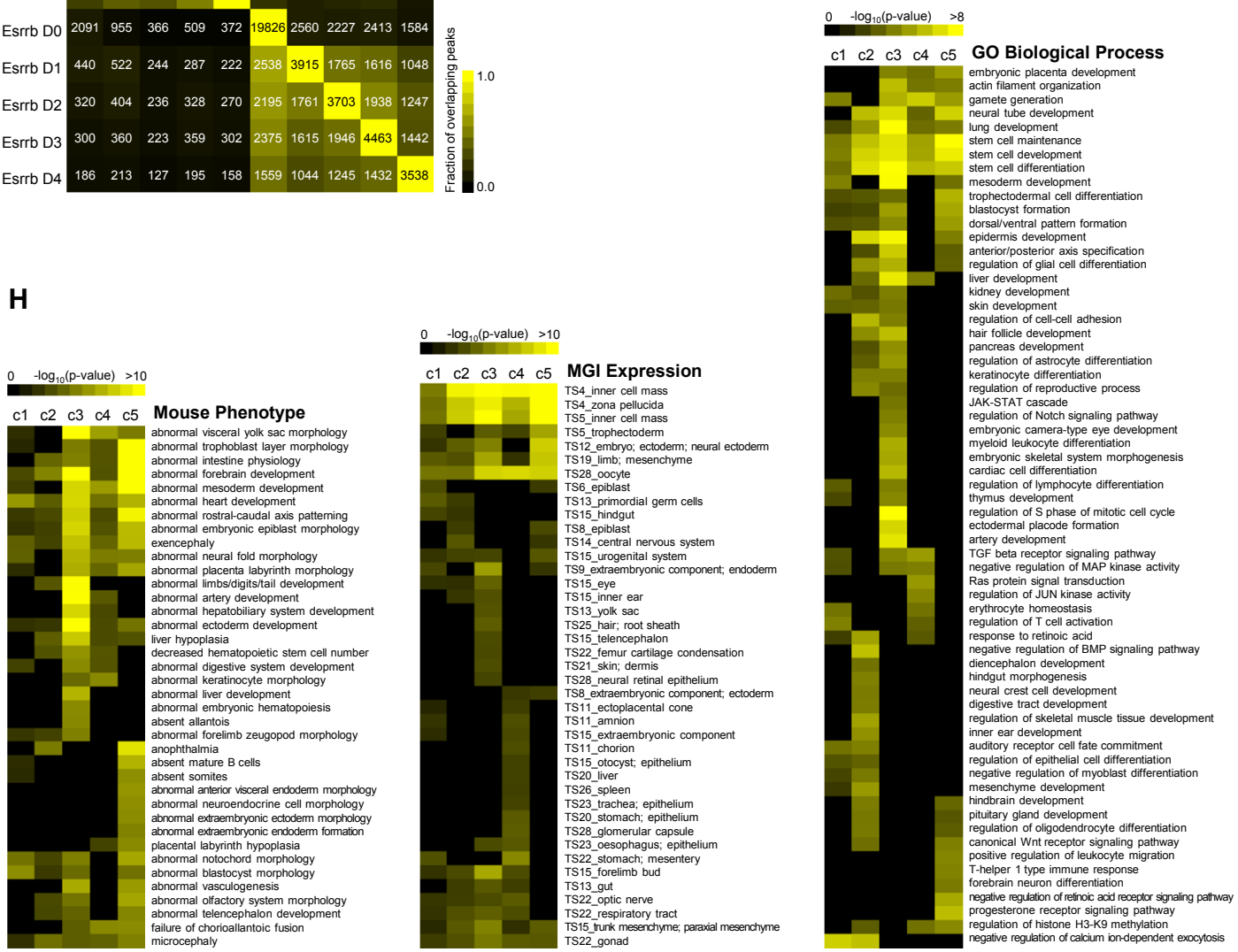



G

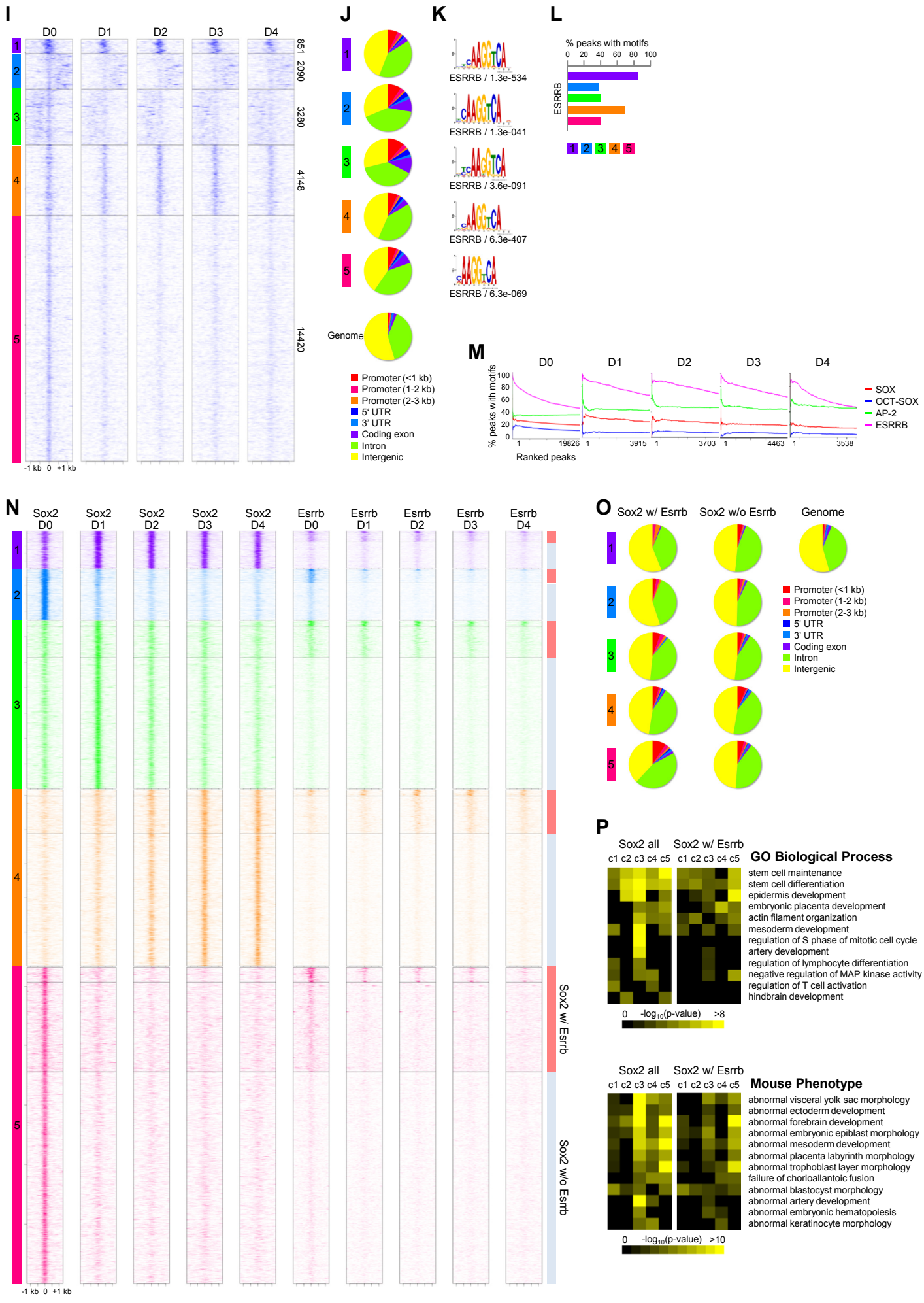
| | Sox2 D0 | Sox2 D1 | Sox2 D2 | Sox2 D3 | Sox2 D4 | Esrrb D0 | Esrrb D1 | Esrrb D2 | Esrrb D3 | Esrrb D4 |
|---|---|---|---|---|---|---|---|---|---|---|
| Sox2 D0 | 5299 | 1063 | 464 | 563 | 449 | 2045 | 439 | 319 | 297 | 185 |
| Sox2 D1 | 1063 | 2699 | 758 | 825 | 668 | 921 | 518 | 403 | 358 | 213 |
| Sox2 D2 | 465 | 760 | 1179 | 697 | 601 | 355 | 243 | 234 | 221 | 126 |
| Sox2 D3 | 562 | 825 | 697 | 2076 | 912 | 494 | 287 | 328 | 355 | 194 |
| Sox2 D4 | 448 | 669 | 600 | 912 | 1807 | 363 | 223 | 270 | 299 | 157 |
| Esrrb D0 | 2091 | 955 | 366 | 509 | 372 | 19826 | 2560 | 2227 | 2413 | 1584 |
| Esrrb D1 | 440 | 522 | 244 | 287 | 222 | 2538 | 3915 | 1765 | 1616 | 1048 |
| Esrrb D2 | 320 | 404 | 236 | 328 | 270 | 2195 | 1761 | 3703 | 1938 | 1247 |
| Esrrb D3 | 300 | 360 | 223 | 359 | 302 | 2375 | 1615 | 1946 | 4463 | 1442 |
| Esrrb D4 | 186 | 213 | 127 | 195 | 158 | 1559 | 1044 | 1245 | 1432 | 3538 |

