npg

ORIGINAL ARTICLE

Oct4 links multiple epigenetic pathways to the pluripotency network

Junjun Ding[1], Huilei Xu[2,*], Francesco Faiola[1,*], Avi Ma'ayan[2], Jianlong Wang[1]

[1]Department of Developmental and Regenerative Biology, Black Family Stem Cell Institute, Mount Sinai School of Medicine, New York, NY 10029, USA; [2]Department of Pharmacology and Systems Therapeutics, Mount Sinai School of Medicine, New York, NY 10029, USA

Oct4 is a well-known transcription factor that plays fundamental roles in stem cell self-renewal, pluripotency, and somatic cell reprogramming. However, limited information is available on Oct4-associated protein complexes and their intrinsic protein-protein interactions that dictate Oct4's critical regulatory activities. Here we employed an improved affinity purification approach combined with mass spectrometry to purify Oct4 protein complexes in mouse embryonic stem cells (mESCs), and discovered many novel Oct4 partners important for self-renewal and pluripotency of mESCs. Notably, we found that Oct4 is associated with multiple chromatin-modifying complexes with documented as well as newly proved functional significance in stem cell maintenance and somatic cell reprogramming. Our study establishes a solid biochemical basis for genetic and epigenetic regulation of stem cell pluripotency and provides a framework for exploring alternative factor-based reprogramming strategies.
Keywords: Oct4; ESCs; pluripotency; self-renewal; epigenetic regulation

Introduction

Pluripotency, the ability of a cell to give rise to all cell types of an organism, is a fundamental characteristic of embryonic stem cells (ESCs). The basis of pluripotency resides in conserved transcriptional regulatory networks [1, 2] and protein interaction networks [3-5] of numerous transcription factors (TFs) and epigenetic regulators, which act together to repress developmental genes and activate stemness genes in ESCs. Oct4, Sox2, and Nanog are well-known key components of the core regulatory network that governs ESC pluripotency, and epigenetic regulators such as the polycomb group proteins, SWI/SNF proteins, and Mi-2/NuRD complex proteins also play important roles in maintaining pluripotency [3]. Understanding the interactions among these pluripotency TFs and epigenetic cofactors is critical for maintaining as well as directly differentiating pluripotent stem cells.

Efforts to decipher the molecular basis for pluripotency of ESCs have culminated in the discovery of a set of reprogramming factors that, when ectopically expressed, directly convert somatic cells to so-called 'induced pluripotent stem cells' (iPSCs) [6]. These reprogramming factors, namely Oct4 and Sox2 in combination with Klf4 and c-Myc [6], or Nanog and Lin28 [7], are also known to be important factors for self-renewal and pluripotency of ESCs. Nanog can promote transfer of pluripotency after cell fusion [8] and ensure direct reprogramming of somatic cells to the pluripotent ground state [7, 9]. Various combinations and replacement of reprogramming factors with other pluripotency factors or small molecules have been achieved in reprogramming different types of somatic cells. However, the reprogramming process is slow and inefficient, and improved reprogramming requires additional epigenetic modifiers, suggesting the existence of epigenetic barriers to somatic cell reprogramming. Oct4 has largely remained an irreplaceable factor with only one exception [10]. In that exceptional case, the orphan nuclear receptor Nr5a2 (also known as Lrh-1), a known Oct4 activator, replaces Oct4 in the derivation of iPSCs from mouse somatic cells and enhances repro-

gramming efficiency partly through Nanog activation [10], consistent with a fundamental role of Nanog and Oct4 in stem cell pluripotency and somatic cell reprogramming. An improved understanding of genetic and epigenetic mechanisms by which the core ESC factors, Nanog and Oct4 in particular, regulate pluripotency should help in designing alternative or improved reprogramming strategies and providing mechanistic insights into somatic cell reprogramming.

Genetic studies [11-13] have defined the homeodomain transcription factor Nanog as the key self-renewal regulator that is essential for early development and for safeguarding the ground-state pluripotency of ESCs. ESCs lacking Nanog exhibit compromised self-renewal and tend to differentiate toward the primitive endoderm lineage [12]. In contrast, enforced expression of Nanog results in enhanced self-renewal at the expense of differentiation propensity [14]. To understand how Nanog functions in regulating self-renewal and maintaining/promoting pluripotency, we have tested and established an *in vivo* biotinylation strategy for affinity purification of protein complexes associated with Nanog, and constructed the first protein interaction network in mouse (m) ESCs (the Nanog interactome) [3]. The Nanog interactome encompasses Oct4 and multiple genetic and epigenetic regulators that individually and combinatorially contribute to stem cell pluripotency [3]. Recent reports suggest that Oct4 is essential for integrating the epigenetic machinery into the pluripotency network. For example, Oct4 cooperates with Nanog and Sox2 to repress *Xist* (X-inactive specific transcript) and thus couples X inactivation reprogramming to pluripotency [15]. Oct4 also interacts with several polycomb group proteins (e.g., Ring1B, Rybp) as part of the Nanog interactome [3] to maintain pluripotency [16]. In addition, Oct4 controls the chromatin architecture of ESCs by directly regulating downstream target genes encoding the H3K9 demethylases Jmjd1a and Jmjd2c, which modulate the H3K9 methylation status of the pluripotency factors Tcl1 and Nanog, respectively, to maintain stem cell identity [17]. Limited studies have been performed to dissect the biochemical basis for Oct4's diverse roles in both genetic and epigenetic regulation of stem cell pluripotency. The most notable ones are two recently published biochemical studies that used FLAG-based affinity purification of Oct4 complexes in mouse embryonic stem cells (mESCs) [4, 5]. These studies resulted in discouragingly few overlapping Oct4 partners [18], and left an open question of whether we have identified the *bona fide* Oct4 interactome in ESCs.

Here we report an extended Oct4 interactome composed of a much larger repertoire of interacting proteins than previously reported [3-5] using an advanced affinity purification approach with demonstrated effectiveness for affinity purification of protein complexes in ESCs [3, 19, 20]. We discovered and confirmed physical association and functional significance of a number of novel Oct4 partners. Our study provides solid biochemical evidence and strong functional validation that Oct4 is critical for epigenetic regulation of stem cell pluripotency. We demonstrate that the Oct4 interactome is connected with multiple chromatin remodeling and epigenetic regulatory protein complexes that are important for stem cell maintenance, pluripotency, and somatic cell reprogramming (iPSC generation).

Results and Discussion

An improved affinity purification strategy

Due to the dosage sensitivity of Oct4 protein for ESC maintenance [21], affinity purification of Oct4-associated proteins via ectopic overexpression of tagged Oct4 posed a limit in identifying Oct4 partners in our previous study [3] and in the two studies [4, 5] that shared a distressfully low number of overlapping partners [18]. Therefore, we decided to further investigate the Oct4 interactome with an improved *in vivo* biotinylation-based affinity purification strategy [19, 20] for purification of Oct4-associated protein complexes in mESCs.

