Cell Stem Cell
Article

Uncovering Early Response of Gene Regulatory Networks in ESCs by Systematic Induction of Transcription Factors

Akira Nishiyama,[1,2] Li Xin,[1,2] Alexei A. Sharov,[1] Marshall Thomas,[1] Gregory Mowrer,[1] Emily Meyers,[1] Yulan Piao,[1] Samir Mehta,[1] Sarah Yee,[1] Yuhki Nakatake,[1] Carole Stagg,[1] Lioudmila Sharova,[1] Lina S. Correa-Cerro,[1] Uwem Bassey,[1] Hien Hoang,[1] Eugene Kim,[1] Richard Tapnio,[1] Yong Qian,[1] Dawood Dudekula,[1] Michal Zalzman,[1] Manxiang Li,[1] Geppino Falco,[1,3] Hsih-Te Yang,[1] Sung-Lim Lee,[1,4] Manuela Monti,[1] Ilaria Stanghellini,[1,5] Md. Nurul Islam,[1] Ramaiah Nagaraja,[1] Ilya Goldberg,[1] Weidong Wang,[1] Dan L. Longo,[1] David Schlessinger,[1] and Minoru S.H. Ko[1,*]
[1]National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, USA
[2]These authors contributed equally to this work
[3]Present address: Istituto di Ricerche Genetiche Gaetano Salvatore Biogem s.c.ar.l., Ariano Irpino 83031, Italy
[4]Present address: College of Veterinary Medicine, Gyeongsang National University, Jinju, Gyeongnam 660701, Republic of Korea
[5]Present address: Reproductive Medicine Unit, S.I.S.Me.R. s.r.l., Bologna 40138, Italy
*Correspondence: kom@mail.nih.gov
DOI 10.1016/j.stem.2009.07.012

SUMMARY

To examine transcription factor (TF) network(s), we created mouse ESC lines, in each of which 1 of 50 TFs tagged with a FLAG moiety is inserted into a ubiquitously controllable tetracycline-repressible locus. Of the 50 TFs, *Cdx2* provoked the most extensive transcriptome perturbation in ESCs, followed by *Esx1*, *Sox9*, *Tcf3*, *Klf4*, and *Gata3*. ChIP-Seq revealed that CDX2 binds to promoters of upregulated target genes. By contrast, genes downregulated by CDX2 did not show CDX2 binding but were enriched with binding sites for POU5F1, SOX2, and NANOG. Genes with binding sites for these core TFs were also downregulated by the induction of at least 15 other TFs, suggesting a common initial step for ESC differentiation mediated by interference with the binding of core TFs to their target genes. These ESC lines provide a fundamental resource to study biological networks in ESCs and mice.

INTRODUCTION

The prevailing paradigm of modern biology states that gene regulatory networks determine the identity of cells, and that their alteration by environmental factors dictates changes of cell identity, i.e., cell differentiation (Davidson, 2006). Analysis of the structure and dynamics of gene regulatory networks is key to the understanding of biological systems but poses a great challenge because of the vast and manifold complexity of regulatory mechanisms.

One possible approach is to carry out a systematic gene perturbation study in order to "reverse engineer" these regulatory networks. Ideally, all the transcription factors would be manipulated one at a time and in different combinations in a variety of cell types. Readout of the impact of such manipulation would be monitored in a variety of ways, including the profiling of all RNA transcripts and proteins. This type of approach is complementary to conventional studies in which systematic gene manipulations have been successfully carried out, but usually with a focus on the phenotype, e.g., cell morphology, growth property, and differentiation markers (Chambers et al., 2003; Ivanova et al., 2006; Pritsker et al., 2006). To this end, mouse ESCs (Evans and Kaufman, 1981; Martin, 1981) are most suitable, because they can be differentiated into a variety of cell types in distinct cell culture conditions in vitro (Murry and Keller, 2008) and can also be developed into animal models to further study the effects of gene perturbation in vivo (Solter, 2006). As a complementary approach to the comprehensive loss-of-function study of all mouse genes (Collins et al., 2007; Skarnes et al., 2004), we aim to generate ESC lines in which a TF can be induced for gain of function in a controlled manner, enabling observations of the network perturbation caused by each TF in a uniform condition across all the ESC lines.

Global gene regulatory networks have been intensively studied in mouse ESCs by expression profiling (Walker et al., 2007), protein complex analysis (Wang et al., 2006), and genome-wide chromatin immunoprecipitation (Loh et al., 2006). The critical roles of three transcription factors—*Pou5f1* (*Oct4* or *Oct3/4*) (Nichols et al., 1998), *Sox2* (Yuan et al., 1995), and *Nanog* (Chambers et al., 2003; Mitsui et al., 2003)—discovered earlier (reviewed in Niwa, 2007; Silva and Smith, 2008) have recently been rationalized by the discovery of core transcriptional regulatory networks between these genes (Boyer et al., 2005; Loh et al., 2006). Furthermore, similar analyses of other key TFs in mouse ESCs have successfully extended the core transcriptional network (Chen et al., 2008; Kim et al., 2008). Obviously, the analysis of many more TFs, including genes that are not expressed in ESCs, is required to explore global TF network(s) that may be outside of the core transcriptional network of *Pou5f1*, *Sox2*, and *Nanog*. To this end, an appropriate mouse embryonic stem cell bank could facilitate a variety of high-throughput, genome-wide analysis methodologies.

Here we describe a strategy for and the establishment of TF-inducible ESC lines, and we show how these ESC lines can

be used in several ways. As a proof of principle for the strategy, we characterize how an exemplary differentiation-inducing TF, *Cdx2*, regulates the global transcriptome and shifts the balance of gene regulatory network toward ESC differentiation.

RESULTS

Generation and Quality Control of Transcription Factor-Inducible Mouse ESC Lines

We analyzed 50 TF genes (~3% of all 1600–2000 TFs encoded in the mouse genome; Kanamori et al., 2004; Messina et al., 2004) to assess the consequences of their induction in ESCs. These genes were selected primarily from a set of high-priority genes involved in critical functions in mouse ESCs and their differentiation, inferred in our previous work (Matoba et al., 2006). About half the genes are expressed in undifferentiated ESCs; the other half are not expressed or are expressed at a low level in undifferentiated ESCs but are induced in differentiating ESCs. We also included genes expressed late in lineage specification (*Ascl1*, *Ascl2*, *Myod1*, *Sox9*, and *Sfpi1*) to see whether TFs can induce their cognate targets without their usual regulatory context for function. Three non-TF genes—*Dppa5a*, *Gadd45a*, and *Tuba1a*—and one empty vector were included as controls. A total of 53 genes and a control used for this work are listed in Figure 1A. To induce a specific TF, we employed the Tet-repressible gene expression system, with the expression cassette integrated at the ROSA26 locus (ROSA-TET locus) (Figure 1B; Figure S1A available online; Masui et al., 2005). This system makes use of the ubiquitous and relatively high expression at the ROSA26 locus (Soriano, 1999). In the absence of doxycycline (Dox), the recombinant ROSA-TET locus expresses a polycistronic transcript for the open reading frame (ORF) of TF and Venus YFP proteins. To facilitate the detailed analyses of individual TFs, we inserted a FLAG tag at the C terminus of all transgenes, making it possible to use FLAG as a universal bait for immunological assays. In order to minimize clone-to-clone variation of gene expression level and to generate these ESCs as a permanent resource for future work, we synchronized passage numbers and performed multistep quality control of these ESCs (Figures 1B, 1F, and 1G; Supplemental Data).

Dox Inducibility of the Transgene in Each of 54 ESC Lines

We first carried out control experiments and found that three medium changes at 3 hr intervals minimized unwanted perturbation associated with commonly used cell passaging while inducing the transgene fully by effectively removing Dox (Figure S2). Indeed, the control ESCs, in which an expression unit without an ORF was inserted into the ROSA-TET locus, showed only a small number of genes differentially expressed (see below). In all transgene induction experiments, we set the last medium change as 0 hr induction.

We also carried out time course DNA microarray analysis of 10 ESC lines (Figure S3) and time course western blot analysis of 17 ESC lines (Figure 1D; Figure S4). Western blots showed that in all examined cases, a transgene-derived protein started to appear by 12 hr after induction and reached a maximum level by 48 hr. DNA microarray analysis showed that whereas the global transcriptome began to change within 24 hr, the expression of the majority of genes changed relatively monotonically until 72 hr (Figure S3). To capture early effects of TF induction, we looked 48 hr after induction for the expression profiling of all 54 ESC lines in the Dox^+ and Dox^- conditions. Except for Dox, all other culture conditions (including LIF) were the same in both Dox^+ and Dox^- conditions. We confirmed that each transgene expressed a protein that was detectable by an antibody against the FLAG tag by western blot and immunohistochemistry (Figures 1D and 1E; Figures S4 and S5). Immunohistochemistry also showed that TF proteins are mainly localized in the nucleus (Figure 1E; Figure S5). The induced level of a transgene was comparable among ESC lines based on the measurement of transcript levels via qRT-PCR (Figure 1C; Figure S6A). To assess the induced level of TFs at the protein level, we also carried out western blot analysis with native antibodies that detect both endogenous and exogenous TFs. As expected for TFs that are already expressed in ESCs, we detected only mild (up to 2- to 3-fold) increases in TF levels (Figures S4C and S4D). For example, the amount of STAT3 protein was induced by 2.7-fold, which was only 3.4-fold higher than that in thymus (Figure S4D). In contrast, CDX2, which is not usually expressed in ESCs, showed an ~80-fold increase, but was only ~2-fold higher than the highest protein level reached in the differentiated trophoblast cells (Figure S4D). These data indicate that the induced levels of TFs in this system are largely within the physiological range of gene dosage.