Fraction of overlapping peaks 1.0 0.0

H



# Figure S4


I
D0
D1
D2
D3
D4
851
2090
3280
4148
14420
-1 kb 0 +1 kb
J
Genome
Promoter (<1 kb)
Promoter (1-2 kb)
Promoter (2-3 kb)
5' UTR
3' UTR
Coding exon
Intron
Intergenic
K
ESRRB / 1.3e-534
ESRRB / 1.3e-041
ESRRB / 3.6e-091
ESRRB / 6.3e-407
ESRRB / 6.3e-069
L
% peaks with motifs
0 20 40 60 80 100
ESRRB
M
% peaks with motifs
Ranked peaks
SOX
OCT-SOX
AP-2
ESRRB
N
Sox2 D0
Sox2 D1
Sox2 D2
Sox2 D3
Sox2 D4
Esrrb D0
Esrrb D1
Esrrb D2
Esrrb D3
Esrrb D4
Sox2 w/ Esrrb
Sox2 w/o Esrrb
O
Sox2 w/ Esrrb
Sox2 w/o Esrrb
Genome
P
Sox2 all
Sox2 w/ Esrrb
c1 c2 c3 c4 c5
GO Biological Process
stem cell maintenance
stem cell differentiation
epidermis development
embryonic placenta development
actin filament organization
mesoderm development
regulation of S phase of mitotic cell cycle
artery development
regulation of lymphocyte differentiation
negative regulation of MAP kinase activity
regulation of T cell activation
hindbrain development
Mouse Phenotype
abnormal visceral yolk sac morphology
abnormal ectoderm development
abnormal forebrain development
abnormal embryonic epiblast morphology
abnormal mesoderm development
abnormal placenta labyrinth morphology
abnormal trophoblast layer morphology
failure of chorioallantoic fusion
abnormal blastocyst morphology
abnormal artery development
abnormal embryonic hematopoiesis
abnormal keratinocyte morphology

# Figure S4

**Figure S4. ChIP-Seq Analysis of Sox2 and Esrrb Binding Sites in Differentiating TSCs, Related to Figure 4**

(A-F) Normalized Sox2 and Esrrb ChIP-seq signals at the *Pou5f1* (A), *Cdx2* (B), *Col4a1* (C), *Sox2* (D), *Nr0b1* (E), and *Esrrb* (F) loci during the differentiation of ZHBTc4 ESCs into TS-like cells. ChIP-seq signals were normalized as levels per million mapped reads.

(G) A heatmap representing the fraction of overlapping peaks in each data set. Peak calling was performed using MACS2 with a q-value cutoff of 0.05. Peak regions were defined as ±100-bp regions around peak summits. The fractions of peaks that overlap at least 1 bp with those of other samples are shown in a heatmap. The numbers of overlapping peaks are shown in the grid. The larger numbers of Sox2 and Esrrb peaks in ESCs may be in part due to their higher expression levels (Figure 1B and Figure S1A).

(H) GREAT ontology enrichments for the peaks of each Sox2 cluster. Only selected GO terms are shown in Figure 4J.

(I) A heatmap of the Esrrb ChIP-seq signals around the merged peak regions. ChIP-seq peaks of each sample were merged and normalized ChIP-seq signals over ±1-kb regions around the merged peak centers were clustered with the k-means algorithm into 5 clusters.

(J) Annotation of ChIP-seq clusters with genomic features. The merged peak centers of each cluster were annotated with genomic features using CEAS.

(K) Motifs found by MEME to be overrepresented in 600 randomly selected ±100-bp regions around the merged peak centers of each cluster (from c1 to c5, top to bottom).