First, we established a transgenic mESC line that expresses only biotinylated Oct4 (^{bio}Oct4) replacing the doxycycline (dox) suppressible Oct4 (^{dox}Oct4) in ZHBTc4 cells [21] via lentivirus infection and dox treatment (Figure 1A). We established four clonal cell lines (dubbed ZO4B1-4) that are equivalent in morphology and stem cell characteristics as shown below representatively for ZO4B4, the line that was used for subsequent affinity purification. In the presence of dox, ZO4B4 ESCs were sustained by ^{bio}Oct4 only (Figure 1B, left lane) and maintained ESC identity manifested by their typical dome-shaped morphology, positive staining for alkaline phosphatase (AP) activity (Figure 1C, top), expression of wild-type levels of the stem cell markers such as Nanog and Sox2 (Figure 1D), and normal ESC clonogenicity (Supplementary information, Figure S1A). In contrast, upon dox withdrawal, ZO4B4 cells expressed both ^{dox}Oct4 and ^{bio}Oct4 (Figure 1B, right lane) that together cause differentiation and concomitant loss of AP activity (Figure 1C, bottom), consistent with the previously reported differentiation phenotype upon Oct4 overexpression [21]. It should be pointed out that the maximum bioOct4 level, although lower than doxOct4 level (Figure 1B), is still within the range that can functionally maintain the ESC state (Figure 1C and 1D) as previously
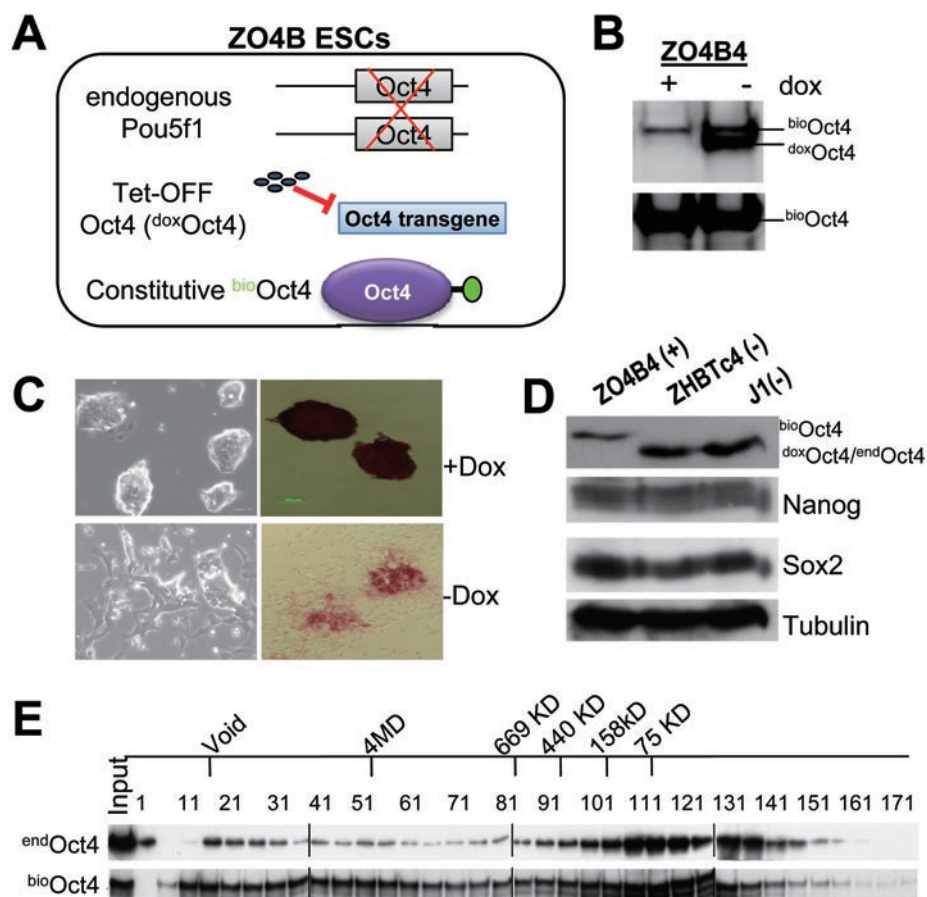
Figure 1 Establishment of *in vivo* biotinylation system for affinity purification of Oct4 protein complexes in modified ZHBTc4 (ZO4B) ESCs. **(A)** Schematic depiction of the modified ZHBTc4 ESCs expressing biotinylated Oct4 protein. **(B)** Western blot analysis of relative expression levels of dox suppressible (^{dox}Oct4) and constitutively active biotinylated Oct4 (^{bio}Oct4) in the presence (+) or absence (−) of dox treatment. Top, anti-Oct4 blot; Bottom, SA-HRP blot. **(C)** Morphology of ZO4B4 ESCs in the presence (top) or absence (bottom) of dox. Images of colonies that are unstained (left panels) or stained for AP activity (right panels) are shown. **(D)** Normal expression levels of Oct4, Nanog, and Sox2 in ZO4B4, ZHBTc4, and J1 ESCs in the presence (+) or absence (−) of dox treatment. **(E)** Size exclusion chromatography (gel filtration) of nuclear extracts of J1 (top) and ZO4B4 (bottom) ESCs. Endogenous Oct4 protein complexes (^{end}Oct4) in wild-type J1 ESCs were detected by anti-Oct4 western blot (top), and the Oct4 protein complexes in ZO4B4 ESCs (^{bio}Oct4) were also detected by anti-Oct4 western blot (bottom).

reported [21]. Second, we utilized conditions that have been optimized for the extraction of the required nuclear components with a significant initial removal of contaminating cytoplasmic and nuclear components that are not required for transcription activity [22]. These include dialysis to low salt (100 mM) and treatment of nuclear extracts with Benzonase to minimize DNA-tethered protein interactions (Supplementary information, Figure S1B) while preserving strong and specific affinity of ^{bio}Oct4 to streptavidin-agarose (SA) beads (Supplementary information, Figure S1C). Third, we reduced the amount of detergent nonidet P40 during affinity purification by 10-fold than previously applied [3, 19].

The combined modifications described above have

greatly improved the preservation of large multi-protein complexes during mESC nuclear extract preparation, as demonstrated by the fractionation of the Oct4 protein complexes on a gel filtration column (Figure 1E). The new gel filtration results indicate that the size of multiprotein complexes associated with both endogenous (^{end}Oct4) and biotinylated Oct4 (^{bio}Oct4) spans over a wide range from ~50 kD to several megadaltons (MD), far greater than initial observations by us [3] and others [5] that the majority of Oct4 was fractionated at approximately its own molecular weight on a gel filtration column due to dissociation of the protein complexes under suboptimal buffer condition. Therefore, our improved purification protocol ensures the preservation of most

Oct4 interactions. Our gel filtration results also show that ^{bio}Oct4 participates in similar complexes as those of endogenous Oct4 (^{end}Oct4) (Figure 1E), indicating functional integrity of ^{bio}Oct4.

