Cell

An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells

Jonghwan Kim,[1,3,4] Jianlin Chu,[1,3] Xiaohua Shen,[1,3] Jianlong Wang,[1,3] and Stuart H. Orkin[1,2,3,4,*]
[1]Department of Pediatric Oncology
[2]Children's Hospital and Dana Farber Cancer Institute, Harvard Stem Cell Institute
[3]Harvard Medical School
[4]Howard Hughes Medical Institute
Boston, MA 02115, USA
*Correspondence: stuart_orkin@dfci.harvard.edu
DOI 10.1016/j.cell.2008.02.039

SUMMARY

Much attention has focused on a small set of transcription factors that maintain human or mouse embryonic stem (ES) cells in a pluripotent state. To gain a more complete understanding of the regulatory network that maintains this state, we identified target promoters of nine transcription factors, including somatic cell reprogramming factors (Oct4, Sox2, Klf4, and c-Myc) and others (Nanog, Dax1, Rex1, Zpf281, and Nac1), on a global scale in mouse ES cells. We found that target genes fall into two classes: promoters bound by few factors tend to be inactive or repressed, whereas promoters bound by more than four factors are largely active in the pluripotent state and become repressed upon differentiation. Furthermore, we propose a transcriptional hierarchy for reprogramming factors and broadly distinguish targets of c-Myc versus other factors. Our data provide a resource for exploration of the complex network maintaining pluripotency.

INTRODUCTION

Pluripotency, the capacity to generate all cell types, is a defining property of embryonic stem (ES) cells, cultured cells derived from the inner cell mass of the mammalian blastocyst (Evans and Kaufman, 1981; Martin, 1981). In addition, ES cells can be maintained in a proliferative state for prolonged periods, the phenomenon of self-renewal. Pluripotency may be imposed on somatic cells following their fusion with ES cells (Cowan et al., 2005), and one transcription factor specifically expressed in ES cells, Nanog, facilitates fusion-induced pluripotency (Silva et al., 2006). Moreover, forced expression of other transcriptional factors (Oct4, Sox2, Klf4, and c-Myc) reprograms mouse fibroblasts to ES-like cells (called iPS cells)(Takahashi and Yamanaka, 2006), the quality of which is enhanced upon selection of cells that express endogenous Oct4 or Nanog (Maherali et al., 2007; Okita et al., 2007; Wernig et al., 2007b). Recently, it has been shown that the same factors reprogram human fibroblasts to a pluripotent state (Park et al., 2007; Takahashi et al., 2007; Yu et al., 2007). How pluripotency is established and maintained in ES cells is of great interest, as an improved understanding of the transcription factors and epigenetic modifications operating in a regulatory network will facilitate both directed programming of ES cells to specific lineages and the reprogramming of somatic cells to an ES-like state.

Until recently, attention has focused almost exclusively on a small set of transcription factors, Oct4, Sox2, and Nanog as "core" pluripotency factors for human or mouse ES cells (Orkin, 2005). Oct4 has long been recognized to be essential in vivo and in vitro for early development and maintenance of pluripotency (Nichols et al., 1998). Indeed, the dosage of Oct4 is crucial: reduced expression permit trophoectoderm development, whereas enhanced expression drives primitive endoderm differentiation (Niwa et al., 2000). Sox2 is a transcriptional partner of Oct4 (Avilion et al., 2003). Rather than directly interacting with Oct4 protein, Sox2 assembles on target regulatory elements with Oct4 to collaborate in transcriptional control. Nanog promotes ES cell self-renewal and alleviates the requirement for Leukemia Inhibitory Factor (LIF) (Chambers et al., 2003; Mitsui et al., 2003). While considerable evidence speaks to the importance of these factors in maintaining the properties of ES cells, evidence also points to the involvement of additional transcription factors in the control of pluripotency (Ivanova et al., 2006; Wang et al., 2006; Zhou et al., 2007).

To account for the unique properties of pluripotent ES or iPS cells at the molecular level, it will be necessary to understand the transcriptional networks responsible for maintaining pluripotency. Studies to this end have entailed the search for additional transcription factors and delineation of a protein-protein interaction network highly enriched for factors involved in the control of pluripotency (Ivanova et al., 2006; Wang et al., 2006), and preliminary global target gene assessment for the initial core factors (Boyer et al., 2005; Loh et al., 2006). Starting with the identification of protein partners of Nanog through protein complexes purification and microsequencing, coupled with iterative affinity purification of interacting proteins, we generated a network that includes additional factors required for maintenance of pluripotency (Wang et al., 2006). Among this latter class, we encountered Sall4 (Sakaki-Yumoto et al., 2006; Wu et al., 2006; Zhang et al., 2006), Dax1 (Niakan et al., 2006), and Rif1

(Loh et al., 2006), factors identified independently by others as involved in maintenance of ES cell pluripotency. The protein network is connected to complexes, such as NuRD remodeling complex and PRC1, implicated in transcriptional repression (Wang et al., 2006). In parallel, other groups have used new methods for global target mapping (ChIP-Chip and ChIP-PET) to predicted target genes regulated by Oct4, Sox2, and Nanog in mouse and human ES cells. These studies revealed combinatorial occupancy of target gene promoters by these core factors and both autoregulatory and feed-forward transcriptional circuits (Boyer et al., 2005; Loh et al., 2006). The discovery of a 4-factor reprogramming set (Takahashi and Yamanaka, 2006) identified Klf4 and c-Myc as additional proteins to be integrated into the network inducing and/or maintaining pluripotency. In parallel, other work has suggested that histone modification signatures, specifically histone 3 lysine 4 and histone 3 lysine 27 trimethylation (H3K4me3 and H3K27me3, respectively), are important in controlling gene regulation in ES cells (Bernstein et al., 2006; Boyer et al., 2006; Lee et al., 2006; Pan et al., 2007; Zhao et al., 2007).

Genome-wide mapping of transcription factor targets by ChIP, combined with microarrays or sequencing methods, is a powerful tool for laying a foundation for understanding transcriptional networks (Iyer et al., 2001; Kim et al., 2005; Loh et al., 2006; Ren et al., 2000; Roh et al., 2004). Expanding the number of transcription factors analyzed by ChIP-based methods should be especially informative in dissecting system level biological processes, as ∼10% of annotated mammalian genes are predicted to encode DNA-binding proteins (Messina et al., 2004). A practical limitation to current ChIP approaches is the availability of suitable "ChIP quality" antibodies.

We report the application of in vivo biotinylation mediated ChIP (bioChIP) to global target mapping (bioChIP-Chip) of an expanded set of factors associated with pluripotency of mES cells (Mito et al., 2005). This approach, which relies on streptavidin affinity capture of tagged proteins, is comparable to conventional ChIP-Chip but circumvents issues related to antibody availability. Using bioChIP-Chip, we have identified target promoters of nine transcription factors, including the somatic cell reprogramming factors, on a global scale. We have constructed an expanded transcriptional regulatory network containing the previously known three core factors, as well as additional factors. Our data argue that differential regulation of target genes correlates with the extent of promoter occupancy by multiple factors. Moreover, we propose a transcriptional hierarchy for the somatic reprogramming factors and broadly distinguish targets of c-Myc versus the other factors. Our data provides a resource with which to probe mechanisms of pluripotency control and differentiation by the complex transcriptional regulatory network in ES or iPS cells.

RESULTS

Global Mapping of Target Genes by Biotin-Mediated Chromatin Immunoprecipitation

Prior genome-wide chromatin immunoprecipitation (ChIP) analyses relying on antibodies have been performed in mouse and human ES cells with core factors (Oct4, Sox2, and Nanog) involved in maintaining pluripotency. Our first goal was to define the targets of a larger set of pluripotency factors with greater consistency in the experimental platform. To this end, we assessed the suitability of streptavidin affinity-capture of in vivo biotin-tagging of proteins (de Boer et al., 2003; Wang et al., 2006) as an alternative to antibody-based ChIP.

Earlier studies proposed application of the biotin-tagging method for ChIP assay (de Boer et al., 2003), and showed its utility in combination with *Drosophila* microarrays (Mito et al., 2005, 2007). However, the feasibility of bioChIP-chip in complex mammalian genomes has not been explored (Figure 1A and Experimental Procedures). We performed bioChIP and conventional ChIP reactions in mES cells for both Nanog and c-Myc (Myc). bioChIP and ChIP samples were hybridized onto Affymetrix mouse promoter arrays with appropriate references to map the target loci of each factor. We compared targets predicted by the two methods for both factors. The majority of targets predicted for each factor by the two methods were shared. 67% and 81% of bioChIP targets of Nanog and Myc, respectively, were identified with the conventional ChIP approach (Figures 1C and 1D). The overall shapes of binding peaks of Nanog and Myc across the genome were nearly identical for the two methods (Figure S1 available online). The correlation of target loci from the two different methods was 0.896 for Nanog (Figure S3, see Experimental Procedures), suggesting that bioChIP is comparable to conventional antibody ChIP.

