LETTERS

A protein interaction network for pluripotency of embryonic stem cells

Jianlong Wang[1], Sridhar Rao[1], Jianlin Chu[1], Xiaohua Shen[1], Dana N. Levasseur[1], Thorold W. Theunissen[1] & Stuart H. Orkin[1,2]

Embryonic stem (ES) cells are pluripotent[1,2] and of therapeutic potential in regenerative medicine[3,4]. Understanding pluripotency at the molecular level should illuminate fundamental properties of stem cells and the process of cellular reprogramming. Through cell fusion the embryonic cell phenotype can be imposed on somatic cells, a process promoted by the homeodomain protein Nanog[5], which is central to the maintenance of ES cell pluripotency[6,7]. Nanog is thought to function in concert with other factors such as Oct4 (ref. 8) and Sox2 (ref. 9) to establish ES cell identity. Here we explore the protein network in which Nanog operates in mouse ES cells. Using affinity purification of Nanog under native conditions followed by mass spectrometry, we have identified physically associated proteins. In an iterative fashion we also identified partners of several Nanog-associated proteins (including Oct4), validated the functional relevance of selected newly identified components and constructed a protein interaction network. The network is highly enriched for nuclear factors that are individually critical for maintenance of the ES cell state and co-regulated on differentiation. The network is linked to multiple co-repressor pathways and is composed of numerous proteins whose encoding genes are putative direct transcriptional targets of its members. This tight protein network seems to function as a cellular module dedicated to pluripotency.

To recover Nanog with its associated proteins, we used a biotinylation/proteomics approach[10]. Nanog complementary DNA bearing an amino-terminal Flag epitope and a short peptide tag that serves as a substrate for *in vivo* biotinylation (Fig. 1a) was expressed at about 20% of the endogenous level in ES cells previously engineered to express *Escherichia coli* biotin ligase BirA (Fig. 1b, left). Low-level expression reduces the likelihood that non-physiological interactions will be identified or that the stoichiometry of protein complexes will be perturbed[11]. Biotinylated Nanog (^{bio}Nanog) is efficiently recovered from ES cell nuclear extracts with streptavidin (SA)-agarose (Fig. 1b right), and is biologically active, because its expression attenuates the dependence of ES cells on leukaemia inhibitory factor (LIF; Supplementary Fig. S1). Cells coexpressing BirA and ^{bio}Nanog are functionally equivalent to wild-type ES cells in stem-cell marker gene expression (Supplementary Fig. S1). On size fractionation of ES cell nuclear extracts, ^{bio}Nanog and endogenous Nanog behave similarly, broadly distributed with an apparent molecular mass of more than 160 kDa to 1 MDa, far greater than that of the predicted polypeptide (34 kDa) (Fig. 1c).

To identify Nanog-associated proteins we used single-step capture on streptavidin beads[10], as well as tandem purification with anti-Flag immunoprecipitation followed by capture with streptavidin. Nuclear extracts of ES cells expressing either BirA alone or BirA plus ^{bio}Nanog were processed in parallel. Total peptides sequenced and proteins identified by whole-lane liquid chromatography–tandem mass spectrometry (LC–MS/MS) are summarized in Fig. 1d (left panel; see also Supplementary Data). With criteria described in the Supplementary Methods, putative Nanog-associated proteins were chosen (Fig. 1d, right panel). The candidates, which are predominantly transcription factors or components of transcriptional complexes, include proteins previously studied in an ES cell context (for example Oct4 or Dax1) and fall into three groups. The first group includes proteins present in at least two of three independent one-step purifications and tandem purification. These proteins (Sall1, Sall4, Rif1, Tif1β, Mybbp, Dax1 and Nac1) are prime interaction candidates. The second group includes proteins (Zfp281, Err2, Elys, Oct4, Zfp198, NF45 and HDAC2) that may be part of unstable or transient complexes that dissociate during tandem purification. Proteins in the third group (REST, Sp1 and Wapl) seem 'masked' in one-step purifications but are recovered in the tandem procedure. Proteins identified by our approach may interact directly with Nanog, or more often indirectly through their association with other components of a protein complex.

We then validated selected Nanog-interacting candidates by performing co-immunoprecipitation experiments. Interactions between Nanog and Dax1, Nac1, Zfp281 and Oct4 were confirmed in heterologous 293T (Supplementary Fig. S2) and ES (Fig. 2a–d) cells transiently transfected with constructs as indicated (see the figure legend for details). In addition, we generated ES cells expressing biotinylated Nac1, Zfp281, Dax1 and Oct4, at or below endogenous protein levels (Supplementary Fig. S3), for affinity capture with anti-streptavidin-agarose and western blotting with anti-streptavidin–horseradish peroxidase (HRP) antibody. We confirmed that Nanog co-purifies with ^{bio}Nac1, ^{bio}Zfp281, ^{bio}Dax1 and ^{bio}Oct4 (Fig. 2e) and that Nac1 co-purifies with ^{bio}Nanog (Fig. 2f). Furthermore, we confirmed the association of Rif1 with endogenous Nanog (Fig. 2g), as well as ectopically expressed Nanog (Fig. 2h, i).

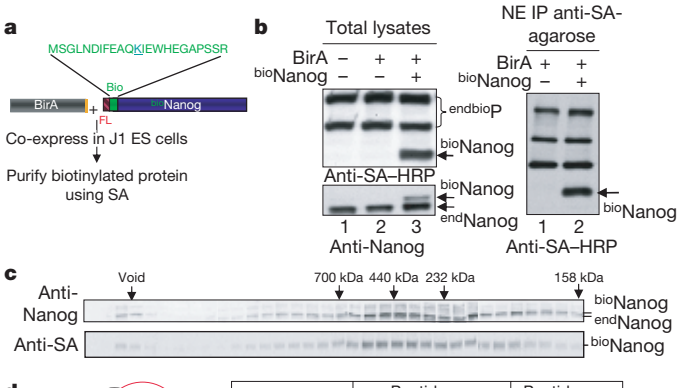
To test the functional relevance of selected components, we employed inhibition of endogenous RNAs by short hairpin RNA (shRNA; Supplementary Fig. S4). If Nanog-associated proteins influence Nanog's transcriptional activity, their loss of function would be expected to affect ES cell pluripotency. We first examined shRNAs directed to Dax1 and Sall4 in ES cells that express green fluorescent protein (GFP) under the control of Oct4 regulatory elements. In both instances a loss of pluripotency was observed, as revealed by diminished Oct4 expression and concomitant derepression of selected lineage-specific markers (Supplementary Fig. S5). Recent conditional gene targeting of Dax1 (ref. 12) and Sall4 (ref. 13) provides independent support for their crucial function in ES cells.

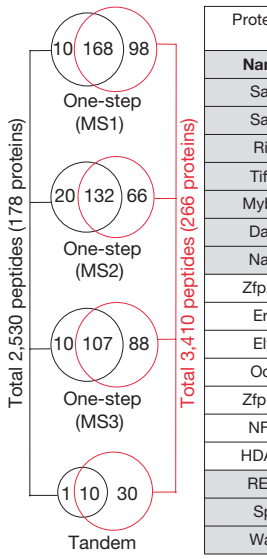
Further functional validation included shRNA inhibition of two Nanog-associated proteins, Nac1 and Zfp281, not previously

[1]Division of Hematology–Oncology, Children's Hospital and Dana Farber Cancer Institute, Harvard Medical School, Harvard Stem Cell Institute, [2]Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA.

examined in ES cells. Nac1 is a BTB-domain containing protein[14,15] related to *Drosophila* bric-a-brac/tramtrack, which prevents inappropriate neural gene expression. Zfp281, the mouse homologue of human zinc-finger protein ZBP99, may regulate cell proliferation, differentiation and oncogenesis[16]. We infected wild-type ES cells with retrovirus containing an shRNA vector carrying a GFP cassette (Supplementary Fig. S4) and isolated GFP^{high} clones harbouring Nac1 and Zfp281 shRNAs. Two clones each with similar extents of RNA knockdown (about 80% for Nac1 and about 70% for Zfp281; Fig. 3a) were studied in detail. Nac1 or Zfp281 knockdown cells maintained on a feeder layer formed small colonies, yet they retained proper ES cell morphology (Fig. 3b, top panels). On passage to gelatin in the presence of LIF, knockdown cells grew poorly and formed disorganized clumps of small cells (Fig. 3b, bottom panels). Striking derepression of primitive endoderm (*Gata6/4*), mesoderm/visceral endoderm (*Bmp2*) and neuroectoderm (*Isl1*) markers was observed, despite normal expression of the stem-cell markers (*Nanog*, *Oct4* and *Rex1*; Fig. 3c). Validation of these findings was achieved through the use of an independent shRNA directed to Zfp281 and by examination of gene-targeted Nac1$^{+/-}$ ES cells (Supplementary Figs S4 and S6). To assess the relevance of Nac1 and Zfp281 with respect to transcriptional control, we examined chromatin occupancy of these proteins at the *Gata6* promoter in wild-type ES cells. As shown in Fig. 3d, Nac1 and Zfp281, as well as Nanog, are highly enriched at the *Gata6* promoter. These data strongly implicate Nac1 and Zpf281, along with Nanog, in the tight control of this crucial target gene and also show that derepression of *Gata6* expression in knockdown cells is not due to off-target effects of shRNAs.