Affinity purification and mass spectrometry identification of Oct4 protein complexes in mESCs

Using the improved affinity purification strategy de-scribed above, we performed three independent SA-mediated affinity purifications followed by mass spectrometry and identified 198 high-confidence Oct4-interacting proteins (Figure 2 and Supplementary information, Table S1). When we compared our dataset with the two recently published Oct4 interactome studies [4, 5], we found a total of 43 proteins that were also discovered by the previous studies to be Oct4-interacting proteins (Figure
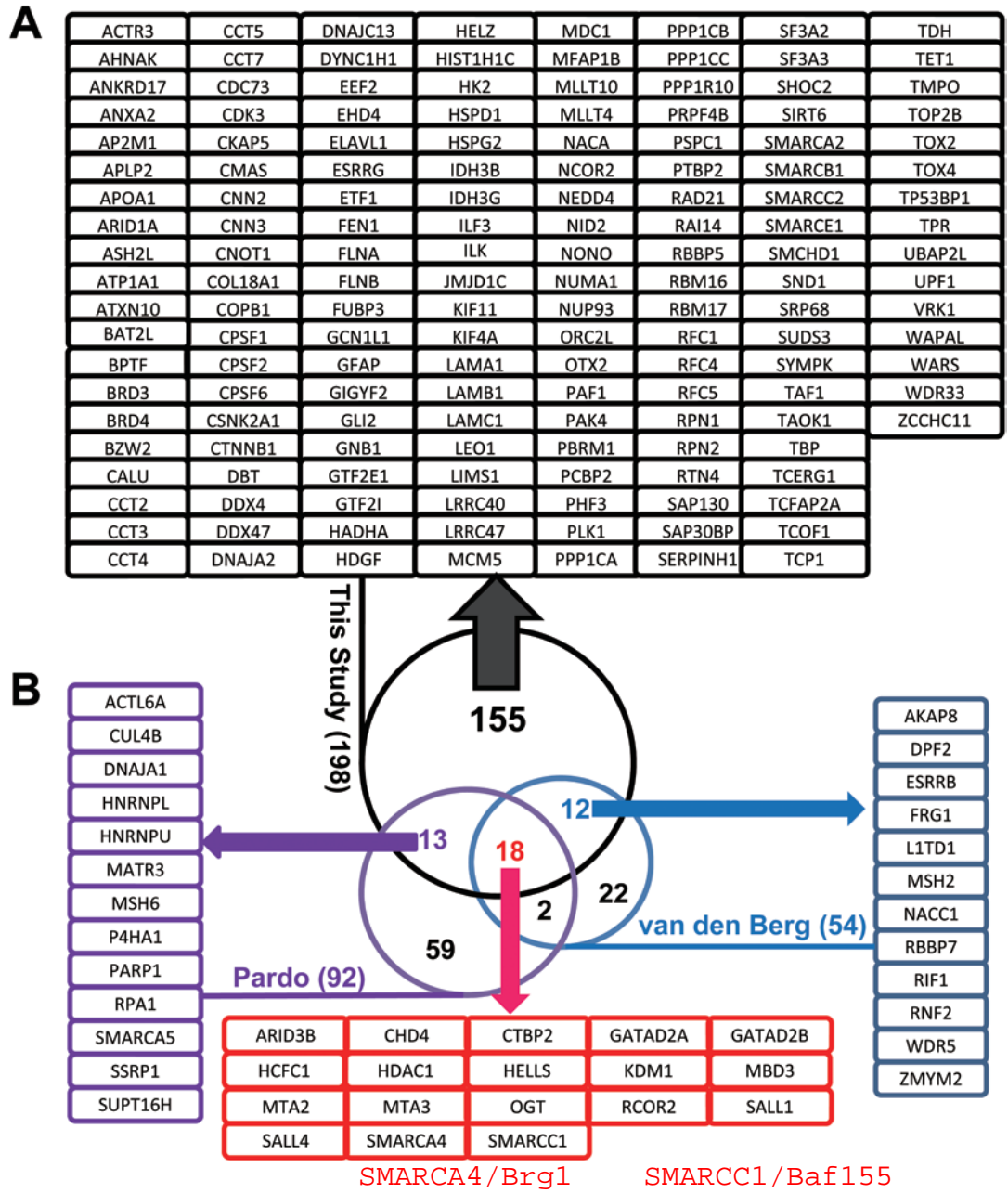

Figure 2 Summary of Oct4-interacting proteins in mouse ESCs. **(A)** List of the 155 novel Oct4-associated proteins from affinity purification in ZO4B4 ESCs. **(B)** Comparison of our Oct4 interactome with the two published Oct4 network studies [4, 5]. The overlapping proteins between our study and the two published studies are listed as blue, purple and red frames.

2B). These 43 common proteins constitute as many as 56% (30/54) of the Oct4 partners from one study [5] and 34% (31/92) from the other [4]. In particular, our Oct4 interactome identified 18 out of the 20 (90%) overlapping Oct4-interacting proteins from the two prior studies [4, 5]. These results suggest that our improved *in vivo* biotinylation-based affinity purification captured the majority of high-confidence Oct4-interacting proteins. More importantly, we uncovered 155 novel Oct4-interacting proteins that were not present in the previous low affinity FLAG tag-based studies [4, 5] (Figure 2A). To justify and substantiate the existence of such additional Oct4 interactors than previously reported [4-5], we found that several studies have individually documented the interactions of Oct4 with CTNNB1 [23, 24] and with multiple components of the PAF1 complex (Leo1, Cdc73, Paf1) [25, 26] and COMPASS-like protein complex (Rbbp5) [27], all of which are among the 155 novel Oct4-interacting proteins we identified in this study (Figure 2A). Our coimmunoprecipitation and immunoprecipitation (CoIP/IP) data also confirmed physical association of Oct4 with additional novel partners including Ash2l, Kif11, and Ppp1cc (Figure 4B). We also performed Oct4 antibody-based immunoprecipitation. However, like studies done by others using the native Oct4 antibody for affinity purification [4, 28], we were able to confirm only a limited number of endogenous interactions due to the low affinity/specificity of the antibody (Supplementary information, Table S2). These data highlight the effectiveness and advantages of our *in vivo* biotinylation-based affinity purification approach as well as the validity of additional Oct4-interacting partners identified in this study.

