Cold Spring Harbor Symposia on Quantitative Biology

The Transcriptional Network Controlling Pluripotency in ES Cells

S.H. Orkin, J. Wang, J. Kim, et al.

Cold Spring Harb Symp Quant Biol 2008 73: 195-202 originally published online November 6, 2008
Access the most recent version at doi:10.1101/sqb.2008.72.001

The Transcriptional Network Controlling Pluripotency in ES Cells

S.H. ORKIN, J. WANG, J. KIM, J. CHU, S. RAO, T.W. THEUNISSEN, X. SHEN, AND D.N. LEVASSEUR
Department of Pediatric Oncology, Dana Farber Cancer Institute and Children's Hospital Boston, Howard Hughes Medical Institute, Harvard Stem Cell Institute, Boston, Maassachusetts 02115

Embryonic stem (ES) cells are capable of continuous self-renewal and pluripotential differentiation. A "core" set of transcription factors, Oct4, Sox2, and Nanog, maintains the ES cell state, whereas various combinations of factors, invariably including Oct4 and Sox2, reprogram somatic cells to pluripotency. We have sought to define the transcriptional network controlling pluripotency in mouse ES cells through combined proteomic and genomic approaches. We constructed a protein interaction network surrounding Nanog and determined gene targets of the core and reprogramming factors, plus others. The expanded transcriptional network we have constructed forms the basis for further studies of directed differentiation and lineage reprogramming, and a paradigm for comprehensive elucidation of regulatory pathways in other stem cells.

Stem cells are distinguished by two properties: self-renewal and the potential to differentiate. Our understanding of the mechanisms underlying these features in vertebrate stem cells is incomplete. Whereas powerful genetic approaches in invertebrates permit delineation of the cellular components required for self-renewal and differentiation, as well as for interactions of stem cells with their niches, analogous strategies are not readily applicable to vertebrate stem cells. In considering this landscape, we chose to focus on mouse ES cells (Evans and Kaufman 1981; Martin 1981) as a tractable system in which to apply biochemical and genetic methods with the goal of providing a comprehensive description of the networks controlling self-renewal and pluripotency. Although ES cells may represent a "special case" given their origin and the artificial manner in which they are maintained in culture, they provide a convenient source of unlimited, quite homogeneous, self-renewing stem cells for biochemical studies. Moreover, the facility with which ES cells may be modified by gene-targeting or other loss-of-function approaches (e.g., si/shRNA inhibition) permits functional assessment of the contribution of specific components to the pluripotent state. Hence, in the work summarized here, our goal has been to identify in a comprehensive, relatively unbiased manner the proteins critical for maintenance of pluripotency and self-renewal and to delineate how they act individually and together in these processes. These approaches should lead to general concepts that may be applied to other stem cell systems and suggest methods for improved reprogramming of somatic cells to pluripotency (Takahashi and Yamanaka 2006; Takahashi et al. 2007) or to alternative fates.

ITERATIVE AFFINITY PROTEIN PURIFICATION FOR GENERATION OF A PLURIPOTENCY PROTEIN–PROTEIN INTERACTION NETWORK: STRATEGIC CONSIDERATIONS

To provide an initial point of reference in our analysis, we elected to focus on the transcription factor Nanog in light of its capacity to drive mouse ES cell self-renewal in the absence of leukemia inhibitory factor (LIF) in the medium (Chambers et al. 2003; Mitsui et al. 2003) and facilitate fusion-induced cellular reprogramming (Silva et al. 2006). In addition, loss of Nanog leads to cellular differentiation, specifically along the primitive endodermal pathway. As an initial hypothesis, we entertained the possibility that Nanog, as a critical regulatory factor, might interact physically with other proteins that participate in maintenance of pluripotency (Fig. 1). If so, purification of Nanog with its associated proteins would serve as a tool for the discovery of novel proteins involved in pluripotency and/or connect Nanog to already recognized proteins (such as Oct4 or Sox2). At one extreme, proteins involved in pluripotency might be "concentrated" within a subnetwork among all cellular proteins (Fig. 2) (Dezso et al. 2003). In this manner, the proteins would "talk" to