The broad pattern of Nanog on size fractionation, and the identification and validation, both physically and functionally, of associated proteins suggest that Nanog is a component of multiple protein complexes. By performing similar proteomic analyses on several of the Nanog-associated proteins, as well as another stem-cell marker protein Rex1 (ref. 17), we sought to identify additional previously unknown proteins and components shared between complexes and to develop a protein network (Fig. 4a). The size distributions of complexes containing Oct4, Dax1, Nac1, Zfp281 and Rex1 overlap that of Nanog (Supplementary Fig. S7). We employed ES cell lines expressing ^{bio}Dax1, ^{bio}Nac1, ^{bio}Zfp281, ^{bio}Oct4 and ^{bio}Rex1 (Fig. 2e, and Supplementary Fig. S3) for the





Protein ID	Peptide nos (MS1+MS2+MS3)	Peptide nos (Tandem)
Nanog	4+7+8	4
Sall1	3+2+1	1
Sall4	4+2+14	8
Rif1	12+5+5	3
Tif1β	11+16+13	15
Mybbp	8+14+15	2
Dax1	0+1+2	1
Nac1	2+0+2	2
Zfp281	1+2+4	0
Err2	7+2+1	0
Elys	1+2+1	0
Oct4	2+0+1	0
Zfp198	0+2+5	0
NF45	4+0+1	0
HDAC2	2+1+0	0
REST	0+0+1	12
Sp1	0+1+0	1
Wapl	0+0+1	1

Figure 1 | Protein complexes containing Nanog protein in mouse ES cells. **a**, Strategy for biotinylation of Nanog in ES cells. Flag (FL) and 23-amino-acid biotin (Bio) tags are indicated. SA, streptavidin. **b**, Western blot analyses of expression of biotinylated Nanog (^{bio}Nanog) as well as endogenous Nanog (^{end}Nanog) (left), and binding of ^{bio}Nanog to SA beads (right). IP, immunoprecipitation. **c**, Western blots of fractions using anti-Nanog (top) and anti-SA–HRP antibodies (bottom). **d**, Summary of LC–MS/MS results from three independent one-step purifications and a tandem purification. Total peptides sequenced and proteins identified in BirA (black circles) and ^{bio}Nanog (red circles) samples are indicated. Components of Nanog complexes selected with the criteria described in Supplementary Methods are listed in the table.

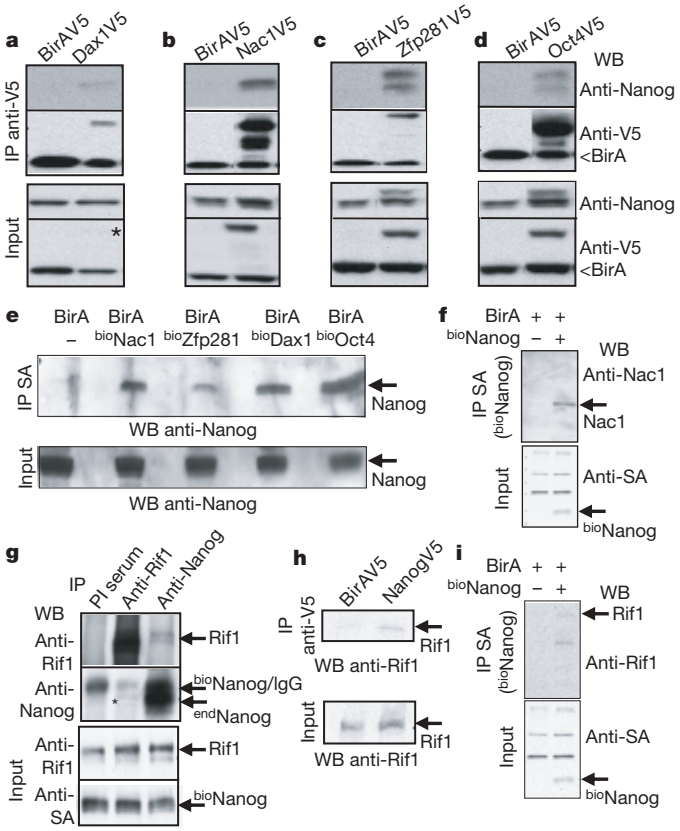


Figure 2 | Confirmation of Nanog association by co-immunoprecipitation in ES cells. **a**–**d**, Co-immunoprecipitation of Nanog and Dax1 (**a**), Nac1 (**b**), Zfp281 (**c**) and Oct4 (**d**) in ES cells expressing BirA without (**a**, **b**) or with (**c**, **d**) ^{bio}Nanog. The asterisk marks a faint band in **a**. IP, immunoprecipitation. **e**, **f**, Co-immunoprecipitation of Nanog and Nac1, Zfp281, Dax1 and Oct4 in ES cells expressing ^{bio}Nac1, ^{bio}Zfp281, ^{bio}Dax1, ^{bio}Oct4 (**e**) and ^{bio}Nanog (**f**). **g**–**i**, Co-immunoprecipitation of Nanog and Rif1 in ES cells. In **g**, nuclear extracts from ES cells expressing BirA and ^{bio}Nanog were incubated with preimmune (PI) serum, anti-Rif1 antibody, and anti-Nanog antibody, respectively, followed by capture with protein-A/G agarose. The asterisk indicates endogenous Nanog immunoprecipitated by anti-Rif1 antibody. Note that ^{bio}Nanog migrates together with the IgG band. Antibodies for IP and western blot (WB) analyses are indicated. In **h** and **i**, experiments similar to those in **a** and **f** were performed except that different constructs and antibodies were used.

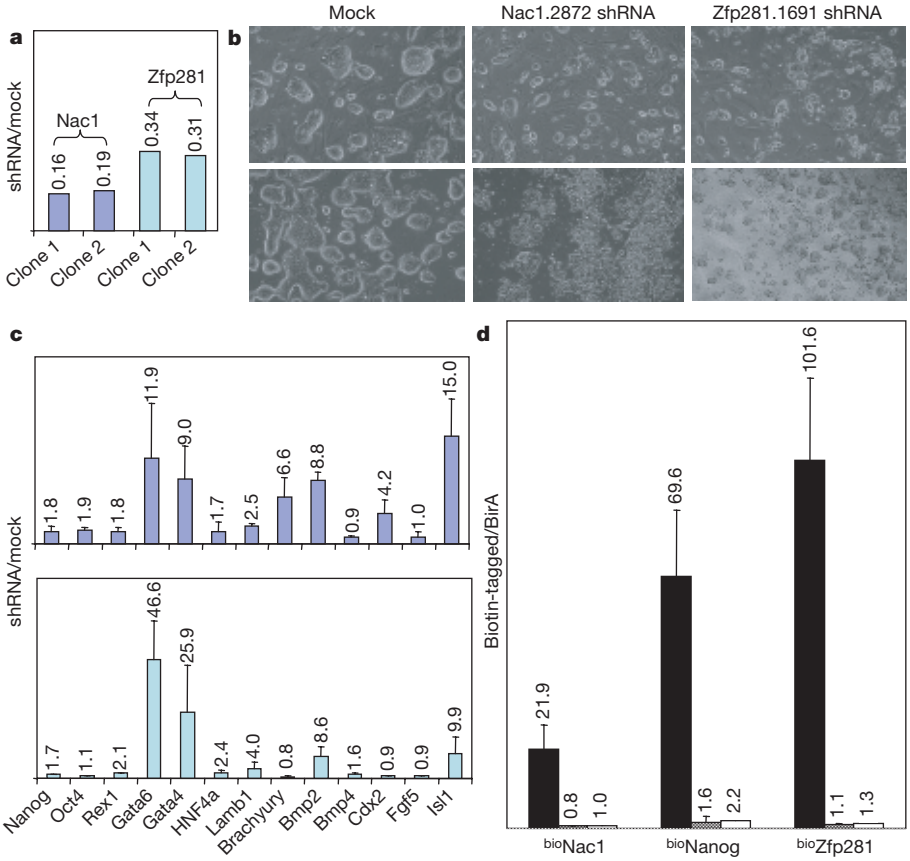


Figure 3 | Functional validation by RNA-mediated interference. **a**, Quantitative reverse transcriptase–polymerase chain reaction (RT–PCR) analysis of Nac1 and Zfp281 gene expression (two clones each) after shRNA-mediated knockdown in wild-type ES cells. **b**, Morphology of ES cells infected with virus containing empty shRNA vector (Mock; left), Nac1 shRNA (middle) or Zfp281 shRNA (right). ES cells were grown either on feeders (top) or on gelatin in the presence of LIF (bottom). **c**, Quantitative RT–PCR analysis of stem-cell markers as well as lineage-specific markers as indicated in Nac1.2872 knockdown cells (top) and Zfp281.1691 knockdown cells (bottom). **d**, Quantitative real-time chromatin immunoprecipitation of biotin-tagged proteins on *Gata6* (black bars) as well as *Gapdh* (grey bars) and β_2 microglobulin (white bars) promoter sequences (as negative controls). Results are shown as means and s.e.m. from two clones in **c** and three independent experiments in **d**.