Like the published studies [4, 5, 28], we also failed to uncover with high confidence the two well-known Oct4 partners Nanog and Sox2. The Nanog protein has been speculated to be resistant to tryptic digest [5], and the Nanog-Oct4 interaction has also been deemed to be so weak that it can only be detected with crosslinking [29]. The Sox2-Oct4 interaction may also be weak and/or stabilized by DNA binding to the Oct-Sox sequence. This would explain why Oct4 interaction with Sox2 was only confirmed by one previous study [5], but not by the other [4] or this study. It is important to point out that we did uncover Sox2 in one of the purifications (with 2 peptides), but it was not included in the final candidate list based on our stringent selection criteria. We observed interaction of Oct4 with some low-expression proteins such as Otx2, which is highly enriched in epiblast stem cells and human ESCs [30]. We postulate that Oct4 interaction with such low-abundance proteins may be necessary for ESCs to be primed for lineage specification.

We further validated our Oct4 interactome (Figure 2) using additional approaches. First, we found that our one-step purification approach allowed us to identify 39 proteins (green and yellow circles in Supplementary information, Figure S2A) of an Oct4 interaction network previously constructed by an iterative tagging strategy of eight pluripotency factors [3-5] (big circles in Supplementary information, Figure S2A). Importantly, we were able to capture the Oct4 partners that were previously reported to be indirectly associated with Oct4 (yellow circles in Supplementary information, Figure S2A) via other pluripotency factors (large circles in Supplementary information, Figure S2A). Second, we confirmed that components of Oct4 protein complexes participate in common pathways and are co-regulated in controlling self-renewal and pluripotency of ESCs in both Oct4 depletion [31] and embryoid body differentiation assays [32] (Supplementary information, Figure S2B). In particular, many well-documented self-renewal regulators and factors important for early development and pluripotency of mESCs were downregulated (Supplementary information, Figure S2B). Expression of genes encoding a smaller subset of Oct4-interacting proteins either remained unchanged or increase with time, representing factors that may have additional functions during differentiation and cell fate determination (Supplementary information, Figure S2B). Third, GO analyses revealed that the Oct4 interactome is over-represented for DNA and chromatin-binding factors (Supplementary information, Figure S3A), and correspondingly, enriched in biological processes such as transcription, regulation of transcription, and chromosome/chromatin organization (Supplementary information, Figure S3B). KEGG pathway analysis revealed that the Oct4 interactome is enriched for factors involved in DNA mismatch repair, DNA replication, and cancer (Supplementary information, Figure S3C). These data are consistent with Oct4's pluripotency transcription factor status and highlight its potential roles in epigenetic regulation of ESC pluripotency. In addition, we also found enrichment for the "focal adhesion" pathway that is active in pluripotent ESCs and activated by the Yamanaka factors during reprogramming [33], consistent with the essential role of Oct4 in iPSC generation.

Taken together, these results indicate that the Oct4 interactome is composed of a much larger repertoire of interacting proteins, than previously reported [4, 5] with demonstrated as well as implicated roles in stem cell maintenance and somatic cell reprogramming. Therefore, our study should provide a much richer resource for discovery of novel self-renewal regulators as well as reprogramming factors.
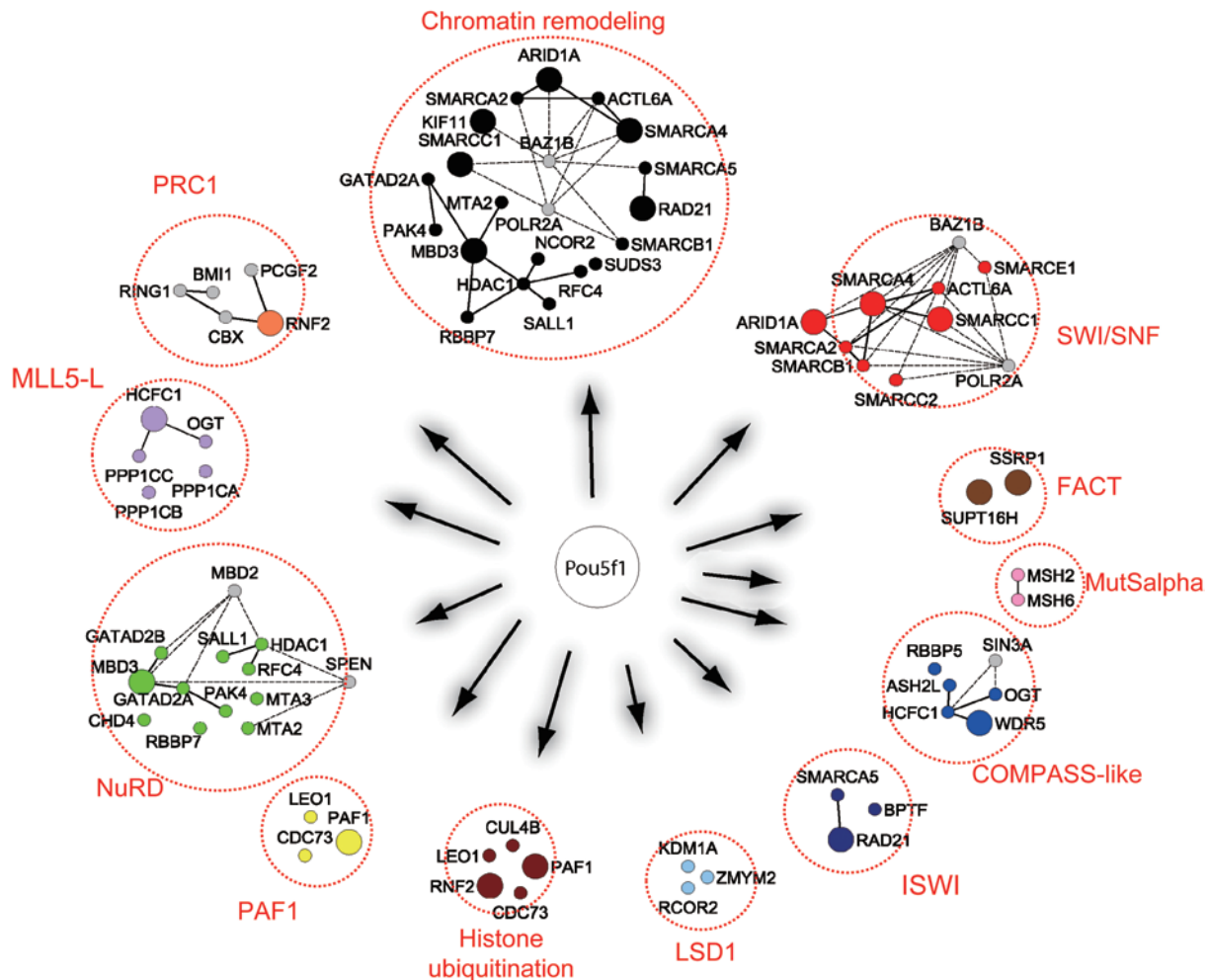
Figure 3 Connection of Oct4 with multiple epigenetic regulatory pathways. Summary of major protein complexes (red dotted circles) associated with the Oct4 interactome. Big filled circles with color are Oct4-interacting proteins that are also the RNAi hits from published genome-wide RNAi studies (see also Figure 5). Small filled circles with color are Oct4-interacting proteins identified in this study. Other components of the complexes not present in the Oct4 interactome are shown in gray small circles.