On comparison with previously published mES ChIP-PET data (Loh et al., 2006), we observed ∼60% overlap in promoter targets (58% and 65% of ChIP-PET targets were also defined as bioNanog and Nanog antibodyChIP-chip targets, respectively). However, given that ChIP-PET predicted only 434 promoter targets (337 comparable RefSeq promoters) for Nanog in contrast to our ChIP data (1284 bioChIP targets and 1742 conventional ChIP targets, Figure 1C), we asked whether the lower number reflects partial coverage, perhaps due to the depth of sequencing and/or tiling array repeat masking. We tested conventional antibody ChIP material on one of the seven Affymetrix whole genome mouse arrays (Mouse Tiling 2.0R F Array) that covers ∼15% of the entire genome and ∼14% of well annotated genes and found 223 promoter targets (≥ 50% number of ChIP-PET promoter targets) with a similar proportion of nonpromoter targets as predicted by ChIP-PET (12.8%). In addition, we identified 137 genes in common in our bioChIP and published human Nanog ChIP data that are not represented in ChIP-PET (Table S3)(Boyer et al., 2005; Loh et al., 2006). Taken together, these data imply that the ChIP-PET data set may lack targets due to inadequate depth of sequencing (Euskirchen et al., 2007).

We proceeded to determine the global target promoters for nine transcription factors in ES cells, including previously examined core factors (Nanog, Oct4, and Sox2)(Boyer et al., 2005; Loh et al., 2006), somatic cell reprogramming factors (Klf4 and Myc, in addition to Oct4 and Sox2) (Maherali et al., 2007; Okita et al., 2007; Takahashi and Yamanaka, 2006; Wernig et al., 2007b), and protein-interacting partners of Nanog and Oct4 (Dax1, Nac1, Zfp281 and Rex1) (Wang et al., 2006). In each instance, we employed cells expressing a subendogenous level of the respective biotin-tagged factor, so as to avoid perturbing the existing network (Figure S2). The levels at which exogeneous proteins

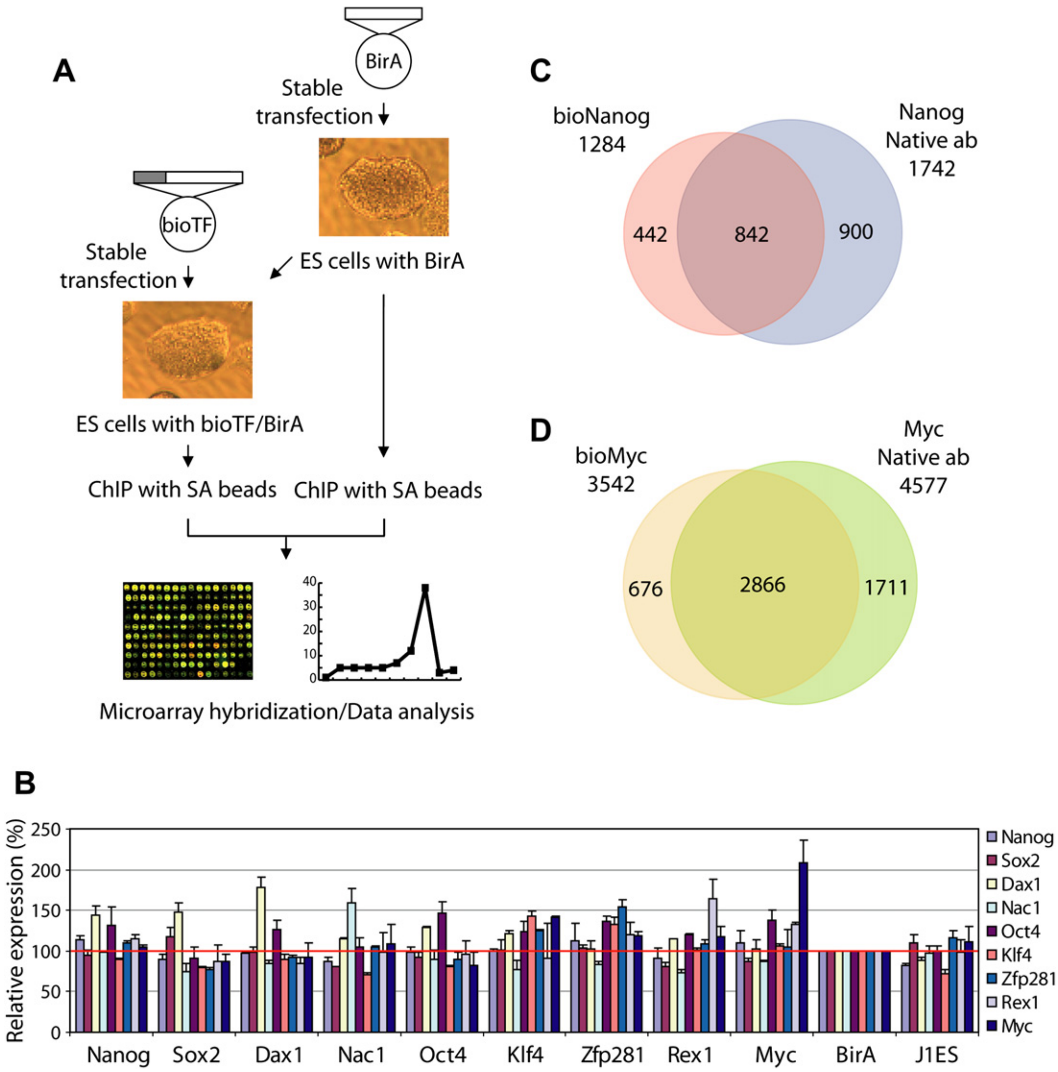


Figure 1. Strategy of In Vivo Biotinylation-Mediated Chromatin Immunoprecipitation and Microarray

(A) Schematic representation of biotin-mediated chromatin immunoprecipitation. The gray bar represents BirA target sequence (MSGLNDIFEAQKIEWHE-GAPSSR).

(B) Expression analysis of nine genes using cell lines expressing biotin-tagged proteins. Biotin-tagged cell lines are indicated on the horizontal axis and transcript levels are presented as color bars. Error bars represent standard deviation from three independent repeats.

(C and D) Overlap of target promoters between bioChIP-chip and conventional ChIP-chip experiments for Nanog (C) and Myc (D). Predicted overlap may be underestimated due to a fixed statistical threshold (see Figures 2 and S3).

were expressed fail to elicit subsequent change in the transcript levels of the nine factors (Figure 1B) (Wang et al., 2006).

Before analyzing global targets, we performed additional validation experiments to assess whether low expression of a biotin-tagged protein might perturb chromatin occupancy by untagged proteins. Accordingly, we performed conventional Nanog antibody ChIP reactions using wild-type mES cells, mES cells expressing BirA alone, and mES cells expressing BirA plus tagged versions of Dax1, Oct4, and Nanog. Figure 2 shows that the overall patterns of Nanog binding peaks among these different cell lines are indistinguishable. Target correlations across the cell lines were also very strong (most were > 0.960 and the correlation between bioChIP data and antibody ChIP data across multiple cells is > 0.880, Figure S3). These data exclude significant effects of subendogenous levels of expressed protein on factor occupancy. To exclude artifacts due to biotinylation of endogeneous proteins by expressed BirA, we performed ChIP-chip using BirA expressing cells with input genomic DNA as a reference (Figure 2). We observed only 18 specific peaks among all promoter regions. Thus, non-specific effects due to biotinlyation of endogenous DNA-binding proteins are insignificant compared with the number of targets for each factor (>500).

Taken together, our findings demonstrate that the bioChIP-chip method is a valid alternative to the conventional ChIP-chip method.