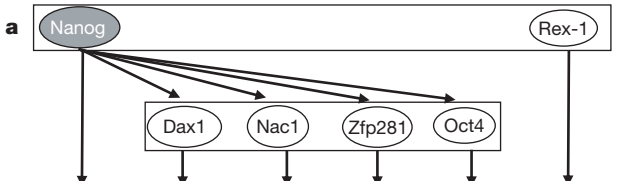
affinity purification of complexes, as described for ^{bio}Nanog. Proteins were identified with the criteria applied for the selection of Nanog-associated proteins and are summarized in Fig. 4a. Multiple proteins are shared between the different complexes. A protein interaction map (or mini-interactome)[18] summarizes these observations (Fig. 4b). Several features suggest that the network is specialized to establish and/or maintain ES cell pluripotency.

First, the network is remarkably enriched for proteins that have been shown, or are likely, to be required individually for controlling the survival or differentiation of the inner cell mass, or aspects of early development. These include 15 proteins (Fig. 4b, green circles) with known phenotypic alleles in published knockout studies or communicated to us (Supplementary Table S1), and 4 proteins (blue circles) from shRNA-mediated inhibition described previously[19] or in this study (Fig. 3, and Supplementary Figs S5 and S6). Three other proteins (yellow circles) have known phenotypes that are not consistent with an early requirement in development (Supplementary Table S1). Thus, more than 80% of proteins for which knockout or knockdown studies have been performed seem essential for early development and/or ES cell properties. To approximate the 'expected' fraction of essential genes, we curated 300 random transcription factor genes and found that less than 15% of those expressed in ES cells had knockout phenotypes consistent with an early requirement (data not shown).

Second, most genes encoding the proteins within the network are co-regulated, and specifically downregulated, on ES cell differentiation. In Fig. 4c, a rank order depicts the expression of transcripts for the network proteins relative to Rex1. For comparison, a 'control' set of randomly chosen gene transcripts was similarly analysed (Supplementary Fig. S8). We chose Rex1 as a comparator because it is abruptly and extensively downregulated on the differentiation of ES cells. Co-regulation of genes in the network is consistent with their participation in a common cellular function or pathway.

Third, in reinforcement of this hypothesis, we find that many (at least 56%) of the genes encoding proteins of the mini-interactome have been reported as putative Nanog and/or Oct4 targets in recent ChIP-on-chip[20] or ChIP-PET[19] analyses of chromatin occupancy in human or mouse ES cells (Supplementary Table S2). Thus, 'downstream' gene targets of Nanog and Oct4 also serve as 'upstream' effectors in the pluripotency network. Although this feature may stabilize aspects of the ES cell state, it also renders the network vulnerable to attack at multiple points rather than at very few[21].

Fourth, the network is linked to several cofactor pathways, which are largely involved in mediating transcriptional repression. These include the histone deacetylase NuRD (P66β and HDAC2), polycomb group (YY1, Rnf2 and Rybp) and SWI/SNF chromatin remodelling (BAF155) complexes. Of potential interest, Rex1 and Oct4, as contrasted with Nanog or its closest partners (for example Dax1, Sall4 and Nac1), are associated with polycomb components. Nanog is linked to HDAC/NuRD through its association with Sall1/4 (ref. 22) and Nac1 (ref. 23). The involvement of multiple co-repressor complexes provides both a method of regulating different sets of target genes and a fail-safe mechanism to prevent differentiation



Nanog complexes	Dax1 complexes	Nac1 complexes	Zfp281 complexes	Oct4 complexes	Rex1 complexes
Nanog					**Rex1**
Oct4		Oct4		**Oct4**	Oct4
Sall1		Sall1		RYBP	YY1
Nac1		**Nac1**	EWS	EWS	Nac1
Zfp281	Zfp609	Zfp281	**Zfp281**	Zfp219	Etl1
Zfp198		Ari3a			Rai14
Elys		Arid3b		Arid3b	Prmt1
NF45		Btbd14a		NF45	Rnf2
Wapl					Wapl
HDAC2	PELO			HDAC2	
REST	REST	P66β		P66β	
Sp1	Sp1	Wdr18		Sp1	Wdr18
Dax1	**Dax1**	BAF155	Cdk1	Cdk1	BAF155
Err2	Err2		Requiem	Requiem	Requiem
Sall4	Sall4	Sall4	Sall4	Sall4	
Tif1β	Tif1β				Tif1β
Rif1	(Rif1)	(Rif1)	(Rif1)		
Mybbp		(Mybbp)			

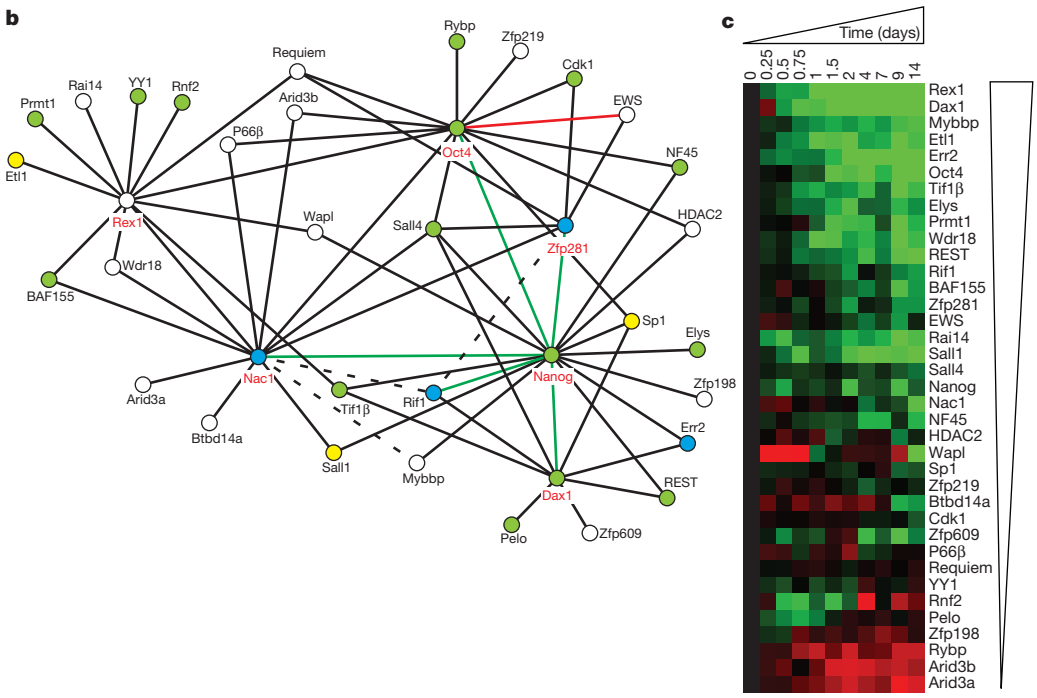


Figure 4 | A protein interaction network in ES cells. a, Schematic depiction of the strategy for mapping protein–protein interaction network in ES cells. Starting with the primary tagged baits Nanog and another stem-cell marker protein Rex1, interaction partners of both Nanog and Rex1 were identified. The validated interactors such as Dax1, Nac1, Zfp281 and Oct4 served as secondary tagged baits for the identification of tertiary interacting partners. Selected high-confidence components of multiprotein complexes are listed in the table. Proteins shared by multiple tagged baits are shaded similarly. Proteins in parentheses are present predominantly in tagged samples with some peptide presence in one of the controls during one-step purifications. **b**, Depiction of the features of the interactome. Proteins with red labels are tagged baits for affinity purification. Green lines indicate interactions confirmed by co-immunoprecipitation and the red line indicates interaction confirmed in the literature[27]. Green circles indicate proteins whose knockout results in defects in proliferation and/or survival of the inner cell mass or other aspects of early development; blue circles indicate proteins whose reduction by RNA-mediated interference results in defects in self-renewal and/or differentiation of ES cells; yellow circles are proteins whose knockout results in later developmental defects; white circles denote proteins for which no loss-of-function data are available. **c**, Co-regulation of proteins in the network on differentiation. Microarray data were curated from the literature[28] and are presented in rank order relative to Rex1.

along different lineages, which is a requisite for maintaining the pluripotent state.

Last, the multiple interactions between critical factors also predict that proteins within the network influence the function of other members, either positively or negatively, and that the stoichiometry within complexes is important for maintaining pluripotency. This latter conclusion is consistent with dosage-dependent effects for Oct4 (ref. 24) and Nanog[25]. Given the numerous gene targets of the proteins within the network, it is likely that subsets of targets are regulated by different complexes. We have shown that *Gata6* promoter sequences are bound by Nac1, Zfp281 and Nanog. Presumably, proteins within the network, such as Dax1 and Sall4, also bind other targets with Nanog, Oct4 and/or Sox2 (among other proteins). Accordingly, the combinatorial control of target genes will be far greater in complexity than suggested by recent chromatin occupancy studies[19,20].

Starting from a central regulator of transcription in ES cells, namely Nanog, followed by iterative tagging and purification of Nanog-associated proteins, we have developed a protein interaction network. Taken together, the extraordinary enrichment for essential genes, the coexpression of most genes, and their roles as both targets and effectors indicate that this mini-interactome might serve as a functional module committed to maintaining ES cell pluripotency. Whereas proteins of the network cooperate to maintain pluripotency, their interdependence renders the cell susceptible to rapid loss of pluripotency on the downregulation, or inactivation, of any one of many components. How differences between mouse and human ES cells are reflected in the network, and whether components of the network that are not strictly specific to ES cells are employed by other types of stem cells for controlling self-renewal versus differentiation, are topics worthy of future study. Last, the network provides a framework for exploring the combinations of factors that may permit faithful reprogramming of differentiated cells to an ES cell state[26].