(L) Occurrence of the ESRRB motif in peaks of each cluster. The motif was mapped to the ±200-bp regions around the merged peak centers of each cluster and the percentages of peaks containing at least one sequence for the motif are shown.

(M) Occurrence of the motifs in the ranked Esrrb peaks during differentiation of ESCs into TSCs. The peak summits of each sample were ranked by the $-\log_{10}$(q-value) for peak call. The cumulative percentage of peaks (summit ±200 bp) containing at least one sequence for a given motif are plotted against ranks of peaks. The SOX, OCT-SOX, and AP-2 motif are motifs discovered in the Sox2 ChIP-seq data (Figure 4F).

(N) A heatmap of the Sox2 and Esrrb ChIP-seq signals around the merged peak regions of Sox2. Each Sox2 peak cluster shown in Figure 4D was further split with the k-means algorithm into 3 subclusters only based on Esrrb ChIP-seq signals. The first and second subclusters of each major cluster showing strong and moderate Esrrb binding, respectively, were considered as co-occupied with Esrrb.

(O) Annotation of Sox2 subclusters with genomic features.

(P) GREAT ontology enrichments for the peaks of each Sox2 cluster co-occupied with Esrrb. For comparison, the enrichments for all the Sox2 peaks of each cluster (Figure 4J) are also shown.