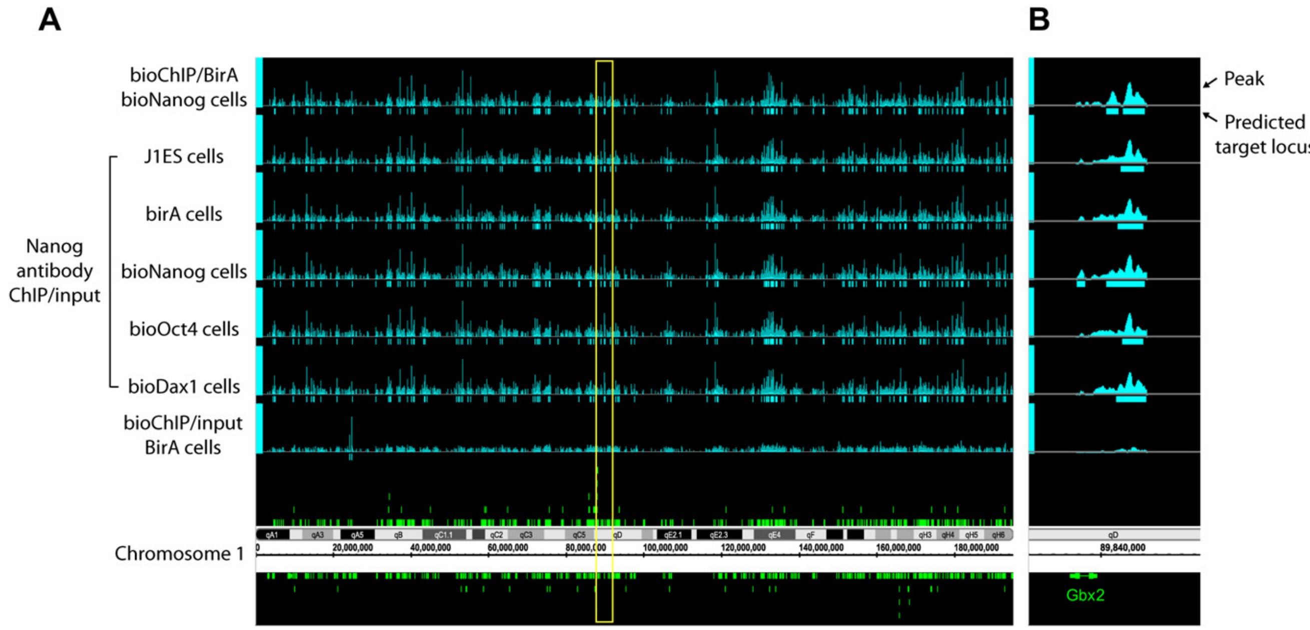


Figure 2. Chromosomal View of Nanog Occupancy Detected by bioChIP-Chip and Conventional ChIP-Chip

(A) Comparison of Nanog binding patterns using multiple cell lines is displayed using Affymetrix Integrated Genome Browser. In addition to bioNanog ChIP (top), antibody ChIP-chip data from control cell lines (J1 ES and BirA expressing cells) and cells expressing ectopic biotin-tagged protein (bioNanog, bioDax1 and bio-Oct4 cells) are tested. Nonspecific biotinylation by BirA enzyme was also tested (bottom). Yellow box indicates the chromosomal loci harboring Gbx2.

(B) Representative view of Nanog occupancy at the Gbx2 upstream promoter.

Promoter Occupancy of Nine Transcription Factors in mES Cells

In addition to the nine transcription factors noted above, we mapped two histone modifications, H3K4me3 and H3K27me3, by antibody ChIP-chip. Information from ~8 kb upstream and 2 kb downstream of 19,253 well-characterized transcription start site (TSS) of RefSeq genes from UCSC genome browser was used for analysis (see Experimental Procedures) (Kuhn et al., 2007).

The number of target promoters occupied by each factor is shown in Figure 3A. A compilation of target genes for each factor and binding peak positions are provided in Table S1 and Table S2, respectively. As expected, the number of targets occupied by the different factors varies greatly (Figure 3A). Notably, Myc occupies many more target promoters (18% of all promoters) than the other factors. This result is in accordance with prior observations in other cell types (Fernandez et al., 2003; Li et al., 2003; Mao et al., 2003). We also found that approximately 50% and 10% of promoters bear H3K4me3 and H3K27me3 marks, respectively (discussed below). We observed that the vast majority of binding sites for each factor were in close proximity to the TSS (Figure 3B), and more than a third of mouse promoters were occupied by at least one of the nine transcription factors we tested (6632 promoters, Figure 3C).

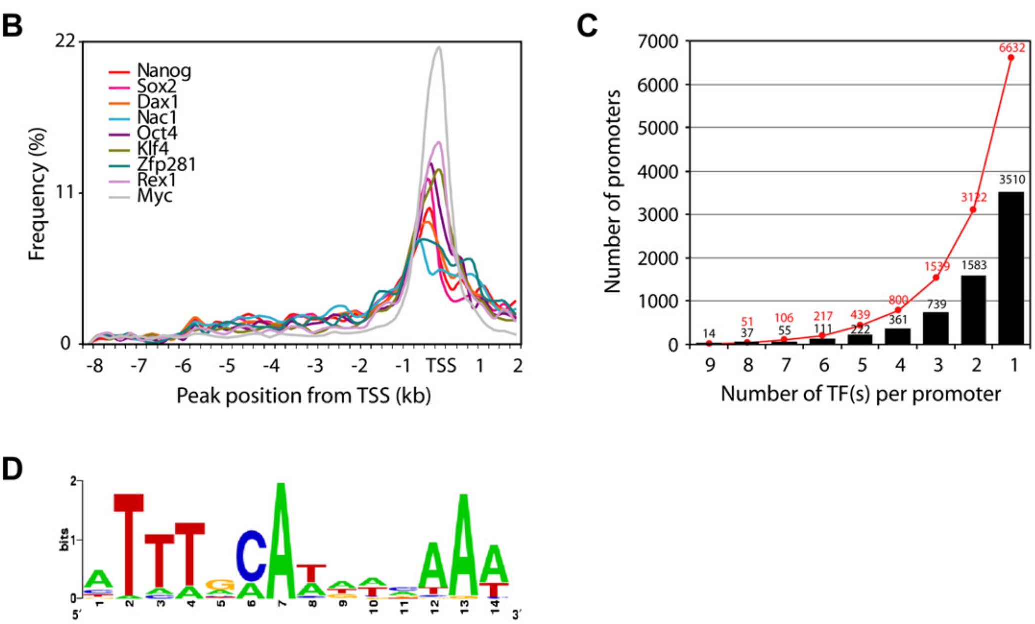
A previous study of hES cells showed that Nanog, Oct4, and Sox2 share many targets (353 genes) (Boyer et al., 2005). Surprisingly, our bioChIP-chip data reveal that many more promoters are co-occupied by multiple factors. We have observed that > 100 promoters are occupied by at least seven factors, and ~800 promoters are occupied by at least four of the nine factors examined (Figure 3C). More interestingly, actual binding loci of multiple factors within the target promoters are virtually coincident, suggesting that factors work as protein complexes or within compact cis-regulatory elements when multiple factors occupy the same target locus (Figure S4). We also observed numerous target loci occupied by fewer factors (Figure S4). To define a consensus motif that might be utilized by the multiple factors, we tested ± 100bp genomic sequence information from the center position of predicted common target loci using MEME (Bailey et al., 2006). Interestingly the consensus motif (ATTTGCAT) predicted from MEME (e-value 1.4E-50, Figure 3D) was similar to sequences previously predicted by different algorithms as Oct4 or Sox2-Oct4 target sequences (Loh et al., 2006; Macisaac et al., 2006). Targets in common among our data, human ChIP-Chip data (Boyer et al., 2006), and ChIP-PET data (Loh et al., 2006) are summarized in Table S3.

We also validated predicted target loci by quantitative ChIP-PCR using primer pairs specific to the predicted target loci that are occupied by either multiple, or fewer, factors with various MAT p values (see Experimental Procedures and Table S4) to confirm that our target cutoff was appropriate to minimize false positives. Most of target loci we tested for each factor show substantial enrichment over the BirA control (Figure S5). With these results and additional quantitative PCR confirmation, we estimate an average false positive rate of < ~5%. These estimates are comparable to those in previous studies in which antibodies were employed (Boyer et al., 2006; Lee et al., 2006). While our global dataset may miss some authentic targets due to our cutoff criteria, few irrelevant loci are likely to be present.

Among the 6632 targets bound in aggregate by the nine factors, 50% are occupied by only one of the nine factors

A

	Nanog	Sox2	Dax1	Nac1	Oct4	Klf4	Zfp281	Rex1	Myc	K4me3	K27me3
Targets	1284	819	1754	804	783	1790	601	1543	3542	9451	2046



E

	Nanog	Sox2	Dax1	Nac1	Oct4	Klf4	Zfp281	Rex1	Myc	K4me3	K27me3
Nanog	1.000										
Sox2	0.425	1.000									
Dax1	0.368	0.406	1.000								
Nac1	0.344	0.341	0.331	1.000							
Oct4	0.285	0.358	0.245	0.258	1.000						
Klf4	0.228	0.255	0.316	0.230	0.237	1.000					
Zfp281	0.142	0.166	0.172	0.201	0.154	0.160	1.000				
Rex1	0.128	0.128	0.139	0.111	0.134	0.154	0.103	1.000			
Myc	0.110	0.084	0.128	0.089	0.148	0.202	0.096	0.238	1.000		
K4me3	0.128	0.121	0.165	0.120	0.130	0.220	0.110	0.210	0.437	1.000	
K27me3	0.013	0.027	0.012	-0.013	0.027	0.024	0.047	-0.020	-0.089	-0.096	1.000

Correlation: 1.000, 0.400, 0.350, 0.300, 0.250, 0.200, 0.150, 0.100, 0.050, -0.100

Figure 3. Summary of Nine Transcription Factor Occupancy and Histone Modification Status

(A) Number of target promoters bound by each factor or associated with H3K4 or H3K27 trimethylation.