METHODS

***In vivo* biotinylation, affinity purification and mass spectrometry.** J1 ES cells were transfected with a plasmid construct expressing BirAV5his followed by selection with G418 ($300\,\mu g\,ml^{-1}$). Clones were picked and expanded for western blotting by using anti-V5–HRP (Invitrogen). A BirA-expressing ES clone was then used for transfection with plasmid constructs containing biotin (or Flag–biotin)-tagged cDNAs. ES cells expressing tagged cDNA were identified by western blotting with anti-streptavidin–HRP of the total lysates or nuclear extracts. One-step affinity purification with streptavidin-agarose was performed as described[10]. In brief, nuclear extracts from ES cells expressing BirA and biotin-tagged cDNA were incubated with streptavidin-agarose in a buffer containing 1 × Tris-buffered saline (TBS), 350 mM NaCl and 0.3% Nonidet P40. Binding was performed at 4 °C for 1 h to overnight on a rocking platform, followed by six washes in binding solution. Bound material was eluted by boiling for 5 min in Laemmli buffer, fractionated on a 10% SDS-polyacrylamide gel, and subjected to LC–MS/MS sequencing and data analysis (see Supplementary Methods for details).

For tandem purification, nuclear extracts were incubated overnight with anti-M2 Flag-agarose (Sigma) at 4 °C on a rocking platform. On day 2, after extensive washing with 1 × TBS, bound materials were eluted by using Flag peptide (Sigma). The pooled eluate was then used for streptavidin-agarose capture as described above.

Details of other methods are provided in Supplementary Materials.

Received 10 April; accepted 27 September 2006.
Published online 8 November 2006.

1. Martin, G. R. 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 (1981).
2. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292,** 154–156 (1981).
3. Donovan, P. J. & Gearhart, J. The end of the beginning for pluripotent stem cells. *Nature* **414,** 92–97 (2001).
4. Prelle, K., Zink, N. & Wolf, E. Pluripotent stem cells—model of embryonic development, tool for gene targeting, and basis of cell therapy. *Anat. Histol. Embryol.* **31,** 169–186 (2002).
5. Silva, J., Chambers, I., Pollard, S. & Smith, A. Nanog promotes transfer of pluripotency after cell fusion. *Nature* **441,** 997–1001 (2006).
6. Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113,** 631–642 (2003).
7. Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113,** 643–655 (2003).
8. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95,** 379–391 (1998).
9. Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17,** 126–140 (2003).
10. de Boer, E. *et al.* Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc. Natl Acad. Sci. USA* **100,** 7480–7485 (2003).
11. Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440,** 637–643 (2006).
12. Niakan, K. K. *et al.* Novel role for the orphan nuclear receptor Dax1 in embryogenesis, different from steroidogenesis. *Mol. Genet. Metab.* **88,** 261–271 (2006).
13. Sakaki-Yumoto, M. *et al.* 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 (2006).
14. Mackler, S. A., Homan, Y. X., Korutla, L., Conti, A. C. & Blendy, J. A. The mouse nac1 gene, encoding a cocaine-regulated Bric-a-brac Tramtrac Broad complex/Pox virus and Zinc finger protein, is regulated by AP1. *Neuroscience* **121,** 355–361 (2003).
15. Mackler, S. A. *et al.* NAC-1 is a brain POZ/BTB protein that can prevent cocaine-induced sensitization in the rat. *J. Neurosci.* **20,** 6210–6217 (2000).
16. Law, D. J., Du, M., Law, G. L. & Merchant, J. L. ZBP-99 defines a conserved family of transcription factors and regulates ornithine decarboxylase gene expression. *Biochem. Biophys. Res. Commun.* **262,** 113–120 (1999).
17. Thompson, J. R. & Gudas, L. J. Retinoic acid induces parietal endoderm but not primitive endoderm and visceral endoderm differentiation in F9 teratocarcinoma stem cells with a targeted deletion of the Rex-1 (Zfp-42) gene. *Mol. Cell. Endocrinol.* **195,** 119–133 (2002).
18. Batagelj, V. & Mrvar, A. Pajek—program for large network analysis. *Connections* **21,** 47–57 (1998).
19. Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38,** 431–440 (2006).
20. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122,** 947–956 (2005).
21. Albert, R., Jeong, H. & Barabasi, A. L. Error and attack tolerance of complex networks. *Nature* **406,** 378–382 (2000).
22. Lauberth, S. M. & Rauchman, M. A conserved twelve amino acid motif in sall1 recruits nuRD. *J. Biol. Chem.* **281,** 23922–23931 (2006).
23. Korutla, L., Wang, P. J. & Mackler, S. A. The POZ/BTB protein NAC1 interacts with two different histone deacetylases in neuronal-like cultures. *J. Neurochem.* **94,** 786–793 (2005).
24. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genet.* **24,** 372–376 (2000).
25. Hatano, S. Y. *et al.* Pluripotential competence of cells associated with Nanog activity. *Mech. Dev.* **122,** 67–79 (2005).
26. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
27. Lee, J., Rhee, B. K., Bae, G. Y., Han, Y. M. & Kim, J. Stimulation of Oct-4 activity by Ewing's sarcoma protein. *Stem Cells* **23,** 738–751 (2005).
28. Perez-Iratxeta, C. *et al.* Study of stem cell function using microarray experiments. *FEBS Lett.* **579,** 1795–1801 (2005).

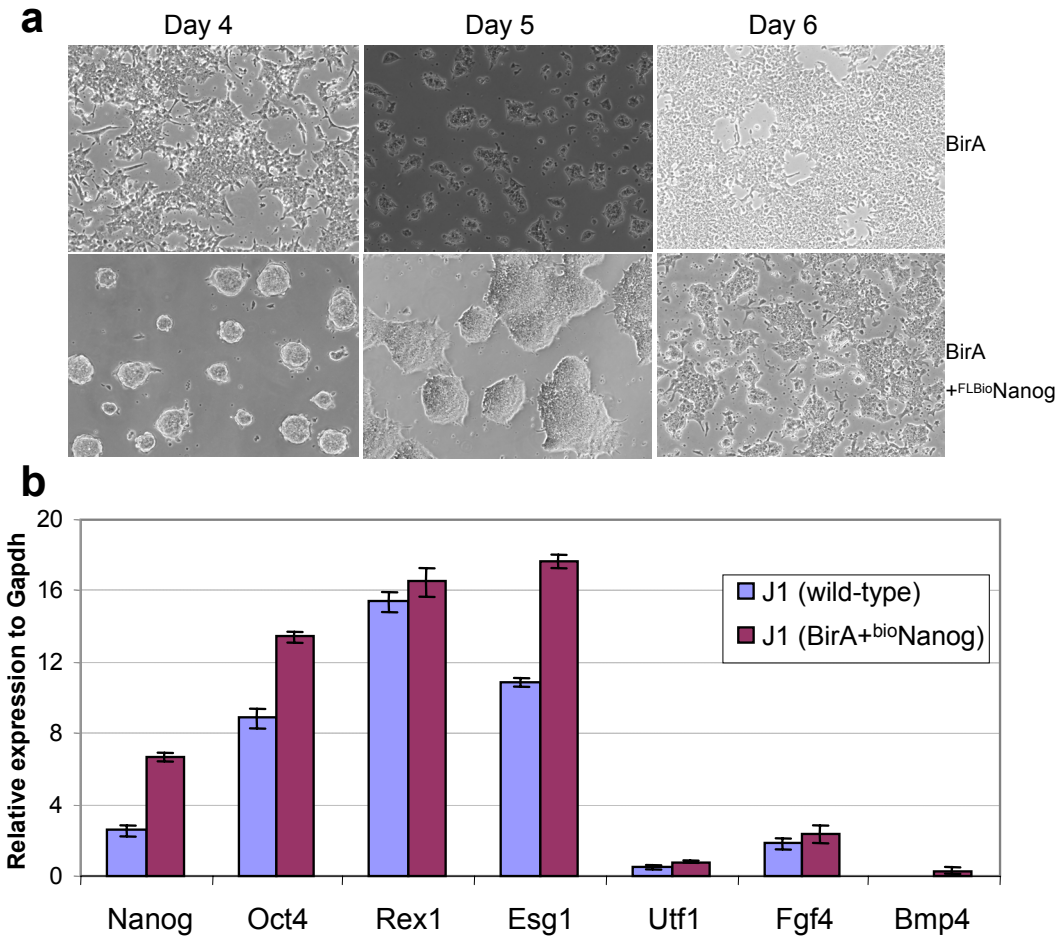
Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Cantor for biotin tagging constructs and for advice on affinity purification and chromatography; Y. Fujiwara and M. Kaku for technical assistance and for Oct4–GFP ES cells; T. De Lange for Rif1 antibody; S. Mackler for Nac1 antibody; S. Lowe and S. Elledge for shRNA vectors; M. Vidal and A.-L. Barabasi for advice and discussion on networks; and R. Tomaino and S. Gygi for performing LC–MS/MS and for providing advice and assistance in data collection and analysis. S.H.O. is an Investigator of the HHMI.

