Cell

Resource

Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells

Xi Chen,[1,2,6] Han Xu,[3,6] Ping Yuan,[1] Fang Fang,[1,2] Mikael Huss,[4] Vinsensius B. Vega,[3] Eleanor Wong,[5] Yuriy L. Orlov,[4] Weiwei Zhang,[1,2] Jianming Jiang,[1,2] Yuin-Han Loh,[1,2] Hock Chuan Yeo,[4] Zhen Xuan Yeo,[4] Vipin Narang,[3] Kunde Ramamoorthy Govindarajan,[3] Bernard Leong,[3] Atif Shahab,[3] Yijun Ruan,[5] Guillaume Bourque,[3] Wing-Kin Sung,[3] Neil D. Clarke,[4] Chia-Lin Wei,[5,*] and Huck-Hui Ng[1,2,*]

[1]Gene Regulation Laboratory, Genome Institute of Singapore, Singapore 138672

[2]Department of Biological Sciences, National University of Singapore, Singapore 117543

[3]Computational and Mathematical Biology

[4]Computational and Systems Biology Group

[5]Genome Technology and Biology Group

Genome Institute of Singapore, Singapore 138672

[6]These authors contributed equally to this work

*Correspondence: weicl@gis.a-star.edu.sg (C.-L.W.), nghh@gis.a-star.edu.sg (H.-H.N.)

DOI 10.1016/j.cell.2008.04.043

SUMMARY

Transcription factors (TFs) and their specific interactions with targets are crucial for specifying gene-expression programs. To gain insights into the transcriptional regulatory networks in embryonic stem (ES) cells, we use chromatin immunoprecipitation coupled with ultra-high-throughput DNA sequencing (ChIP-seq) to map the locations of 13 sequence-specific TFs (Nanog, Oct4, STAT3, Smad1, Sox2, Zfx, c-Myc, n-Myc, Klf4, Esrrb, Tcfcp2l1, E2f1, and CTCF) and 2 transcription regulators (p300 and Suz12). These factors are known to play different roles in ES-cell biology as components of the LIF and BMP signaling pathways, self-renewal regulators, and key reprogramming factors. Our study provides insights into the integration of the signaling pathways into the ES-cell-specific transcription circuitries. Intriguingly, we find specific genomic regions extensively targeted by different TFs. Collectively, the comprehensive mapping of TF-binding sites identifies important features of the transcriptional regulatory networks that define ES-cell identity.

INTRODUCTION

Embryonic stem (ES) cells are derived from early preimplantation embryos, and they can be maintained for extended periods of time in culture through self-renewing division (Smith, 2001). These cells are pluripotent, as they retain the ability to differentiate into many, and perhaps all, cell lineages. The ability to generate transgenic mouse ES cells through homologous recombination has revolutionized biological research through the creation of genetically altered animals (Thomas and Capecchi, 1986). In addition, human ES cells can potentially serve as an inexhaustible source of cells for the derivation of clinically useful cells for regenerative medicine and cell-based therapy.

Mouse ES cells were first isolated in 1981 from mouse blastocysts (Smith, 2001). Maintenance of the self-renewing state of mouse ES cells requires the cytokine leukemia inhibitory factor (LIF). The binding of LIF to its receptor activates STAT3 through phosphorylation (Niwa et al., 1998). LIF alone, however, is not sufficient to maintain ES cells, as their maintenance requires the presence of fetal calf serum. Bone morphogenetic proteins (BMPs) appear to be key serum-derived factors that act in conjunction with LIF to enhance the self-renewal and pluripotency of mouse ES cells (Ying et al., 2003). The binding of BMP4 to its receptors triggers the phosphorylation of Smad1 and activates the expression of members of the *Id* (inhibitor of differentiation) gene family. As ES cells overexpressing *Id*s can self-renew in the absence of BMP4, it is proposed that induction of *Id* expression is the critical contribution of the BMP/Smad pathway. Hence, the LIF and BMP signaling pathways play a central role in the maintenance of a pluripotential stem cell phenotype.

Besides these signaling pathways, which sense the presence of extrinsic growth factors in the environment, intrinsic factors such as transcription factors (TFs) are also essential for specifying the undifferentiated state of ES cells. Oct4, encoded by *Pou5f1*, is a POU domain-containing TF known to be essential to ES cells and early embryonic development (Boiani and Schöler, 2005; Nichols et al., 1998; Smith, 2001). Oct4 interacts with Sox2 (an HMG-containing TF), and genome-wide mapping of OCT4 and SOX2 sites in human ES cells shows that they cotarget multiple genes (Boyer et al., 2005). The *cis*-regulatory element to which the Sox2-Oct4 complex is bound consists of neighboring *sox* (5′-CATTGTA-3′) and *oct* (5′-ATGCAAAT-3′) elements (Loh et al., 2006). Recent works indicate that Oct4 and Sox2, along with c-Myc and Klf4, are sufficient for reprogramming fibroblasts to induced pluripotent stem cells (iPS),

which are functionally similar to ES cells (Maherali et al., 2007; Okita et al., 2007; Takahashi and Yamanaka, 2006; Wernig et al., 2007). Hence, these TFs can exert a dominant role in reconstructing the transcriptional regulatory network of ES cells. A third well-studied TF in ES cells is Nanog. Nanog is a homeodomain-containing TF that can sustain pluripotency in ES cells even in the absence of LIF (Chambers et al., 2003; Mitsui et al., 2003). Other transcriptional regulators are required as well to maintain ES cells. Recent work has begun to identify new components of the transcriptional regulatory network required for the maintenance of pluripotency. Through genetic studies, Esrrb and Zfx have been shown to regulate the self-renewal of ES cells (Ivanova et al., 2006; Loh et al., 2006; Galan-Caridad et al., 2007).

Despite the critical roles of transcriptional regulators in the maintenance of mouse ES cells, detailed knowledge of their in vivo targets is lacking. The targets of the downstream effectors of key signaling pathways are poorly studied, and the targets of many of the TFs in ES cells have not been defined. The manner in which the different transcriptional circuitries are integrated is also not clear. Elucidation of the transcriptional regulatory networks that are operating in ES cells is fundamental for understanding the molecular nature of pluripotency, self-renewal, and reprogramming.

In this study, we use chromatin immunoprecipitation coupled with massively parallel short-tag-based sequencing (ChIP-seq) (Barski et al., 2007; Johnson et al., 2007; Mikkelsen et al., 2007; Robertson et al., 2007) to map the in vivo binding loci for 13 sequence-specific TFs and 2 transcription coregulators in living mouse ES cells. Intriguingly, these TFs are wired to the ES-cell genome in two major ways. The first cluster includes Nanog, Oct4, Sox2, Smad1, and STAT3. The second cluster consists of c-Myc, n-Myc, Zfx, and E2f1. The coactivator p300 is predominantly recruited to dense binding loci involving proteins found in the first type of cluster. Our analysis also reveals that highly dense binding loci involving these factors have characteristic features of enhanceosomes. ES-cell-specific gene expression is associated with binding of many of the factors studied. Based on these associations between binding and expression, we have constructed a transcriptional regulatory network model that integrates the two key signaling pathways with the intrinsic factors in ES cells.

RESULTS

Mapping of In Vivo Binding Sites of 13 Transcription Factors by Using ChIP-seq

Whole-genome binding sites of 13 sequence-specific transcription factors (TFs) were profiled in mouse ES cells by using the ChIP-seq approach (Barski et al., 2007; Johnson et al., 2007; Mikkelsen et al., 2007; Robertson et al., 2007). Nanog, Oct4, Sox2, Esrrb, and Zfx are known regulators of pluripotency and/or self-renewal. Smad1 and STAT3 are key components of the signaling pathways mediated by BMP and LIF, respectively. Tcfcp2l1 is preferentially upregulated in ES cells but has uncharacterized DNA-binding properties and function (Ivanova et al., 2006). E2F1 is best known for its role in regulating cell-cycle progression and has also been shown to associate extensively with promoter regions (Bieda et al., 2006). Klf4 and Myc TFs are reprogramming factors that are also implicated in the maintenance of the undifferentiated state of ES cells (Cartwright et al., 2005; Jiang et al., 2008). CTCF is required for transcriptional insulation (Kim et al., 2007). Through mapping the binding sites of these 13 TFs, we seek to investigate the binding behavior of these factors and to uncover insights into how they are wired in the ES-cell genome. Here, chromatin immunoprecipitation (ChIP) with specific antibodies against these TFs was used to enrich the DNA fragments bound by these TFs, followed by direct ultra-high-throughput sequencing with the Solexa Genome Analyzer platform. Genomic regions defined by multiple overlapping DNA fragments derived from the ChIP enrichments were considered as putative binding sites. We used Monte Carlo simulations to determine the minimal number of overlapping ChIP fragment reads required to distinguish true binding from nonspecific, randomly expected overlaps. Regions with overlapping ChIP DNA counts higher than the threshold were further filtered by removing peaks that were also found in the negative control (anti-GFP ChIP) library. ChIP-qPCR validations were carried out on randomly selected sites with different ''intensities'' (i.e., ChIP tag counts within the defined overlap region) to further refine the threshold used. Based on the ChIP-qPCR analyses, we determined that the specificity of binding site determination was greater than 95% for the majority of the libraries. We identified between 1,126 and 39,609 TF-binding sites (TFBSs) for the 13 factors. As examples, the binding profiles for all 13 factors at the *Pou5f1 (Oct4)* and *Nanog* gene loci are shown in Figure 1. The specificity of the antibodies, the depth and coverage of sequencing, ChIP-qPCR validation data, and detailed bioinformatics analyses can be found in the Supplemental Data available with this article online.