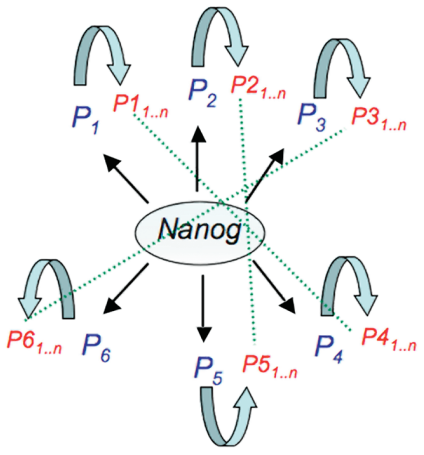

Figure 1. Iterative identification of Nanog-associated proteins. Proteins that associate with Nanog in protein complexes are depicted as P1....P6. Subsequent isolation of complexes containing P1....P6 in an iterative fashion leads to identification of secondary interaction proteins, e.g., $P1_{1...n}$. In some instances (*dashed lines*), the same protein will be recovered in different protein complexes.

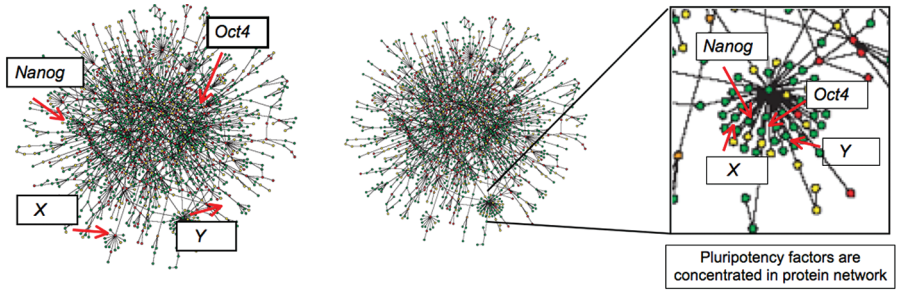

Figure 2. Alternative organization of pluripotency factors within the cellular proteome. (*Left*) Pluripotency factors Nanog and Oct4, as well as unknown factors X and Y, do not associate in shared protein complexes within the cellular proteome (indicated by the "*hairball*"). (*Right*) Pluripotency factors are concentrated within shared protein complexes.

one another to create a highly interactive network. At the other extreme, however, Nanog or other pluripotency factors would not associate within shared protein complexes but rather converge downstream in the control of critical target genes (Fig. 2).

We considered several alternative experimental strategies for identifying Nanog-associated proteins. Yeast two-hybrid screening using an ES cDNA library would provide a convenient, high-throughput discovery approach (Li et al. 2004). However, the nonphysiologic nature of this method, taken together with underrepresentation of DNA-binding transcription factors in large-scale two-hybrid screens, argued against this strategy. Instead, we favored purification of protein complexes under relatively physiologic conditions to approximate the in vivo setting as closely as possible. At the time we initiated our experiments, antibodies to the Nanog protein were not readily available. Hence, we chose to engineer into expressed Nanog protein an affinity tag suitable for protein purification. As part of this strategy, we also envisioned iterative purification of protein complexes, because Nanog-associated proteins could be used as "hooks" for the isolation of their associated proteins. In this iterative fashion, a protein network could be extended from a central point, the Nanog protein, to many other proteins. Hence, the ease with which affinity purification could be applied to purification of numerous proteins was given high priority.

With these issues in mind, we chose to use in vivo biotinylation of proteins, coupled with streptavidin affinity capture (Wang et al. 2006). Although the method was described more than 10 years ago, Strouboulis and colleagues revived its use (de Boer et al. 2003), as illustrated by one-step purification of complexes containing the GATA-1 transcription factor (Rodriguez et al. 2005). In this approach, the *Escherichia coli* biotin ligase (*BirA)* gene is stably introduced in a suitable expression vector into a host cell. Subsequently, an expression vector harboring the cDNA of interest with a short biotin ligase substrate tag is also stably transferred to the *BirA*-expressing host cell. Following expression of polypeptide, the substrate tag is biotinylated by *BirA.* To facilitate tandem affinity purification, we engineered a FLAG epitope tag into the substrate. After either streptavidin bead capture or tandem FLAG-antibody immunoprecipitation followed by streptavidin capture, samples are subjected to whole-lane liquid chromatography–tandem mass spectrometry (LC-MS/MS). Putative associated proteins are revealed by the peptide sequences obtained. Validation of protein association can be accomplished by conventional immunoprecipitation experiments.