(B) Relative position of chromosomal target loci of each factor to the TSS.

(C) Number of common targets of multiple factors. *y* axis represents the number of target promoters occupied by transcription factor(s). Red dots represent the accumulated number of target promoters.

(D) Predicted consensus binding motif of multiple factor target loci using MEME.

(E) Correlation between each factor targets and hierarchical cluster of nine factors based on their target similarity.

(Figure 3C). Clustering of the nine transcription factors based on their target correlations (Figure 3E) shows that of the nine tested factors, Nanog, Sox2, Dax1, Nac1, Oct4, Klf4, and Zfp281 exhibit overall similarity in their targets. In contrast, targets of Myc and Rex1 segregate to a distinct cluster (Figure 3E, cluster). Functional classification of the presumptive targets of each factor using the PANTHER classification tool also demonstrates separation of factors in two classes (Mi et al., 2007). In general, target genes of each of the tested factors are enriched in genes involved in nucleic acid metabolism and transcriptional control. Interestingly, targets of Myc or Rex1 are implicated in protein metabolism, rather than in developmental processes, whereas targets of the other factors are enriched in genes for developmental processes (Figure S6).

Histone Modification Signatures

Core pluripotency factors are involved in both gene activation and repression in ES cells (Boyer et al., 2005; Loh et al., 2006). Mechanisms that account for this differential regulation are not understood. To address this question, we performed supervised clustering (Figure 4A, see Experimental Procedures) to reveal the relationship, if any, between targets of various combinations of transcription factors and corresponding H3K4me3 and H3K27me3 marks, as well as gene expression profiles.

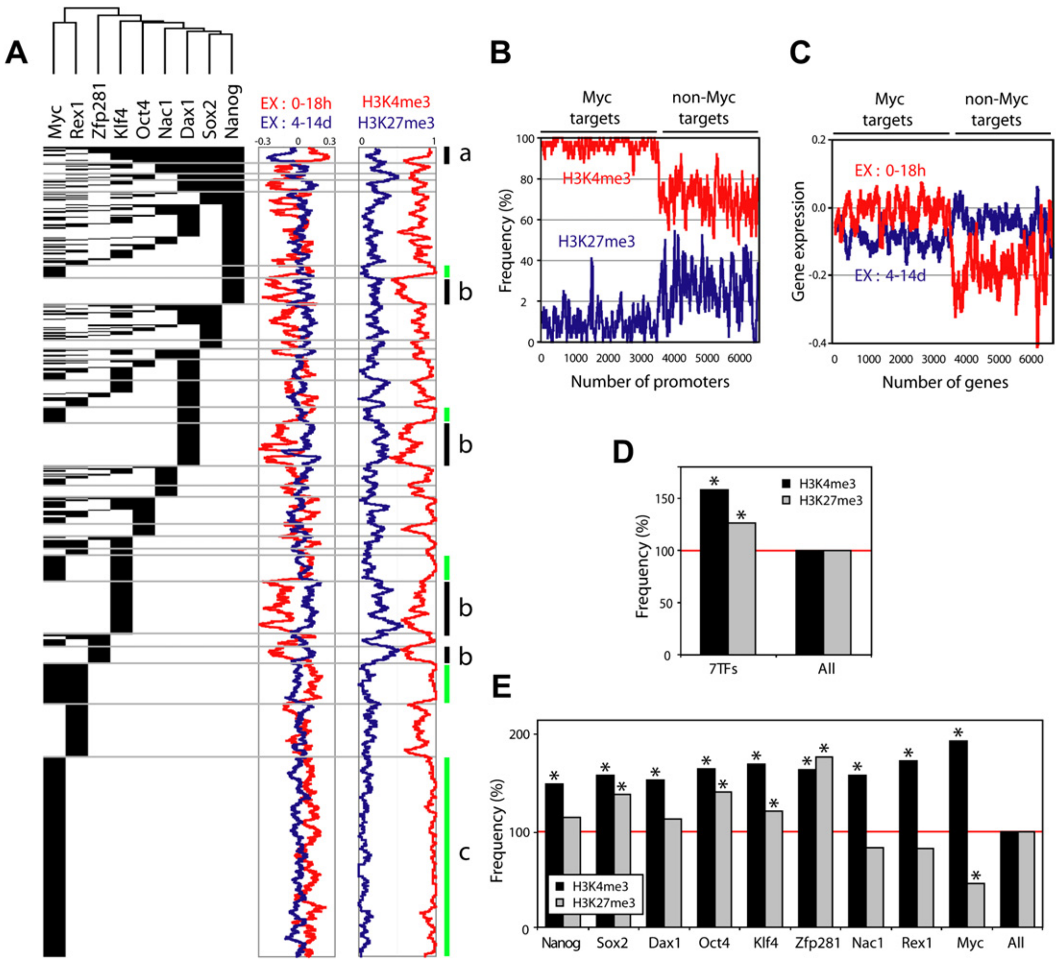


Figure 4. H3K4me3 and H3K27me3 Status and Factor Occupancy of the Promoters

(A) A supervised cluster image showing 6632 target promoters occupied by different factor combinations (see Experimental Procedures). Corresponding H3K4me3 (red) and H3K27me3 (blue) histone marks (presence: 1; absence: 0) as well as gene expression profiles (log2) upon J1 ES cell differentiation (0–18 hr: red, 4–14 days: blue, see Experimental Procedures) are shown as moving window averaged lines (bin size 50 and step size 1). Bar "a" represents the promoters occupied by multiple factors including at least Nanog, Sox2, Dax1, Nac1, and Oct4 (left panel) and corresponding gene expression changes upon differentiation (middle panel) as well as their histone marks (right panel). Bars "b" represent the clusters of promoters occupied by a single factor Nanog, Dax1, Klf4, and Zfp281, respectively (see Figure 5E and 5H). Green lines (bars "c") represent Myc target promoters with corresponding gene expression profiles and histone mark status.

(B) H3K4me3 (red line) and H3K27me3 (blue line) status for Myc target promoters.

(C) Expression profiles of Myc target genes at different time points upon differentiation (0–18 hr: red, 4–14 days: blue). Total 6632 target genes of any of nine factors are shown, and moving window average (bin size 50 and step size 1) was applied (B and C).

(D) Factor target promoters are both H3K4me3 and H3K27me3 rich over all promoters. "7TFs" represent the targets of any of seven factors (Nanog, Sox2, Dax1, Nac1, Oct4, Klf4, and Zfp281), and "All" represents all promoters. Asterisk indicates hypergeometric probability < 0.0001.

(E) Histone marks on the target promoters of each factor. Asterisk indicates hypergeometric probability < 0.0001.

The H3K4me3 and H4K27me3 marks of Myc target promoters exhibit a unique distribution in comparison with promoters bound by the other factors (Figure 4A, green bars c). 96% and 5% of Myc target promoters bear H3K4me3 and H3K27me3, respectively (Figure 4B). A previous study suggested a relationship between Myc occupancy and various histone marks (Guccione et al., 2006). Our data confirm this observation on a genome-wide scale, and establish the correlation within ES cells. As anticipated by these histone marks, Myc target genes are more frequently expressed as compared with targets of other factors in ES cells (Figure 4C). Our findings provide evidence that Myc occupancy is associated with large-scale, global alteration of chromatin at Myc targets, and that such effects are qualitatively different from those associated with the core pluripotency factors.

To pursue these correlations, we examined the relationship of the target promoters of seven factors with the H3K4me3 and H3K27me3 marks. We removed Myc and Rex1 from this analysis, because the predicted targets of these factors reveal functional segregation. Interestingly, the predicted target promoters are enriched overall in both H3K4me3 and H3K27me3 marks as compared with all promoters (58% and 26% respectively, Figure 4D). However, closer examination of the correlation of targets of individual factors with these histone marks reveals three different classes (Figure 4E). Target promoters of Nanog, Sox2, Dax1, Oct4, and Klf4 bear enriched marks for both H3K4me3 and H3K27me3. However, predicted targets of Zfp281 show considerable enrichment for the repressive H3K27me3 mark, consistent with a role of Zfp281 in gene repression. Similar to target promoters of Myc, Rex1 and Nac1 targets show less H3K27me3 marks, indicating possible roles of these factors in gene activation.

Recent data with hES cells showed that most H3K27me3 marks overlap H3K4me3 marks (Pan et al., 2007; Zhao et al., 2007). We have observed somewhat less extensive overlap as 35% of H3K27me3 marks (725 among 2046) overlap with H3K4me3 marks. The apparent quantitative difference relates to the threshold level used in assigning target genes. As we reduce the stringency of target selection, we observe a 39% increase in genes with H3K27me3 marks (2046 to 2843). Bivalent signatures increased 138% (725 to 1729) and 61% of H3K27me3 marks then lie within H3K4me3 marks on the promoters. Our results are in accord with the prior observation that H3K27me3 signals are in general weaker than H3K4me3 signals (Pan et al., 2007; Zhao et al., 2007).