Motif Analyses of TFBSs

To determine the in vivo sequence specificity of these TFs, we derived the consensus sequence motifs by using a de novo motif-discovery algorithm (as described by Loh et al., 2006). Sequences (±100 bp) from the top 500 binding peaks were selected from each factor, repeats were masked, and the program Weeder (Pavesi et al., 2001) was used to find overrepresented sequences. Because of the high resolution in defining the binding sites offered by the high sequence depth coverage, overrepresented motifs could be uncovered from 12 of the 13 factors (excluding E2f1) (Figure 2). Consistent with our previous study, we obtained a *sox-oct* composite element consisting of a Sox2-binding site consensus (5′-CATTGTT-3′) and a canonical Oct4-binding sequence (5-′ATGCAAAT-3′) adjacent to one another from both the Oct4 and Sox2 data sets (Loh et al., 2006). The presence of a common motif suggests that the Sox2 and Oct4 heterodimer is the functional binding unit. Interestingly, the de novo-predicted matrices for Nanog- and Smad1-bound sequences resemble the *sox-oct* joint motif. This reflects the frequent co-binding of Nanog and Smad1 with Sox2 and Oct4. It is noteworthy that the Nanog motif reported previously (Loh et al., 2006) can be found by using another motif-discovery algorithm, NestedMICA (see Figure S5 available online). The binding consensus sequences identified for Klf4, Esrrb, CTCF, c-Myc, n-Myc, STAT3, and Zfx are closely related to the binding sequences reported previously (Ehret et al., 2001; Galan-Caridad

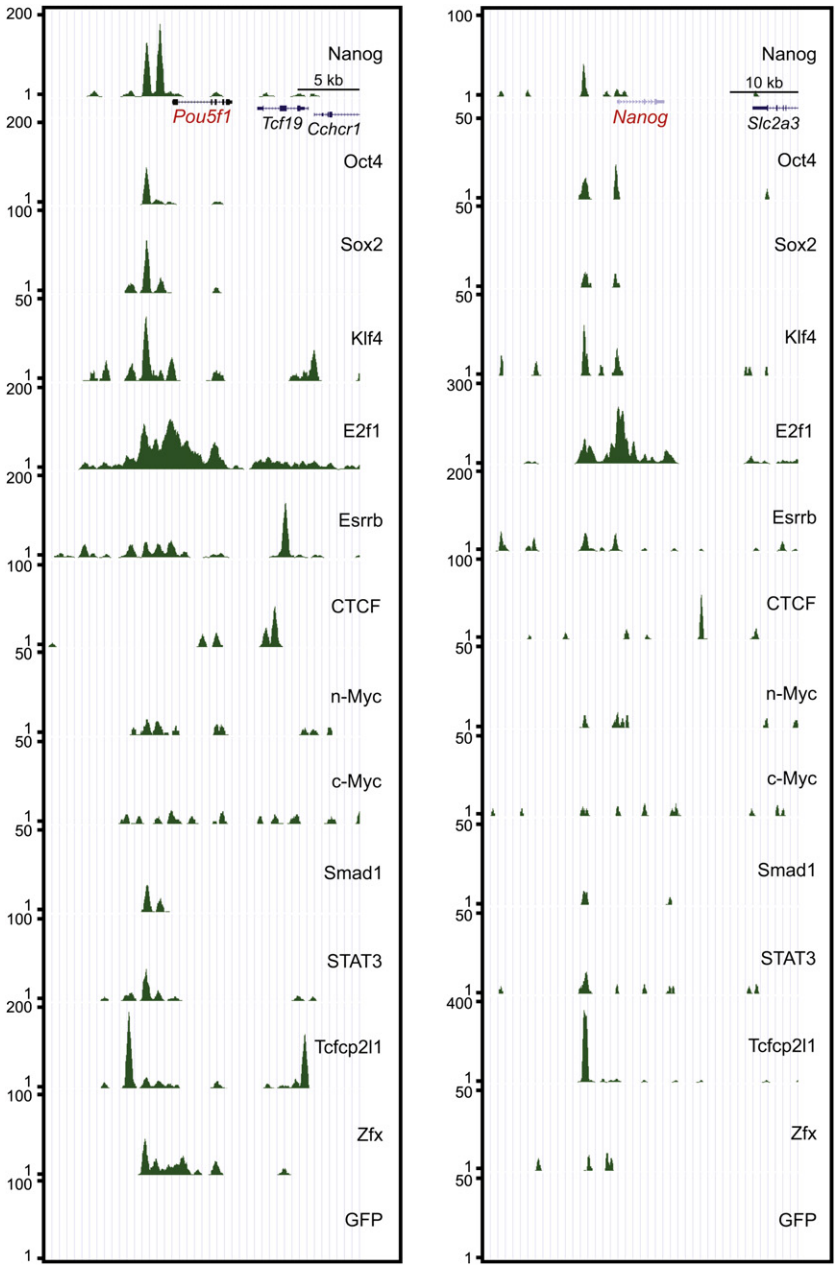

Figure 1. Genome-Wide Mapping of 13 Factors in ES Cells by Using ChIP-seq Technology

TFBS profiles for the sequence-specific transcription factors and mock ChIP control at the *Pou5f1* and *Nanog* gene loci are shown.

et al., 2007; Kaczynski et al., 2003; Kim et al., 2007; Pettersson et al., 1996; Zeller et al., 2006). Hence, we showed that sequence motifs can be identified from the in vivo-bound sites.

A Subset of Multiple Transcription-Factor-Binding Loci as ES-Cell Enhanceosomes

Upon close examination of the binding profiles from these 13 TFs, we found that a subset of binding sites was bound by many of these TFs. To investigate their biological relevance, we first determined the significance of such enrichments of TFBSs (see Supplemental Experimental Procedures). Peak sites within 100 bp were iteratively clustered to define multiple transcription factor-binding loci (MTL) (see Table S8). The number of these MTL, plotted as a function of the number of different TFs in the MTL, is shown in Figure 3A. Loci bound by 4 or more TFs are highly significant ($p < 0.001$, Figure 3A), and there is a total of 3583 such MTL. Of these, 1440 loci (40.2%) were found in the intergenic regions, and the remaining loci were spread between promoter regions (1334 loci, 37.2%) and within gene regions (809 loci, 22.6%). Less than 20% of the clusters with 7 or more TFs are found at promoter regions (yellow columns, Figure 3B), compared with 40% of the clusters that have fewer than 5 TFs. Hence, the co-occurrence of TFBSs within the MTL is not mainly due to their occurrence at promoters.

To further dissect the composition of the MTL, we examined the co-occupancy of different factors found in the 3583 MTL. Among the 13 TFs, Nanog, Sox2, Oct4, Smad1, and STAT3 (blue box, Figure 4A) tend to co-occur quite often, as do members of a second, distinct group comprised of n-Myc, c-Myc, E2f1, and Zfx (green box, Figure 4A). In addition to these two high-level groupings of TFs, we find it useful to define four groups of MTL based on the presence or absence of binding sites of (i) Oct4, Sox2, or Nanog and (ii) c-Myc or n-Myc. The Nanog-Oct4-Sox2 clusters (binding observed for Nanog, Oct4, or Sox2, but not for n-Myc or c-Myc) constitute 43.4% of the 3583 MTL (orange sector, Figure 4B). The Myc-specific clusters (n-Myc or c-Myc, but not Nanog, Oct4, or Sox2) make up 32.9% of the MTL (light-blue sector, Figure 4B).