In the application of this method, we have been careful to choose ES cell clones that express low levels of the exogenous cDNA, because high-level expression of critical proteins could affect the protein network itself and the behavior of ES cells (Wang et al. 2006). Indeed, in most instances, we have expressed exogenous proteins at well below their endogenous level. As noted below, ES cells expressing tagged proteins may also be used to identify DNA targets of the respective proteins. Hence, a single cellular platform can be used for protein interaction and target gene analyses (Kim et al. 2008).

A PROTEIN INTERACTION NETWORK SURROUNDING NANOG

We first isolated protein complexes containing Nanog and then proceeded to tag several of the partner proteins in an iterative fashion, each time microsequencing the recovered proteins (Wang et al. 2006). Large-scale purifications were performed with both one-step (streptavidin capture alone) or tandem (FLAG-immunoprecipation followed by streptavidin capture) in order to identify as many associated proteins as possible. Stringent criteria were applied for selection of candidate interacting proteins (Wang et al. 2006).

The protein interaction network surrounding Nanog is depicted in Figure 3. Tagged proteins used as "baits" are shown in red. Lines connect proteins that were present together in isolated complexes. Given that proteins may be brought into complexes through secondary protein interactions, rather than by direct interaction with the tagged protein itself, this representation does not specify the number or variety of actual protein complexes. We suspect that there are many protein complexes, often containing shared components.

Several features of the protein interaction network are notable. First, Nanog is, indeed, connected to other critical pluripotency factors through its associated proteins. For example, Oct4 is recovered in Nanog-associated complexes. Concurrent or subsequent studies have also identified additional associated proteins, including Dax1, Sall4,

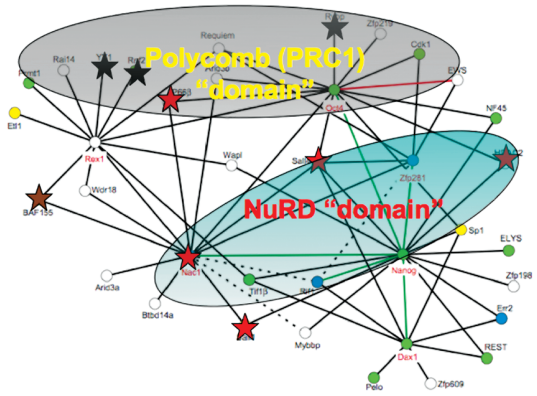

Figure 3. Protein interaction network surrounding Nanog. (*Red*) Proteins tagged by biotin and used for protein purification and peptide microsequencing, including Nanog, Oct4, Nac1, Dax1, Zfp281, and Rex1. (*Green circles*) Genes that are essential for early mouse development or maintenance of the ES cell state. (*Yellow circles*) Proteins that are dispensable. (*Red, black, brown stars*) HDAC/NuRD, PRC1, and Swi/Snf components, respectively. (Modified from Wang et al. 2006.)

Rif1, REST, Zfp281, Nac1, and Errb, among others (Loh et al. 2006; Niakan et al. 2006; Wang et al. 2006; Wu et al. 2006; Zhang et al. 2006; Singh et al. 2008), as being important within an ES cell context. Thus, our findings argue that pluripotency factors are highly concentrated within a subdomain of all proteins within ES cells (Fig. 2). This organization provides a simple rationale for the dose-dependent action of many of these factors, as higher-order protein complexes assemble, disassemble, and presumably compete for shared components.

Second, proteins within the Nanog interaction network associate with various components of chromatin remodeling or corepression complexes. Nanog and its immediate partners associate with histone deacetylases (HDACs) and the NuRD remodeling complex. Oct4 and its more immediate partners associate with components of the PRC1 polycomb complex. In addition, at least one connection to the Swi/Snf ATPase-dependent remodeling complex is evident. These relationships provide the means by which cell-specific factors within ES cells are linked to rather general modulators of transcription, largely implicated in transcriptional repression. Silencing differentiation-promoting genes is an essential role of pluripotency factors. The protein interaction network suggests pathways through which repression is maintained.