Previously identified clusters of gene promoters devoid of H3K4me3 and H3K27me3 marks on their promoters in hES cells (Guenther et al., 2007; Zhao et al., 2007) are also not bound by any of nine factors we tested (Figure S7). Presumably, the mechanism of repression is unique, as H3K27 methylation is one of the principal histone modification marks associated with gene repression.

Regulation of Target Gene Expression by Transcription Factor Occupancy

A striking observation emerges from supervised clustering analysis in considering potential mechanisms that might account for differential regulation of transcription factor targets in ES cells. The genes whose promoters are occupied targets by multiple factors, including Nanog, Sox2, Dax1, Nac1, Oct4, and Klf4, are generally active in ES cells, and repressed upon differentiation (Figure 4A, bar a and Figure S8). On the other hand, the clusters of genes that are inactive or repressed in ES cells, but are expressed upon differentiation, are comprised largely of those gene promoters bound by a single factor (for example, Nanog, Dax1, Klf4, or Zfp281) as shown in Figure 4A, bars b. This observation suggests that the roles of the pluripotency factors are sensitive to their immediate context. A single factor may bind to targets that are "poised" and inactive, or may act to repress its targets, presumably in association with corepression complexes, whereas it may participate in gene activation when bound to a promoter region in concert with other pluripotency regulators. Prior protein network analysis revealed multiple connections between several core factors and repressive chromatin remodeling complexes, including NuRD and Polycomb (PRC1) (Wang et al., 2006).

To pursue this observation, we classified target genes based on the co-occupancy of transcription factors on their promoters. Since Myc and Rex1 have distinct sets of targets (Figures 3E and 4A), we focused on six other factors for further analysis (Nanog, Sox2, Dax1, Nac1, Oct4 and Klf4; Zfp281 was excluded due to less target gene overlap). We observed significant differences between common targets of all 6 factors and targets of single factors in their gene expression profiles, as revealed by gene set enrichment analysis (GSEA) (Subramanian et al., 2005) (Figures 5A–5E). The majority of common targets of six factors are highly active (Figure 5A). Among targets bound by fewer factors, both active and repressed genes are nearly balanced (Figures 5C and S8). Targets occupied by any single factor were predominantly inactive or repressed in ES cells (Figure 5D) and this is even more apparent with the targets of Nanog, Dax1, Klf4, and Zfp281 as shown in Figure 4A, bars b (Figure 5E, 1TF*). The relationship between target promoter occupancy and gene expression level is in excellent accordance with the observed histone marks. The common target promoters of 6 factors show an 80% increase of the H3K4me3 signature and a 60% decrease of H3K27me3 signature, as compared to all promoters. On the other hand, unique targets of only one factor exhibited an increased level of the H3K27me3 signature (Figure 5H, 1TF and 1TF*).

Our findings argue that pluripotency factors act in a highly combinatorial fashion to activate or maintain expression of a subset of target genes, while they are inactive or function more often to repress genes when acting alone, or with only one or few other factors. Distinguishing the "on"-"off" state of targets based on the extent of promoter occupancy may provide a mechanism by which a relatively small set of factors controls two complementary aspects of transcription required for maintenance of pluripotency. While pluripotency factors hold differentiation-promoting genes in check, they must also function together to drive expression of genes encoding proteins required for self-renewal.

In accord with this interpretation, the predicted targets of multiple factors and single factors differ in gene ontology (GO) categories (Figures 5I and 5J). The six factor common target genes are implicated more frequently in developmental processes than targets occupied by fewer factors. The enrichment for genes involved in developmental processes is correlated positively with the number of bound factors from 3–6 (Figure 5I).

Furthermore, we tested roles of each factor in their target gene regulation to determine if any single factor is more associated with either gene activation or repression. Surprisingly, except for Myc and Rex1, all the remaining factors occupy promoters of both nonexpressed and expressed genes (Figure S9). An example is shown in Figure 5F. Among all Nanog target promoters, genes are roughly equally expressed or repressed, whereas among Nanog-only targets, genes are predominantly inactivated or repressed (Figure 5G). This observation is common to all the other factors, except Myc and Rex1 (Figure S10).

Expansion of Core Transcriptional Regulatory Network in ES Cells

Previous studies suggested that transcription factors in ES cells participate in several transcriptional regulatory circuits, including autoregulation, feed-forward regulation and interconnectivity (Boyer et al., 2005). To explore this further, we visualized transcriptional interconnectivity of the nine factors we tested (Figure 6A). Our data describe highly intertwined, complex regulatory circuits exhibiting all three regulatory mechanisms. In addition to autoregulatory mechanisms involving Nanog, Oct4, and Sox2, we observe that Dax1 and Klf4 also display potential autoregulatory loops. Among these five genes, Nanog, Oct4, Sox2, and Dax1 are

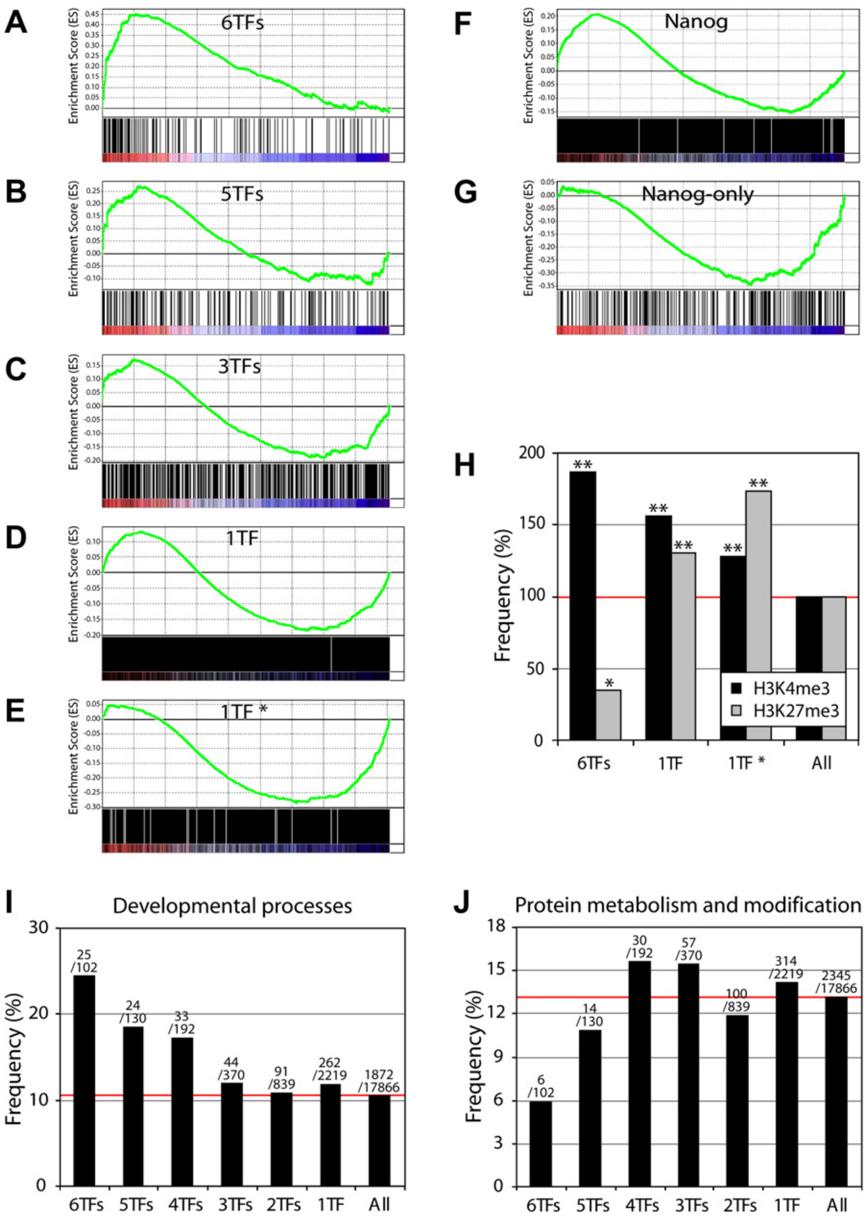


Figure 5. Target Gene Expression and Transcription Factor Occupancy on Their Target Promoters

(A–E) GSEA analyses showing the relationship between target gene expression and factor occupancy. Target promoters were classified based on the number of co-occupying factors and corresponding gene expression upon differentiation was tested. Common targets of six factors (A) are enriched in active genes in ES cells, whereas single-factor only targets are more repressed. (D) "1TF*" represents a subset of "1TF," which includes promoters solely occupied by either Nanog, Dax1, Klf4, or Zfp281 as described in Figure 4A, bars "b" (E).