Consistent with the pairwise co-occurrence shown in Figure 4A, 87.4% of Smad1- and 56.8% of STAT3-binding sites within MTL were associated with the Nanog-Oct4-Sox2-specific MTL (orange sector, Figure 4C). This indicates that Smad1 and STAT3 share many common target sites with Nanog, Oct4, and Sox2 and reflects a point of convergence of the two key signaling pathways (via Smad1 and STAT3) with the core circuitry defined by Nanog, Oct4, and Sox2 (Boyer et al., 2005). This is consistent with a previous study showing the link between Nanog and the LIF pathway (Chambers et al., 2003). A total of 56.9% of Esrrb- and 41.9% of Klf4-binding sites within MTL were found in the Nanog-Oct4-Sox2-specific MTL. Indeed, Esrrb has been shown to reside in the same complex as Nanog (Wang

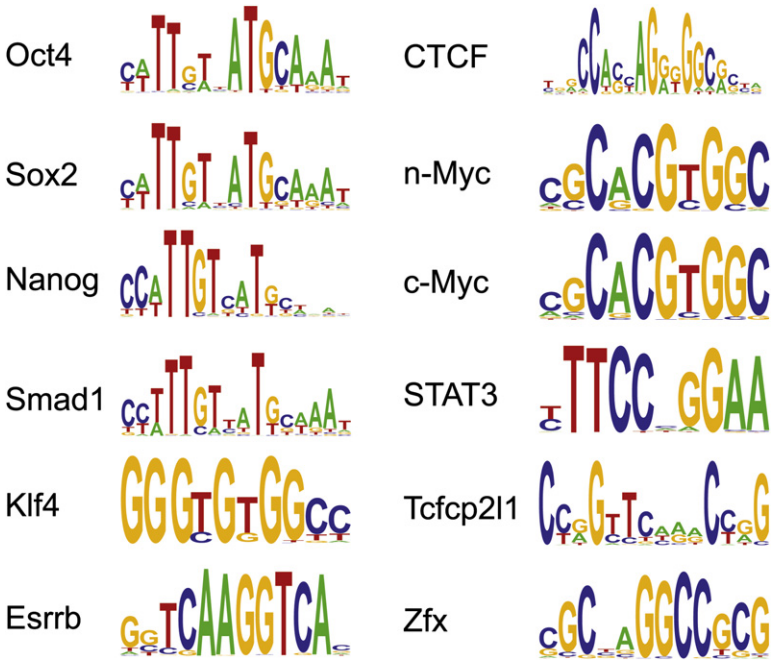

Figure 2. Identification of Enriched Motifs by Using a De Novo Approach

Matrices predicted by the de novo motif-discovery algorithm Weeder.

et al., 2006). In contrast, the co-occurrence of Zfx, CTCF, and E2f1 was skewed toward the Myc-specific cluster (light-blue sector, Figure 4C).

As the majority of the Nanog-Oct4-Sox2-specific MTL are found outside of promoter regions (91.2%), we assayed genomic sequences from this MTL cluster type for enhancer activity. A total of 25 genomic fragments from the Nanog-Oct4-Sox2 cluster and 8 genomic fragments from the Myc cluster were cloned downstream of a luciferase reporter. The genomic fragment was placed 2 kb away from the minimal *Pou5f1* promoter used to drive the *luciferase* gene. These constructs were transfected into ES cells and 293T cells, and luciferase activity was measured. Remarkably, all 25 constructs with genomic fragments spanning Nanog-Oct4-Sox2 clusters showed robust ES-cell-specific enhancer activity (Figure 4D). A total of 21 of the constructs were even more active than a *Nanog* enhancer positive control. In contrast, the control constructs with genomic fragments from the Myc cluster were either not active or showed very weak ES-cell-specific enhancer activity.

Combinatorial binding of TFs to enhancers can impart transcriptional synergy (Struhl, 2001). To address the relationships between Oct4, Smad1, and STAT3, we perturbed the binding of these factors through RNAi or growth factor withdrawal. Depletion of Oct4 led to a reduction in Smad1 and STAT3 binding (Figures 4E and 4F). The alteration of Smad1 and STAT3 binding occurs specifically on Oct4, Smad1, and STAT3 co-bound sites and was not due to a reduction in Smad1 and STAT3 levels (data not shown). We also performed the reciprocal experiments of withdrawing LIF or BMP4 from the culture media. LIF withdrawal reduced STAT3 binding to its targets, whereas BMP4 withdrawal reduced Smad1 binding to its targets (Figures S3L and S3M). Perturbation of the two signaling pathways, however, did not affect the binding of Oct4 (Figure 4G). This indicates that Oct4 is pivotal in stabilizing the nucleoprotein complex and establishes a hierarchy of regulatory interactions between Oct4, STAT3, and Smad1. The mechanism for Oct4-dependent STAT3 and Smad1 binding is not clear. It is possible that Oct4 may interact with STAT3 or Smad1 to facilitate their interactions with chromatin.

In summary, through the global binding sites of TF profiling, we uncovered over 3000 genomic regions densely bound by TFs. The Nanog-Oct4-Sox2 cluster exhibits features of enhanceosomes by enhancing transcription from a distance and shows extensive co-occupancy with Smad1 and STAT3. Importantly, we showed that Oct4 is required for the binding of Smad1 and STAT3, suggesting that Oct4 plays a pivotal role in stabilizing the TF complex.

p300 Is Recruited to the Nanog-Oct4-Sox2 Cluster

To further assign functionality to the MTL, we determined the locations of transcriptional coactivator p300 by using ChIP-seq (Figures S6 and S7; Table S11). p300 is a histone acetyltransferase commonly found at enhancer regions (Heintzman et al., 2007; Ogryzko et al., 1996). Genome-wide mapping of a chromatin regulator like p300 has the potential to reveal the DNA-binding factor(s) responsible for recruiting the regulator to specific sites in the genome (Birney et al., 2007). We also profiled the locations of another chromatin regulator, Suz12, to serve as a control (Figures S6 and S7; Table S11).

Strikingly, p300 was found to co-occur with the Nanog-Oct4-Sox2 cluster type (Figure 5A). Most p300-binding sites are associated with 3–6 other TFs, up to as many as 9 in one case (Figure 5B). The composition of these p300-containing clusters is highly diverse, but, typically, they include one or more of the factors Nanog, Oct4, or Sox2, followed, at lower probability, by Smad1, Esrrb, Klf4, Tcfcp2l1, and STAT3 (Figure 5B). In contrast to p300, Suz12 did not show strong association with any of the 13 TFs (data not shown). Using the de novo motif-discovery algorithm Weeder, we were able to uncover an enriched motif from p300-enriched sequences that resembles the *sox-oct* composite element (Figure 5C). The association of p300 with Nanog-Oct4-Sox2 clusters was validated for 12 sites by using ChIP-qPCR. As controls, 12 Myc-bound MTL that lack p300 as determined by the ChIP-seq assay also lack p300 as determined by ChIP-qPCR. (Figure S8). These data suggest that Oct4, Sox2, and Nanog are recruiting p300 to the genomic sites. To test this hypothesis, we depleted Oct4, Sox2, or Nanog by RNAi and checked for p300 binding. Our ChIP result showed that p300 binding was reduced by Oct4, Sox2, or Nanog depletion

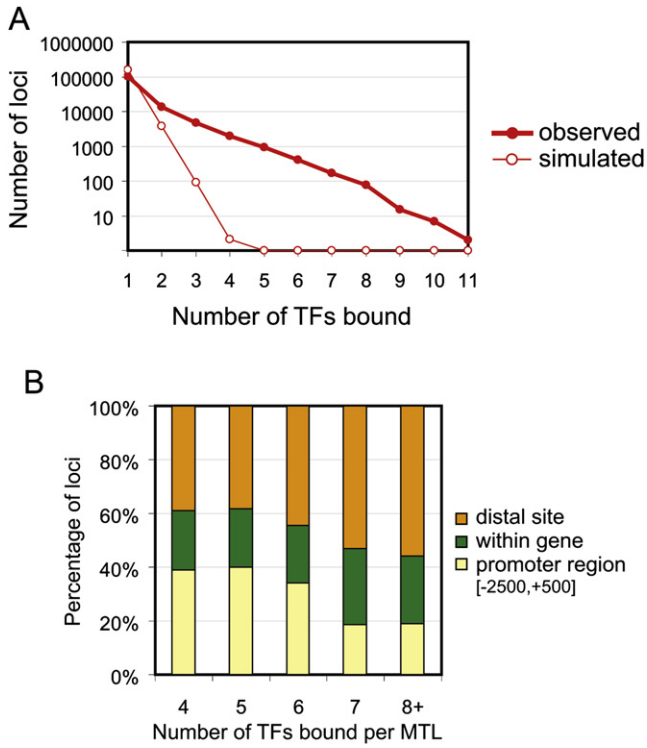

Figure 3. Multiple Transcription Factor-Binding Loci

(A) Plot of the number of transcription factors (TFs) bound per co-bound locus. The distribution of randomly occurring co-bound loci is obtained by simulation. (B) Distribution of clusters with different numbers of co-bound TFs. Promoter regions are defined as sequences 2500 bp upstream and 500 bp downstream of TSS (Heintzman et al., 2007).