Global Patterns of Gene Expression in Response to Induction of TFs

To assess the extent of changes in global gene expression patterns, we first combined all the new microarray data obtained from 54 ESC lines with previous microarray data that we had obtained from ESCs differentiating into three cell lineages (Aiba et al., 2009). The data sets were fully compatible because the same microarray platform was used. Principal component analysis (PCA) of all 304 microarray data sets revealed that the transcriptome state of all the TF-inducible ESCs, even after 48 hr of TF induction (Dox^-), did not shift away from the zone where undifferentiated pluripotent ESCs were clustered (Figure 2A; see Table S1 for data on expression changes for all genes). This indicates that transcriptome changes measured 48 hr after TF induction reflect the early effects of the TF in undifferentiated or nearly undifferentiated ESCs, but not in more differentiated cell types. Consistent with the PCA, most of the TF-induced ESCs showed no significant morphological changes at 48 hr (data not shown). However, after 7 days of continuous induction of TFs, most of the examined ESC lines showed morphological changes indicative of differentiation (Figure S7).

Interestingly, even these early effects of transcriptome changes often revealed incipient trajectories of differentiation, as shown as clusters in the "heatmap" (Figure 2B; see Table S2 for data of individual genes). This became more evident when these expression profiles were compared to the microarray data of ESCs differentiating into specific lineages (Aiba et al., 2009) and those of mature tissues and organs (Su et al., 2002). For example, ESC lines with *Sox2*, *Pou5f1*, *Nrip1*, and *Ascl1* showed the greatest similarity to epiblast/neural cells; ESC lines with *Ascl2*, *Cdx2*, *Eomes*, and *Esx1* showed the greatest similarity to trophoblast cells; and ESC lines with *Gata3* showed

A

Gene Symbol	Aliases	Gene Description
Control		
Aes	Grg5	amino-terminal enhancer of split
Ascl1	Mash1	achaete-scute complex homolog 1 (Drosophila)
Ascl2	Mash2	achaete-scute complex homolog 2 (Drosophila)
Atf3	LGR-21	activating transcription factor 3
Cbx8	Pc3	chromobox homolog 8 (Drosophila Pc class)
Cdx2		caudal type homeo box 2
Ctnnb1	Catnb, Mesc	catenin (cadherin associated protein), beta 1
Dlx3		distal-less homeobox 3
Dnmt3b		DNA methyltransferase 3B
Dppa5a	Esg1	developmental pluri-potency associated 5A
Eed		embryonic ectoderm development
Elf1		E74-like factor 1
Eomes		eomesodermin homolog (Xenopus laevis)
Esx1		extraembryonic, spermatogenesis, homeobox 1
Etv3		ets variant gene 3
Fem1b		feminization 1 homolog b (C. elegans)
Foxj2		forkhead box J2
Gadd45a		growth arrest and DNA-damage-inducible 45 alpha
Gata3		GATA binding protein 3
Id1		inhibitor of DNA binding 1
Klf4		Kruppel-like factor 4 (gut)
Mef2c		myocyte enhancer factor 2C
Msc		musculin
Mybl2	B-Myb, Bmyb	myeloblastosis oncogene-like 2
Myc	c-myc	myelocytomatosis oncogene
Mycn	N-myc	v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian)
Myod1	MyoD	myogenic differentiation 1
Nanog		Nanog homeobox
Nr0b1	Dax1	nuclear receptor subfamily 0, group B, member 1
Nr2f2	COUP-TF2	nuclear receptor subfamily 2, group F, member 2
Nr5a2	LRH-1	nuclear receptor subfamily 5, group A, member 2
Nrip1	RIP140	nuclear receptor interacting protein 1
Otx2		orthodenticle homolog 2 (Drosophila)
Pou5f1	Oct4, Oct3, Oct3/4	POU domain, class 5, transcription factor 1
Rhox6	Psx1	reproductive homeobox 6
Rxra	Nr2b1	retinoid X receptor alpha
Sall4	Tex20	sal-like 4 (Drosophila)
Sfpi1	Dis1, PU.1, Spi-1	SFFV proviral integration 1
Smad1		MAD homolog 1 (Drosophila)
Smad4		MAD homolog 4 (Drosophila)
Smad7		MAD homolog 7 (Drosophila)
Sox2		SRY-box containing gene 2
Sox9		SRY-box containing gene 9
Stat3		signal transducer and activator of transcription 3
Suz12		suppressor of zeste 12 homolog (Drosophila)
T	Bra	brachyury
Tcea3		transcription elongation factor A (SII), 3
Tcf3		transcription factor 3
Tcf4		transcription factor 4
Tuba1a		tubulin, alpha 1A
Whsc2	NELFA	Wolf-Hirschhorn syndrome candidate 2 (human)
Zfand3	Tex27	zinc finger, AN1-type domain 3
Zscan4c	Gm397	zinc finger and SCAN domain containing 4C

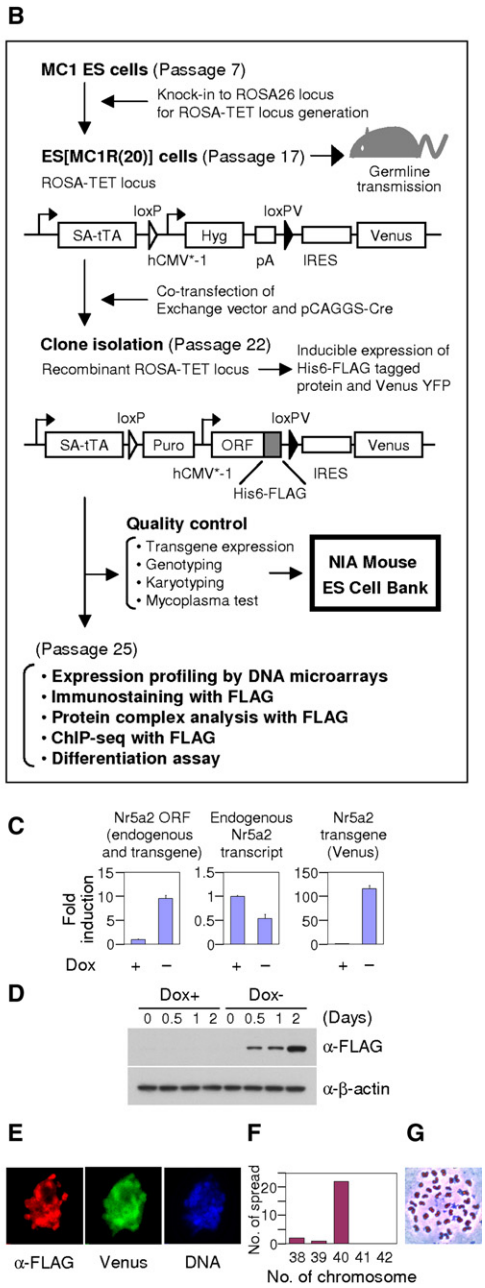

Figure 1. Strategy to Establish and Quality-Control ESC Lines

(A) List of ESC lines generated and analyzed in this study.

(B) Schematic diagram for the strategy. A parental ESC line was named ES[MC1R(20)], which stands for MC1 ESCs, ROSA-TET locus [R], and clone number 20. Each ESC line was named by adding the name of a transgene and the clone number. For example, the ESC line that was generated by integrating the *Aes* gene was named ES[MC1R(20):tetAes(24)]. For brevity, ESC lines are simply referred by the name of a transgene (e.g., Aes) throughout the text and figures.

(C–G) Representative results for quality control of the ESC line ES[MC1R(20):tetNr5a2(7)].

(C) qRT-PCR analysis of transcript levels of the ESCs cultured for 48 hr in the presence (+) or absence (−) of doxycycline: (left) transcripts measured by a primer pair for ORF (both endogenous and transgene *Nr5a2*); (middle) transcripts measured by endogenous *Nr5a2*-specific primer pair; (right) transcripts measured by a primer pair for Venus (representing a transgene). Values are shown as fold-induction compared with Dox^+ condition. Data are presented as means ± SEM.

(D) Time-course analysis of exogenous (i.e., a transgene-derived) NR5A2 protein expression by western blotting with an antibody against FLAG (top) and β-actin (bottom).

(E) Localization of the exogenous NR5A2 protein by immunostaining with FLAG antibody (left) and localization of Venus fluorescence (middle) and DNA (right).

(F) Karyotypes of ES[MC1R(20):tetNr5a2(7)] showing 88% euploidy.

(G) A representative picture of the metaphase spread. See Supplemental Experimental Procedures for information on other clones.