(F and G) Nanog targets are both active and repressed in ES cells (F), however targets only occupied by Nanog are repressed (G).

(H) Common target promoters of six factors (Nanog, Sox2, Dax1, Nac1, Oct4, and Klf4) are enriched for H3K4me3 marks and reduced for H3K27me3 marks. Promoters occupied by only one factor show an increase in H3K27me3 marks (1TF and 1TF*). Double asterisk indicates hypergeometric probability < 0.0001, and single asterisk indicates hypergeometric probability = 0.006.

(I and J) Genes of multiple factor targets (at least 4TFs) are enriched in developmental processes. Percentages of gene hit against total number of genes are indicated on the *y* axis, and actual numbers of genes are also shown (hit/total).

target hubs of at least 4 of the nine tested factors (and four of six factors: Nanog, Sox2, Dax1, Nac1, Oct4, and Klf4).

Combining transcription regulatory networks and protein interaction networks facilitates a comprehensive understanding of differential gene expression regulation in complex genomes (Walhout, 2006). We asked if the interconnectivity of nine factors might be useful to expand the core transcriptional network by combining target data with protein-protein interaction data. We merged our transcriptional regulatory network with the protein interaction network we previously reported (Figure 6B) (Wang et al., 2006). The initial protein network is comprised 35 proteins, the majority of which are essential to pluripotency and/or early development. Surprisingly, promoters of 77% of the protein network genes are occupied by at least one of the nine factors tested (27 of 35, p value $< 2.4 \times 10^{-7}$). Eleven of 35 genes are occupied by at least four factors (of any nine factors, p value $< 9.1 \times 10^{-8}$). Nine of 35 are occupied by at least 4 of 6 factors (Nanog, Sox2, Dax1, Nac1, Oct4, and Klf4, p value $< 5.3 \times 10^{-8}$). In Figure 6B, target interactions are depicted with the size of each circle reflecting the degree of factor co-occupancy of the promoter of the gene encoding each factor. In this manner, we identify additional target hubs (by four of nine factors), including Dax1, Rest, Rif1, Rex1, Sall4, Rybp, Sall1, Ewsr1, and SP1 within the interactome, in addition to previously accepted target hubs (Nanog, Oct4, and Sox2) (Figure 6B). Independent evidence demonstrating the importance of several of these factors (Sall4 [Sakaki-Yumoto et al., 2006; Zhang et al., 2006], Rif1 [Loh et al., 2006], Rybp [Pirity et al., 2005]) in maintaining pluripotency is consistent with this network architecture. Taken together, our data reveal that the actual core factor set in ES cells is larger and more highly interconnected than previously suspected (Loh et al., 2006; Wang et al., 2006; Zhang et al., 2006).

In addition to the target hubs within the network, we have identified many additional targets of multiple factors. These targets are highly likely to be important in self-renewal and lineage commitment (Jeong et al., 2001). Table 1 lists DNA-binding (or chromatin-associated) proteins whose promoters are occupied by multiple factors (at least five of six factors, Nanog, Dax1, Sox2,

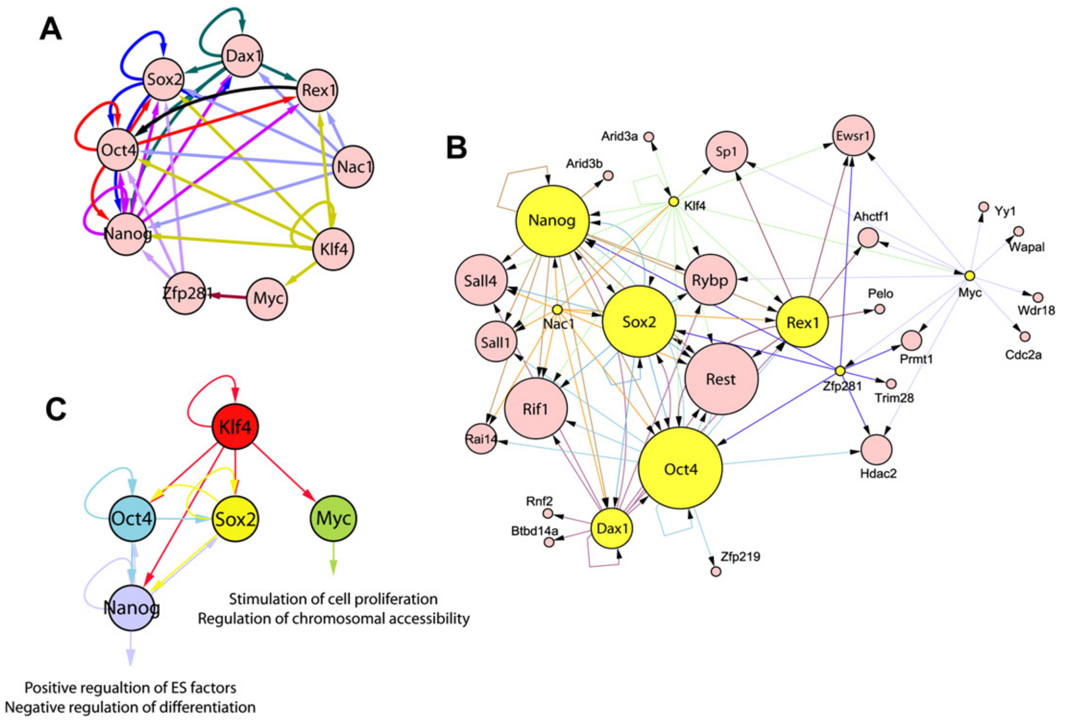


Figure 6. Expanded Transcriptional Regulatory Network and Regulatory Circuit within Four Somatic Cell Reprogramming Factors and Nanog

(A) Transcriptional regulatory circuit within nine factors. Five factors (Nanog, Oct4, Sox2, Dax1, and Klf4) show autoregulatory mechanism.

(B) Expanded transcriptional regulatory network showing target hubs of multiple factors within the previously identified protein interaction network. Yellow circles represent nine factors examined. The size of each circle reflects the degree of factor co-occupancy. Arrowhead indicates the direction of transcriptional regulation (A–C). Sox2, Klf4, and Myc were not in the original protein interaction network (Wang et al., 2006).

(C) Transcriptional regulatory circuit within four somatic cell reprogramming factors and Nanog.

Nac1, Oct4, and Klf4). The predicted targets encode a set of proteins involved in regulation of development decisions, signaling pathways, and chromatin remodeling. Although functional assessment is necessary to determine the roles of many of these proteins in pluripotency and self-renewal, validation of several target genes (Nanog, Oct4, Rest, Sall4, Sox2, Rex1) and identification of several others within the protein interaction network (Wang et al., 2006) make it highly likely that many others in this set will be shown subsequently to be functionally relevant. Moreover, among targets encoding non-DNA associated proteins, many are important in an ES cell context, including Tcl1, which participates in the PI3K/Akt signaling pathway and promotes ES cell proliferation (Ivanova et al., 2006); Il6st (gp130), which is involved in the LIF/STAT3 pathway (Ernst et al., 1996; Yoshida et al., 1994); and Bmp4, a critical signaling molecule for early differentiation and ES cells. Indeed, Fbxo15, the locus first employed as a marker for somatic reprogramming (Takahashi and Yamanaka, 2006), is also a multifactor target gene.

Regulatory Network within Four Somatic Cell Reprogramming Factors

Fibroblasts of either mouse or human origin can be reprogrammed to a pluripotent ES-like state (iPS cells) upon forced expression of three or four (or more) factors, including Klf4, Oct4, Sox2, Nanog, and c-Myc (Park et al., 2007; Takahashi et al., 2007; Takahashi and Yamanaka, 2006; Wernig et al., 2007b; Yu et al., 2007). iPS cells appear highly similar to conventional ES cells. The regulatory relationships among the reprogramming factors are, therefore, of particular interest in accounting for the potency of this cocktail of factors. The transcriptional hierarchy within the original 4 reprogramming factors is depicted in Figure 6C. In addition to previously identified feed-forward regulation within Oct4, Sox2, and Nanog, our results argue that Klf4 is an upstream regulator of larger feed-forward loops containing Oct4, Sox2, and other common downstream targets, such as Nanog, and also occupies the c-Myc promoter. In addition to the histone marks of Myc targets (discussed above), our findings from target categorization and the predicted regulatory network also support distinct functions of Myc in ES cells. These functions are likely to include positive regulation of proliferation, negative regulation of differentiation, and regulation of chromosomal accessibility of other factors, as previously suggested (Niwa, 2007). The findings described above regarding histone marks associated with Myc occupancy provide experimental evidence in support of these inferences.