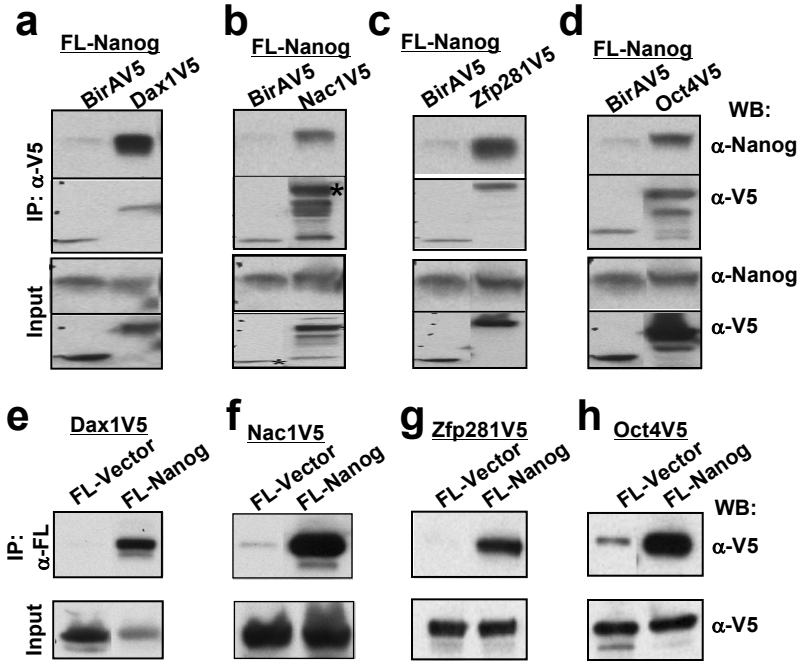
Author Contributions S.R. and J.C. contributed equally to this study. J.W. and S.H.O. conceived and initiated the study. J.W., J.C., X.S., D.N.L. and T.W.T. performed the experiments. S.R. and J.W. analysed data and bioinformatics of the network. J.W. and S.H.O. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.H.O. (stuart_orkin@dfci.harvard.edu).