# Figure S5

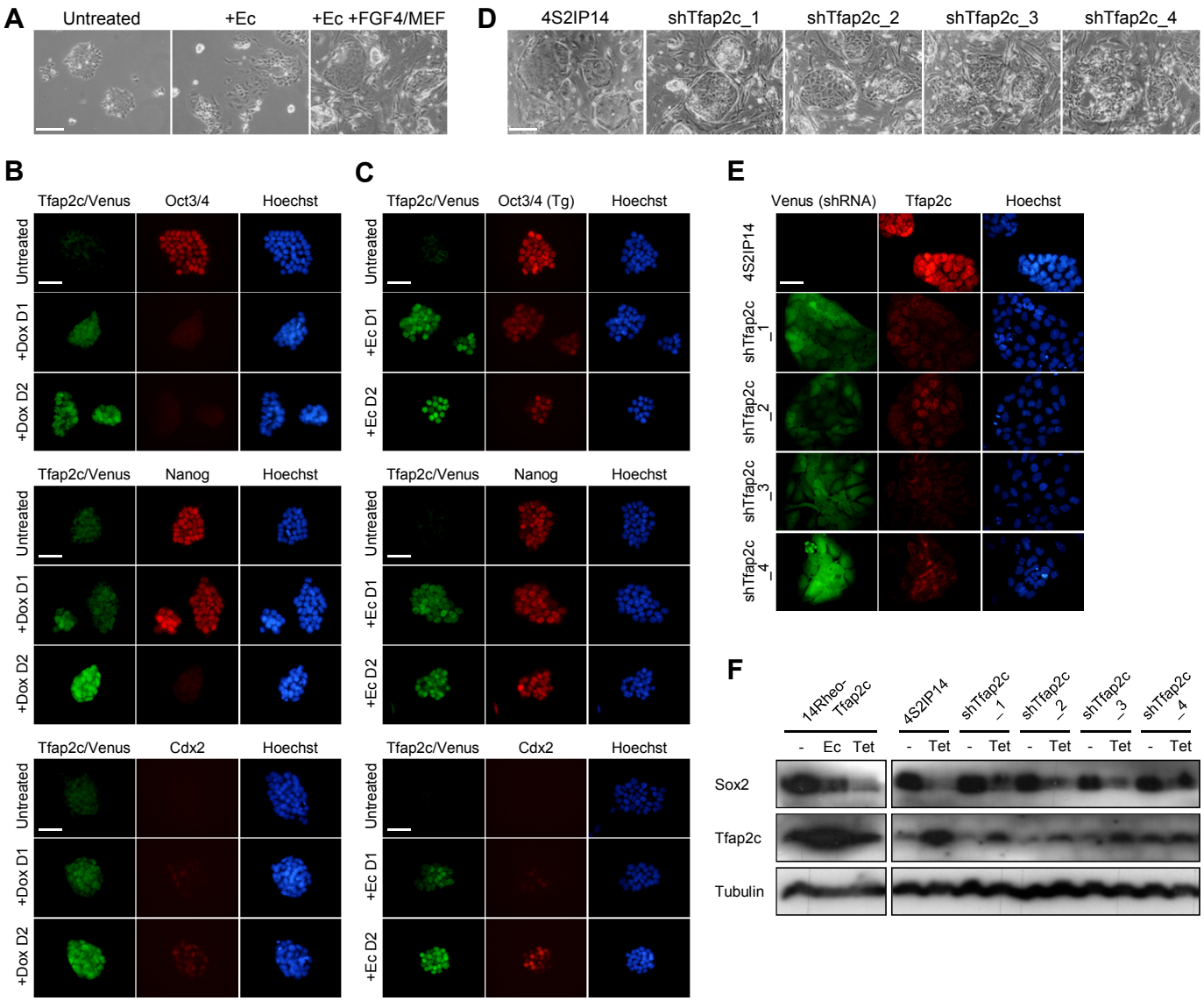



**Figure S5. Trophoblast Differentiation Induced by *Tfap2c* in ESCs, Related to Figure 5**

(A) Morphological alterations of ZHBTc4:*Sox2-IRES-pac* (4S2IP14)-derived 14RheoTfap2c ESCs four days after induction of *Tfap2c*. Expression of the transgene encoding *Tfap2c* (NM_009335) tagged with *IRES-Venus* was induced by the addition of an ecdysteroid agonist (Ec) in the absence or presence of FGF4/MEFs. Scale bar, 200 μm.

(B) Immunostaining for Oct3/4, Nanog and Cdx2 after induction of *Tfap2c* in EB5 ESCs using a Tet-on system. Preferential suppression of *Oct3/4* expression by *Tfap2c* was observed. Scale bars, 100 μm.

(C) Immunostaining for Oct3/4, Nanog and Cdx2 after induction of *Tfap2c* with Ec in 4S2IP14 ESCs, in which *Oct3/4* is constitutively expressed from the Tet-off transgene. Scale bars, 100 μm.

(D) Photomicrographs of 4S2IP14 ESCs stably expressing miR-155-based shRNA against *Tfap2c* three days after induction of trophoblast differentiation by *Oct3/4* knockout. Scale bar, 200 μm.

(E) Immunostaining of 4S2IP14 ESCs stably expressing miR-155-based shRNA against *Tfap2c* two days after induction of trophoblast differentiation by *Oct3/4* knockout. Expression of *Venus*, which is encoded by the same transcript as *miR-155-shRNA*, was detected using anti-GFP antibody (GF090R, Nacalai Tesque). Note that nuclear staining of Tfap2c was weakly detected in the shTfap2c_1- and _2- expressing cells, but not in the shTfap2c_3- and _4- expressing cells. Scale bar, 100 μm.

(F) Western blot analysis two days after *Tfap2c* induction (Ec) or *Oct3/4* knockout (Tet) in 14RheoTfap2c ESCs (left), and two days after *Oct3/4* knockout (Tet) in *Tfap2c* knockdown 4S2IP14 ESCs (right).