Third, an unexpectedly high fraction of proteins within the network are essential either for early mouse development and/or for maintenance of ES cell pluripotency. The vast majority of proteins are required based on either knockout or knockdown studies. Thus, the network, although highly interconnected, is also susceptible to breakdown through loss of any one of many components.

Fourth, consistent with this last point, our studies (as well as concurrent work by others) establish several "new" proteins beyond the pluripotency core factors Oct4, Sox2, and Nanog that must be considered in a broader accounting of transcriptional control in ES cells. These proteins include Dax1, Sall4, REST, Rif1, Zfp281, Errb, and Nac1. For example, depletion of these components by short hairpin RNA (shRNA) leads to differentiation of ES cells. In the studies leading to construction of the initial Nanog protein interaction network, neither the core factor Sox2 nor the reprogramming factor Klf4 was identified. Subsequent work in our laboratory has linked these proteins to the Nanog network through affinity purification of Sox2 and Klf4 complexes (J. Chu and S.H. Orkin, unpubl.). Thus, the majority of transcription factors contributing to the maintenance of pluripotency in ES cells are contained within the broad Nanog protein interaction network. Consistent with this finding is the observation that most of the proteins within the network are down-regulated upon differentiation of ES cells.

Finally, taken together, the proteomic data suggest that the Nanog protein interaction network operates as a cellular "module" dedicated to pluripotency in ES cells. A priori, this might not have been anticipated. On reconsideration, however, the connection of the critical regulators within a subdomain of cellular proteins reflects parsimony in the evolution of pluripotency control.

IDENTIFICATION OF TARGET GENES OF THE PLURIPOTENCY NETWORK

The advent of chromatin immunoprecipitation (ChIP)-on-Chip, ChIP-PET, and ChIP-Sequencing (ChIP-Seq) methods now permits global identification of DNA targets of transcription factors. Initial work by other investigators indicated that the core pluripotency factors Nanog, Oct4, and Sox2 each bind several hundred putative target loci and also cooccupy many gene promoters (Boyer et al. 2005; Loh et al. 2006). For example, Boyer et al. (2005) reported that Nanog, Oct4, and Sox2 cooccupy approximately 350 target genes in human ES cells. In addition, each binds to its own regulatory sequences and those of other core members, leading to feed-forward and autoregulatory circuits. In our studies, we sought to develop a more comprehensive view of the transcriptional circuitry by determining the putative direct targets of additional pluripotency factors from the Nanog interaction network.

In an effort to maximize consistency in the experimental platform, we first evaluated the suitability of biotinylated factors expressed in our bank of tagged ES cell lines for ChIP-on-Chip analyses using streptavidin bead capture in place of conventional immunoprecipitation (Kim et al. 2008). Standard ChIP-on-Chip and biotin–ChiP-on-Chip analyses of promoter arrays for Nanog and c-Myc were comparable. Moreover, we demonstrated that low-level expression of tagged Nanog does not affect the sensitivity or range of targets identified by conventional ChIP-on-Chip. Because it is often challenging to identify suitable quality antibodies for ChIP-on-Chip studies, use of the biotin–ChIP-on-Chip method provides a convenient alternative to the conventional approach. The extraordinary avidity of streptavidin–biotin interactions also allows for the use of more stringent washing conditions. As a consequence, we believe that the sensitivity and specificity of biotin–ChIP-on-Chip should exceed that of other methods. The platform should be readily

applicable to the emerging ChIP-Seq strategies that are especially promising.

In our studies, we initially determined promoter-binding targets for nine proteins within the ES cell protein network. This set includes the core pluripotency factors (Nanog, Oct4, and Sox2), the "Yamanaka reprogramming set" (Oct4, Sox2, c-Myc, and Klf4) (Takahashi and Yamanaka 2006), and others (Dax1, Rex1, Zfp281, and Nac1). We chose to use promoter arrays, rather than whole-genome arrays, because most binding within the +8- to –2-kb window of the Affymetrix array (relative to the transcriptional start site [TSS]) was observed within a few hundred base pairs of the TSS. Moreover, although characterization of distant binding sites might eventually be of interest, it is difficult to assign specific target genes with far-distant binding events. Stringent threshold criteria were applied to choose putative targets. Direct ChIP analyses demonstrated a false-positive assignment rate of <5%. Table 1 lists the numbers of putative target promoters for each of the factors studied.