Oct4 links to multiple epigenetic pathways for pluripotency of ESCs

When using the Genes2Networks tool [34] to analyze the features of our Oct4 interactome, it is striking that multiple epigenetic regulatory complexes are associated with Oct4 (Figure 3). In addition to the known repressor complexes that are also present in the Nanog interactome (i.e., the PRC1, NuRD, and SWI/SNF complexes) [3], we uncovered in the Oct4 interactome components of regulatory protein complexes such as LSD1, FACT, COMPASS-like, MLL5-L, MutSalpha, ISWI, and PAF1 complexes (Figure 3). Of note, the common 18 candidate proteins from all three studies (Figure 2B) include major components of the NuRD, SWI/SNF, and LSD1

complexes, as well as several other factors that have not been previously studied (Supplementary information, Figure S4), indicating the importance of these epigenetic regulatory pathways and factors for stem cell function. The potential physical connection with the pluripotency network and functional significance of these epigenetic regulatory complexes for stem cell maintenance and/or somatic cell reprogramming are supported by multiple independent studies. For example, it is well established that SWI/SNF chromatin-remodeling complexes are important for both stem cell pluripotency [35, 36] and reprogramming [37]. A genome-wide RNAi study [25] and two targeted knockdown studies [27, 38] have confirmed Wdr5 and Hcfc1, two subunits of the COMPASS-like

complex, to be critical for stem cell maintenance as well as for efficient iPSC generation [27]. The FACT complex is a conserved chromatin-remodeling complex implicated in DNA replication, basal and regulated transcription, and DNA repair. A knockout study demonstrated an essential role of the subunit Ssrp1 for blastocyst growth and survival [39], and a biochemical study showed the physical association of FACT with Chd1, a chromatin-remodeling factor recently shown to be essential for open chromatin and pluripotency of ESCs and for reprogramming somatic cells to the pluripotent state [40]. Genome-wide RNAi studies identified both core proteins (Ssrp1 and Supt16h) of this complex as important factors for stem cell maintenance [41]. The LSD1 complex demethylates mono- and di-methylated H3K4, and LSD1-null ESCs exhibit defects in differentiation [42, 43]. A more recent study showed that Rcor2, a component of the LSD1 complex, regulates ESC properties and substitutes for Sox2 in iPSC generation [44]. The ISWI complex has been demonstrated to be required for proper stem cell differentiation. Its component Smarca5 is important for ICM survival [45], whereas Bptf is important for cell fate specification and pluripotency [46]. The MLL5-L protein complex can methylate H3K4, and the knockdown study on its major component Hcfc1 [38] confirmed functional importance of this complex in stem cell maintenance.

To further validate the potential physical association of Oct4 with these factors, we examined co-fractionation of these proteins with Oct4-associated protein complexes on a size exclusion column and measured the size of multiprotein complexes of Oct4 and its associated proteins by western blot (Figure 4A). We confirmed that Ring1B (Rnf2), Supt16h, Ssrp1, Msh2, Msh6, and Ppp1cc all form large multiprotein complexes ranging from ~70 kD to several MD, similar to or overlapping with what has been observed for the Oct4 protein complexes (red rectangle in Figure 4A). We have already confirmed co-fractionation of Oct4 with several components of the COMPASS-like protein complex (Wdr5, Rbbp5 and Ash2l) [27]. By contrast, components of LSD1 (Zmym2), chromatin-remodeling (Kif11) complexes and the transcription factor Arid3b participated in much larger protein complexes (~4 MD) that also partly overlap with Oct4 protein complexes (blue rectangle in Figure 4A). We then performed CoIP/IP and confirmed interactions of Oct4 with components of many of these epigenetic regulatory protein complexes including Supt16h (FACT), Zmym2 (LSD1), Ring1B/Rnf2 (PRC1), Msh2/6 (MutSalpha), Ash2l (COMPASS-like), Kif11 (chromatin remodeling) and Ppp1cc (MLL5-L) (Figure 4B). Interaction of Oct4 with the COMPASS-like complex through Wdr5 was independently confirmed and its functional

significance in controlling stem cell identity and pluripotency was demonstrated in our recent study [27]. In addition, we also confirmed physical association of Oct4 with Arid3b (Figure 4B), a transcription factor that may play a role in chromatin structure modification [47]. Taken together, these data support the physical association of Oct4 with multiple chromatin remodeling and epigenetic regulatory protein complexes in ESCs.

Functional validation of Oct4 interactome for stem cell pluripotency and somatic cell reprogramming

We assessed the functional significance of our Oct4 interactome by analyzing the presence of factors in the interactome that are also positive hits from published genome-wide RNAi studies on stem cell maintenance [41, 48-50]. We found that 31 proteins (P-value ~2 × 10^{-9} from Fisher exact test) from our candidate list were also positive hits in the RNAi studies (Figure 5A and 5B, details are summarized in Supplementary information, Figure S5), suggesting that our Oct4 interactome is significantly enriched for factors required for stem cell maintenance. Notably, many of these positive RNAi hits in our Oct4 interactome are also components of chromatin regulatory complexes (big circles in Figure 3), supporting the important roles of these epigenetic regulatory protein complexes in stem cell maintenance.