DISCUSSION

Here, we demonstrate the utility of in vivo biotinylation of tagged proteins and streptavidin affinity capture to identify global

Table 1. Examples of DNA Binding Proteins that Are Common Targets of Multiple Transcription Factors: At Least Five of Six Factors, Nanog, Dax1, Sox2, Nac1, Oct4, and Klf4

Symbol	Accession Number	Gene Name
6030445D17Rik	NM_177079	Riken cDNA 6030445d17 gene
Ankrd10	NM_133971	Ankyrin repeat domain 10
Asxl1	NM_001039939	Additional sex combs like 1 (drosophila)
Cbx1	NM_007622	Chromobox homolog 1 (drosophila hp1 beta)
Cbx7	NM_144811	Chromobox homolog 7
Cdx1	NM_009880	Caudal type homeo box 1
Chd9	NM_177224	Chromodomain helicase dna binding protein 9
Dido1	NM_175551	Death inducer-obliterator 1
E2f4	NM_148952	E2F transcription factor 4
Evx1	NM_007966	Even skipped homeotic gene 1 homolog
Fubp3	NM_001033389	Far upstream element (fuse) binding protein 3
Gbx2	NM_010262	Gastrulation brain homeobox 1
Grhl3	NM_001013756	Grainyhead-like 3 (drosophila)
H2afx	NM_010436	H2A histone family, member x
Hist1h2an	NM_178184	Hypothetical protein 1190022l06
Hist1h3i	NM_178207	Histone 1, h3g
Hnrpdl	NM_016690	Heterogeneous nuclear ribonucleoprotein d-like
Hoxb13	NM_008267	Homeo box b13
Jarid2	NM_021878	Jumonji, at rich interactive domain 2
Klf2	NM_008452	Kruppel-like factor 2 (lung)
Klf9	NM_010638	Kruppel-like factor 9
Max	NM_008558	Max protein
Mllt6	NM_139311	Myeloid/lymphoid or mixed lineage-leukemia translocation to 6 homolog (drosophila)
Msh6	NM_010830	Muts homolog 6 (e. coli)
Msx2	NM_013601	Homeo box, msh-like 2
Mybl2	NM_008652	Myeloblastosis oncogene-like 2
Myst2	NM_177619	Myst histone acetyltransferase 2
Mzf1	NM_145819	Myeloid zinc finger 1
Nanog	NM_028016	Nanog homeobox
Nkx2-2	NM_010919	Nk2 transcription factor related, locus 2 (drosophila)
Otx2	NM_144841	Orthodenticle homolog 2 (drosophila)
Pax6	NM_013627	Paired box gene 6
Phc1	NM_007905	Polyhomeotic-like 1 (drosophila)
Pou5f1	NM_013633	Pou domain, class 5, transcription factor 1
Rarg	NM_001042727	Retinoic acid receptor, gamma
Rax	NM_013833	Retina and anterior neural fold homeobox
Rbbp5	NM_172517	Riken cDNA 4933411j24 gene
Rest	NM_011263	Re1-silencing transcription factor
Rnf12	NM_011276	Ring finger protein 12
Sall4	NM_175303	Testis expressed gene 20
Sox13	NM_011439	Sry-box containing gene 13
Sox2	NM_011443	Sry-box containing gene 2
Spic	NM_011461	Spi-c transcription factor (spi-1/pu.1 related)
T	NM_009309	Brachyury
Tbx3	NM_198052	T-box 3
Tcea3	NM_011542	Transcription elongation factor a (sii), 3
Tcfap2c	NM_009335	Transcription factor ap-2, gamma
Tcfcp2l1	NM_023755	Riken cDNA 4932442m07 gene
Tgif	NM_009372	TG interacting factor
Trib3	NM_144554	Induced in fatty liver dystrophy 2
Trib3	NM_175093	Induced in fatty liver dystrophy 2
Trp53bp1	NM_013735	Transformation related protein 53 binding protein 1
Zfp13	NM_011747	Zinc finger protein 13
Zfp206	NM_001033425	Zinc finger protein 206
Zfp36l1	NM_007564	Zinc finger protein 36, c3h type-like 1
Zfp42	NM_009556	Zinc finger protein 42
Zfp704	NM_133218	Zinc finger protein 704
Zic2	NM_009574	Zic finger protein of the cerebellum 2
Zic5	NM_022987	Zinc finger protein of the cerebellum 5

targets of multiple factors involved in the transcriptional control of pluripotency in ES cells. Our approach provides a degree of consistency in the experimental platform generally not attainable in ChIP-Chip experiments that rely on diverse antibodies of unknown specificity and sensitivity. We suggest that the bioChIP-Chip method may serve as a useful tool for assessing the quality of native antibodies, given that there is no simple a priori method for determining the suitability of a given antibody for ChIP procedures. As cell lines expressing tagged proteins may also be employed to study protein-protein interactions, the generation of two independent data-rich resources can be achieved with a single cell line "reagent" and similar procedures.

Our studies not only suggest a more comprehensive and complex view of the pluripotency network in ES cells than prior work (Boyer et al., 2005; Loh et al., 2006), but also provide additional insights into specific regulatory features and an extensive database for further exploration. First, by mapping promoter occupancy of nine factors, including the original four somatic cell reprogramming factors, we have uncovered remarkable combinatorial binding at many targets: 800 gene promoters are bound by four or more transcription factors of those tested (Figure 3C).

Second, whereas numerous targets are shared by an extended set of pluripotency factors (Oct4, Sox2, Nanog, Klf4, Dax1, Zfp281, and Nac1), the targets of c-Myc (and also Rex1) largely fall into a different cluster (Figures 3E and 4A). Third, we have discovered a striking correlation between the number of bound factors and the likelihood that a target gene is expressed in wild-type ES cells and then repressed on differentiation (Figures 4A and 5). These observations provide a means for direct involvement of these factors in promoting self-renewal by activating expression of those genes (including the pluripotency factors themselves) and simultaneously inhibiting expression of differentiation-promoting genes. One possibility is that the pluripotency factors individually serve as weak activators, and that multi-factor binding augments activator function. A priori, the converse situation might have applied; that is, multifactor binding would predominantly be associated with repression. Despite the presence of pluripotency factors in protein complexes with corepressor components (Wang et al., 2006), our findings are inconsistent with this possibility. Another notable observation regarding histone marks is that the Polycomb targets are largely different from common targets of multiple core transcription factors. We demonstrate that common targets of multiple factors are active in ES cells and their histone marks show distinct patterns (Figure 5H). Fourth, by combining target promoter occupancy data with our prior protein interaction network, we identified additional regulatory hubs, defined as those gene promoters bound by multiple factors (Figure 6B). These new hubs include Sall4, Rif1, Rest, and Dax1, all of which have been shown to be important for ES cell properties in independent studies. Fifth, our studies suggest a hierarchy within the four somatic reprogramming factors, such that Klf4 serves as an upstream regulator of feed-forward circuits involving Oct4 and Sox2, as well as more downstream effectors (e.g., Nanog), and is predicted to regulate c-Myc based on promoter occupancy (Figure 6C).

In addition to the inferences gained regarding the pluripotency factors themselves, our data provide insight into how c-Myc differs from these core factors in regulating its targets. In addition to sharing few targets, c-Myc occupancy is associated with striking differences in associated histone marks (Figures 4A, 4B, and 4E), and with enrichment for expressed genes (Figure 4C). These findings are consistent with the view that c-Myc occupancy is associated with broad changes in chromatin accessibility. This unique target regulation by Myc may account for its capacity to enhance reprogramming, while also being dispensable as an exogenous factor (Nakagawa et al., 2007; Wernig et al., 2007a).

The discovery of a class of predicted targets bound by multiple (>4) pluripotency transcription factors (Figures 4A, 5A, and 5B and Table 1) is of interest as these genes are largely expressed in ES cells and repressed on differentiation. As this class includes several genes within the protein interaction network (e.g., Nanog, Oct4, Rest, Sall4, Sox2), it is likely that additional genes within this set, that have not as yet been evaluated for potential roles in ES cells, will prove critical to the maintenance of pluripotency. The recognition that human skin cells can be reprogrammed to iPS cells with the same factors that are active in mouse cells (Park et al., 2007; Takahashi et al., 2007; Yu et al., 2007) provides strong evidence in favor of common networks in pluripotent mouse and human cells, despite differences in the growth factor requirements and behavior of cultured mouse and human ES cells. Our data constitute a framework for further exploration of the complex transcriptional network dedicated to establishment and preservation of pluripotency.

EXPERIMENTAL PROCEDURES

ES Cell Lines and Culture

Mouse J1 ES cell lines were maintained in ES medium as described previously (Wang et al., 2006). Briefly, cells were maintained in ES medium (DMEM; Dulbecco's modified Eagle's medium) supplemented with 15% fetal calf serum, 0.1 mM β-mercaptoethanol, 2 mM L-glutamine, 0.1 mM nonessential amino acid, 1% of nucleoside mix (100× stock, Sigma), 1000 U/ml recombinant leukemia inhibitory factor (LIF; Chemicon) and 50 U/ml Penicillin/Streptomycin. Further details are documented in Supplemental Data.