Several general conclusions emerge from review of the data. First, the target loci shared by the principal pluripotency factors (excluding Rex1) are largely distinct from those bound by c-Myc. Second, c-Myc targets tend to be largely expressed in ES cells, rather than either expressed or "off" (or repressed), as is the case for pluripotency factor targets. Consistent with this observation, we find that targets of c-Myc are highly associated with the active H3K4me3 chromatin mark. As such, we speculate that c-Myc binding is associated with global effects on chromatin accessibility, a finding that may account for its role in facilitating somatic cell reprogramming to a pluripotent state. Third, closer inspection of target loci of the pluripotency factors reveals a striking association between the number of factors bound to a promoter region and the likelihood of target gene expression in undifferentiated ES cells. Remarkably, approximately 800 targets are bound by four or more of the nine factors we analyzed, and 450 targets are bound by five or more (Fig. 4). These "multifactor" binding loci tend to be expressed rather than "off" or repressed in ES cells, and then turned off on differentiation. In marked contrast, loci that are bound by less than four factors tend to be silent in ES cells and then expressed upon differentiation. This correlation is particularly striking for target loci bound by only a single factor. Thus, the extent of factor binding appears to correlate with gene expression in ES cells and also segregates target loci into two broad classes.

Table 1. Numbers of promoters occupied by transcription factors in ES cells

Protein	Number of promoters
Nanog	1284
Sox2	819
Dax1	1754
NacI	804
Oct4	783
Klf4	1790
Zfp281	601
Rex1	1543
Myc	3542

Data from Kim et al. (2008 [© Elsevier]).

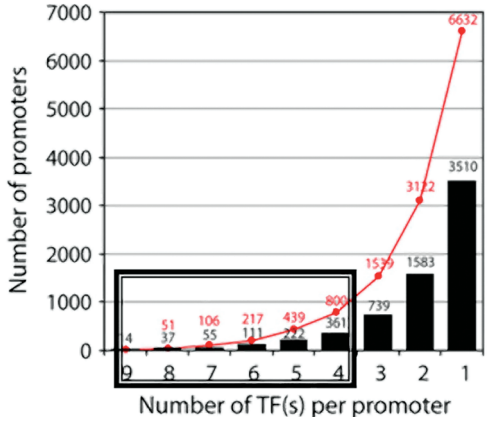

Figure 4. Distribution of numbers of factors bound to target proteins. Multifactor target genes (*boxed*) are defined as those promoters bound by more than four transcription factors among the nine tested. Approximately 800 target loci are represented in the multifactor category. (Modified from Kim et al. 2008 [© Elsevier].)

Finally, the merging of the protein interaction network with target gene identification highlights "hubs" (Dezso et al. 2003), including Nanog, Oct4, Sox2, REST, Sall4, Rif1 among others (Fig. 5). The regulatory circuit is highly intertwined and the effectors of the network (i.e., the pluripotency factors) are themselves direct targets. A full understanding of pluripotency necessitates consideration of more than the "core" set of factors (Oct4, Nanog, and Sox2).

DUAL ROLE OF PLURIPOTENCY FACTORS

Our analysis of target loci of the pluripotency factors provides a logic for their dual action in maintaining the ES cell state (Fig. 6) (Orkin 2005). On the one hand, the pluripotency factors must prevent expression of differentiation-promoting genes. For example, it appears that GATA-6, which is essential for primitive endoderm gene expression, is under direct repression by Nanog, likely in concert with two of its partners (Zfp281 and Nac1) (Wang et al. 2006). The differentiation-promoting targets tend to be occupied by a limited number of the pluripotency factors. Indeed, the extent to which the lack of target gene expression is due to active repression versus insufficient factor loading to achieve transcriptional activation is uncertain. In parallel, pluripotency factors provide a positive stimulus for self-renewal and pluripotency, in part through maintenance of their own expression by autoregulatory and cross-regulatory interactions, but also through activation of additional targets. The multifactor gene targets fall within this broad category and reflect a dominant action of the pluripotency factors. Among this class are numerous transcription factors, the majority of which have not been studied in an ES cell context (Table 2). We speculate that the set of multifactor target loci is highly enriched for additional proteins that participate in maintenance of pluripotency. Further work will be required to validate this prediction.