Somatic cell reprogramming by defined factors is a slow epigenetic process that entails a gradual reconstruction of the pluripotency network in somatic cells. In this regard, it is noteworthy that, as an essential reprogramming factor, Oct4 physically connects with multiple epigenetic regulatory complexes via protein-protein interactions (Figure 3). Importantly, components of many such complexes including SWI/SNF chromatin-remodeling complex (Smarcc1, Smarca4, Arid1a) [37], COMPASS-like complex (Wdr5) [27] and LSD1 complex (Rcor2) [44] have recently been proved to be important factors in facilitating reprogramming. Oct4, Sox2, and Nanog as the core pluripotency factors as well as key reprogramming factors are interrelated at both the protein-protein interaction level [3] and at the transcriptional expression level [2]. Remarkably, we found that ~53% (104/198, P-value < 10^{-15}, Fisher's test, two-sided) of the genes encoding Oct4 interactome proteins are direct targets of Oct4 (blue arrows in Figure 6), and ~66% (130/198, P-value < 10^{-15}, Fisher's test, two-sided) of them are targets of at least one of the three core pluripotency/reprogramming factors, reinforcing the notion that the core ESC factors co-occupancy and feedback regulation are important features of the pluripotency network in ESCs. More importantly, promoters of over 21% (42/198, P-value < 10^{-9}, Fisher's test, two-sided) of the genes encod-
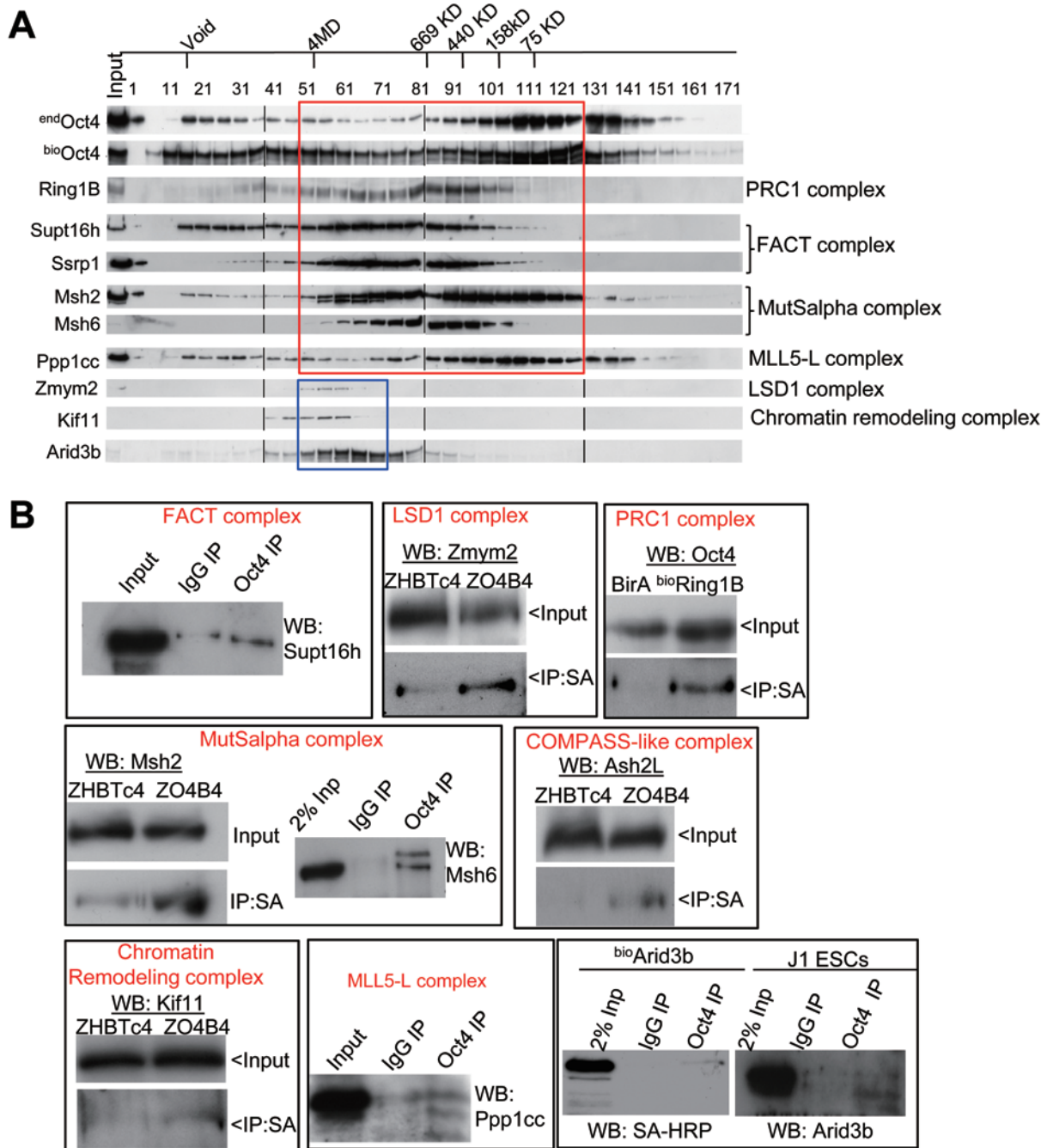
Figure 4 Validation of physical association of Oct4 with multiple epigenetic regulators. **(A)** Size exclusion chromatography analysis of Oct4-associated protein complexes. Nuclear extracts were fractionated in a DuoFlow BioLogic System and eluates were analyzed by western blotting with indicated antibodies to evaluate co-fractionation of endogenous endOct4 and biotinylated bioOct4 with PRC1, FACT, MutSalpha, MLL5-L, LSD1, Chromatin-remodeling complexes, as well as Arid3b. Migration of molecular markers is indicated above the panels, antibodies for WB are shown on the left and the complex name on the right. The two