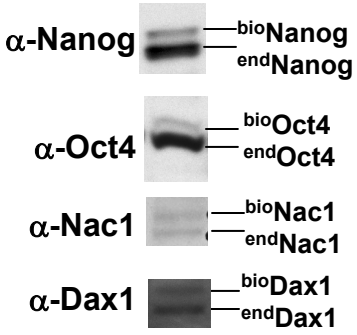
Supplementary Figure S1



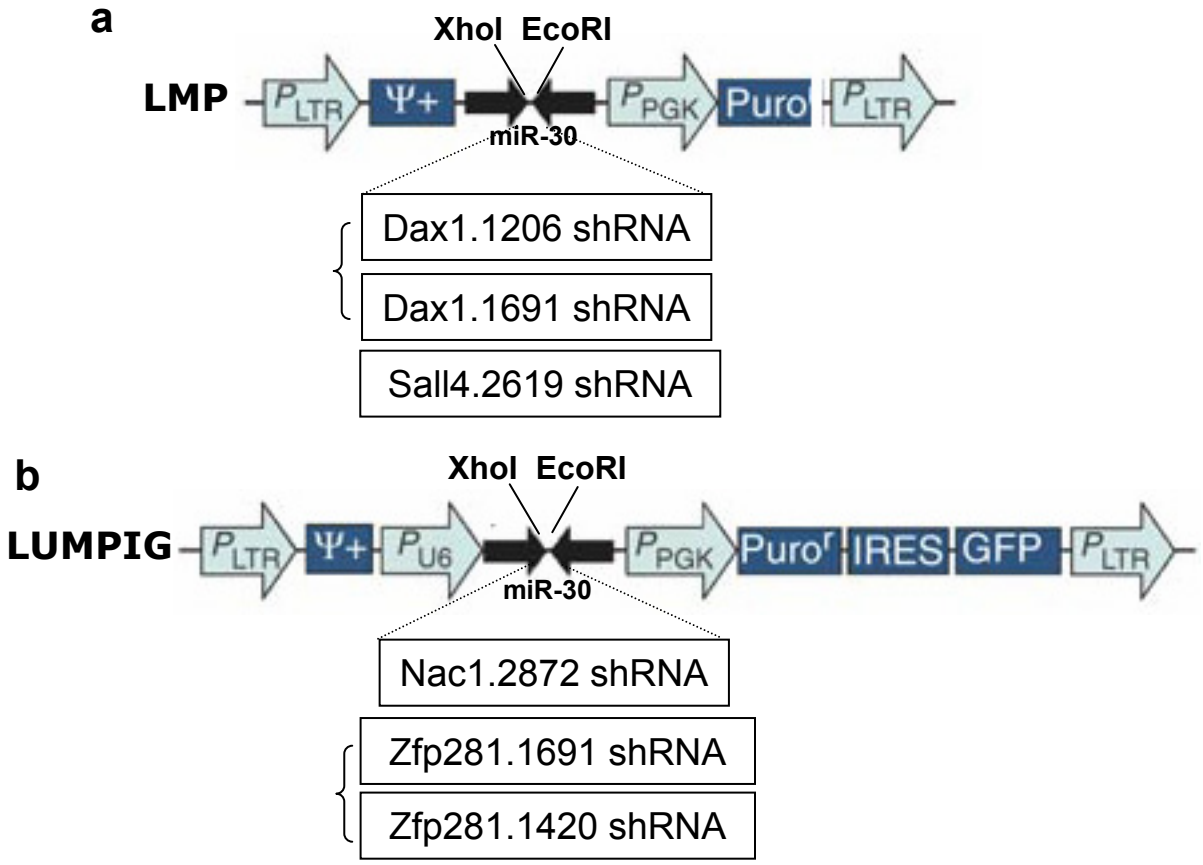
Supplementary Figure S2



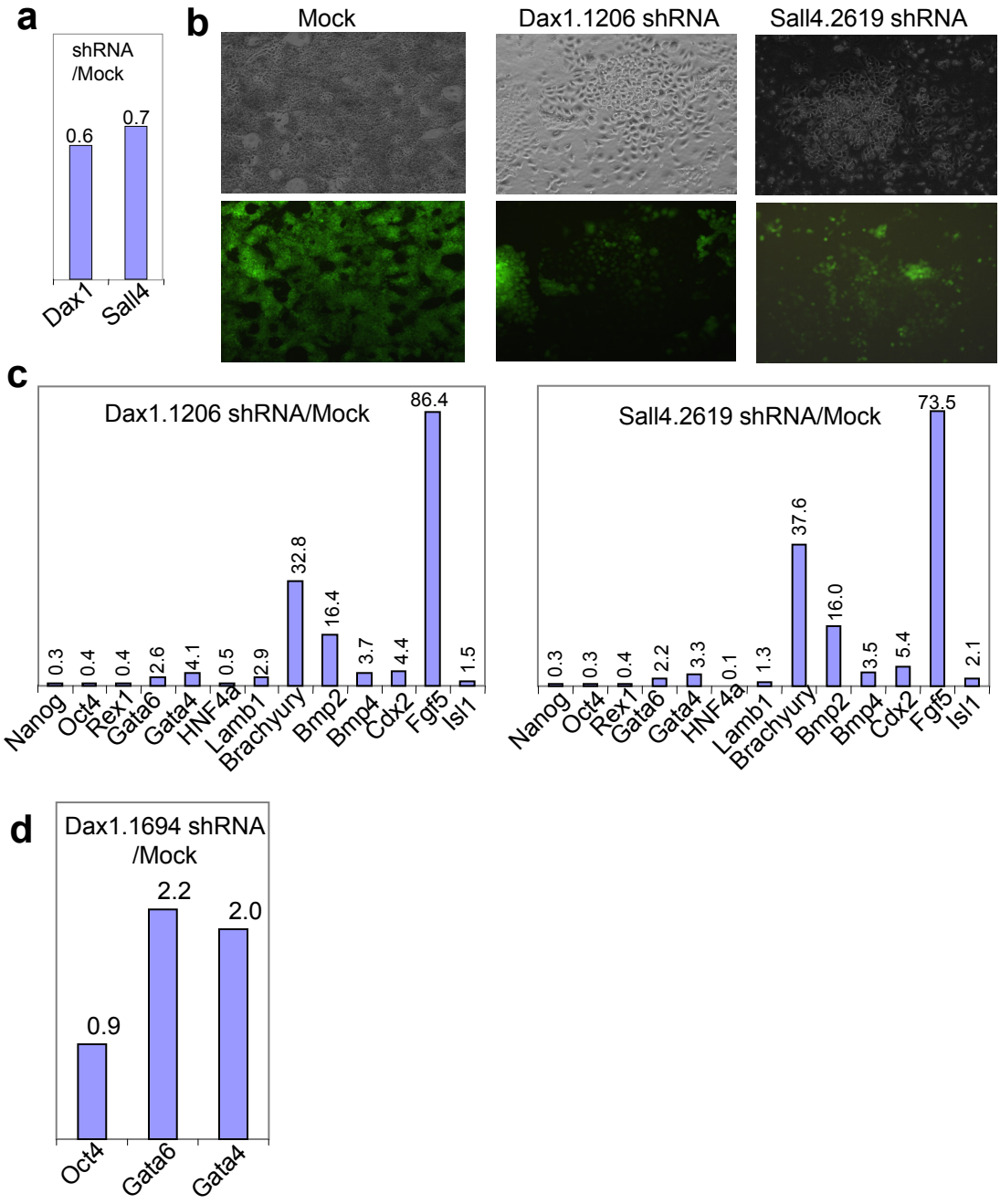
Supplementary Figure S3



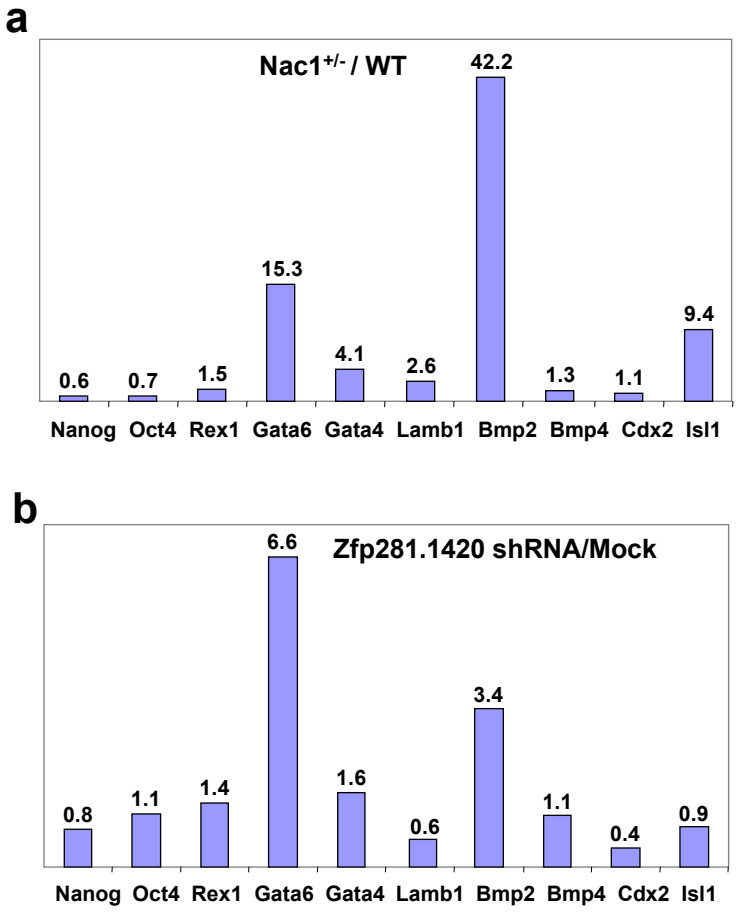
Supplementary Figure S4



Supplementary Figure S5

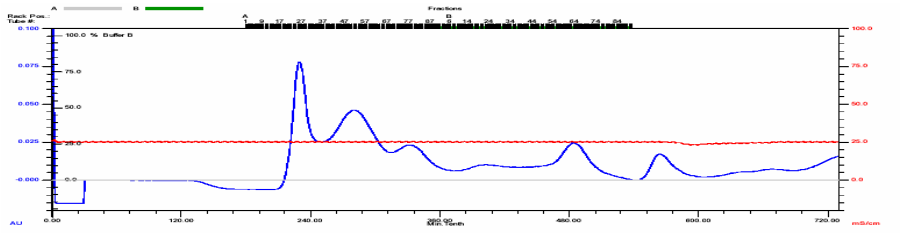


Supplementary Figure S6

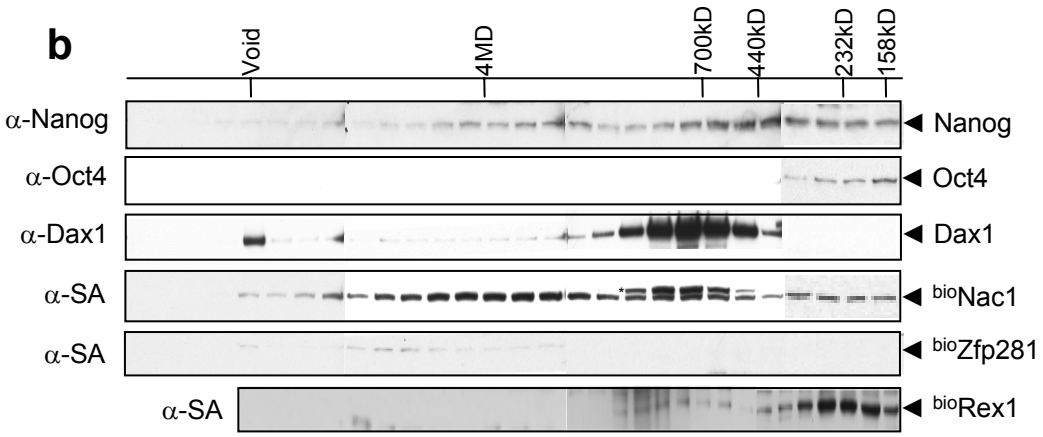


Supplementary Figure S7

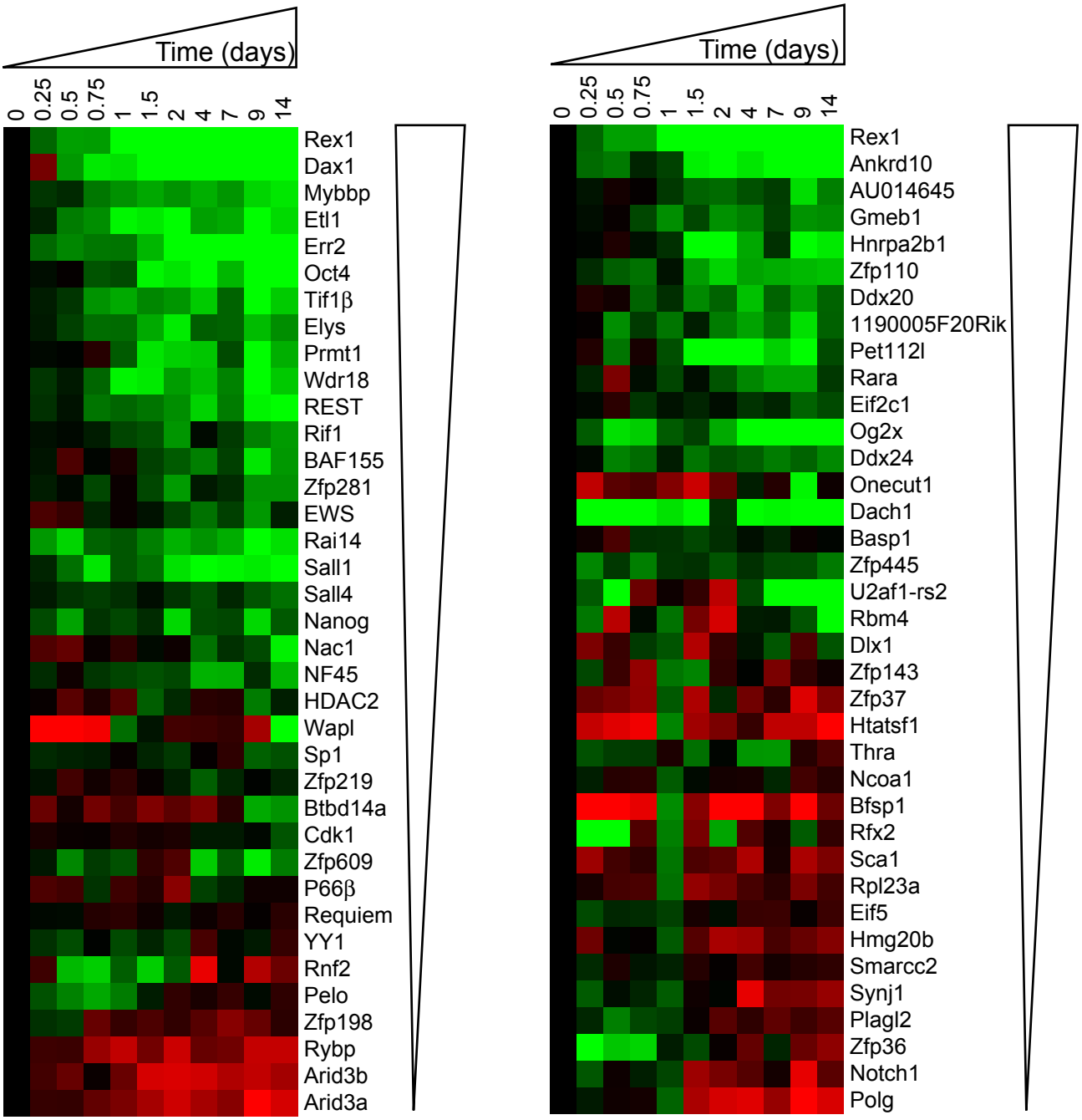
a



b



Supplementary Figure S8



Supplementary Data

Supplementary Methods

Plasmid construction. Plasmids for bacterial biotin ligase BirA and biotin tagging, and shRNA-mir expression were as described[1,2]. Plasmids for cDNA epitope tagging (pEF1αV5his series) were purchased from Invitrogen. All constructs were verified by DNA sequencing.

ES cell culture and transfection. ES cell lines (J1 or CJ7) were maintained in ES medium (DMEM; Dulbecco's modified Eagle's medium) supplemented with 15% fetal calf serum, 10^{-4} M 2-mercaptoethanol, 2 mM L-glutamine, 0.1 mM non-essential amino acid, 1% of nucleoside mix (100X stock, Sigma), 1000U/ml recombinant leukemia inhibitory factor (LIF; Chemicon) and 50 U/ml Penicillin/Streptomycin. Transfection of ES cells was performed according to manufacturer's protocol (Roche's FuGene 6).