# Figure S6

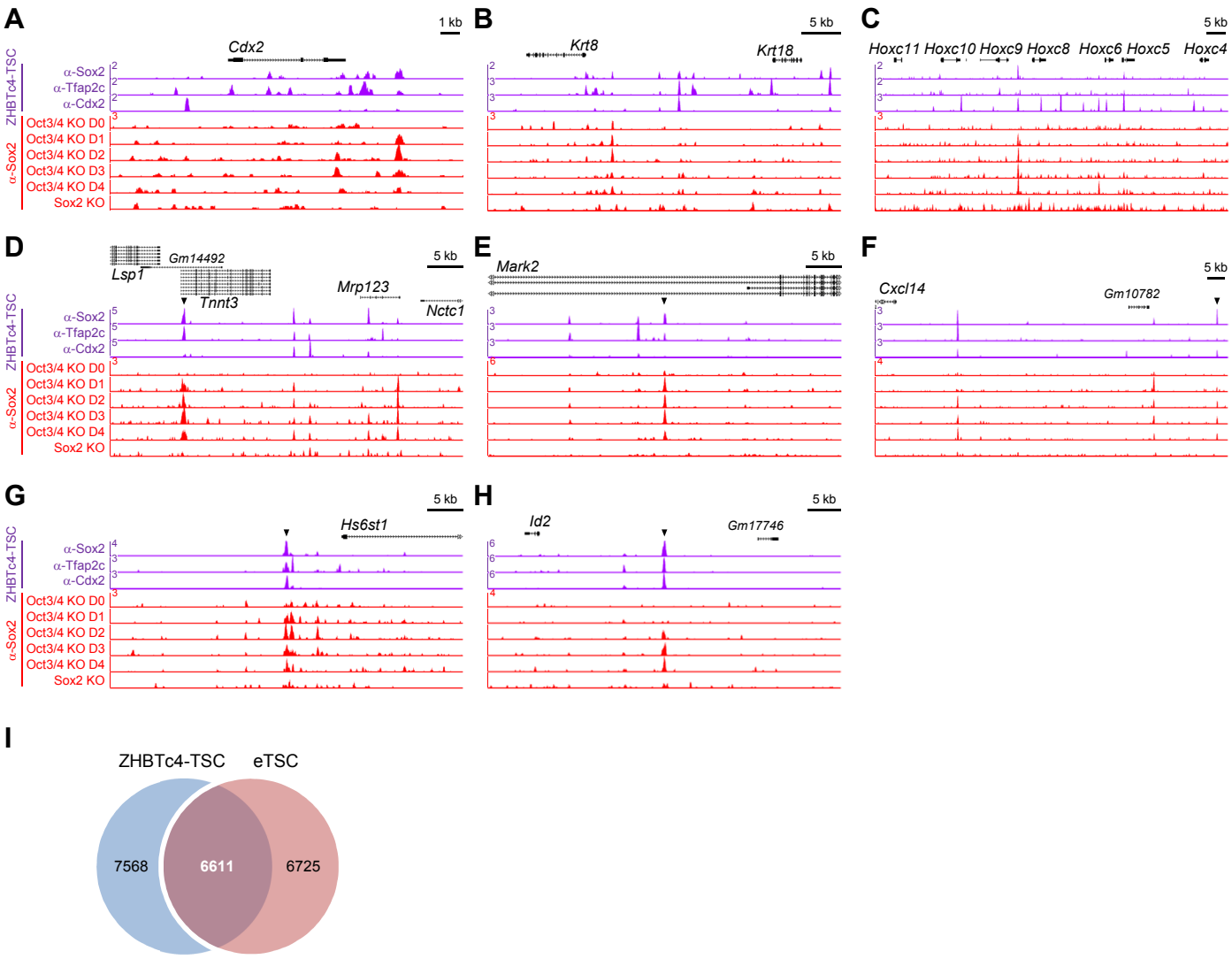



**Figure S6. ChIP-Seq Analysis of Sox2, Tfap2c, and Cdx2 Binding Sites in TSCs, Related to Figure 6**

(A-H) Normalized ChIP-seq signals at the *Cdx2* (A), *Krt8/Krt18* (B), *Hoxc11-Hoxc4* (C), *Tnnt3* (D), *Mark2* (E), *Cxcl14/Gm10782* (F), *Hs6st1* (G), and *Id2/Gm17746* (H) loci in ZHBTc4-TSCs and during the differentiation of ZHBTc4 ESCs into TS-like cells. The regions analyzed by ChIP-PCR in Figure 5 are shown (arrowheads).

(I) A venn diagram showing the number of overlapping peaks of Sox2 in ZHBTc4-TSCs and eTSCs. Peak regions were defined as ±100-bp regions around peak summits.

## SUPPLEMENTAL EXPERIMENTAL PROCEDURES

### Generation of Stable Cell Lines with Constitutive and Inducible Expression or Knockdown

To obtain stable transfectants, expression constructs driven by the CAG promoter (Niwa et al., 1991) were introduced into ESCs and TSCs by electroporation or using the *piggyBac* transposon system (Guo et al., 2009), followed by selection, subcloning, and screening for transgene expression. For inducible expression using a Tet-on system, the *piggyBac* transposon vectors for the expression of rtTA2$^{S}$-M2 and a Tet-responsive transgene followed by an *IRES-Venus* cassette were co-transfected and the selected clones were treated with 1 μg/ml Dox. For inducible expression using an EcR-based system, the *piggyBac* transposon vectors for the expression of GAL4-EcR, VP16-RXR (from the RheoSwitch Inducible Expression System, New England Biolabs) and a GAL4-responsive transgene followed by an *IRES-Venus* cassette were co-transfected and the selected clones were treated with a non-steroidal ecdysteroid agonist, diacylhydrazine (GenoStat Ligand, 500 nM, Millipore). For stable knockdown, miR-155-based shRNA expression vectors containing the stem-loop sequence derived from the *miR-155* precursor followed by an *IRES-Venus* cassette under the control of the CAG promoter were constructed. The following target sequences were selected using BLOCK-iT$^{TM}$ miR RNAi Select (Invitrogen): shTfap2c_1, 1223-1243; shTfap2c_2, 979-999; shTfap2c_3, 352-372; shTfap2c_4, 678-698 (nucleotide positions respective to the translation initiation site of NM_009335).