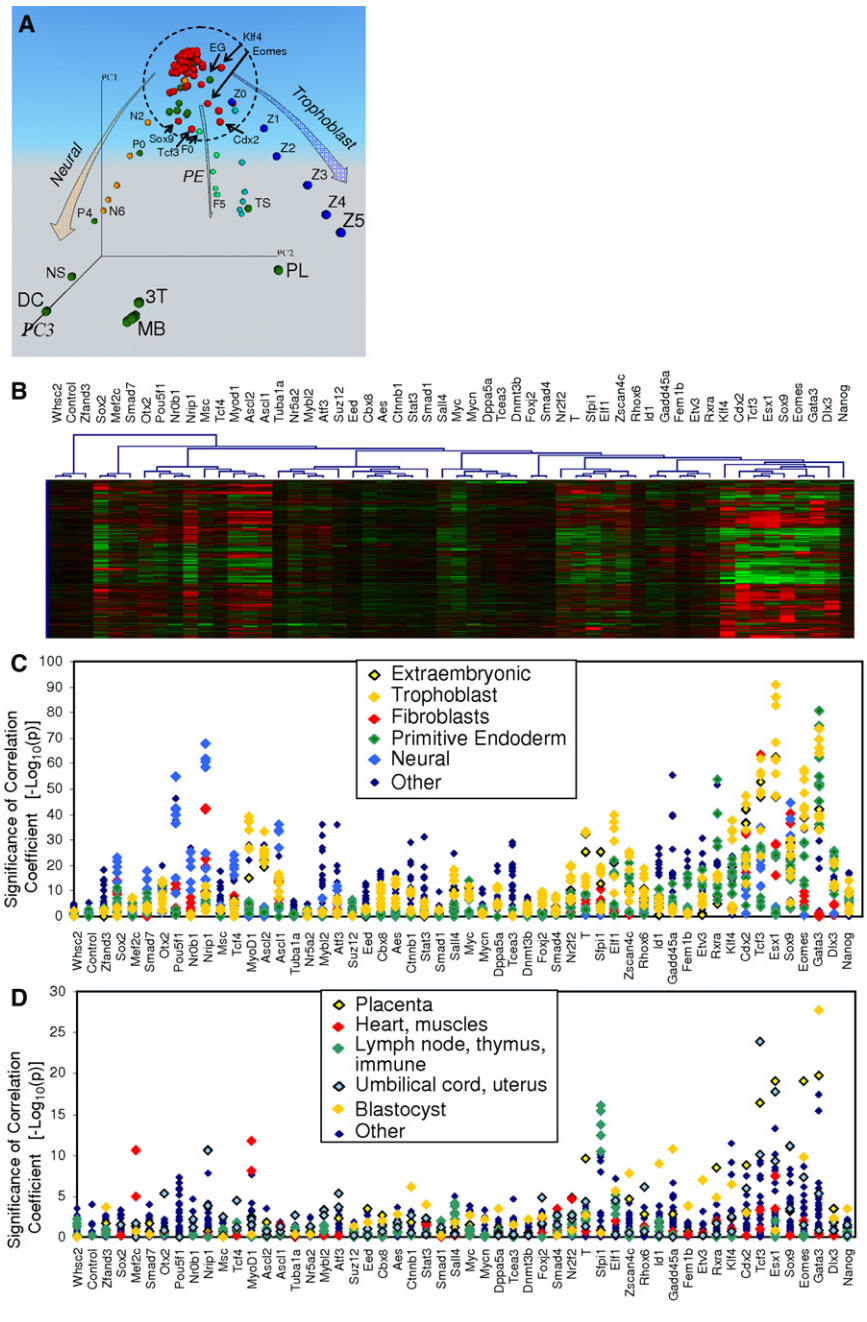

Figure 2. Global Expression Profiles of TF-Inducible ESC Lines

(A) Principal component analysis (PCA) of gene expression profiles of 152 different cell types: 54 TF-inducible ES lines with induced overexpression of various TFs (48 hr in Dox$^-$, marked red), the same 54 TF-inducible ES lines (48 hr in Dox$^+$, marked red), and 44 different cell types, which represent ESCs differentiating into three cell lineages (trophoblast, primitive endoderm [PE], and neural, marked blue, green, and yellow, respectively). All cell lines with induced TFs show gene expression profiles (encircled) very similar to that of undifferentiated ESCs, although a few TFs caused some changes toward differentiation (shown by arrows). The explanation of PCA and the details of these 44 cell types are given in the previous publication (Aiba et al., 2009). Only representative cell types are labeled: *Klf4* (Dox$^-$), *Sox9* (Dox$^-$), *Tcf3* (Dox$^-$), *Cdx2* (Dox$^-$), and *Eomes* (Dox$^-$). The trophoblast lineage is represented by Z0–Z5 (ESCs differentiating into trophoblast cells from day 0 to day 5 after repressing *Pou5f1* expression), TS (trophoblast stem cells), and PL (E13.5 placenta). The PE lineage is represented by F0–F5 (embryonal carcinoma cells differentiating into primitive endoderm from day 0 to day 5). The neural lineage is represented by N2–N6 (ESCs differentiating into neural lineage from day 2 to day 6), P0–P4 (embryonal carcinoma cells differentiating into neural cells), NS (neural stem/progenitor cells), and DC (differentiated neuron and glia). 3T, NIH 3T3 fibroblast cells; MB, mouse embryo fibroblast cells.

(B) A heatmap showing the results of hierarchical clustering analysis of all the microarray data (54 ESC lines). Only the top 3000 genes whose expression are most significantly altered are used for the analysis. Both genes and ESC lines are clustered according to their similarity of global gene expression patterns measured by Pearson correlation of log-transformed expression values. The list of genes and their expression change for this heatmap is given in Table S2.

(C and D) Significance of correlations between gene expression response to the induction of TFs in TF-inducible ESC lines (data from this paper) and gene expression in published data sets (Aiba et al., 2009; Su et al., 2002). Gene expression in published data sets was log-transformed and centered; the mean value was subtracted for each gene.

(C) Cell types in the data set for trajectories of early lineage differentiation (Aiba et al., 2009): Extraembryonic (TS, PL); trophoblast (Z2–Z5); fibroblasts (3T, MB, MD, ST); primitive endoderm (F2–F5, G1–G5); neural/primitive ectoderm (N2–N6, NS, DC); and other (E1–E5, EG, F0, F1, G0, IF, IN, N0, N1, P0, P4, TG, Z0, Z1).

(D) Tissues in the GNF database (Su et al., 2002): placenta; heart and muscles (skeletal); lymph node, thymus, immune (B220$^+$ B cells, CD4$^+$ T cells, CD8$^+$ T cells); umbilical cord, uterus; blastocyst; and other (adipose tissue, adrenal gland, amygdala, bladder, bone, bone marrow, brown fat, cerebellum, cerebral cortex, digits, dorsal root ganglia, dorsal striatum, embryo day 10.5, embryo day 6.5, embryo day 7.5, embryo day 8.5, embryo day 9.5, epidermis, eye, fertilized egg, frontal cortex, hippocampus, hypothalamus, kidney, large intestine, liver, lung, mammary gland [lact], medial olfactory epithelium, olfactory bulb, oocyte, ovary, pancreas, pituitary, preoptic, prostate, salivary gland, small intestine, snout epidermis, spinal cord lower, spinal cord upper, spleen, stomach, substantia nigra, testis, thyroid, tongue, trachea, trigeminal, vomeralnasal organ).

the greatest similarity to primitive endoderm cells (Figure 2C). Similarly, even TFs that are known to function in late lineage specification induced expression profiles that correspond to those late-stage differentiated cells. For example, ESCs with *Myod1* and *Mef2c* showed the greatest similarity to heart and muscle tissues, and ESCs with *Sfpi1* showed the greatest similarity to lymph node, thymus, and immune cells (Figure 2D). The results were generally consistent with previously published

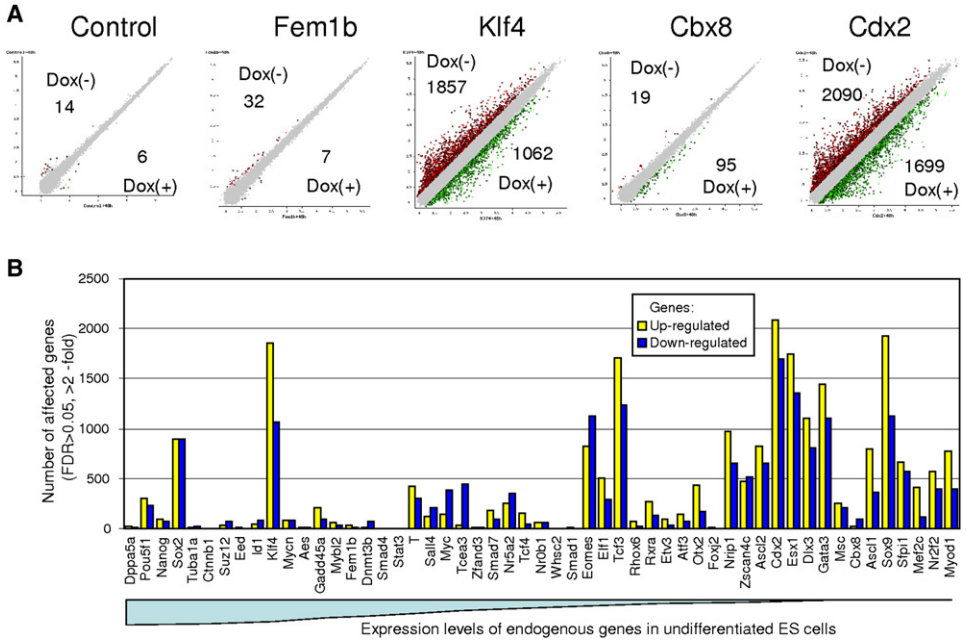

Figure 3. Extent of Transcriptome Perturbation by TFs and Pair-wise Comparison of Expression Changes

(A) Scatter-plots comparing expression profiles of representative ESC lines between Dox$^+$ and Dox$^-$ conditions. Red spots represent genes that show higher than 2-fold expression (upregulated) and green spots represent genes that show lower than 2-fold expression (downregulated) in Dox$^-$ condition than in Dox$^+$ condition with statistical significance of FDR < 0.05. The number of up- and downregulated genes are also shown.

(B) The number of genes whose expression was affected significantly (FDR < 0.05 and expression changes > 2-fold) by induction of various TFs in ESCs. ESC lines are ordered according to the expression levels of endogenous TF gene in undifferentiated ESCs based on published RNA-Seq data (Table S11; Cloonan et al., 2008).

functions of these TFs: *Cdx2*, *Esx1*, *Ascl2*, and *Eomes* (Simmons and Cross, 2005); *Sox2* and *Ascl1* (Diez del Corral and Storey, 2001); and *Myod1* and *Mef2c* (Naya and Olson, 1999).

Next, genes whose expression was affected by induction of TFs were identified via pair-wise statistical comparison (FDR < 0.05 and expression changes > 2-fold) between microarray data for the same clone in Dox$^+$ and Dox$^-$ conditions (Figure 3A; Figures S6B and S8). Some TFs (e.g., *Cdx2*, *Esx1*, *Gata3*, *Klf4*, *Sox9*, and *Tcf3*) caused substantial changes in the transcriptome, whereas other TFs (e.g., *Fem1b* and *Cbx8*) had little effect (Figure 3B; Tables S3 and S4). For the most part, induction of TFs that were already present in ESCs (ESC lines on the left side of Figure 3B) had a smaller effect on the gene expression profile than induction of differentiation-related genes that have low endogenous expression in ESCs (ESC lines on the right side of Figure 3B). It seems reasonable that the greater the fold-induction of a TF, the greater the global perturbation of transcriptome. However, induction of *Klf4* and *Sox2* (and to some extent *Pou5f1*) resulted in a strong response even though they were already expressed in ESCs and thus showed low fold-induction of their expression levels, indicating that these TFs have an unusually dose-sensitive and potent regulatory role.