(Figure 5D). Previous work has shown that c-Myc interacts with p300 and mediates the recruitment of p300 to the *hTERT* promoter (Faiola et al., 2005). In ES cells, we did not observe global recruitment of p300 to Myc sites. Depletion of c-Myc by RNAi, however, did not affect p300 recruitment to these sites (data not shown). The data suggest that p300 could be a general factor that is recruited to enhancers (Heintzman et al., 2007), and we conclude that p300 recruitment is promoted by Oct4, Sox2, and Nanog.

Combinatorial Binding of Transcription Factors Is Associated with ES-Cell-Specific Expression

Next, we sought to establish the correlation between TF occupancies and gene expression. A commonly employed approach for assigning target genes to a TF is to associate TFBSs with genes based on proximity. However, the relevant threshold for proximity could be different for different TFs. For that reason, we developed an approach to cluster genes based on TF-binding data (see the Supplemental Experimental Procedures). For each pair of TF and gene, we assigned an association score based on the genomic location of the binding site that is closest to the transcription start site (TSS). This association score is based on the distribution of the nearest site-to-TSS distances in the genome and is thus different for, and characteristic of, each TF. A higher score implies a higher chance of the gene being the target of the TF (Figure S9). This avoids an arbitrary threshold. Based on the association scores for all TFs, we performed k-means clustering to define five classes of genes that are associated with a similar set of TFs (Figure 6A).

Class I genes are enriched in binding sites for Nanog, Oct4, Sox2, Smad1, and STAT3 (Figure 6B). Class II genes are bound heavily by c-Myc and n-Myc. Class III genes show enrichment (more than 1-fold) in binding by n-Myc, Klf4, Esrrb, Tcfcp2l1, Zfx, and E2f1. Class IV genes are highly enriched in Suz12-bound genes, whereas class V genes are deficient in all TFs. In total, 48% of genes are deficient in TF binding by the 13 TFs (class IV and class V). We note that E2f1 and Suz12 localization is essentially mutually exclusive, suggesting that Polycomb repressor complexes inhibit the binding of E2f1 to its target sites. Next, we analyzed the expression level of the genes found in each class by using a published microarray data set for undifferentiated ES cells (Ivanova et al., 2006), and we found that the expression level is highest for genes in class II, followed by genes in class I and class III. Genes in class IV and class V are either not expressed or are expressed at a very low level in ES cells (Figure 6C; Supplemental Results and Discussion).

To further characterize the gene-expression profiles of each class, we used a microarray data set that interrogated the transcriptome dynamics of retinoic acid (RA)-induced differentiation (Ivanova et al., 2006). The genes in this data set were divided into three categories (see Supplemental Experimental Procedures): genes upregulated in ES cells, nondifferentially expressed genes, and genes downregulated in ES cells. Class I genes constitute less than 10% of the nondifferentially expressed genes and genes downregulated in ES cells (compare the red columns in Figure 6D). This compares to 24% of the genes in the upregulated category. The percentage of class II genes is only 12% among nondifferentially expressed genes, but 36% in the upregulated set (compare the blue columns in Figure 6D). Hence, class I and class II genes are 2.7 ($p = 8.14E-52$)- and 2.9 ($p = 1.28E-91$)-fold enriched, respectively, in genes upregulated in ES cells. In contrast, class IV and class V genes are underrepresented in this set. Class III is slightly enriched in genes that are downregulated in ES cells, but not enriched in genes that are preferentially upregulated in ES cells. As a validation, we compared the five classes with another independent microarray data set (Zhou et al., 2007), and similar results were obtained (Figure S10). In summary, our global analysis showed that 60% of genes upregulated in ES cells are from class I and class II. Most importantly, the results demonstrate that gene clustering based on TF occupancies has the potential to predict ES-cell-specific gene expression. This suggests that the TF-binding patterns of these two groups are relevant in specifying ES-cell-specific expression. In summary, we demonstrate that combinatorial binding patterns of TFs have greater predictive power for ES-cell-specific expression.

Constructing a Regulatory Network Defining ES-Cell-Specific Expression

The self-renewing state of undifferentiated ES cells is characterized by the expression of genes specifically upregulated in this cell type. We sought to construct a regulatory network that specifies ES-cell-specific expression by using binding sites of

transcriptional regulators under the undifferentiated state. In order to infer regulatory interactions, we made use of published expression profiling data that compared undifferentiated with differentiating ES cells. The rationale is that 9 (Nanog, Oct4, Sox2, Klf4, n-Myc, c-Myc, Esrrb, Zfx, Tcfcp2l1) out of the 13 TFs we studied are known to be downregulated upon differentiation or in differentiated cell types (Ivanova et al., 2002, 2006; Ramalho-Santos et al., 2002). Two sets of published experiments were used to define genes that are differentially expressed during differentiation (Ivanova et al., 2006; Zhou et al., 2007). The use of two independently generated data sets minimizes biases in gene-expression differences that are due to different ways of differentiating ES cells.

The regulatory interaction between a TF and its target gene is first defined for an individual TF by intersecting the rank-ordered bound genes (based on the total number of sequence tags associated with binding site peaks) and the rank-ordered differentially expressed genes (see Supplemental Data for method). The thresholds for defining top-ranked genes in the two lists were determined empirically by maximizing the number of genes in the intersection subject to two constraints: there had to be at least twice as many genes in the intersection as the number expected by chance, and the null model (that there are no genes in excess) had to be rejected with $p < 10^{-3}$. This method allows us to make use of the unique features of our binding data sets (signal intensity and unbiased survey) and avoid the use of a single cutoff for all data sets.

A network model based on the 13 TFs as depicted in Figure 7 reveals both anticipated and unanticipated aspects of the relationships between these TFs. Consistent with previous studies, this model shows regulatory feedback loops for Oct4, Sox2, and Nanog (Boyer et al., 2005; Chew et al., 2005; Loh et al., 2006). An interesting feature of this network is the interconnectivity among 11 of the 13 TFs being profiled.

DISCUSSION

The Repertoire of Binding Sites in the Mammalian Genome Revealed by Global Mapping of Transcription-Factor-Binding Sites

Ultra-high-throughput sequencing technology through massively parallel short-read sequencing has recently been developed for mapping TFBSs and histone-modification profiles in mammalian cells (Barski et al., 2007; Johnson et al., 2007; Mikkelsen et al., 2007; Robertson et al., 2007). In this study, we performed a large-scale mapping study of multiple TFs in mammalian cells.

Genome-wide mapping studies reveal abundant binding sites for different TFs in mammalian cells (Bieda et al., 2006; Birney et al., 2007; Cawley et al., 2004). Although advances in mapping technologies allow for comprehensive and unbiased disclosure of the repertoire of binding sites, it is difficult to determine which sites are functional regulatory elements that influence transcription. It is possible that a sizeable fraction of these binding sites are nonfunctional and are the consequence of biological noise (Struhl, 2007). It is also important to note that ChIP experiments may capture indirect TF-DNA interactions through protein-protein interaction. The strength of our study lies in the concurrent survey of the locations of multiple TFs in a single cell type. Our data show that there are genomic regions extensively co-occupied by TFs (TF colocalization hotspots). These are more likely to represent functionally important sites.