co-fractionation patterns of Oct4 with other factors are indicated with red and blue rectangles. **(B)** Validation of physical interactions between Oct4 and factors associated with multiple protein complexes. For validation of Oct4 interaction with Zmym2, Msh2, Ash2l, and Kif11, nuclear extracts from ZHBTc4 (control) or ZO4B4 (bioOct4) cells were incubated with SA-beads as described in Materials and Methods. Precipitated samples were analyzed by western blotting with indicated antibodies. For IP with anti-Oct4 antibody, nuclear extracts were prepared from J1 ESCs (for Supt16h, Msh6, Ppp1cc, Arid3b) or bioArid3b ESCs, and endogenous Oct4 complexes were immunoprecipitated with an Oct4 antibody followed by western blot with indicated antibodies or SA-HRP. The interaction of Ring1B with Oct4 was validated by SA-IP of nuclear extracts from bioRing1B, and control BirA ESCs followed by western blot with anti-Oct4.
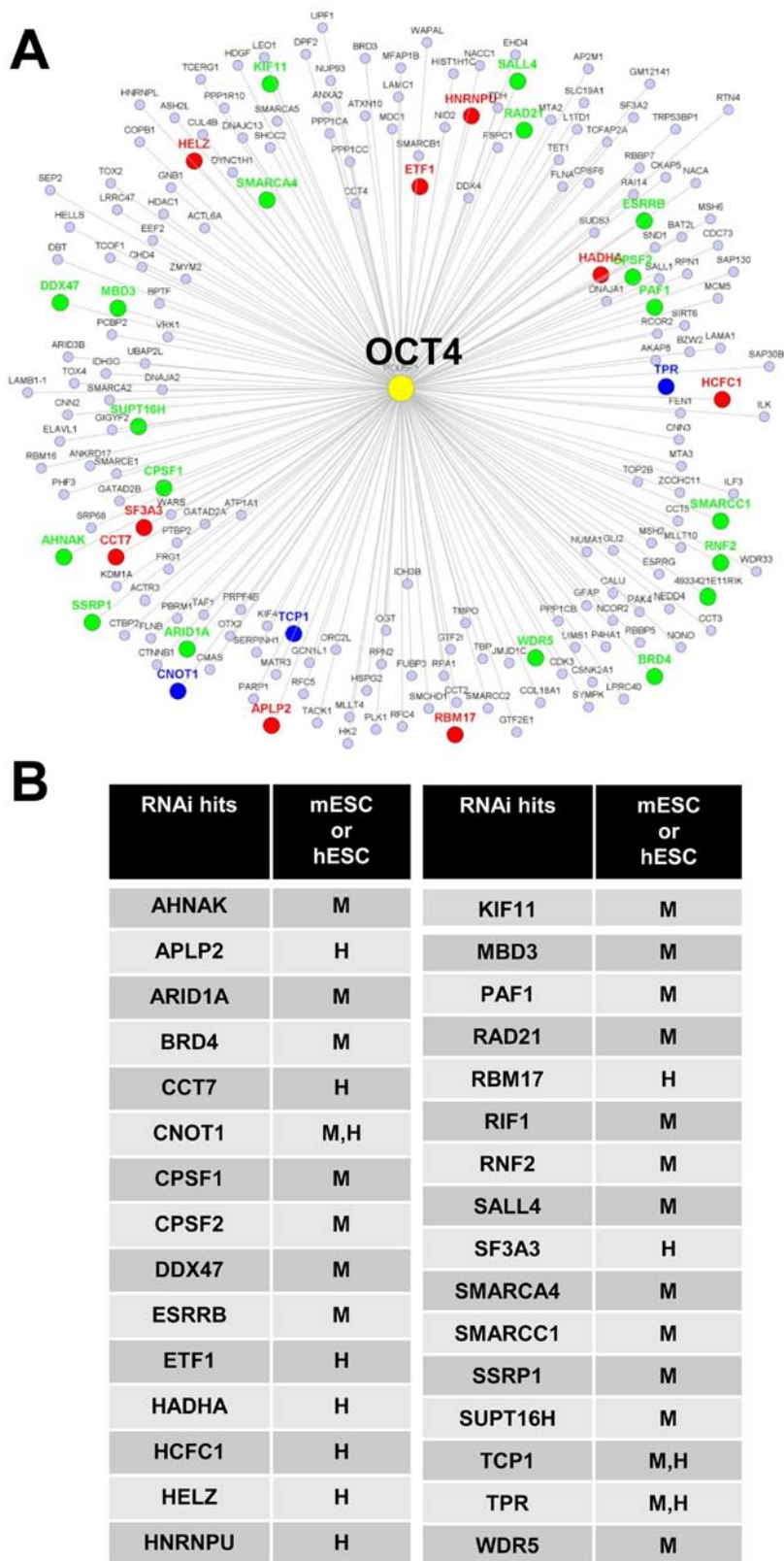
Figure 5 Functional validation of the Oct4-interacting proteins by RNAi. **(A)** The Oct4 interactome is enriched for positive RNAi hits required for stem cell maintenance. Green and red circles denote RNAi studies in mESCs [25, 41, 48] and hESCs [50], respectively. Blue circles denote RNAi studies in both mESCs and hESCs. The *P* value is ~2 × 10^{-9} from Fisher exact test. **(B)** The list of Oct4-interacting proteins whose encoding genes are also the positive RNAi hits shown in **A**.
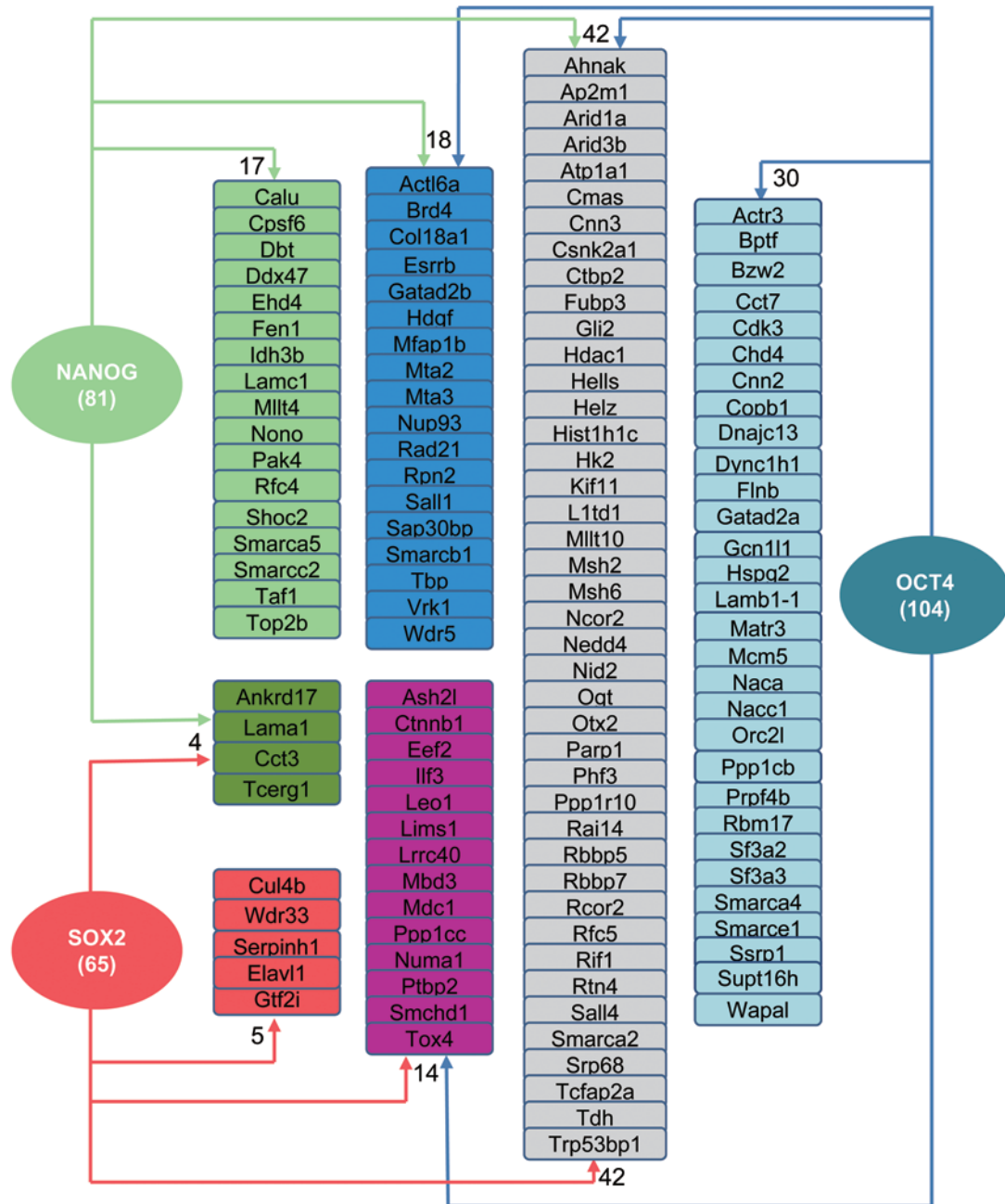
Figure 6 Enrichment of the Oct4 interactome for factors whose encoding genes are downstream targets of the core pluripotency factors Nanog, Oct4, and Sox2. We utilized previously published genome-wide binding data for Sox2, Oct4, and Nanog in mESCs [1-2] to generate the protein-DNA regulatory network. Target genes are represented as smooth-edged rectangles and upstream transcription factors are represented as filled ovals. Arrows emanating from each upstream transcription factor refer to its binding to the indicated genes. The total number of genes occupied by each of the core factors is indicated.