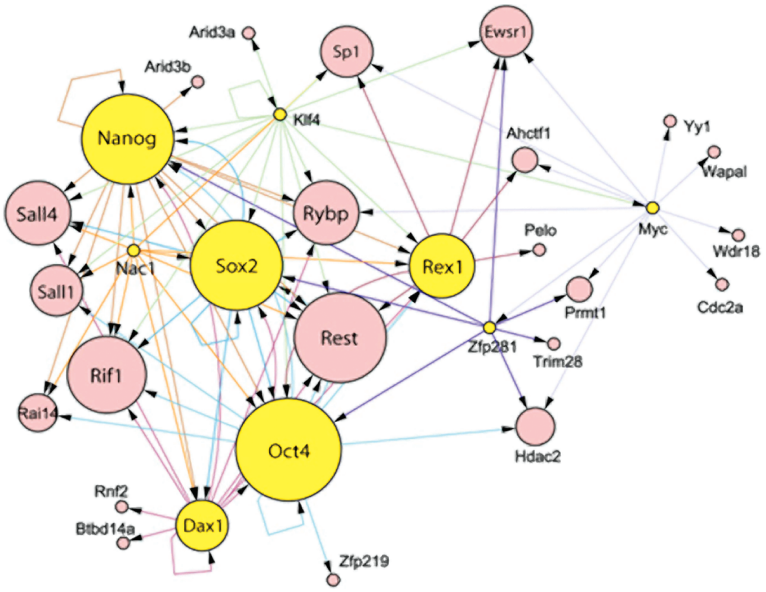

Figure 5. Expanded transcriptional regulatory network showing target hubs of multiple factors within the protein interaction network (Fig. 3). (*Yellow circles*) Nine transcription factors for which target loci were identified. The size of each circle reflects the extent of factor cooccupancy. (*Arrowheads*) Directions of transcriptional regulation. (Reprinted, from Kim et al. 2008 [© Elsevier].)

INSIGHTS INTO NANOG PROTEIN FUNCTION

Besides identifying new pluripotency regulators, it is of particular interest to characterize how the known factors participate in transcriptional control. To this end, we have examined how the Nanog protein functions. Among homeodomain proteins, Nanog is most closely related to the Nk2 subfamily. Members of this class often homodimerize through the homeodomain (HD). Outside of the HD, Nanog bears little resemblance to other homeodomain proteins. In size fractionation of nuclear extracts, we had observed that Nanog polypeptide (34 kD) distributes into two broad regions, one corresponding to large protein complexes (–2 MDa) and another approximating the mass of a Nanog dimer (Wang et al. 2006). Through study of tagged forms of Nanog, we demonstrated that Nanog monomers assemble into dimers (Wang et al. 2008). However, dimerization is mediated through a tryptophan-rich (WR) subregion of the carboxy-terminal CD domain, which had previously been associated with *trans*-activation potential. In contrast to the NK2 proteins, the HD of Nanog does not mediate dimer formation. To assess the role of dimer formation in protein interactions and the function of Nanog, we generated tethered dimers, based on the pioneering studies of Wold and her colleagues (Neuhold and Wold 1993). We established that Nanog interacts with other pluripotency proteins (e.g., Oct4, Sall4, Zfp281, and Dax1) principally as a dimer. Furthermore, Nanog dimers promote self-renewal of ES cells in the absence of LIF. These findings suggest that Nanog dimer formation constitutes a critical point of control in ES cell pluripotency. We anticipate mechanisms that might shift the equilibrium of dimers to monomers and thereby inactivate Nanog. Recently, caspase cleavage of Nanog has been described as a control point for ES cell differentiation (Fujita et al. 2008). Thus, as supported by independent studies of Chambers (Mullin et al. 2008), Nanog serves to fine-tune the pluripotent state.

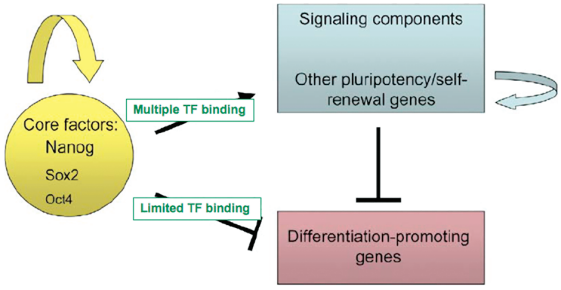

Figure 6. Dual roles of pluripotency factors in ES cells. (Modified from Orkin 2005 [© Elsevier].)