It is also of interest to note that a recent study has profiled the binding sites of nine TFs in mouse ES cells (Kim et al., 2008). The major distinctions between the two studies are the coverage of the survey and the method for capturing the TF-DNA complexes. This study surveyed the whole-genome binding sites of TFs by using antibodies recognizing endogenous proteins, whereas Kim and colleagues used promoter DNA microarrays to investigate the occupancies of biotin-tagged proteins. The two studies also focused on a slightly different set of TFs. For example, this study analyzed the locations for the downstream effectors of key signaling pathways (Smad1, STAT3), self-renewal regulators (Zfx, Esrrb), insulator-binding proteins (CTCF), and two transcriptional coregulators (p300, Suz12). The ChIP-seq method reveals a richer repertoire of binding sites, as it can identify binding sites found at repetitive elements. Future work that integrates these data sets should provide insights into the usage of sites identified by the two studies.

ES-Cell-Specific Enhanceosomes

An enhanceosome is a nucleoprotein complex composed of distinct sets of TFs bound directly or indirectly to enhancer DNA (Thanos and Maniatis, 1995). The density of TFs occurring on this short segment of DNA is high compared to more "modular" enhancers that have less dense binding clusters occurring over a longer segment of genomic DNA (Arnosti and Kulkarni, 2005). The virus-inducible enhancer of the interferon-β gene (*IFN-β*) is a prototypical enhanceosome. This 55 bp enhancer is bound by the p50 and p65 subunits of NF-kB, ATF-2, IRF-3, IRF-7, c-Jun, and the architectural TF HMGA. An atomic model for the complex of eight of these factors on the DNA has been constructed based on three crystal structures (Panne et al., 2007). The basis for cooperativity is unlikely to be mediated through protein-protein interactions, as these structures reveal limited contact between the TFs. It is proposed that the binding of these eight TFs on a composite DNA interface creates a continuous surface for recruiting coactivators such as p300 (Merika et al., 1998; Wathelet et al., 1998).

Our genome-wide mapping study reveals genomic regions with features of enhanceosomes. First, the binding sites are densely clustered within relatively compact genomic segments. It is of interest to note that the densest binding locus we identified is the distal enhancer of *Pou5f1*. This region (Chew et al., 2005) was bound by 11 TFs. Second, we showed that 25 of these genomic regions act as enhancers when placed downstream of the luciferase reporter. Third, they are associated with the H3K4me3 mark, which is one of the signatures of active genomic regions. Fourth, our p300 ChIP-seq analysis revealed on a global scale the recruitment of this coactivator to the Nanog-Oct4-Sox2 cluster, but not the Myc cluster. Importantly, we showed that the recruitment of p300 is dependent on Oct4, Sox2, and Nanog.

In higher eukaryotes, transcriptional enhancers play important roles in integrating multiple signaling pathways to achieve activation of specific genes. By profiling multiple TFBSs on the

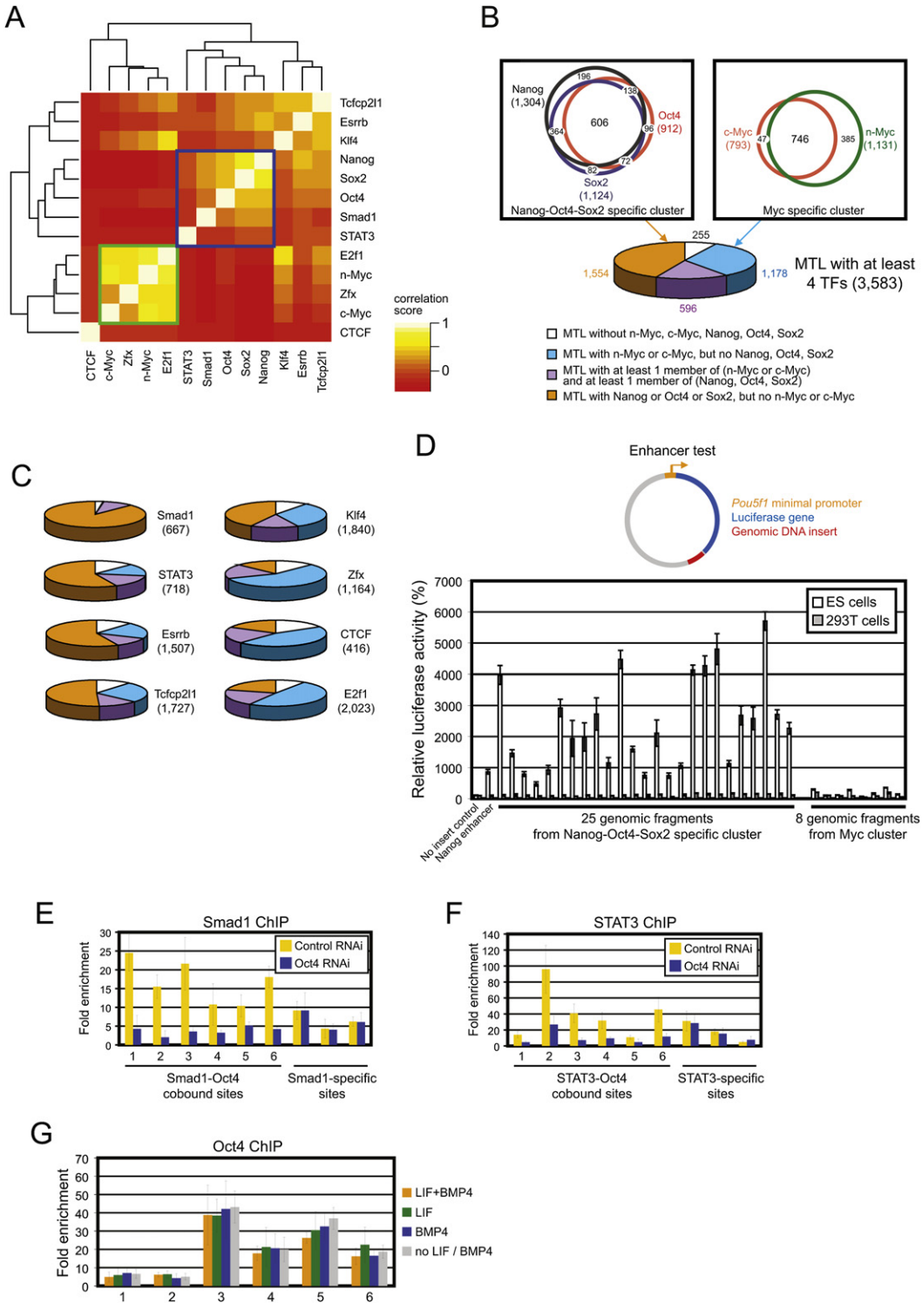

Figure 4. Multiple Transcription Factor-Binding Loci Associated with Nanog, Oct4, Sox2, Smad1, and STAT3 as ES-Cell Enhanceosomes

(A) Co-occurrence of transcription factor (TF) groups within MTL. Colors in the heat map reflect the colocalization frequency of each pair of TFs in MTL (yellow means more frequently colocalized, red means less). TFs have been clustered along both axes based on the similarity in their colocalization with other factors. (B) Dissection of the TF makeup within MTL. Two major clusters exist within the 3583 MTL. The first group (orange sector) consists of Oct4, Nanog, or Sox2, but not n-Myc and c-Myc. The second group (light-blue sector) consists of n-Myc or c-Myc, but not Oct4, Nanog, and Sox2. The purple sector is a mixture of the first two groups (orange and light-blue sectors).

whole-genome scale, we discovered extensive colocalization of multiple TFs on selective sites in the ES-cell genome.

Wiring of the ES-Cell Genome

LIF has long been known to be essential for the derivation or maintenance of mouse ES cells (Smith, 2001). Beside LIF, other factors in fetal calf serum could be essential for the self-renewal of mouse ES cells. Smith and coworkers have identified BMPs as growth factors that work in conjunction with LIF to promote self-renewal (Ying et al., 2003). The addition of BMP4 to chemically defined media leads to the phosphorylation of Smad1 in ES cells. As constitutive expression of the *Id* genes bypasses the BMP4 or fetal calf serum requirement for maintenance of ES cells, the *Id* genes are implicated as downstream targets of the BMP/Smad signaling pathway (Ying et al., 2003). ES cells can be passaged without differentiation with LIF and BMP4, indicating that the pathways induced by these two ligands are sufficient to maintain stem cells. Importantly, we showed here that the binding of STAT3 and Smad1 to genomic sites is dependent on the LIF and BMP pathways, respectively, confirming the importance of these TFs as effectors of the signal transduction pathways that maintain pluripotency in ES cells (Figures S3L and S3M). Until the present study, the roles of transcriptional regulatory proteins downstream of these signaling pathways have not been well defined in the context of ES-cell transcriptional regulatory networks. STAT3 had been shown to bind to the *Nanog* enhancer (Suzuki et al., 2006), but there were no known targets of Smad1.