### RT-PCR Analysis

Total RNA was isolated using QuickGene RNA cultured cell HC kit S (Fuji Film). First-strand cDNA was synthesized with ReverTra Ace reverse transcriptase (TOYOBO). Real-time PCR was performed using SYBR Green Realtime PCR Master Mix (TOYOBO). Expression levels were normalized to *Gapdh* and presented as fold change over control samples. Each experiment was performed in technical triplicate. Error bars indicate SD of three replicates. Similar results were obtained in at least two independent experiments. All primers are listed in the Table S6.

### Immunofluorescence Analysis

Immunostaining was performed using the following primary antibodies with dilutions: mouse monoclonal anti-Cdx2 (CDX2-88, BioGenex), 1:1,000; rabbit monoclonal anti-Cdx2 (3977, Cell Signaling Technology), 1:200; rabbit anti-Eomes (AB9618, Millipore), 1:1000; mouse monoclonal anti-Esrrb (H6705, Perseus Proteomics), 1:1,000; rabbit anti-Sox2 [raised by ourselves (Masui et al., 2007)], 1:1,000; goat anti-Sox2 (GT15098, Neuromics), 1:1000; rabbit anti-Tfap2c (sc-8977, Santa Cruz), 1:250; goat anti-Elf5 (sc-9645, Santa Cruz), 1:100; rabbit monoclonal anti-phosphorylated ERK (4370, Cell Signaling Technology), 1:100; mouse monoclonal

anti-Oct3/4 (sc-5279, Santa Cruz), 1:1,000; rabbit anti-Nanog (A300-397A, Bethyl Laboratories), 1:1000; rat monoclonal anti-Nanog (14-5761, eBioscience), 1:250.

**Production of TSC Chimeras**

Two reverted TSC clones and parental EGFP-$TS_{3.5}$ line were injected into E3.5 ICR blastocysts (10-20 TSCs/blastocyst) as described previously (Oda et al., 2009). Injected blastocysts were transferred into the uteri of 2.5 days post-coitum (dpc) pseudopregnant ICR female mice and analyzed at 13.5 dpc. For histological analysis, placentas were fixed with 4% paraformaldehyde for overnight at 4°C, infiltrated with sucrose, embedded in OCT compound and frozen at -80°C. Cross sections were prepared and analyzed for EGFP expression by fluorescent microscopy. They were subsequently stained with hematoxylin-eosin and micrographs of the same field were taken.

**Microarray Analysis**

Total RNA was extracted from ZHBTc4 ESCs treated with Tet, ZHBTc4-TSCs, EGFP-$TS_{3.5}$ TSCs, or 2TS22C-ESCs and -TSCs treated with Tet cultured in ESC or TSC media on MEF feeders. For each dataset, the experiments were done in triplicate. cDNA synthesis and cRNA labeling reactions were performed using the Affymetrix GeneChip WT Terminal Labeling Kit and the Ambion WT Expression Kit according to the manufacturer's protocols. Affymetrix high-density oligonucleotide arrays for *Mus musculus* (GeneChip Mouse Exon 1.0 ST Array) were hybridized, stained, and washed according to the Expression Analysis Technical Manual (Affymetrix). The expression values were summarized and normalized by the robust multiarray analysis and quantile normalization method using by Expression Console software (Affymetrix). The microarray data are available at the GEO (GSE28455). For classification of the 11 cell types, PCA was performed using the prcomp function in R (http://www.r-project.org/). For identification of genes affected by Sox2 knockout, the less stringent cutoff (q-value <0.1, log2 ratio > 0.25) was applied to maximize sensitivity while maintaining reasonable specificity. For functional annotation of the gene sets, DAVID (Huang da et al., 2009) was used. Heatmaps showing the enrichment were generated using Cluster 3.0 and Java Treeview. GSEA (Subramanian et al., 2005) was performed to assess enrichment of the gene sets. FDR was estimated by gene set permutation tests. Previously published microarray data for CGR8 ESCs undergoing differentiation through embryoid body formation (Schulz et al., 2009) and $TS_{3.5}$ eTSCs differentiating into mature trophoblast cell types upon withdrawal of FGF4 and MEF-conditioned medium (Ralston et al., 2010) were obtained from the EBI ArrayExpress (E-TABM-672) and the GEO (GSE12985), respectively. Enrichment was determined by comparing the expression levels in the undifferentiated ESCs or TSCs with those in

the rest of the samples.