OCT4 DEPENDENCE OF CHROMATIN STRUCTURE WITHIN THE NANOG LOCUS

Multiple lines of evidence indicate that Oct4 has a central role in establishing and maintaining the pluripotent state. Besides the recognized dose-dependent role of Oct4 in ES cells, its inclusion in all somatic cell reprogramming "cocktails" to date suggests that its functions cannot be readily replaced by other factors. Moreover, the consensus binding motif predicted for multifactor binding targets very closely resembles an Oct4-binding sequence (Kim et al. 2008). Thus, it is likely that Oct4 protein bound to DNA provides a docking site for other pluripotency factors. The *Nanog* gene lies within a phylogenetically conserved chromosomal region that encodes several other genes that are expressed in early development, including *Aicda, Apobec1, GDF3*, and *Dppa3* (also known as *Stella* and *PGC7*). We have hypothesized that this extended *Nanog* locus may provide a window into regulation of pluripotency-related genes. Using a high-throughput quantitative chromatin profiling approach,

Table 2. 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 no.	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 β)
Cbx7	NM_144811	Chromobox homolog 7
Cdx1	NM_009880	Caudal type homeobox 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	Homeobox b13
Jarid2	NM_021878	Jumonji, at rich interactive domain 2
Klf2	NM_008452	Krüppel-like factor 2 (lung)
Klf9	NM_010638	Krüppel-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	Homeobox, 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, γ
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, γ
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	Zinc finger protein of the cerebellum 2
Zic5	NM_022987	Zinc finger protein of the cerebellum 5

Reprinted from Kim et al. (2008 [© Elsevier]).

we identified multiple potential regulatory elements over more than 160 kb, as reflected by DNase-I-hypersensitive sites (Levasseur et al. 2008). ChIP assays reveal the cooccupancy of DNase-I-hypersensitive regions by Oct4 and other pluripotency factors, including Nanog and Zfp281. Activity of these regions in conventional enhancer assays suggests that they are likely to function as authentic regulatory elements in situ. Chromatin conformation capture (3C) assay also indicates that the Nanog proximal promoter contacts hypersensitivity sites as far as 150 kb away. Importantly, these long-range interactions are sensitive to depletion of Oct4, indicating that Oct4 is critical for maintenance of the structure of the extended Nanog chromatin region (Fig. 7). We speculate that Oct4 serves an analogous role at many other critical gene targets in ES cells.

CONCLUSIONS

Through our studies (Wang et al. 2006; Kim et al. 2008) and those of others (Boyer et al. 2005; Ivanova et al. 2006; Loh et al. 2006), the transcription factors and their direct targets responsible for maintaining ES cells in a self-renewing, pluripotent state are being uncovered in a comprehensive manner. The panoply of factors individually required for pluripotency is remarkable. Although Oct4, Sox2, and Nanog have earned respect as "core" factors, it is still unknown how many of the other factors that are just being studied can drive LIF-independent self-renewal and/or substitute to other factors in somatic cell reprogramming experiments. Although the identification of transcription factors and targets critical for pluripotency is a powerful strategy for discovery of biologically relevant genes and proteins for more in-depth analysis, the use of ES cell protein or transcriptional networks as tools for prediction of reprogramming factors or how networks change on cellular differentiation has yet to be fully exploited. Realization of the value of these networks may require development of new computational methods to model changes on a global scale. The current efforts are a first step in that direction.

ACKNOWLEDGMENTS

J.W. was supported in part by a pilot grant from the Harvard Stem Cell Institute. S.H.O. is an Investigator at the Howard Hughes Medical Institute.