Dissecting Gene Regulatory Networks: The Example of Cdx2

As a proof of principle for the utility of the ESC lines, we report our study of Caudal type homeobox 2 (*Cdx2*), which was selected because of its exceptionally strong effect on the transcriptome (Figure 3B) and its unique role in mouse embryo development. *Cdx2* is the earliest differentiation marker in the embryo and is highly expressed in the trophectoderm lineage (Strumpf et al., 2005), and the balance between *Pou5f1* and *Cdx2* expression shifts cell fate during preimplantation development (Niwa et al., 2005).

Time course western blot analysis showed that the product of the transgene (CDX2-FLAG) was induced to a high level within 0.5 days after removal of Dox, reaching a maximum by 48 hr and remaining very high until day 5, with a slight reduction by day 7 (Figure 4A). Antibodies against CDX2, which recognize both endogenous and exogenous CDX2, showed similar expression patterns (Figure 4A). Colony formation assays followed by alkaline phosphatase staining showed that ESCs and colonies became very flat and lost alkaline phosphatase staining, indicating that *Cdx2* induction alone caused differentiation of ESCs by day 7 of induction (Figure 4B). Differentiation of the *Cdx2*-inducible ESC line to trophoblast cells was confirmed by positive immunostaining with trophoblast markers CDC42 (Natale and Watson, 2002) and ITGA7 (Klaffky et al., 2001) (Figure 4C). These data are thus consistent with the previous report of the induction of trophoblast cells from ESCs by *Cdx2* overexpression (Niwa et al., 2005). Taken together, these data confirm that transgene *Cdx2*-FLAG was induced by Dox, was translated properly, and was functional as CDX2 protein.

Based on the DNA microarray analysis, 2090 genes were upregulated and 1699 were downregulated by 48 hr of Cdx2 induction (Figure 3A). These genes include not only direct targets

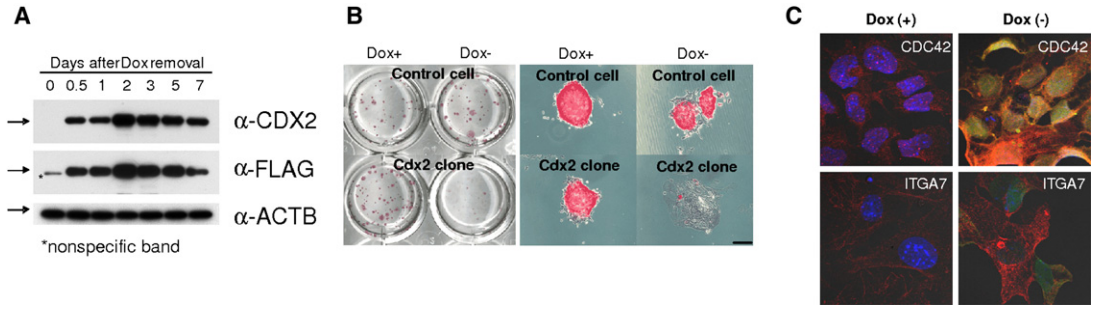

Figure 4. Analysis of the CDX2-Inducible ESC Line

(A) Time-course analysis of CDX2 protein expression by western blotting (day 0, 0.5, 1, 2, 3, 5, and 7 after removal of Dox). Antibody against CDX2 protein recognizes both endogenous and exogenous (i.e., transgene-derived) CDX2 protein. Antibody against FLAG recognizes only exogenous CDX2 protein. Antibody against ACTB is used for the loading control.

(B) Alkaline phosphatase activity was tested in *Cdx2* overexpressing (bottom) and control (top) cells with or without Dox for 6 days in culture (over view, left; magnified, right).

(C) Cdx2-overexpressing cells induce trophectoderm markers CDC42 (top) and integrin alpha 7 (ITGA7; bottom) by 6 days.

of CDX2, but also indirect targets that are regulated by the direct targets. To identify direct targets of CDX2 at the genomic level, we applied ChIP-Seq to *Cdx2*-inducible ESCs 48–60 hr after induction. At this time, the ESCs did not yet show signs of differentiation; thus, the ChIP-Seq results reflect *Cdx2* function at the very start of ESC differentiation. ChIP-western confirmed that FLAG-IP pulled down cross-linked DNA-CDX2 protein complexes in the Dox^{-} condition (Figure 5A). Sequencing of FLAG-ChIP DNAs produced 17.59 million 36-nucleotide tags that were mapped to the latest mouse genome sequence (mm9, NCBI/NIH). We found that 5.72 million tags matched to the genome with $\leq$2 nucleotide mismatches; the remaining tags either did not match to the genome or matched to more than 5 different locations. We found a total of 59,098 peaks with at least 6 tags (Table S5), of which 15,855 had at least 10 tags. Significant peaks of tags (>9) were observed at 3152 loci (genes) within 15 kb upstream and downstream of the transcription start sites (TSS) (Table S6). However, only 38 genes had peaks in the promoter regions (<300 bp upstream or downstream from the TSS), which indicates that CDX2 binds mostly to more distant regulatory regions (300–15,000 bp upstream or downstream of the TSS). Figure 5B shows a typical example of peaks. Analysis of the CDX2-ChIP target sequence by CisFinder software (Sharov and Ko, 2009; available at http://lgsun.grc.nia.nih.gov/CisFinder/) indeed identified one main motif and five additional variant motifs (Figure 5C). CDX2 binds mainly to a [T/C][A/C]ATAAA[A/G] motif and to its direct repeat (Figure 5D). The major motif matched the CDX2-binding motif identified by direct binding in an oligonucleotide assay (Berger et al., 2008).

It is conceivable that some binding sites of TF may not be involved in the regulation of gene expression. Therefore, we used our recently published approach to identify "functional" direct targets of TF by combining TF binding information from ChIP-Seq and gene expression changes caused by TF induction (Sharov et al., 2008). The method compares binding score distributions in genes that responded to TF manipulation with those in control genes that did not respond to TF manipulation (see details in Supplemental Experimental Procedures, Table S7). Of the 3152 genes with CDX2-binding sites, the analysis revealed a total of 337 functional target genes that satisfied statistical criteria ($p < 0.1$ and FDR < 0.6) (Table S8). Of these genes, 334 were upregulated after the induction of *Cdx2* gene, and only 3 were downregulated. Functional CDX2-target genes included Hox genes and other differentiation-associated genes, consistent with CDX2 function as an inducer of ESC differentiation (Figure 5E; Table S8). The GO annotations of CDX2 target genes are also available (Table S9). Selected target genes were also further validated by ChIP-qPCR (Figure 5F). Consistent with the ChIP-Seq data, we observed notably high enrichment of CDX2 at promoter regions of Hox genes by ChIP-qPCR validation (Figure 5F).

To see the correlation between the up- or downregulated genes and CDX2-binding sites, we ranked all 25,030 genes according to the changes caused by *Cdx2* induction and estimated the proportion of genes with CDX2-binding sites in a sliding window of 500 genes (Figure 5G; Tables S6 and S7). Interestingly, genes upregulated after *Cdx2* induction were strongly enriched in genes with CDX2 binding, but no enrichment was observed among downregulated genes. This implies that upregulation (i.e., positive regulation) of downstream genes is mediated by direct binding of CDX2, whereas downregulation of downstream genes is not. To gain further insights into a possible mechanism for downregulation by CDX2, we used published ChIP-Seq data for 13 TFs (Table S10; Chen et al., 2008). For each TF, we estimated the proportion of genes with its binding sites in their distal regulatory regions among genes that were up- or downregulated by CDX2 and compared this with the proportion of genes with binding sites among "control" genes unresponsive to CDX2 (response < 1.25-fold). Strikingly, a list of genes downregulated by the induction of Cdx2 was enriched in genes that carry binding sites for POU5F1, SOX2, NANOG, STAT3, and SMAD1 (Figure 5H). Because it is known that POU5F1, SOX2, and NANOG form a core transcriptional network (Boyer et al., 2005; Chen et al., 2008; Jaenisch and Young, 2008; Kim et al., 2008; Loh et al., 2006), we tentatively call them, as a group, pluripotency-associated transcription factors (pTFs). When we plotted the proportion of genes with binding sites of at least two pTFs against the changes of expression caused by *Cdx2* induction (Figure 5G; Table S7), we found that genes downregulated after *Cdx2* induction were strongly enriched in