**ChIP-Seq and Data Analysis**

Chromatin was prepared from ZHBTc4 ESCs cultured in TSC medium supplemented with Tet for 0 to 4 days on MEFs, or ZHBTc4- and EGFP-$TS_{3.5}$ TSCs cultured in TSC medium on MEFs. Sequencing libraries were prepared according to the Illumina protocols and sequenced on the Genome Analyzer IIx or HiSeq 1500 (Illumina) as single-end 50-bp reads. Sequence reads were aligned to the mouse reference genome (mm9) using Bowtie2 (Langmead and Salzberg, 2012) with default parameters. Peak Calling was performed using MACS2 (Zhang et al., 2008) with input DNA as a control and with a q-value cutoff of 0.05. The signals were normalized by sequencing depth of million mapped reads with the --SPMR option. The peaks that overlap with the Sox2 peaks in *Sox2* knockout ESCs (2TS22C +Tet) were removed. To calculate the number of overlapping peaks, ±100-bp regions around peak summits were considered to avoid a bias toward longer peak regions. K-means clustering and visualization of ChIP-seq signals around the peak regions were performed using the heatmapr script from the Cistrome package (Liu et al., 2011). Annotation of peak regions with genomic features around the RefSeq genes were performed using CEAS (Shin et al., 2009). De novo motif discovery was performed using the MEME Suite (Bailey et al., 2009). The sequences of ±100-bp regions around the 600 randomly selected merged peak centers or around the top 1,000 ranked peak summits that were masked for repetitive elements using RepeatMasker were obtained from the UCSC Genome Browser. The overrepresented motifs identified by MEME were then searched against the JASPAR CORE and UniPROBE databases using TOMTOM. MAST was used for mapping of the motifs to ±200-bp regions around each peak center/summit with a p-value cutoff of 1e-4. Functional annotation of genes associated with peak regions was performed using GREAT (McLean et al., 2010) to integrate distal binding events, with the default association rule (5+1 kb basal, up to 1 Mb extension) and the whole mouse genome as background. For GSEA, each peak center was assigned to RefSeq genes whose TSSs are located within a 10-kb distance using the peak2gene script from the CEAS package. Enrichment was determined by comparing the expression levels in *Sox2* knockout ESCs or TSCs with those in their respective controls, or the levels in ZHBTc4 ESCs differentiating to TSCs (+Tet D1-D4) with those in the control (D0). The ChIP-Seq data are available at the GEO (GSE28455).

**Co-Immunoprecipitation Analysis**

Co-immunoprecipitation was performed using Dynabeads Co-Immunoprecipitation Kit (Invitrogen) following the manufacturer's instructions. Briefly, EGFP-$TS_{3.5}$ cells were treated with or without 1%

formaldehyde for 10 min, and lysed in buffer containing 100 mM NaCl, protease and phosphatase inhibitor cocktails. The whole cell lysate was immunoprecipitated with anti-Sox2 (raised by ourselves) or anti-Tfap2c antibodies covalently coupled to beads. The immunoprecipitated samples were analyzed by Western blot with anti-Sox2 (GT15098, Neuromics) and anti-Tfap2c antibodies.

**SUPPLEMENTAL REFERENCES**

Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development *136*, 1063-1069.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y.*, et al.* (2011). Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol *12*, R83.

Niwa, H., Yamamura, K., and Miyazaki, J. (1991). Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene *108*, 193-199.

Oda, M., Tanaka, S., Yamazaki, Y., Ohta, H., Iwatani, M., Suzuki, M., Ohgane, J., Hattori, N., Yanagimachi, R., Wakayama, T.*, et al.* (2009). Establishment of trophoblast stem cell lines from somatic cell nuclear-transferred embryos. Proc Natl Acad Sci U S A *106*, 16293-16297.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.