Total lysate and nuclear extract preparation, gel filtration chromatography and Western blot analyses. Total lysates and nuclear extracts were prepared as described[3]. Gel filtration chromatography was performed according to manufacturer's manual (Amersham Biosciences). Briefly, nuclear extracts (10~20 mg) were applied to a S400 gel filtration column, fractions were collected and fractionated on a SDS-polyacrylamide gel followed by Western blotting.

For Western blot analysis, aliquots of total lysates or nuclear extracts (10~20 μg) were fractionated on a SDS-polyacrylamide gel and electroblotted onto PVDF membrane. Antibody incubation and chemiluminescence detection were performed according to manufacturer's instruction (Amersham Biosciences). Anti-Nanog, anti-Oct4 and anti-Dax1 antibodies were purchased from Santa Cruz Biotechnology, anti-streptavidin-HRP from Amersham, anti-Streptavidin agarose and anti-V5-HRP from Invitrogen, anti-V5 agarose and anti-M2 Flag antibody from Sigma, anti-Nac1 was a generous gift from Scott Mackler at University of Pennsylvania, anti-Rif1 was a gift from Titia DeLange at Rockefeller University.

Coimmunoprecipitation. 293T cells were transiently co-transfected with plasmids expressing FLAG-tagged Nanog and V5his-tagged candidate gene constructs. The BirAV5his construct was used as a negative control. Two days after transfection, total lysates or nuclear extracts were prepared and incubated with anti-V5-agarose or anti-M2 FLAG agarose overnight. On day 2, unbound material was washed away, and bound material eluted by boiling in Laemmli buffer and subjected to Western blot analyses.

A similar coimmunoprecipitation procedure was applied to J1 ES cells except that only V5his-tagged candidate gene constructs (along with BirAV5his) were used for transient transfection. Coimmunoprecipitation using anti-SA agarose was performed as described for affinity purification and mass spectrometry in the main text except that the eluate from SA capture was used for Western blot analyses.

RNAi and real-time PCR analysis. Oligonucleotides for shRNAs were designed using RNAi Central (RNAi Codex) as described[2,4]. Nac1 and Zfp281 shRNAs were cloned into

retroviral vector LUMPIG (LTR-U6-Mir30-PuroIresGFP) as described[2]. Dax1 and Sall4 shRNAs were cloned into the same vector from which the IresGFP cassette was removed (see Supplementary Fig. S4). Retrovirus generation and infection were performed according to manufacturer's manual (Clontech). Briefly, ES cells (either Oct4-GFP reporter line or wt CJ7 line) were infected with retrovirus expressing shRNAs, sorted for GFP^{high} cells or selected with puromycin (2 μg/ml). Cells were photographed and harvested at day 4 for Q RT-PCR analysis.

Total RNA was isolated using RNAeasy kit (Qiagen). First strand cDNA was synthesized using iScript cDNA synthesis kit (Biorad). Q PCR was performed on an iCycler iQ Thermal Cycler (Biorad) with SyBr Green PCR master mix (Biorad). The average threshold cycle (Ct) for each gene was determined from triplicate reactions and the levels of gene expression were normalized to β-actin. Data were presented as expression levels in knockdown cells relative to those in mock infected cells (shRNA/mock). Sequences for shRNAs can be found at http://codex.cshl.org and PCR primers are listed in Supplementary Table 2.

Protein sequence analysis by LC-MS/MS. Excised gel bands were cut into approximately 1 mm^3 pieces. Gel pieces were then subjected to a modified in-gel trypsin digestion procedure[5]. Gel pieces were washed and dehydrated with acetonitrile for 10 min. followed by removal of acetonitrile. Pieces were then completely dried in a speed-vac. Rehydration of the gel pieces was with 50 mM ammonium bicarbonate solution containing 12.5 ng/μl modified sequencing-grade trypsin (Promega, Madison, WI) at 4ºC. After 45 min., the excess trypsin solution was removed and replaced with 50 mM ammonium bicarbonate solution to just cover the gel pieces. Samples were then placed in a 37ºC room overnight. Peptides were later extracted by removing the ammonium bicarbonate solution, followed by one wash with a solution containing 50% acetonitrile and 5% acetic acid. The extracts were then dried in a speed-vac (~1 hr). The samples were then stored at 4ºC until analysis.

On the day of analysis the samples were reconstituted in 5~10 μl of HPLC solvent A (2.5% acetonitrile, 0.1% formic acid). A nano-scale reverse-phase HPLC capillary column was created by packing 5 μm C18 spherical silica beads into a fused silica capillary (100 μm inner diameter x ~12 cm length) with a flame-drawn tip[6]. After equilibrating the column each sample was loaded via a Famos auto sampler (LC Packings, San Francisco CA) onto the column. A gradient was formed and peptides were eluted with increasing concentrations of solvent B (97.5% acetonitrile, 0.1% formic acid).

As each peptide was eluted they were subjected to electrospray ionization and then they entered into an LTQ linear ion-trap mass spectrometer (ThermoFinnigan, San Jose, CA). Eluting peptides were detected, isolated, and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide. Peptide sequences (and hence protein identity) were determined by matching protein or translated nucleotide databases with the acquired fragmentation pattern by the software program, Sequest (ThermoFinnigan, San Jose, CA) (http://fields.scripps.edu/sequest/).

Selection criteria for exclusion/inclusion of proteins as candidates. First, all endogenously biotinylated proteins and associated enzymes, as previously described[1], are excluded; second, most proteins present both in BirA and ^{bio}Nanog samples during single-step purifications are excluded, except those that were present only in ^{bio}Nanog samples,

but not the BirA control, upon tandem purification. Third, for the proteins specific to ^{bio}Nanog samples, only those with ≥ 2 peptides from at least two independent purifications (either two singles or one single and one tandem) were included. Lastly, proteins with documented membrane, cytoplasmic, or mitochondrial localization were excluded.

Quantitative Chromatin Immunoprecipitation (Q-ChIP): Murine J1 embryonic stem cells expressing BirA and biotinylated proteins (Nac1, Nanog and Zfp281) were grown to 40-70% confluence. Cells were chemically crosslinked by the addition of one-tenth volume of a 10% formaldehyde solution for 10 minutes at room temperature. Cells were rinsed twice with 1xPBS and harvested with 0.25% trypsin for 5 min. followed by immediate inactivation with FBS. Cells were washed, lysed and sonicated to nucleosomal-sized fragments averaging 400-500 bp. Sonication was done with a Misonic Sonicator 3000 by delivering 10 pulses of 10 sec. with 30 sec. pauses between bursts. Samples were kept on ice throughout the sonication period. The sample was then centrifuged at 13000 RPM for 10 min. at 4^o to remove membrane debris and aliquotted into 1x10^7 cell/ml fractions for immediate immunoprecipitation with 1ml of material. The supernatant was pre-cleared with 40 μl of Protein A Dynabeads (Dynal-Invitrogen) for 30 min. and allowed to immunoprecipitate overnight with 40 μl of Streptavidin Dynabeads (Dynal-Invitrogen). Beads were washed with 2% SDS followed by three washes with buffer containing 100mM Tris pH 8.0, 500mM LiCl, 1% NP-40 and 1% deoxycholic acid. All washes were for 10 min. at room temperature while rotating. Elution was employed by adding 0.3M NaCl and crosslinks were reversed by overnight heating at 65^o. Supernatant was then harvested and spun for 7500 RPM for 2 min. at room temperature to remove residual beads and the sample was then treated with RNase A and proteinase K. The sample was then purified using a Qiagen miniprep column. Eluted ChIP was stored in aliquots at -20^o or processed immediately for quantitative real-time PCR.

Quantitative real-time PCR was performed on ChIPs with a Biorad iCycler and a 25 μl SYBR Green reaction using the following parameters: 95^{o}C for 3 min. followed by 40 cycles of (95^{o}C 20 sec, 60^{o}C 30 sec, 68^{o}C 30 sec.). The amount of each amplification product was determined relative to a standard curve of input chromatin. The dissociation curve for each primer pair confirmed a single reaction product. Reactions were performed in triplicate using samples from three independent cell harvests, and data are represented as the mean (± SEM). Data are shown as fold enrichment of biotin tagged protein binding compared to a control J1 ES cell line expressing the BirA biotin ligase gene alone. The primers used are listed in Supplementary Table 2.

Bioinformatics and gene expression analyses: Microarray data[7] were obtained for a time course experiment in which J1 ES cells were differentiated *in vitro* and RNA harvested for microarray analysis (see reference for details). All 41 candidate genes were analyzed in the microarray data set. Replicates were averaged and normalized to the initial expression level. For genes in which there were multiple probes, the probe with the best data fit (by linear regression) was chosen. Genes were then rank ordered based upon their correlation with Rex1, and displayed using Java Tree-View.

Supplementary Figure Legends

Figure S1 | Cells coexpressing BirA and ^{bio}Nanog are functionally equivalent to wildtype ES cells. a, FLAG-tagged and biotinylated Nanog (^{bio}Nanog) is biologically active. ES cells expressing BirA without (top) and with (bottom) tagged Nanog were grown in ES medium with LIF on day 1, LIF was withdrawn on day 2, and cellular morphology was monitored from days 4-6. **b**, Stem cell marker gene expression in wt J1 ES cells and J1 ES cells coexpressing BirA and ^{bio}Nanog.