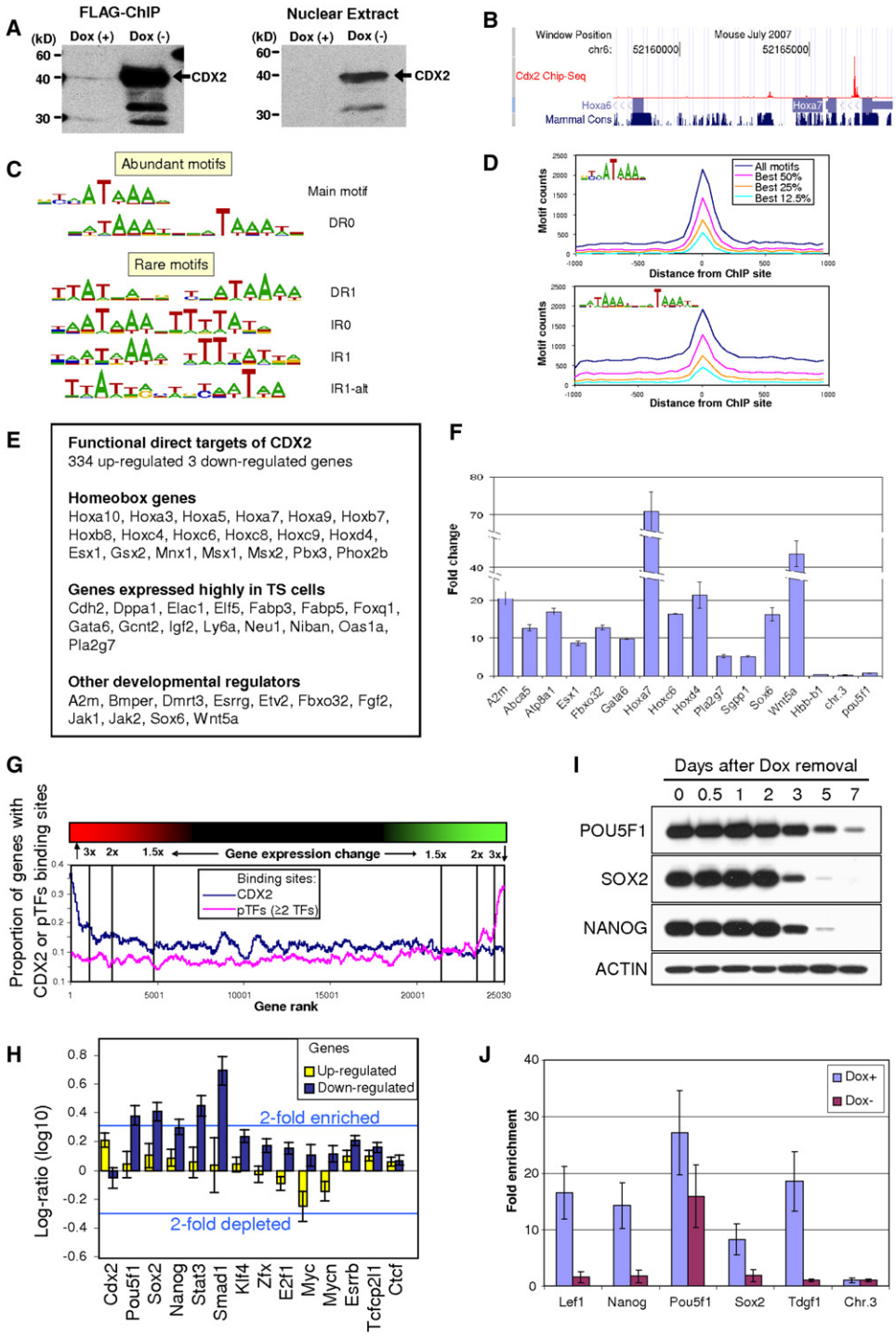

Figure 5. ChIP-Seq Analysis of Cdx2-Inducible ESC Line

(A) Chromatins were prepared from *Cdx2*-inducible ESCs cultured for 48–60 hr in the Dox$^+$ and Dox$^-$ conditions. Chromatin immunoprecipitation (ChIP) was carried out by using anti-FLAG M2 affinity gel. ChIP product was tested by western blotting with FLAG antibody. Nuclear extract from ESCs cultured for 48–60 hr in Dox$^+$ and Dox$^-$ condition was used for the western blot.

(B) CDX2 ChIP-Seq peaks in the *Hoxa7* gene region. UCSC Mouse Mm9 browser view of *Hoxa7* gene locus after mapping CDX2 ChIP-Seq tags locations in the wiggle format. CDX2 ChIP-Seq peaks are shown in red.

(C) CDX2-binding motifs identified with CisFinder via 200 bp sequences centered at ChIP sites.

(D) Overrepresentation of CDX2 binding motifs in ChIP sites. Genomic sequences (2000 bp) centered at CDX2 binding sites found by ChIP-seq were extracted from the genome and searched for the occurrence of CDX2 motifs. Binding motif was characterized by the position-frequency matrix (PFM) generated with CisFinder software (see the text). Motif fit was evaluated by log-likelihood method assuming false positive rate of 1 per 10 Kb of a random sequence.

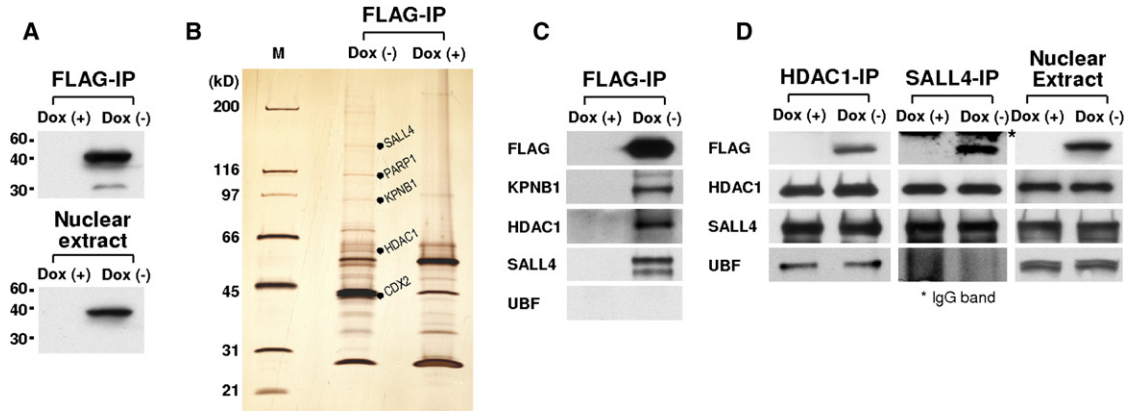

Figure 6. Analysis of the CDX2 Protein Complex Pulled down by FLAG Immunoprecipitation

(A) Confirmation of immunoprecipitation with FLAG antibody. Immunoprecipitates and nuclear extract were used for western blotting with FLAG antibody.

(B) A silver-stained SDS gel showing the presence of CDX2 major band and other distinct bands. Anti-FLAG M2 affinity gel was used for IP of CDX2 protein complex from Cdx2-inducible ESCs. Nuclear extracts were prepared from Cdx2-inducible ESCs cultured for 48–60 hr in the Dox^+ and Dox^- condition. Some bands are marked with protein names identified by the mass spectrometry of the IP products. M, markers.

(C) Mass-spectrometry result was verified by IP-western assay. Anti-FLAG M2 affinity gel was used to immunoprecipitate (IP) CDX2 protein complex from the nuclear extracts from Cdx2-inducible ESCs cultured for 48–60 hr in the Dox^+ and Dox^- conditions. IP products were tested by western blotting with antibodies against FLAG, KPNB1, HDAC1, and SALL4. Antibody against UBF was used as a control.

(D) Reverse IP carried out with antibodies against either HDAC1 or SALL4. IP products were tested by western blotting with antibodies against FLAG, HDAC1, SALL4, and UBF. Nuclear extracts were also used as controls. Control UBF was detected in HDAC-IP sample as reported previously, but not detected in the SALL4-IP sample.

genes with pTF binding, but no enrichment was observed among upregulated genes.

We initially considered that CDX2 might first downregulate transcription of *Pou5f1*, *Sox2*, or *Nanog*, in turn resulting in the reduction of either POU5F1, SOX2, or NANOG protein, and consequently the downregulation of pTF-target genes. However, ChIP-Seq data showed that CDX2 did not bind to enhancer/promoters of *Pou5f1*, *Sox2*, and *Nanog*. Furthermore, when we tested the protein level of these TFs with time course western blot analysis of POU5F1, SOX2, and NANOG after CDX2 induction in ESCs, we found that the levels of POU5F1, SOX2, and NANOG protein did not change by day 2, when pTF-target genes were already significantly downregulated (Figure 5I, microarray data). These findings make it unlikely that CDX2 first downregulated *Pou5f1*, *Sox2*, and *Nanog*, leading to subsequent downregulation of pTF-target genes. Instead, CDX2 seems to interfere with the binding of POU5F1, SOX2, and/or NANOG to enhancers/promoters of their target genes. To further investigate this possibility, we selected Pou5f1 as an example and carried out ChIP-qPCR analysis of POU5F1-target genes by using POU5F1 antibody in Cdx2-inducible ESCs in the Dox^+ (i.e., in the absence of CDX2) or in the Dox^- (i.e., in the presence of CDX2) conditions (Figure 5J). The results indeed showed that the induction of CDX2 caused reduced binding of POU5F1 to its target sequence in downstream genes (*Lef1*, *Tdgf1*, *Sox2*, *Nanog*, and to some extent, *Pou5f1* itself) (Figure 5J). Taken together, the data strongly suggest that CDX2 upregulates direct target genes by directly binding to their regulatory regions, but downregulates genes by interfering with the binding of pTFs to the regulatory regions of pTF-target genes.

To investigate a possible mechanism by which CDX2 might interfere with the binding of pTFs to their target genes, we used a FLAG antibody to isolate a putative protein complex from the Cdx2-manipulated ESCs 48–60 hr after CDX2 induction (Dox^-) (Figure 6A). A silver-stained SDS-PAGE gel showed a series of discrete bands that were not observed in a control sample isolated from the Cdx2-manipulated ESCs in Dox^+ conditions (Figure 6B). The silver-stained gel also indicated

(E) Functional CDX2-target genes.

(F) Cdx2 ChIP-Seq result was verified by qPCR. Primers flanking a promoter region of *Hbb-b1* and *Pou5f1* as well as a gene desert region in chromosome 3 were used as negative controls. Primers flanking of *Actb* gene promoter were used for normalization. The relative enrichment of CDX2 binding was indicated as fold change. Error bars: ±SEM.

(G) Blue line: Relationship between gene expression changes caused by Cdx2 induction (x axis) and the proportion of genes with CDX2 binding sites identified with a sliding window of 500 genes (y axis). Red line: Relationship between gene expression changes caused by Cdx2 induction (x axis) and the proportion of genes with pTF binding sites identified with a sliding window of 500 genes (y axis) (more than two TFs out of POU5F1, SOX2, NANOG, STAT3, and SMAD1; data from H).

(H) Enrichment of genes with binding sites for each of 14 TFs among genes whose expression was upregulated or downregulated by the induction of Cdx2. The enrichment was measured as a log ratio (log10) of the proportion of genes with binding sites among sets of genes that were upregulated (or downregulated) by >2-fold and the proportion of genes with binding sites among the set of control genes that changed <1.25-fold. The error bar corresponds to the standard error (SE) of the log ratio.

(I) Time course analysis of endogenous POU5F1, SOX2, NANOG, and ACTB protein expression by western blotting with antibodies against each protein. Cdx2-inducible ESC line was cultured for 0, 0.5, 1, 2, 3, 5, and 7 days in the Dox^- condition.

(J) ChIP-qPCR analysis for POU5F1 binding on its target genes in Cdx2-inducible ESCs. Primers flanking a gene desert region in chromosome 3 were used as a negative control. The relative enrichment of POU5F1 binding was represented as a fold change. Error bars: ±SEM.