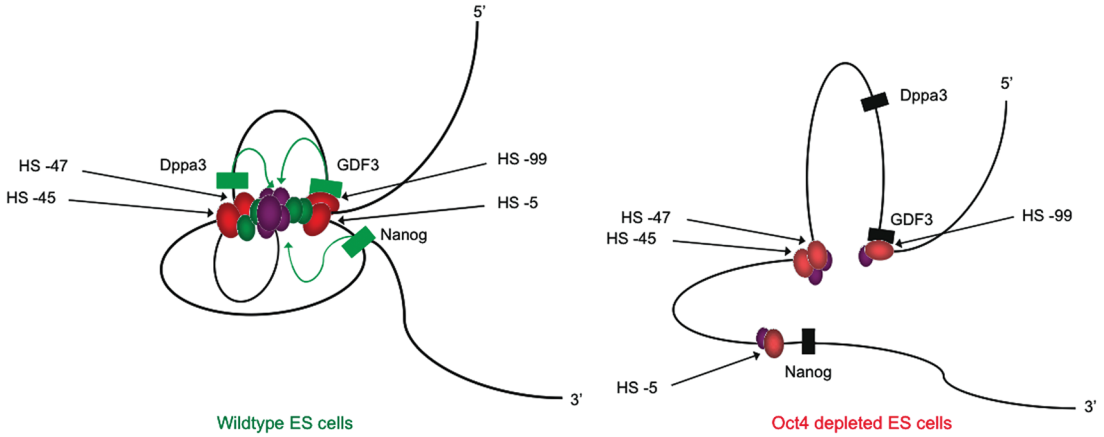

Figure 7. Proposed conformation of the extended Nanog locus as determined by 3C analysis. (*Red ovals*) Distant regulatory elements. (*Green and black rectangles*) Nanog locus genes indicating active and repressed states, respectively. The model depicts DNA-bound factors within the proximal promoters of GDF3, Dppa3, and Nanog initiating contact (*green arrows*) with an active transcriptional node formed by RNA polymerase II (*large purple oval*); accessory DNA-binding or bridging transcription factors p300, zfp281, Nac1, and CTCF (*smaller purple ovals*); and essential DNA-bound transcription factors Oct4 and Nanog (*green ovals*). (Reprinted from Levasseur et al. 2008.)

REFERENCES

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.

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.

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.* **100:** 7480–7485.

Dezso, Z., Oltvai, Z.N., and Barabasi, A.L. 2003. Bioinformatics analysis of experimentally determined protein complexes in the yeast *Saccharomyces cerevisiae*. *Genome Res.* **13:** 2450–2454.

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

Fujita, J., Crane, A.M., Souza, M.K., Dejosez, M., Kyba, M., Flavell, R.A., Thomson, J.A., and Zwaka, T.P. 2008. Caspase activity mediates the differentiation of embryonic stem cells. *Cell Stem Cell* **2:** 595–601.

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.

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.

Levasseur, D.N., Wang, J., Dorschner, M.O., Stamatoyannopoulos, J.A., and Orkin, S.H. 2008. Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells. *Genes Dev.* **22:** 575–580.

Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D., Chesneau, A., Hao, T., et al. 2004. A map of the interactome network of the metazoan *C. elegans*. *Science* **303:** 540–543.

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

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

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.

Mullin, N.P., Yates, A., Rowe, A.J., Nijmeijer, B., Colby, D., Barlow, P.N., Walkinshaw, M.D., and Chambers, I. 2008. The pluripotency rheostat Nanog functions as a dimer. *Biochem. J.* **411:** 227–231.

Neuhold, L.A. and Wold, B. 1993. HLH forced dimers: Tethering MyoD to E47 generates a dominant positive myogenic factor insulated from negative regulation by Id. *Cell* **74:** 1033–1042.

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.

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

Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K.E., Guyot, B., Heck, A.J., Vyas, P., de Boer, E., Grosveld, F., and Strouboulis, J. 2005. GATA-1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J.* **24:** 2354–2366.

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

Singh, S.K., Kagalwala, M.N., Parker-Thornburg, J., Adams, H., and Majumder, S. 2008. REST maintains self-renewal and pluripotency of embryonic stem cells. *Nature* **453:** 223–227.

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.

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.

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.

Wang, J., Levasseur, D.N., and Orkin, S.H. 2008. Requirement of Nanog dimerization for stem cell self-renewal and pluripotency. *Proc. Natl. Acad. Sci.* **105:** 6326–6331.

Wu, Q., Chen, X., Zhang, J., Loh, Y.H., Low, T.Y., Zhang, W., 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.

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.