Consistent with a previous study implicating *Id* genes as downstream targets of the BMP4 pathway (Ying et al., 2003), we identified a MTL (bound by Smad1, Oct4, Sox2, Nanog, Klf4, E2f1, Esrrb, and Tcfcp2l1) 1.5 kb upstream of the TSS of *Id3*. Strikingly, the majority (97.3%, 649/667) of Smad1 at the MTL is associated with Nanog, Oct4, or Sox2. STAT3 (72.5%, 521/718) is also predominantly localized with Nanog, Oct4, or Sox2 within the MTL.

The multiple TFBS maps provide us with the opportunity to examine the mode of targeting genes by these factors on a global scale. E2f1 binds to ~50% of all genes, almost all of which fall into what we call classes I, II, and III (Figure 6A). Genes in these three classes (I, II, and III) are expressed at higher levels in ES cells than are genes in the other classes (Figure 6C), and class I and class II are enriched in genes that are expressed at higher levels in ES cells than in differentiating cells. Roughly 50% of all genes, those in classes IV and V, are not enriched in TF binding (Figure 6A). These TF-deficient genes are not expressed or are expressed at a low level. A fraction of these genes are bound by Suz12, suggesting that Suz12 plays a role in preventing TF occupancy and in silencing these genes (Boyer et al., 2006; Lee et al., 2006). However, a larger fraction of the TF-deficient genes are not bound by Suz12. It is possible that the chromatin structure of these genes is not permissive to the binding of TFs.

In summary, the genome-wide maps of TFs and coregulators demarcate different gene compartments in the ES-cell genome. The densely co-occupied sites represent key regions of potential functional importance and will assist in the identification of new regulators of self-renewal, pluripotency, and reprogramming. We demonstrate that the two key signaling pathways are integrated to the Oct4, Sox2, and Nanog circuitries through Smad1 and STAT3. Our data also provide a framework for modeling gene expression and understanding the transcriptional regulatory networks in pluripotent cells.

EXPERIMENTAL PROCEDURES

Cell Culture and Transfection

E14 mouse ES cells, cultured under feeder-free conditions, were maintained in Dulbecco's modified Eagle's medium (DMEM, GIBCO-BRL), with 15% heat-inactivated ES-qualified fetal bovine serum (FBS, GIBCO-BRL), 0.055 mM β-mercaptoethanol (GIBCO-BRL), 2 mM L-glutamine, 0.1 mM MEM nonessential amino acid, 5000 U/ml penicillin/streptomycin, and 1000 U/ml LIF (Chemicon). 293T cells were cultured in DMEM with 10% FBS and maintained at 37°C with 5% CO_2. For serum-free cell cultures, feeder-free E14 mouse ES cells were plated onto gelatin-coated plates in ESGRO Complete Basal Medium (Chemicon) supplemented with 10 ng/ml LIF (Chemicon) and 50 ng/ml BMP4 (Sigma). Cells were passaged every 2–3 days using accutase (GIBCO-BRL).

Oct4, *Sox2*, and *Nanog* shRNA constructs were designed as described previously (Loh et al., 2006). Transfection of shRNA was performed by using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. For the ChIP assay, 35 μg plasmids were transfected into ES cells on 150 mm plates. Puromycin (Sigma) selection was introduced 1 day after transfection at 1.0 μg/ml, and the cells were crosslinked and harvested 48 hr after transfection.

Luciferase Assay

For the 25 Nanog-Oct4-Sox2 cluster fragments and 8 Myc cluster fragments tested for enhancer activity, the fragments (~300 bp) were amplified from genomic DNA and cloned into BamHI and SalI sites of pGL3-*Pou5f1* pp vector (a *Pou5f1* minimal promoter driving *luciferase*) and sequence verified. E14 mouse ES cells or 293T cells were transfected with these reporter constructs by using Lipofectamine 2000 (Invitrogen). A *Renilla* luciferase plasmid (pRL-SV40 from Promega) was cotransfected as an internal control. Cells were harvested 36 hr after transfection, and the luciferase activities of the cell lysates were measured by using the Dual-luciferase Reporter Assay System (Promega).

ChIP Assay

A ChIP assay was carried out as described previously (Loh et al., 2006). Briefly, cells were crosslinked with 1% formaldehyde for 10 min at room temperature,

(C) The occurrence of the other TFs (Smad1, STAT3, Esrrb, Tcfcp2l1, Klf4, Zfx, CTCF, and E2f1) within the 3583 MTL. The color legend is the same as in (B).
(D) Genomic fragments associated with the Nanog-Oct4-Sox2 cluster show enhancer activity. A genomic fragment of ~300 bp (shown in red) was cloned downstream of a luciferase reporter (shown in blue) driven by minimal *Pou5f1* promoter (shown in orange). These reporter constructs were transfected into ES cells or 293T cells to determine ES-cell-specific enhancer activity. The loci tested for enhancer activity and primers for cloning these genomic fragments are listed in Table S9. Ten of the 25 Nanog-Oct4-Sox2 loci (40%) are marked with H3K4me3, whereas all 8 of the Myc loci are marked with H3K4me3.
(E) Smad1 occupancy is dependent on Oct4. ChIP assays were performed by using the anti-Smad1 antibody with extracts from ES cells transfected with the control RNAi construct (yellow bar) or the *Oct4* RNAi construct (blue bar).
(F) STAT3 occupancy is dependent on Oct4. ChIP assays were performed by using the anti-STAT3 antibody with extracts from ES cells transfected with the control RNAi construct (yellow bar) or the *Oct4* RNAi construct (blue bar).
(G) Oct4 occupancy is not dependent on the LIF and BMP pathways. ChIP assays were performed by using the anti-Oct4 antibody with extracts from ES cells treated with LIF+BMP4 (orange bar), LIF alone (green bar), BMP4 alone (blue bar), or no LIF and BMP4 (gray bar). Coordinates and q-PCR primers of the loci tested in (E)–(G) are listed in Table S10. All data are presented as mean ± SEM.