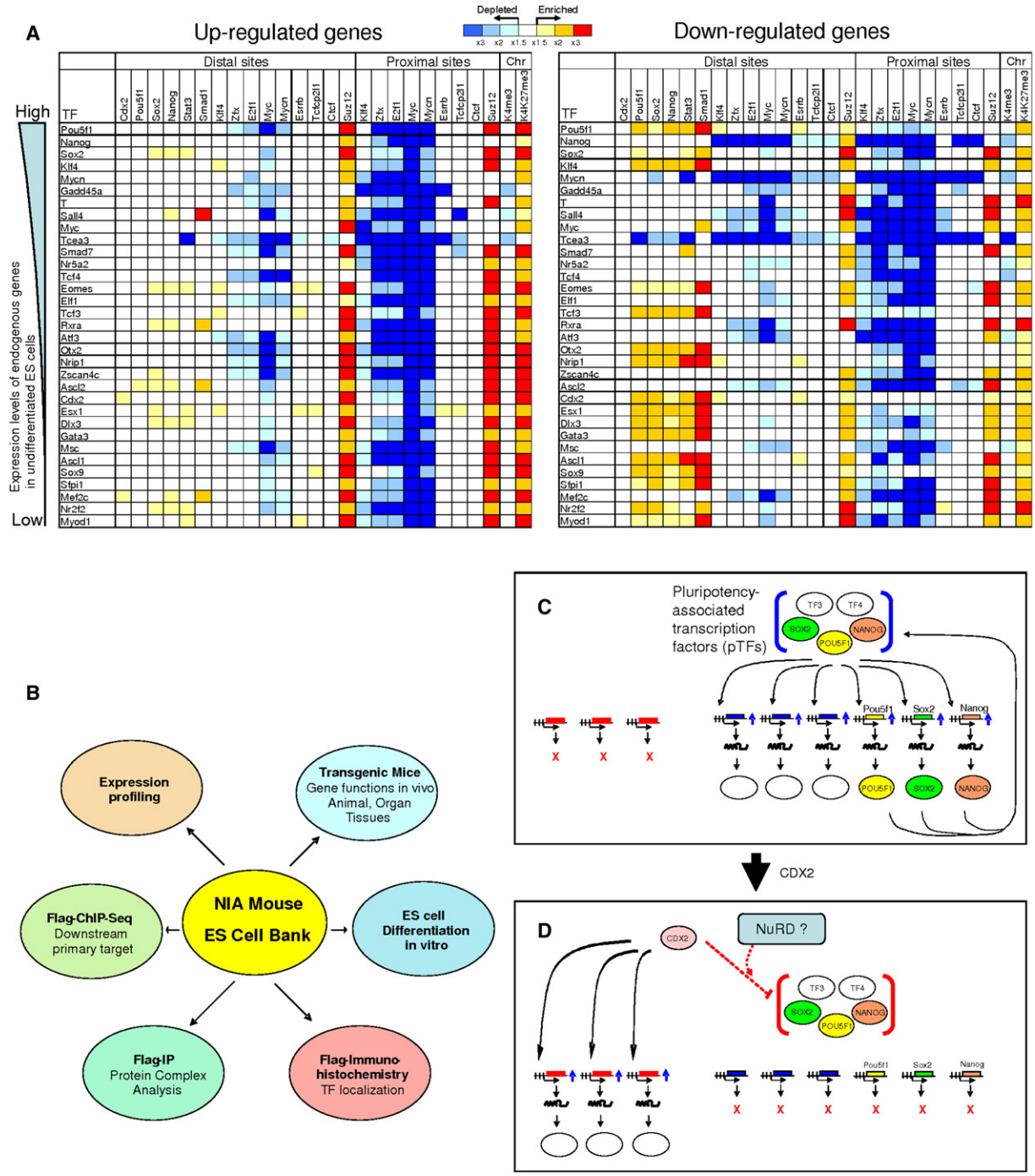

Figure 7. Compendium Analysis of TF-Binding Loci and Expression Profiles after TF Induction

(A) Relative abundance of genes with TFBS (transcription factor binding sites) in distal (0.3–15 Kb upstream and downstream from the TSS) and proximal (<300 bp upstream and downstream from the TSS) portions of the promoter, as well as genes with chromatin modifications within 1 Kb from the TSS, in sets of genes upregulated or downregulated (>2-fold changes of gene expression, but at least 200 genes in each group) by the induction of TFs (shown in the first column), compared to the control set of unaffected genes. CDX2, POU5F1, SOX2, NANOG, STAT3, and SMAD1 bind mostly to distal sites, and the number of binding sites in proximal promoters was not sufficient, and thus was not included for analysis. TFs are ordered according to the expression level of endogenous genes in ESCs from abundant to those that are not expressed in ESCs based on published RNA-Seq data (Table S11; Cloonan et al., 2008). Cells are color-coded based on the overrepresentation or underrepresentation of genes with TFBS compared to the control group of genes that did not respond to the induction of TF (<1.25-fold change). Cells are not color-coded if the difference in the proportion of genes with TFBS is not statistically significant. Data on TFBS and chromatin modifications (Chr) in promoters of genes were compiled from our ChIP-Seq experiment with CDX2 (Figure 5) and published data (H3K4me3 and H3K27me3 data from Mikkelsen et al. [2007]; the rest of the data from Chen et al. [2008]). K4me3 indicates genes marked with H3K4me3; K4K27me3 indicates genes marked with both H3K4me3 and H3K27me3 ("bivalent domains").

that a significant fraction of CDX2 was present in a free form. Mass spectrometric analysis of the immunoprecipitated protein complex revealed a number of proteins matched by multiple peptides (Figure S9). Based on the relatively high number of peptide matches, the following protein groups are likely to be the components of CDX2 complex: (1) NuRD (nucleosomal remodeling and histone deacetylase) complex, including HDAC1, MBD3, and CHD4 (Denslow and Wade, 2007); (2) SALL4; (3) PARP1; and (4) KPNB1 (Importin-1-beta), a protein known for its function in nuclear transport (Lange et al., 2007).

The presence of HDAC1, SALL4, and KPNB1 in the CDX2-associated complex was validated by IP-western blotting with antibodies against these proteins (Figure 6C). To test whether SALL4 is a component of NuRD-CDX2 complex, we carried out a reverse-IP assay with antibodies against HDAC1 and SALL4, respectively (Figure 6D). Western blotting results confirmed the interaction between SALL4 and HDAC1 both in the absence of CDX2 (Dox^+) and in the presence of CDX2 (Dox^-). By contrast, UBF, used as a control, was present in the HDAC1-IP sample but absent in the SALL4-IP sample. HDAC1 and SALL4 were present at similar levels in the nuclear extract from Dox^+ and Dox^- cells. Taken together, these data indicate that CDX2 associates with NuRD and SALL4 in Cdx2-induced ESCs.

Compendium Analysis of TF-Binding Sites in Genes Affected by the Induction of TFs

To gain further information about global TF regulatory networks in ESCs, we extended the analysis done for CDX2 to the other 32 TFs that caused significant changes in the expression of >150 genes (Figure 3B). Gene sets that were up- or downregulated by each induced TF were examined for overrepresentation of genes with binding sites of various TFs and with various chromatin modifications based on ChIP-Seq data published for ESCs with sufficient tag numbers (Figure 7A; Table S10; Chen et al., 2008; Mikkelsen et al., 2007). The compendium analysis revealed three global features of gene regulatory networks in ESCs.

First, like CDX2, lists of genes downregulated by the induction of at least 15 other TFs were enriched for those with binding sites for pTFs (i.e., POU5F1, SOX2, NANOG, STAT3, and SMAD1) (Figure 7A; Figures S10 and S11). To confirm this finding, we did a similar analysis with TF binding sites identified with ChIP-chip methodology (Kim et al., 2008). The results (Figure S12) were consistent with our previous analysis and showed that pTFs include three additional TFs: *Nr0b1* (*Dax1*), *Nac1*, and *Zfp281*. pTF targets include genes that are commonly associated with ESC pluripotency: *Nr5a2*, *Fgf4*, *Lrrc2*, *Foxd3*, *Klf2*, *Nr0b1*, *Tcea3*, *Tdgf1*, *Zfp42*, *Aire*, *Phc1*, *Mycn*, *Sox2*, *Jmjd1a*, *Jarid2*, *Nanog*, *Spp1*, *Myc*, *Nodal*, *Dppa3*, *Trim24*, *Zic3*, *Sall4*, *Dppa5a*, *Rest*, *Lefty1*, *Lefty2*, *Mybl2*, and *Pou5f1* (see Figure S11 for a full gene list).

Second, genes with binding sites for the polycomb gene *Suz12* were enriched in the lists of genes that responded to the induction of nearly all the TFs. We looked at the genes previously identified as having "bivalent" chromatin domains in their promoters, characterized by a combination of H3K4me3 and H3K27me3 marks (Azuara et al., 2006; Bernstein et al., 2006; Roh et al., 2006), because it is known that *Suz12* is associated with bivalent domains (Boyer et al., 2006; Lee et al., 2006). With published data (Table S10; Mikkelsen et al., 2007), we found that both genes upregulated and genes downregulated by nearly all the TF inductions were enriched for those with bivalent domains (Figure 7A). Previously, a bivalent domain has been attributed to genes that are "poised" or "primed," indicating that the expression levels of these genes are low or none, but that the gene is ready to be activated immediately (Azuara et al., 2006; Bernstein et al., 2006). Therefore, it is not unusual to see that upregulated genes fall into the category of genes marked with bivalent domains. However, current models do not anticipate that many of the genes that are downregulated also fall into bivalent domains. Intrigued by the downregulation of genes with bivalent domains, we first searched for genes with relatively high expression in ESCs (>30% of maximum expression level) in our earlier compendium microarray data of differentiating ESCs (Aiba et al., 2009) and found 460 such genes with bivalent domains (Figure S13). Among them, 280 genes were downregulated more than 2-fold after induction of some TFs (Figure S13). To validate whether these genes were indeed downregulated during differentiation, we examined the changes of expression by microarray and qRT-PCR for five genes from this list during ESC differentiation into trophoblasts. In all genes examined, expression was indeed downregulated by hundreds of folds (Figure S13E).