Chromatin Immunoprecipitation and Antibodies

ChIP reactions and bioChIP reactions were performed as described previously with minor modifications (Kim et al., 2005). For bioChIP reactions, streptavidin beads (Dynabeads MyOne Streptavidin T1) were used to precipitate chromatin, and 2% SDS was applied for one of the washing steps. Further details are described in Supplemental Data. At least three biological replicates were performed in each case. Detailed procedure, list of antibodies, list of primers used for ChIP-PCR validation are available in Supplemental Data.

Microarray and Data Processing

Ligation-mediated PCR was performed to amplify ChIP samples as described previously (Ren et al., 2000). Microarray hybridizations were performed on the Affymetrix GeneChip Mouse promoter 1.0R arrays and Model-based Analysis of Tiling-array (MAT) was applied to predict the target loci (Johnson et al., 2006). Further details are available in Supplemental Data.

Histone Modification Signatures

The significance of enrichment or depletion of H3K4me3 or H3K27me3 signature on the promoters occupied by each factor (Figure 4E) or multiple factors (Figure 4D and Figure 5H) was evaluated by hypergeometric distribution test (asterisk or double asterisk).

SUPPLEMENTAL DATA

Supplemental Data include Supplemental Experimental Procedures, ten figures, one table, Supplemental References, and three Excel spreadsheets and can be found with this article online at http://www.cell.com/cgi/content/full/132/6/1049/DC1/.

ACKNOWLEDGMENTS

We thank the Gene Expression Microarray Core at the DFCI for sample processing. J.K. is a Howard Hughes Medical Institute Research Associate. S.H.O. is an Investigator of the Howard Hughes Medical Institute.

Received: October 24, 2007
Revised: January 7, 2008
Accepted: February 25, 2008
Published: March 20, 2008

REFERENCES

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

Bailey, T.L., Williams, N., Misleh, C., and Li, W.W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic. Acids Res. *34*, 369–373.

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.

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.

Cowan, C.A., Atienza, J., Melton, D.A., and Eggan, K. (2005). Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. Science *309*, 1369–1373.

de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F., and Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. Proc. Natl. Acad. Sci. USA *100*, 7480–7485.

Ernst, M., Oates, A., and Dunn, A.R. (1996). Gp130-mediated signal transduction in embryonic stem cells involves activation of Jak and Ras/mitogen-activated protein kinase pathways. J. Biol. Chem. *271*, 30136–30143.

Euskirchen, G.M., Rozowsky, J.S., Wei, C.L., Lee, W.H., Zhang, Z.D., Hartman, S., Emanuelsson, O., Stolc, V., Weissman, S., Gerstein, M.B., et al. (2007). Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies. Genome Res. *17*, 898–909.

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

Fernandez, P.C., Frank, S.R., Wang, L., Schroeder, M., Liu, S., Greene, J., Cocito, A., and Amati, B. (2003). Genomic targets of the human c-Myc protein. Genes Dev. *17*, 1115–1129.

Guccione, E., Martinato, F., Finocchiaro, G., Luzi, L., Tizzoni, L., Dall'Olio, V., Zardo, G., Nervi, C., Bernard, L., and Amati, B. (2006). Myc-binding-site recognition in the human genome is determined by chromatin context. Nat. Cell Biol. *8*, 764–770.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77–88.

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.

Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature *409*, 533–538.

Jeong, H., Mason, S.P., Barabasi, A.L., and Oltvai, Z.N. (2001). Lethality and centrality in protein networks. Nature *411*, 41–42.

Johnson, W.E., Li, W., Meyer, C.A., Gottardo, R., Carroll, J.S., Brown, M., and Liu, X.S. (2006). Model-based analysis of tiling-arrays for ChIP-chip. Proc. Natl. Acad. Sci. USA *103*, 12457–12462.

Kim, J., Bhinge, A.A., Morgan, X.C., and Iyer, V.R. (2005). Mapping DNA-protein interactions in large genomes by sequence tag analysis of genomic enrichment. Nat. Methods *2*, 47–53.

Kuhn, R.M., Karolchik, D., Zweig, A.S., Trumbower, H., Thomas, D.J., Thakkapallayil, A., Sugnet, C.W., Stanke, M., Smith, K.E., Siepel, A., et al. (2007). The UCSC genome browser database: update 2007. Nucleic Acids Res. *35*, 668–673.

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.

Li, Z., Van Calcar, S., Qu, C., Cavenee, W.K., Zhang, M.Q., and Ren, B. (2003). A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. Proc. Natl. Acad. Sci. USA *100*, 8164–8169.

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.

Macisaac, K.D., Gordon, D.B., Nekludova, L., Odom, D.T., Schreiber, J., Gifford, D.K., Young, R.A., and Fraenkel, E. (2006). A hypothesis-based approach for identifying the binding specificity of regulatory proteins from chromatin immunoprecipitation data. Bioinformatics *22*, 423–429.

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.

Mao, D.Y., Watson, J.D., Yan, P.S., Barsyte-Lovejoy, D., Khosravi, F., Wong, W.W., Farnham, P.J., Huang, T.H., and Penn, L.Z. (2003). Analysis of Myc bound loci identified by CpG island arrays shows that Max is essential for Myc-dependent repression. Curr. Biol. *13*, 882–886.

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.

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.

Mi, H., Guo, N., Kejariwal, A., and Thomas, P.D. (2007). PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. Nucleic Acids Res. *35*, D247–D252.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. Nat. Genet. *37*, 1090–1097.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2007). Histone replacement marks the boundaries of cis-regulatory domains. Science *315*, 1408–1411.

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.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2007). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat. Biotechnol. *26*, 101–106.

Niakan, K.K., Davis, E.C., Clipsham, R.C., Jiang, M., Dehart, D.B., Sulik, K.K., and McCabe, E.R. (2006). Novel role for the orphan nuclear receptor Dax1 in embryogenesis, different from steroidogenesis. Mol. Genet. Metab. *88*, 261–271.

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

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

Orkin, S.H. (2005). Chipping away at the embryonic stem cell network. Cell *122*, 828–830.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell *1*, 299–312.

Park, I.H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2007). Reprogramming of human somatic cells to pluripotency with defined factors. Nature *451*, 141–146.

Pirity, M.K., Locker, J., and Schreiber-Agus, N. (2005). Rybp/DEDAF is required for early postimplantation and for central nervous system development. Mol. Cell. Biol. *25*, 7193–7202.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Genome-wide location and function of DNA binding proteins. Science *290*, 2306–2309.

Roh, T.Y., Ngau, W.C., Cui, K., Landsman, D., and Zhao, K. (2004). High-resolution genome-wide mapping of histone modifications. Nat. Biotechnol. *22*, 1013–1016.

Sakaki-Yumoto, M., Kobayashi, C., Sato, A., Fujimura, S., Matsumoto, Y., Takasato, M., Kodama, T., Aburatani, H., Asashima, M., Yoshida, N., et al. (2006). The murine homolog of SALL4, a causative gene in Okihiro syndrome, is essential for embryonic stem cell proliferation, and cooperates with Sall1 in anorectal, heart, brain and kidney development. Development *133*, 3005–3013.

Silva, J., Chambers, I., Pollard, S., and Smith, A. (2006). Nanog promotes transfer of pluripotency after cell fusion. Nature *441*, 997–1001.

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

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

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.

Walhout, A.J. (2006). Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. Genome Res. *16*, 1445–1454.

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.

Wernig, M., Meissner, A., Cassady, J.P., and Jaenisch, R. (2007a). C-Myc is dispensable for direct reprogramming of mouse fibroblasts. Cell Stem Cell *2*, 10–12.

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

Wu, Q., Chen, X., Zhang, J., Loh, Y.H., Low, T.Y., Zhang, W., Sze, S.K., Lim, B., and Ng, H.H. (2006). Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. J. Biol. Chem. *281*, 24090–24094.

Yoshida, K., Chambers, I., Nichols, J., Smith, A., Saito, M., Yasukawa, K., Shoyab, M., Taga, T., and Kishimoto, T. (1994). Maintenance of the pluripotential phenotype of embryonic stem cells through direct activation of gp130 signalling pathways. Mech. Dev. *45*, 163–171.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917–1920.

Zhang, J., Tam, W.L., Tong, G.Q., Wu, Q., Chan, H.Y., Soh, B.S., Lou, Y., Yang, J., Ma, Y., Chai, L., et al. (2006). Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. Nat. Cell Biol. *8*, 1114–1123.

Zhao, X.D., Han, X., Chew, J.L., Liu, J., Chiu, K.P., Choo, A., Orlov, Y.L., Sung, W.-K., Shahab, A., Kuzetsov, V.A., et al. (2007). Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. Cell Stem Cell *1*, 286–298.

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.