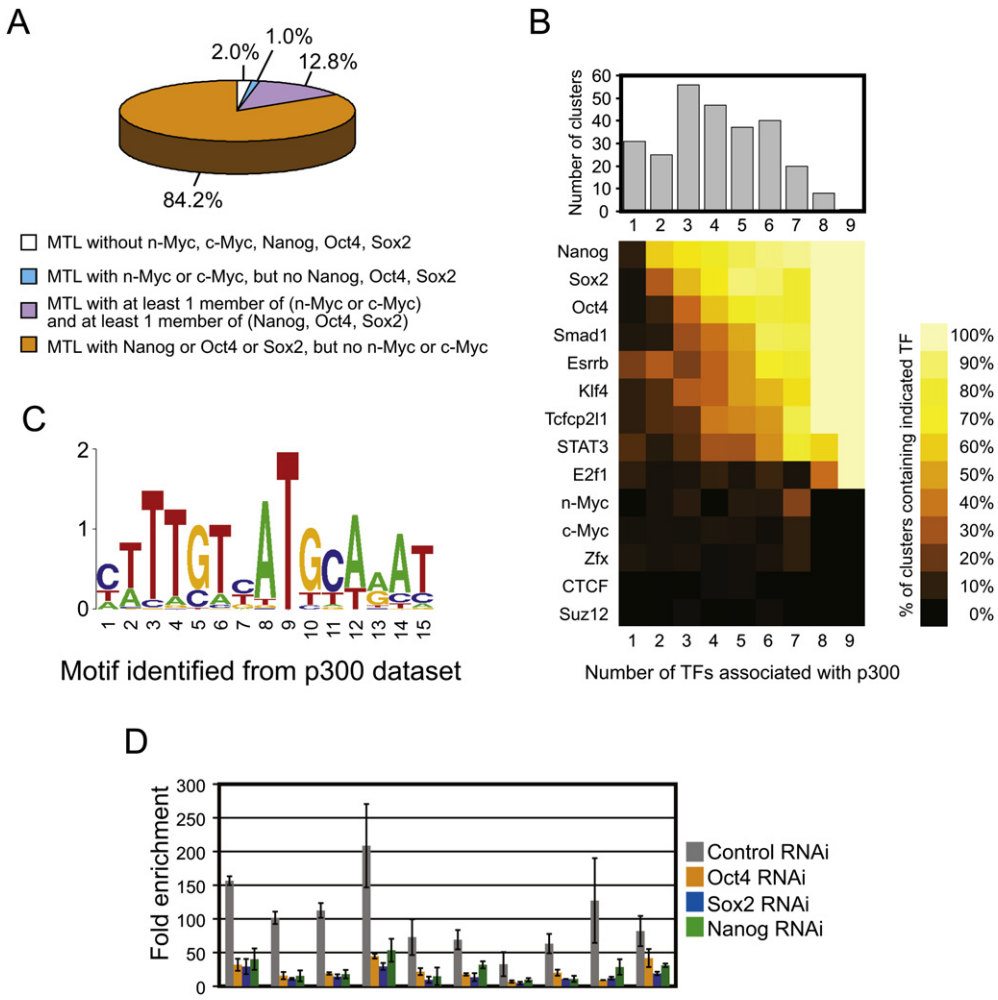

Figure 5. p300 Is Recruited to the Nanog-Oct4-Sox2 Cluster

(A) p300 is associated with the Nanog-Oct4-Sox2 cluster, but not with the Myc cluster. The pie chart shows the occurrence of p300 in different MTL types. The color legend is the same as for Figure 4B.

(B) Size distribution and composition of binding-site groups containing p300. (Top) Histogram showing the number of binding-site groups of different sizes. Size, here, refers to the number of non-p300 transcription factors (TFs) that have binding sites in the same group. (Bottom) Composition of p300-containing binding-site groups for different group sizes. Composition is expressed in terms of the percentage of p300-containing groups that contain the indicated TF. For example, Nanog, Sox2, and Oct4 are each found in 70% or more of the p300-containing clusters that have five other factors bound, whereas Smad1, Esrrb, Klf4, Tcfcp2l1, and STAT3 are each found at a frequency of ~30%–50%.

(C) Motif predicted by the de novo motif-discovery algorithm Weeder.

(D) Recruitment of p300 is dependent on Oct4, Sox2, and Nanog. ChIP assays were performed by using the anti-p300 antibody with extracts from ES cells transfected with the control RNAi construct (gray bar), the *Oct4* RNAi construct (orange bar), the *Sox2* RNAi construct (blue bar), or the *Nanog* RNAi construct (green bar). Coordinates of loci and q-PCR primers are listed in Table S10. The level of p300 was not altered after RNAi depletion of these TFs (data not shown). Data are presented as mean ± SEM.

and formaldehyde was then inactivated by the addition of 125 mM glycine. Chromatin extracts containing DNA fragments with an average size of 500 bp were immunoprecipitated by using the antibodies shown in Table S1. All ChIP experiments were repeated at least three times.

Computational Analyses

To identify the MTL, a list of genomic sites co-bound by any of the 13 TFs was generated. Two binding regions were clustered if their centers were 100 bp apart at most. This clustering procedure was done iteratively to form the largest possible clusters, forming what we call MTL. ChIP-seq data sets for p300 and Suz12 were also generated to determine where these coregulators are recruited with respect to the TFs profiled. Distances from one coregulator site to the nearest TFBSs were then calculated. Pairs of sites within 50 bp of one another were considered to belong to the same group. We computed the Pearson correlation coefficient for each pair of such colocalization vectors and used it as a similarity measure to cluster these factors. To associate binding site information with gene expression, we computed an association score for each pair of gene and TF based on the relative distance to the TSS of the gene. We then performed k-mean clustering on an association matrix to group the genes with similar TF association. Gene groups by this method were then analyzed with a previously published RA-induced differentiation data set (Ivanova et al., 2006).

Two published sets of gene-expression experiments were used in combination with the ChIP-seq data reported here to obtain a set of genes that are

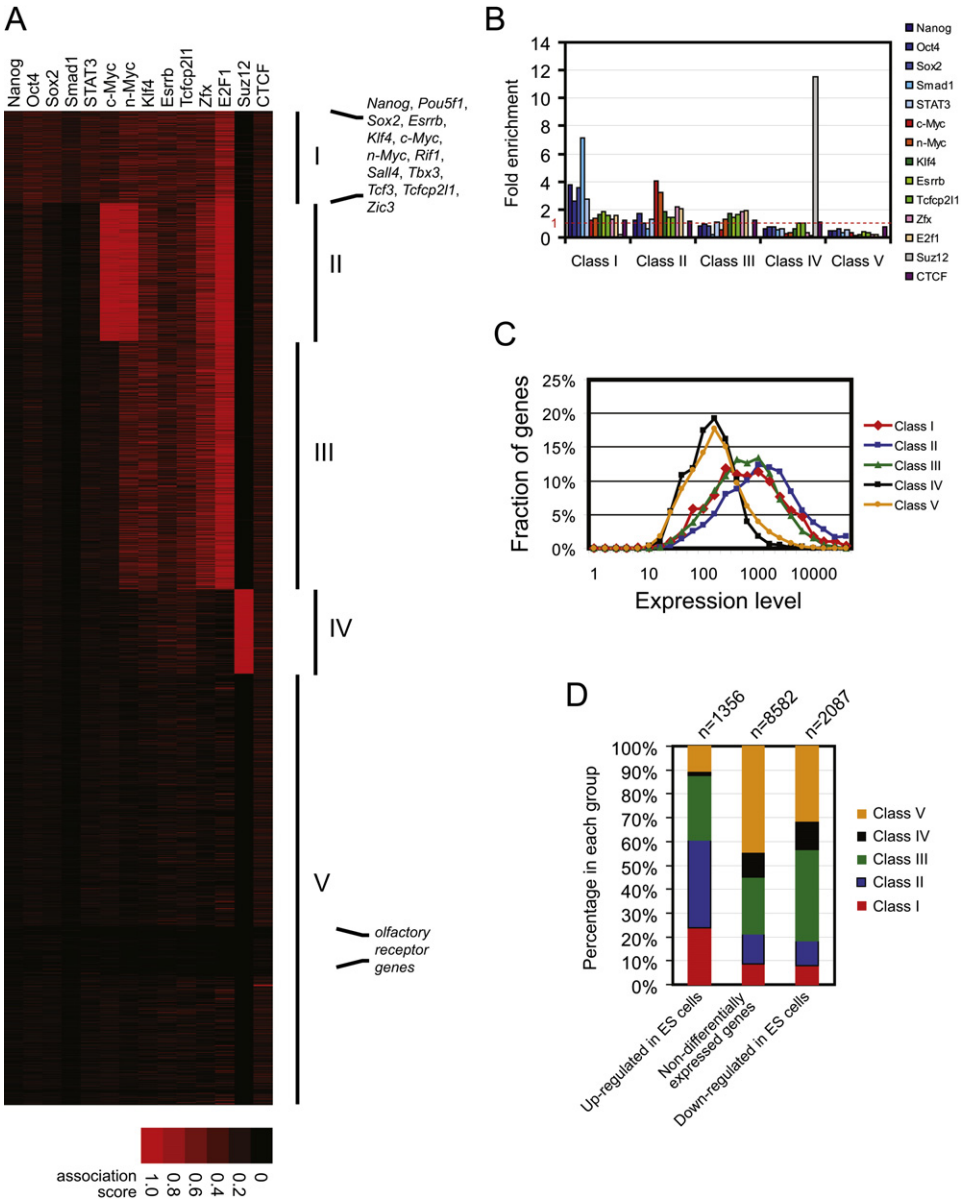

Figure 6. Association between Transcription Factor Binding and Gene Expression in ES Cells

(A) Heatmap showing five classes of genes obtained from k-means clustering based on the transcription factor (TF)-gene association score. In this analysis, we included a Suz12 ChIP-seq data set to explore the potential association of Suz12 and the other 13 TFs.

(B) Enrichment of TFs in the five classes. The y axis represents the ratio of the average TF-gene association score for the group to the average association score for all genes.

(C) Histogram of the levels of gene expression for genes found in each of the five classes.

(D) Proportion of different classes of genes found in differentially (up- or downregulated in ES cells) and nondifferentially expressed genes in a published expression data set (Ivanova et al., 2006).

enriched in direct transcriptional targets (Ivanova et al., 2006; Zhou et al., 2007). For a given TF, we scored and ranked each gene based on the number and "intensity" of ChIP-seq-defined binding sites. For a given expression change ranking and a given TF-binding ranking, we used responder analysis to determine the significance of association between binding and expression, as well as to define gene sets that are at least 2-fold enriched in direct targets. Regulatory targets were inferred from the intersection of top-ranked bound genes and top-ranked differentially expressed genes. Further details are available in the Supplemental Data.