Third, genes with binding sites of MYC, MYCN, E2F, and ZFX in the proximal regulatory regions were strongly depleted in both up- and downregulated gene lists in almost all the TF-induced ESCs (Figure 7A). Similar results were obtained with MYC binding sites from another report (Figure S12; Kim et al., 2008). Some of these genes are already maximally expressed and therefore cannot be upregulated further. It is not clear why expression of these genes is not effectively downregulated after manipulation of TFs. In any case, binding sites of MYC, MYCN, E2F, and ZFX seem to mark genes that are refractory to the induction of TFs.

DISCUSSION

Initial analyses of ~3% (50) of all TF genes have provided a glimpse of the structure and dynamics of global gene regulatory networks as well as proof-of-principle that this experimental system provides potentially universal tools and resources for gain-of-function analyses of transcription factors (TFs) in vitro and in vivo (Figure 7B). The ESC lines reported here will be freely

(B) Potential applications of TF-inducible ESC Bank, for which a proof-of-principle has been shown in this paper.
(C) A model for ESCs in undifferentiated state. Cdx2-target genes (red boxes, e.g., Hoxa7; Figure 5F) are not actively transcribed. Pluripotency-associated transcription factors (pTFs, e.g., POU5F1, SOX2, and NANOG) are present and bind to the regulatory regions of pTF-target genes (blue boxes), resulting in the active transcription of these genes. pTF-target genes include genes encoding pTFs, which thus form positive feedback loops and maintain the levels of pTFs.
(D) A model for ESCs in the early phase after the forced induction of Cdx2. CDX2 protein binds directly to the regulatory region of Cdx2-target genes, which begin to be actively transcribed. CDX2 suppresses the binding of pTFs to the regulatory regions of pTF-target genes and shut downs the transcription of pTF-target genes.

available to the research community, which could facilitate rapid accumulation and comparison of a variety of information on these standardized ESCs.

Some Notable Biological Findings from the Study

One of the striking observations is the difference between TFs in terms of their relative impact on the ES transcriptome (Figure 3). Some TFs can cause a huge perturbation, whereas others cause almost no change. Of particular interest, TFs with a high impact on the transcriptome include the four TFs (*Klf4*, *Pou5f1*, *Sox2*, and *Myc*) that have been successfully used to convert fibroblast cells into iPS cells (Takahashi and Yamanaka, 2006). This may indicate that formation of iPS cells correlates with the capacity of these TFs to perturb the transcriptome dramatically. Overall, the systematic study of TFs reveals important behaviors that would not be immediately evident in traditional phenotype-driven screens.

Interestingly, TFs that are known to function in late lineage specification (*Ascl1*, *Ascl2*, *Myod1*, *Sox9*, and *Sfpi1*) induced expression profiles that overlapped substantially with late-stage differentiated cells within 48 hr of TF overexpression (Figures 2C and 2D). The data suggest that undifferentiated ESCs may be in a permissive or susceptible state, in which forced induction of single TFs can make relevant changes in the transcriptome, regardless of whether usual TF partners or regulatory context are in place. These data seem to be consistent with the idea that the chromatin of ESCs is less restricted and more open, so changes of TF level alone can cause critical transcriptome changes (Meshorer and Misteli, 2006; Niwa et al., 2005; Silva and Smith, 2008). Whether or not this feature is specific to ESCs should be further tested by examining other cell types, such as fibroblasts.

There are nevertheless categories of genes whose response is modulated by structural or epigenetic cues. For example, genes upregulated by various TFs are enriched in genes with bivalent domains (H3K4me3+H3K27me3) and depleted in genes with binding sites of MYC, MYCN, E2F, and ZFX in promoters (Figure 7A). The same trend was observed for genes that were downregulated by these TFs, although the number of genes was much smaller (only up to 10% of all ~3000 genes with bivalent domains). Thus, genes with bivalent domains may form a dynamic network that can be rapidly up- or downregulated by changes in expression of TFs, whereas genes with binding sites of MYC, MYCN, E2F, and ZFX in promoters tend to maintain the status quo in their expression level responses to TF changes.

Modes of Gene Regulation by CDX2

By current thinking, the maintenance of expression of ES-specific genes is governed by the transcriptional network of pTFs (especially, *Pou5f1*, *Sox2*, and *Nanog*) (Boyer et al., 2005; Chen et al., 2008; Jaenisch and Young, 2008; Kim et al., 2008; Loh et al., 2006). These genes form a positive feedback loop to maintain their own expression levels while at the same time regulating other ESC-specific genes (Figure 7C). Our data imply that CDX2 induction causes the widespread downregulation of pTF-target genes. However, the data further indicate that CDX2 does not inactivate the transcription of pTFs by directly binding to their regulatory regions; rather, CDX2 interferes with the binding of pTFs to the regulatory regions of pTF-target genes (Figure 7D). Such an effect at the protein complex level would facilitate swift adaptation for ESCs to begin or commit to differentiation pathways. Because at least some of the pTFs (POU5F1, SOX2, and NANOG) are also pTF targets, the protein levels of POU5F1, SOX2, and NANOG would eventually decline, and the differentiation process would pass a point of no return and become irreversible.

It has been well established that NuRD is involved in gene transcriptional repression and chromatin remodeling (Denslow and Wade, 2007). Therefore, NuRD could play a major role in interfering with the bindings of pTFs to their targets. This notion is consistent with previous findings that ESCs lacking MBD3, one of the key components of the NuRD complex, cannot differentiate and remain undifferentiated even under differentiation-inducing conditions (Kaji et al., 2006). Indeed, our mass-spectrometry analysis of CDX2-associated protein complex identified the presence of MBD3. However, the exact mechanism of NuRD actions remains unknown. For example, CDX2 may recruit NuRD to the pTFs, resulting in inactivation of the pTFs. Alternatively, CDX2 may compete with the pTFs (particularly POU5F1; Niwa et al., 2005) for the binding of NuRD; it has been shown that pTFs interact with the NuRD in the absence of CDX2 in undifferentiated ESCs (Liang et al., 2008; Wang et al., 2006). The SALL4-CDX2 association revealed in this study may also be significant, because Sall4 is required to maintain ESC pluripotency and is important for early embryonic cell-fate decisions (Kim et al., 2008; Lim et al., 2008; Yang et al., 2008; Zhou et al., 2007).

Hints of the Mechanism of TF-Mediated Global Downregulation of ES-Specific Genes

We note at least three possible groups of TFs based on their ability to downregulate the pTF-target genes. A first group of TFs shows no strong effects on pTF-target genes. A second group of TFs is exemplified by CDX2 and includes *Esx1*, *Dlx3*, *Gata3*, *Ascl1*, *Sox9*, *Sfpi1*, *Mef2c*, *Nr2f2*, and *Myod1* (Figure 7A). These TFs are not expressed or are expressed at low levels in undifferentiated ESCs. The forced induction of these TFs downregulates direct target genes of pTFs, possibly through the same mechanism as CDX2. This is reasonable, because the differentiation of ESCs requires the downregulation of ES-specific genes, particularly genes involved in the maintenance of pluripotency of ESCs. A third group of TFs includes *Pou5f1*, *Sox2*, and *Klf4*, which are expressed relatively highly in undifferentiated ESCs, but still show a significant perturbation of the global transcriptome after a moderate increase in their expression levels (Figure 7A). Because it has been shown that pTFs (e.g., POU5F1, SOX2, and NANOG) form a protein complex (Liang et al., 2008; Wang et al., 2006), overexpression of one component could distort the stoichiometry of this complex, possibly resulting in a reduced amount of the effective protein complex and leading to the downregulation of direct target genes of pTFs. As an additional consistent observation, it has been reported that *Klf4* regulates downstream genes in a synergistic manner with *Pou5f1* and *Sox2* (Nakatake et al., 2006). This model rationalizes the fact that both Pou5f1 overexpression and repression can downregulate pTF-target genes, as indicated by the fact that both of them could cause a reduction in effective pTF complexes. This peculiar dose-response pattern of gene

expression has previously been called "squelching" (Scholer et al., 1991), but has also been called "bell-shaped" or "inverse bell-shaped" based on the behavior of a large number of genes regulated by *Pou5f1* in DNA microarray studies (Matoba et al., 2006).

Concluding Remarks

At the outset of this project, we reasoned that analyses of a large number of TFs might be requisite to explore global TF network(s) beyond the core transcriptional network of *Pou5f1*, *Sox2*, and *Nanog*. After analyzing large-scale transcriptome changes induced by 50 TFs, the core network remains one of the most conspicuous features of gene expression regulation in ESCs. These data thus reinforce the current paradigm that *Pou5f1*/*Sox2*/*Nanog* are the key regulators maintaining the pluripotent undifferentiated state of ESCs (Jaenisch and Young, 2008; Niwa, 2007; Silva and Smith, 2008). Furthermore, our data have revealed that suppression of pTF activity, which causes widespread downregulation of ES-specific genes, is an early step of TF-induced ESC differentiation. Further mining of the results reported here may provide additional inferences about relevant gene regulatory networks. Carrying out similar experiments for a larger number of regulatory proteins—and ideally all 1600–2000 TFs and additional signaling proteins—should give increasingly complete information to help infer the cybernetic networks in mammalian cells.

EXPERIMENTAL PROCEDURES

Establishment of TF-Inducible ESCs

MC1 mouse ESCs derived from 129S6/SvEvTac were cultured in DMEM with 15% FBS and LIF on feeder cells. Cells were electroporated with linearized pMWROSATcH and selected by hygromycin B. Knock-in for ROSA-TET locus in ES[MC1R(20)] cells was confirmed by Southern blotting. For exchange vectors, PCR-amplified ORFs were subcloned into pZhcSfi that was modified to express His6-FLAG-tagged protein and puromycin-resistant gene. ES[MC1R(20)] cells (passage 17) cultured on feeder cells were cotransfected with a sequence-verified exchange vector and pCAGGS-Cre and selected by puromycin in the presence of Dox. Isolated clones were tested for Venus expression, hygromycin B susceptibility, transgene RNA expression, genotyping for Cre-mediated integration, karyotyping, western blotting with FLAG antibody, and mycoplasma contamination (Supplemental Data). Further details about the ESC lines and how to obtain them can be found in the project website (http://esbank.nia.nih.gov/).