Figure S2 | Confirmation of Nanog association by coimmunoprecipitation (coIP) in heterologous 293T cells. CoIP of Nanog and Dax1 (**a** and **e**), Nac1 (**b** and **f**), Zfp281 (**c** and **g**) and Oct4 (**d** and **h**) in 293T cells. 293T cells were transiently transfected with constructs indicated, total lysates were prepared, incubated with anti-V5-agarose (**a-d**) or anti-M2 FLAG (FL) agarose (**e-h**) overnight. Unbound material was washed away; bound material was eluted and subjected to Western blot analyses using antibodies indicated (top panels). Total lysates as input were also subjected to Western blot analyses with the indicated antibodies (bottom panels). Asterisks indicate the expressed proteins of correct size.

Figure S3 | Biotinylated Nanog, Oct4, Nac1 and Dax1 are expressed at or below endogenous protein levels. Twenty microgram of nuclear extracts from ES cells expressing biotinylated protein were subjected to Western blot analysis using the antibodies indicated. Both the biotinylated and endogenous levels of the protein are indicated. Zfp281 and Rex1, however, cannot be determined for relative expression levels between the biotinylated and endogenous protein due to the lack of reliable antibodies. Nevertheless, clones with medium or low expression levels relative to the endogenously biotinylated protein were selected for studies.

Figure S4 | Schematic depiction of retroviral shRNA constructs for RNAi. a. Vectors for expression of shRNAs targeting Dax1 and Sall4 contain the puromycin selection cassette. **b.** Vectors for expression of shRNAs targeting Nac1 and Zfp281 contain the puromycin resistance gene as well as a GFP marker. Details of the vectors are described in Supplementary Methods.

Figure S5 | Functional validation by RNAi. a, Quantitative RT-PCR analysis of Dax1 and Sall4 gene expression after shRNA-mediated knockdown in Oct4-GFP ES cells. Extent of RNA knockdown is blunted by the presence of undifferentiated cells that express shRNA at low level. **b,** Morphology of ES cells infected with virus containing empty shRNA vector (Mock; left two panels), Dax1 shRNA (middle panels), and Sall4 shRNA (right panels). Differentiation of ES cells expressing Dax1 and Sall4 shRNAs is evident by morphological changes under phase contrast microscopy (top panels) and concomitant silencing of GFP expression under fluorescence microscopy (bottom panels). Note, the mock-infected cells are nearly confluent and maintain GFP expression. **c,** Q RT-PCR analysis of stem cell markers, as well as lineage-specific markers as indicated in Dax1 knockdown cells (left) and Sall4 knockdown cells (right).

Figure S6 | Gel filtration analyses of Nanog, Oct4, Nac1, Zfp281, Dax1, and Rex1 complexes. Nuclear extracts from wt ES cells and ES cells expressing ^{bio}Nac1, ^{bio}Zfp281, and ^{bio}Rex1 were applied to S400 gel filtration column, and fractions (2 ml/fraction) containing protein complexes were collected and subjected to Western blot analyses using antibodies indicated. In the ^{bio}Nac1 samples the upper band (*) is an endogenous biotinylated protein.

Supplementary Table S1
Overlap of the ChIP-on-Chip[8] and ChIP-PET[9] Data with Interactome

List of target genes of Nanog and Oct4 from ChIP-on-Chip[8] and ChIP-PET[9] whose protein products are components of the interactome in this study

Nanog Targets	Oct4 Targets
Nanog^{m,h}	Nanog^{m,h}
Oct4^{m,h}	Oct4^{m,h}
Rybp^{m}	Rybp^{m}
Err2^{m}	Err2^{m}
REST^{m,h}	REST^{m,h}
Rif1^{m,h}	Rif1^{m,h}
Sall1^{m,h}	Sall1^{m,h}
Etl1^{m,h}	Etl1^{m,h}
Zfp281^{m,h}	Zfp281^{h}
Rex1^{h}	Rex1^{m}
Dmrt1^{m,h}	Dmrt1^{h}
Elys^{m}	Zfp219^{m}
Dax1^{m}	
Zfp198^{m}	
Sall4^{m}	
Btbd14a^{m}	
Rai14^{m}	
Arid3b^{m}	
NF45^{h}	
Cdk1^{h}	
Rbbp4^{h}	
Rnf2^{h}	

Note: X$^{m,\ h}$ indicates that gene X identified as targets of Nanog and/or Oct4 in both mouse (m) and human (h) ES cells. Shaded are the targets of both Nanog and Oct4.

Supplementary Table S2
Primers Used for Quantitative Real-Time PCR in This Study

Gene Name	Sequence	Application
Nanog	CAAGGGTCTGCTACTGAGATGCTCTG	RNAi
	TTTTGTTTGGGACTGGTAGAAGAATCAG	
Oct4	CTCCCGAGGAGTCCCAGGACAT	RNAi
	GATGGTGGTCTGGCTGAACACCT	
Rex1	GACAGACTGACCCTAAAGCAAGACGA	RNAi
	GGTATCCGTCAGGGAAGCCATCT	
Gata6	CTTGCGGGCTCTATATGAAACTCCAT	RNAi
	TAGAAGAAGAGGAAGTAGGAGTCATAGGGACA	
Gata4	CTCTATCACAAGATGAACGGCATCAAC	RNAi
	TCTGGCAGTTGGCACAGGAGAG	
HNF4a	GGAGCGTGAGGAAGAACCACATG	RNAi
	TTGTCTACCACACATTGTCGGCTAAA	
Lamb1	GCACAAACCAGAGCCCTACTGTATTG	RNAi
	GTTGAGGGTCTCGTGATAAGGGTCTC	
Brachury	CTGTGACTGCCTACCAGAATGAGGAG	RNAi
	GGTCGTTTCTTTCTTTGGCATCAAG	
Bmp2	CGCAGCTTCCATCACGAAGAAG	RNAi
	TGAGAAACTCGTCACTGGGGACAGA	
Bmp4	CACTGTGAGGAGTTTCCATCACGAAG	RNAi
	GGATGCTGCTGAGGTTGAAGAGGA	
Cdx2	GCGAAACCTGTGCGAGTGGATG	RNAi
	CGGTATTTGTCTTTTGTCCTGGTTTTCA	
Fgf5	CAAAGTCAATGGCTCCCACGAAG	RNAi
	CTACAATCCCCTGAGACACAGCAAATA	
Isl1	GGGATGGGAAAACCTACTGTAAAAGAGA	RNAi
	GTCGTTCTTGCTGAAGCCTATGCTG	
β-Actin	GATCTGGCACCACACCTTCTACAATG	RNAi
	CGTACATGGCTGGGGTGTTGAAG	
Gapdh	CATGTTCCAGTATGACTCCACTCACG	RNAi
	CCAGTAGACTCCACGACATACTCAGCA	
Utf1	GGTTCGCCGCCGCTCTACTG	RNAi
	GCAGGGGCAGGTTCGTCATTTTC	
Esg1	GTGGGTGAAAGTTCCTGAAGACCTGA	RNAi
	TGTGAGACATTCGAGATCCCTGTGG	
Fgf4	AGCGAGGCGTGGTGAGCATCTT	RNAi
	CGTTGTAGTTGTTGGGCAGAAGTATTTCT	
Dax1	GACACCAAAGAGTATGCCTATCTGAA	RNAi
	TCTGGTACTCTCTCTGCATCATCC	
Sall4	GGAGAGAAGCCTTTCGTGTG	RNAi
	CTCTATGGCCAGCTTCCTTC	
Nac1	AATGAAGAGGACGAAGAAGAAGATG	RNAi
	ATGCGACTGCGGGACTCA	
Zfp281	TGAGCCCAGGCACCCA	RNAi
	TGGAGAGGTGAAGACAAGCTGAC	
Gata6	CCGTAGTCCAGTGGGCAGAGAAAC	Q-CHIP
	CGGGCTGCTTTTCCAAATCG	
Gapdh	AAGCTCATGAGGCACAGAATGGTC	Q-CHIP
	TGGGTACATGGTGACTTTCCTAGGC	
β2-microglobulin	TCAGGTCCTAAGTCCTTTTCTGAGTGG	Q-CHIP
	TGGACCCTTGGTGCCCTACTATCTAG	

Supplementary File (Excel: MS1+MS2+MS3+Tandem Nanog-Associated Proteins)
List of all proteins identified by MS from ^{bio}Nanog but not BirA samples

Literature Cited

1. de Boer, E. et al. Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci U S A* **100**, 7480-5 (2003).
2. Dickins, R. A. et al. Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. *Nat Genet* **37**, 1289-95 (2005).
3. Gobert, S. et al. Identification of tyrosine residues within the intracellular domain of the erythropoietin receptor crucial for STAT5 activation. *Embo J* **15**, 2434-41 (1996).
4. Olson, A., Sheth, N., Lee, J. S., Hannon, G. & Sachidanandam, R. RNAi Codex: a portal/database for short-hairpin RNA (shRNA) gene-silencing constructs. *Nucleic Acids Res* **34**, D153-7 (2006).
5. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* **68**, 850-8 (1996).
6. Peng, J. & Gygi, S. P. Proteomics: the move to mixtures. *J Mass Spectrom* **36**, 1083-91 (2001).
7. Perez-Iratxeta, C. et al. Study of stem cell function using microarray experiments. *FEBS Lett* **579**, 1795-801 (2005).
8. Boyer, L. A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947-56 (2005).
9. Loh, Y. H. et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**, 431-40 (2006).