ACCESSION NUMBERS

Our ChIP-seq datasets have been deposited in the GEO database with ID number GSE11431.

SUPPLEMENTAL DATA

Supplemental Data include Supplemental Experimental Procedures, Supplemental Results, Supplemental Discussion, fifteen figures, fourteen tables,

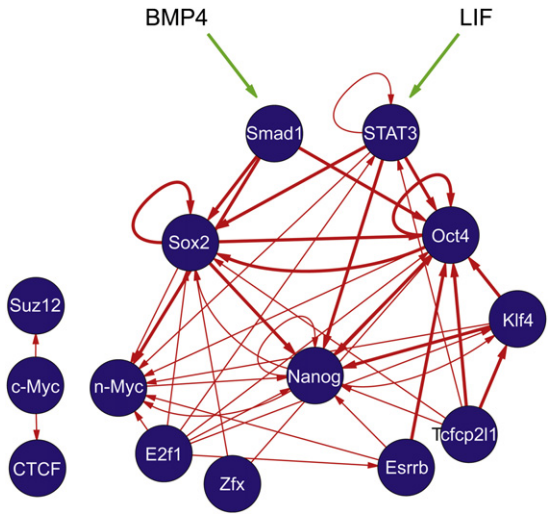

Figure 7. Transcriptional Regulatory Network in ES Cells

Network of regulatory interactions inferred from ChIP-seq binding assays and from gene-expression changes during differentiation. Nodes are ChIP-seq-assayed transcription factors. Arrows point from the transcription factor to the target gene. Two sets of published experiments were used to define genes that are differentially expressed during differentiation (Ivanova et al., 2006; Zhou et al., 2007). Thick arrows represent interactions inferred from binding data and both expression experiments, whereas thin arrows represent interactions inferred from binding data and only one of the expression experiments. Regulatory targets were inferred from the intersection of top-ranked bound genes and top-ranked differentially expressed genes. Thresholds for defining top-ranked genes in the two lists were determined empirically by maximizing the number of genes in the intersection, subject to two constraints: the p value for the enrichment of genes in the intersection had to be 0.001 or better, and there had to be at least twice as many genes in the intersection as expected. All regulatory interactions in this network involve higher-level expression in ES cells and lower-level expression during differentiation. There were no interactions among the factors in this network when regulation in the opposite direction was evaluated. The network was drawn by using Cytoscape.

and Supplemental References and are available at http://www.cell.com/cgi/content/full/133/6/1106/DC1/.

ACKNOWLEDGMENTS

This work is supported by the Agency for Science, Technology and Research (A*STAR) of Singapore, Singapore Stem Cell Consortium grant (to H.H.N) and the NIH ENCODE grant 1R01HG003521-01 (to Y. R and C.L.W). X.C, W.Z, and J.M.J are supported by the National University of Singapore graduate scholarships. We thank Andrew Hutchins, Ching-Aeng Lim, Edwin Chueng and Feng Lin for critical comments on the manuscript. We are grateful to Katty Kuay and Qiu-Li Tan for technical assistance.

Received: January 24, 2008
Revised: March 11, 2008
Accepted: April 21, 2008
Published: June 12, 2008

REFERENCES

Arnosti, D.N., and Kulkarni, M.M. (2005). Transcriptional enhancers: intelligent enhanceosomes or flexible billboards? J. Cell. Biochem. *94*, 890–898.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Bieda, M., Xu, X., Singer, M.A., Green, R., and Farnham, P.J. (2006). Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. Genome Res. *16*, 595–605.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799–816.

Boiani, M., and Schöler, H.R. (2005). Regulatory networks in embryo-derived pluripotent stem cells. Nat. Rev. Mol. Cell Biol. *6*, 872–884.

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.

Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., and Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development *132*, 885–896.

Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., et al. (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell *116*, 499–509.

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.

Chew, J.L., Loh, Y.H., Zhang, W., Chen, X., Tam, W.L., Yeap, L.S., Li, P., Ang, Y.S., Lim, B., Robson, P., et al. (2005). Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. Mol. Cell. Biol. *25*, 6031–6046.

Ehret, G.B., Reichenbach, P., Schindler, U., Horvath, C.M., Fritz, S., Nabholz, M., and Bucher, P. (2001). DNA binding specificity of different STAT proteins. Comparison of in vitro specificity with natural target sites. J. Biol. Chem. *276*, 6675–6688.

Faiola, F., Liu, X., Lo, S., Pan, S., Zhang, K., Lymar, E., Farina, A., and Martinez, E. (2005). Dual regulation of c-Myc by p300 via acetylation-dependent control of Myc protein turnover and coactivation of Myc-induced transcription. Mol. Cell. Biol. *25*, 10220–10234.

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

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

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.

Ivanova, N.B., Dimos, J.T., Schaniel, C., Hackney, J.A., Moore, K.A., and Lemischka, I.R. (2002). A stem cell molecular signature. Science *298*, 601–604.

Jiang, J., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat. Cell Biol. *10*, 353–360.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497–1502.

Kaczynski, J., Cook, T., and Urrutia, R. (2003). Sp1- and Kruppel-like transcription factors. Genome Biol. *4*, 206.

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.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell *128*, 1231–1245.

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.

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.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., et al. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell *1*, 55–70.

Merika, M., Williams, A.J., Chen, G., Collins, T., and Thanos, D. (1998). Recruitment of CBP/p300 by the IFN β enhanceosome is required for synergistic activation of transcription. Mol. Cell *1*, 277–287.

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.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, 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., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev. *12*, 2048–2060.

Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H., and Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell *87*, 953–959.

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313–317.

Panne, D., Maniatis, T., and Harrison, S.C. (2007). An atomic model of the interferon-β enhanceosome. Cell *129*, 1111–1123.

Pavesi, G., Mauri, G., and Pesole, G. (2001). An algorithm for finding signals of unknown length in DNA sequences. Bioinformatics *17* (*Suppl 1*), S207–S214.

Pettersson, K., Svensson, K., Mattsson, R., Carlsson, B., Ohlsson, R., and Berkenstam, A. (1996). Expression of a novel member of estrogen response element-binding nuclear receptors is restricted to the early stages of chorion formation during mouse embryogenesis. Mech. Dev. *54*, 211–223.

Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R.C., and Melton, D.A. (2002). "Stemness": transcriptional profiling of embryonic and adult stem cells. Science *298*, 597–600.

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., et al. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat. Methods *4*, 651–657.

Smith, A.G. (2001). Embryo-derived stem cells: of mice and men. Annu. Rev. Cell Dev. Biol. *17*, 435–462.

Struhl, K. (2001). Gene regulation. A paradigm for precision. Science *293*, 1054–1055.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat. Struct. Mol. Biol. *14*, 103–105.

Suzuki, A., Raya, A., Kawakami, Y., Morita, M., Matsui, T., Nakashima, K., Gage, F.H., Rodriguez-Esteban, C., and Izpisua Belmonte, J.C. (2006). Nanog binds to Smad1 and blocks bone morphogenetic protein-induced differentiation of embryonic stem cells. Proc. Natl. Acad. Sci. USA *103*, 10294–10299.

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.

Thanos, D., and Maniatis, T. (1995). Virus induction of human IFN β gene expression requires the assembly of an enhanceosome. Cell *83*, 1091–1100.

Thomas, K.R., and Capecchi, M.R. (1986). Introduction of homologous DNA sequences into mammalian cells induces mutations in the cognate gene. Nature *324*, 34–38.

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.

Wathelet, M.G., Lin, C.H., Parekh, B.S., Ronco, L.V., Howley, P.M., and Maniatis, T. (1998). Virus infection induces the assembly of coordinately activated transcription factors on the IFN-β enhancer in vivo. Mol. Cell *1*, 507–518.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318–324.

Ying, Q.L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281–292.

Zeller, K.I., Zhao, X., Lee, C.W., Chiu, K.P., Yao, F., Yustein, J.T., Ooi, H.S., Orlov, Y.L., Shahab, A., Yong, H.C., et al. (2006). Global mapping of c-Myc binding sites and target gene networks in human B cells. Proc. Natl. Acad. Sci. USA *103*, 17834–17839.

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.