Transgene Induction and DNA Microarray

ESCs (passage 25) were cultured in the standard LIF$^+$ medium on a gelatin-coated dish through the experiments. Dox was removed through washing three times with PBS at 3 hr intervals and total RNA was isolated by TRIzol (Invitrogen) after 2 days. All procedures for each ES line were done in two independent replications (Figure S14). Cy3-CTP-labeled sample targets were prepared with total RNA by Low RNA Input Fluorescent Linear Amplification Kit (Agilent). A Cy5-CTP-labeled reference target was produced from mixture of Stratagene Universal Mouse Reference RNA and MC1 cells RNA. Targets were hybridized to the NIA Mouse 44K Microarray v3.0 (Agilent, design ID 015087) (Carter et al., 2005). Slides were scanned with Agilent DNA Microarray Scanner. All DNA microarray data are available in Table S1.

Immunoprecipitation

CDX2 complexes were purified with anti-FLAG M2 affinity gel and proteins were eluted by 3XFLAG peptide for mass spectrum analysis or Laemmli's sample buffer for western blotting.

ChIP and Sequencing Analysis

Cross-linked chromatin from *Cdx2*-expressed cells was fragmented by sonication and incubated with anti-FLAG M2 affinity gel. The immunoprecipitate was eluted and reverse cross-linked. Sequence sample preparation, Cluster generation on Cluster Station (Illumina), and sequencing by Genome Analyzer II (Illumina) were performed according to Illumina's manuals. See details in Supplemental Experimental Procedures. ChIP-Seq data are available at GEO/NCBI (GSE16375).

ACCESSION NUMBERS

All DNA microarray data are available at GEO/NCBI (http://www.ncbi.nlm.nih.gov/geo/; GSE16375) and at NIA Array Analysis software (Sharov et al., 2005; http://lgsun.grc.nia.nih.gov/ANOVA/).

SUPPLEMENTAL DATA

Supplemental Data include Supplemental Experimental Procedures, 16 figures, and 12 tables and can be found with this article online at http://www.cell.com/cell-stem-cell/supplemental/S1934-5909(09)00348-8.

ACKNOWLEDGMENTS

We thank Dr. Hitoshi Niwa for providing the Tet-inducible vector system and some cDNA clones and for discussion. This research was supported entirely by the Intramural Research Program of the NIH, National Institute on Aging.

Received: February 3, 2009
Revised: June 10, 2009
Accepted: July 22, 2009
Published: October 1, 2009

REFERENCES

Aiba, K., Nedorezov, T., Piao, Y., Nishiyama, A., Matoba, R., Sharova, L.V., Sharov, A.A., Yamanaka, S., Niwa, H., and Ko, M.S. (2009). Defining developmental potency and cell lineage trajectories by expression profiling of differentiating mouse embryonic stem cells. DNA Res. *16*, 73–80.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., et al. (2006). Chromatin signatures of pluripotent cell lines. Nat. Cell Biol. *8*, 532–538.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Pena-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., et al. (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell *133*, 1266–1276.

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

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

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

Carter, M.G., Sharov, A.A., VanBuren, V., Dudekula, D.B., Carmack, C.E., Nelson, C., and Ko, M.S. (2005). Transcript copy number estimation using a mouse whole-genome oligonucleotide microarray. Genome Biol. *6*, R61.

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

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways

with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

Cloonan, N., Forrest, A.R., Kolle, G., Gardiner, B.B., Faulkner, G.J., Brown, M.K., Taylor, D.F., Steptoe, A.L., Wani, S., Bethel, G., et al. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat. Methods *5*, 613–619.

Collins, F.S., Rossant, J., and Wurst, W. (2007). A mouse for all reasons. Cell *128*, 9–13.

Davidson, E.H. (2006). The Regulatory Genome: Gene Regulatory Networks in Development and Evolution (Burlington, MA: Academic Press).

Denslow, S.A., and Wade, P.A. (2007). The human Mi-2/NuRD complex and gene regulation. Oncogene *26*, 5433–5438.

Diez del Corral, R., and Storey, K.G. (2001). Markers in vertebrate neurogenesis. Nat. Rev. Neurosci. *2*, 835–839.

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

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.

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

Kaji, K., Caballero, I.M., MacLeod, R., Nichols, J., Wilson, V.A., and Hendrich, B. (2006). The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. Nat. Cell Biol. *8*, 285–292.

Kanamori, M., Konno, H., Osato, N., Kawai, J., Hayashizaki, Y., and Suzuki, H. (2004). A genome-wide and nonredundant mouse transcription factor database. Biochem. Biophys. Res. Commun. *322*, 787–793.

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

Klaffky, E., Williams, R., Yao, C.C., Ziober, B., Kramer, R., and Sutherland, A. (2001). Trophoblast-specific expression and function of the integrin alpha 7 subunit in the peri-implantation mouse embryo. Dev. Biol. *239*, 161–175.

Lange, A., Mills, R.E., Lange, C.J., Stewart, M., Devine, S.E., and Corbett, A.H. (2007). Classical nuclear localization signals: definition, function, and interaction with importin alpha. J. Biol. Chem. *282*, 5101–5105.

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

Liang, J., Wan, M., Zhang, Y., Gu, P., Xin, H., Jung, S.Y., Qin, J., Wong, J., Cooney, A.J., Liu, D., et al. (2008). Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. Nat. Cell Biol. *10*, 731–739.

Lim, C.Y., Tam, W.L., Zhang, J., Ang, H.S., Jia, H., Lipovich, L., Ng, H.H., Wei, C.L., Sung, W.K., Robson, P., et al. (2008). Sall4 regulates distinct transcription circuitries in different blastocyst-derived stem cell lineages. Cell Stem Cell *3*, 543–554.

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

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. USA *78*, 7634–7638.

Masui, S., Shimosato, D., Toyooka, Y., Yagi, R., Takahashi, K., and Niwa, H. (2005). An efficient system to establish multiple embryonic stem cell lines carrying an inducible expression unit. Nucleic Acids Res. *33*, e43.

Matoba, R., Niwa, H., Masui, S., Ohtsuka, S., Carter, M.G., Sharov, A.A., and Ko, M.S. (2006). Dissecting Oct3/4-regulated gene networks in embryonic stem cells by expression profiling. PLoS One *1*, e26.

Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat. Rev. Mol. Cell Biol. *7*, 540–546.

Messina, D.N., Glasscock, J., Gish, W., and Lovett, M. (2004). An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. Genome Res. *14*, 2041–2047.

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

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631–642.

Murry, C.E., and Keller, G. (2008). Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. Cell *132*, 661–680.

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

Natale, D.R., and Watson, A.J. (2002). Rac-1 and IQGAP are potential regulators of E-cadherin-catenin interactions during murine preimplantation development. Mech. Dev. *119* (*Suppl 1*), S21–S26.

Naya, F.J., and Olson, E. (1999). MEF2: A transcriptional target for signaling pathways controlling skeletal muscle growth and differentiation. Curr. Opin. Cell Biol. *11*, 683–688.

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

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

Niwa, H., 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.

Pritsker, M., Ford, N.R., Jenq, H.T., and Lemischka, I.R. (2006). Genomewide gain-of-function genetic screen identifies functionally active genes in mouse embryonic stem cells. Proc. Natl. Acad. Sci. USA *103*, 6946–6951.

Roh, T.Y., Cuddapah, S., Cui, K., and Zhao, K. (2006). The genomic landscape of histone modifications in human T cells. Proc. Natl. Acad. Sci. USA *103*, 15782–15787.

Scholer, H.R., Ciesiolka, T., and Gruss, P. (1991). A nexus between Oct-4 and E1A: Implications for gene regulation in embryonic stem cells. Cell *66*, 291–304.

Sharov, A.A., and Ko, M.S. (2009). Exhaustive search for over-represented DNA sequence motifs with CisFinder. DNA Res., in press.

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

Sharov, A.A., Masui, S., Sharova, L.V., Piao, Y., Aiba, K., Matoba, R., Xin, L., Niwa, H., and Ko, M.S. (2008). Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. BMC Genomics *9*, 269.

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

Simmons, D.G., and Cross, J.C. (2005). Determinants of trophoblast lineage and cell subtype specification in the mouse placenta. Dev. Biol. *284*, 12–24.

Skarnes, W.C., von Melchner, H., Wurst, W., Hicks, G., Nord, A.S., Cox, T., Young, S.G., Ruiz, P., Soriano, P., Tessier-Lavigne, M., et al. (2004). A public gene trap resource for mouse functional genomics. Nat. Genet. *36*, 543–544.

Solter, D. (2006). From teratocarcinomas to embryonic stem cells and beyond: A history of embryonic stem cell research. Nat. Rev. Genet. *7*, 319–327.

Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat. Genet. *21*, 70–71.

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.

Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinoso, L.M., Moqrich, A., et al. (2002). Large-scale analysis of the human and mouse transcriptomes. Proc. Natl. Acad. Sci. USA *99*, 4465–4470.

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.

Walker, E., Ohishi, M., Davey, R.E., Zhang, W., Cassar, P.A., Tanaka, T.S., Der, S.D., Morris, Q., Hughes, T.R., Zandstra, P.W., et al. (2007). Prediction and testing of novel transcriptional networks regulating embryonic stem cell self-renewal and commitment. Cell Stem Cell *1*, 71–86.

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

Yang, J., Chai, L., Fowles, T.C., Alipio, Z., Xu, D., Fink, L.M., Ward, D.C., and Ma, Y. (2008). Genome-wide analysis reveals Sall4 to be a major regulator of pluripotency in murine-embryonic stem cells. Proc. Natl. Acad. Sci. USA *105*, 19756–19761.

Yuan, H., Corbi, N., Basilico, C., and Dailey, L. (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes Dev. *9*, 2635–2645.

Zhou, Q., Chipperfield, H., Melton, D.A., and Wong, W.H. (2007). A gene regulatory network in mouse embryonic stem cells. Proc. Natl. Acad. Sci. USA *104*, 16438–16443.