ing the Oct4 interactome proteins (shaded gray in Figure 6) are co-bound by all three core factors Nanog, Oct4 and Sox2. Many of these Oct4 partners are components of multiple epigenetic regulatory complexes, some of which have been proved to be reprogramming factors or facilitators (e.g., Arid1a/chromatin-remodeling complex [37], the Rcor2/LSD1 complex [44], and the Sall4/NuRD complex [51-52]). The remaining factors in this group represent the prime candidate regulators of ESC self-renewal and pluripotency as well as potentially important

factors and/or effectors of somatic cell reprogramming, which is worthy of future investigation.

In summary, we demonstrate in this study that Oct4 forms a much larger protein interaction network than previously reported, and that the Oct4 interactome links multiple epigenetic regulatory pathways to the pluripotency network. Our data support the hypothesis that Oct4 is a central player in genetic and epigenetic regulation of stem cell pluripotency as well as somatic cell reprogramming.

Materials and Methods

Detailed methods on ESC culture, biochemical and bioinformatics analyses are available in Supplementary information, Data S1.

Affinity purification of oct4-associated protein complexes

One-step affinity purification with SA was performed as described [19, 53] with modifications. Briefly, 1 ml of Protein G agarose (Roche Diagnostic) equilibrated in buffer D (20 mM HEPES pH 7.6, 0.2 mM EDTA, 1.5 mM $MgCl_2$, 100 mM KCl, 20% glycerol) containing 0.02% NP40 was added to 3 ml of nuclear extract in 50 ml tubes (BD Falcon) and incubated for 1 h to pre-clear in the presence of 750 units of Benzonase (Novagen). Precleared extract was then transferred to the already equilibrated (with buffer D) SA beads (Invitrogen), and rotated for 6 h at 4 °C. Beads were washed five times for 15 min each with buffer D containing 0.02% NP40 and bound material was eluted by boiling for 5 min in Laemmli buffer, fractionated on a 10% SDS-polyacrylamide gel, stained with the GelCode™ Blue Safe Protein Stain buffer (Thermo), and subjected to whole lane LC-MS/MS sequencing and data analysis.

Mass spectrometry and data analysis

Whole lane LC-MS/MS sequencing and peptide identification were performed at the Taplin Biological Mass Spectrometry Facility at Harvard Medical School. Three biological replicates were performed for ZHBTc4 cells and modified ZHBTc4 (ZO4B4) cells expressing biotinylated Oct4. The detailed procedure for sample process, MS instrumentation, and data analyses have been described in our previous study [3]. Selection criteria for high confidence interaction proteins within the purified complexes are as follows. First, we removed common background proteins such as naturally biotinylated carboxylases and their associated enzymes, as well as some ribosomal proteins as characterized [3], and proteins with documented membrane, cytoplasmic, or mitochondrial localization. Second, for proteins specific to [bio]Oct4 samples, only those with ≥ 2 peptides sequenced from at least two independent purifications were included in the final candidate list. Third, for proteins with peptide sequenced in both [bio]Oct4 and BirA samples, only those with predominant peptide presence in [bio]Oct4 over BirA samples in at least two independent purifications were included.

Acknowledgments

References

1   Chen X, Xu H, Yuan P, *et al*. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 2008; **133**:1106-1117.

2   Kim J, Chu J, Shen X, Wang J, Orkin SH. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 2008; **132**:1049-1061.

3   Wang J, Rao S, Chu J, *et al*. A protein interaction network for pluripotency of embryonic stem cells. *Nature* 2006; **444**:364-368.

4   Pardo M, Lang B, Yu L, *et al*. An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell* 2010; **6**:382-395.

5   Van den Berg DL, Snoek T, Mullin NP, *et al*. An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* 2010; **6**:369-381.

6   Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006; **126**:663-676.

7   Yu J, Vodyanik MA, Smuga-Otto K, *et al*. Induced pluripotent stem cell lines derived from human somatic cells. *Science* 2007; **318**:1917-1920.

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

9   Silva J, Nichols J, Theunissen TW, *et al*. Nanog is the gateway to the pluripotent ground state. *Cell* 2009; **138**:722-737.

10 Heng JC, Feng B, Han J, *et al*. The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. *Cell Stem Cell* 2010; **6**:167-174.

11 Hatano SY, Tada M, Kimura H, *et al*. Pluripotential competence of cells associated with Nanog activity. *Mech Dev* 2005; **122**:67-79.

12 Mitsui K, Tokuzawa Y, Itoh H, *et al*. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 2003; **113**:631-642.

13 Chambers I, Silva J, Colby D, *et al*. Nanog safeguards pluripotency and mediates germline development. *Nature* 2007; **450**:1230-1234.

14 Chambers I, Colby D, Robertson M, *et al*. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 2003; **113**:643-655.

15 Navarro P, Chambers I, Karwacki-Neisius V, *et al*. Molecular coupling of Xist regulation and pluripotency. *Science* 2008; **321**:1693-1695.

16 Endoh M, Endo TA, Endoh T, *et al*. Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* 2008; **135**:1513-1524.

17 Loh YH, Zhang W, Chen X, George J, Ng HH. Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal

in embryonic stem cells. *Genes Dev* 2007; **21**:2545-2557.

18 Lemischka IR. Hooking up with Oct4. *Cell Stem Cell* 2010; **6**:291-292.

19 Kim J, Cantor AB, Orkin SH, Wang J. Use of *in vivo* biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. *Nat Protoc* 2009; **4**:506-517.

20 Wang J, Cantor AB, Orkin SH. Tandem affinity purification of protein complexes in mouse embryonic stem cells using *in vivo* biotinylation. *Curr Protoc Stem Cell Biol* 2009; **Chapter 1**:Unit1B 5.

21 Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* 2000; **24**:372-376.

22 Dignam JD, Lebovitz RM, Roeder RG. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* 1983; **11**:1475-1489.

23 Abu-Remaileh M, Gerson A, Farago M, *et al*. Oct-3/4 regulates stem cell identity and cell fate decisions by modulating Wnt/beta-catenin signalling. *EMBO J* 2010; **29**:3236-3248.

24 Kelly KF, Ng DY, Jayakumaran G, Wood GA, Koide H, Doble BW. beta-catenin enhances Oct-4 activity and reinforces pluripotency through a TCF-independent mechanism. *Cell Stem Cell* 2011; **8**:214-227.

25 Ding L, Paszkowski-Rogacz M, Nitzsche A, *et al*. A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell* 2009; **4**:403-415.

26 Ponnusamy MP, Deb S, Dey P, *et al*. RNA polymerase II associated factor 1/PD2 maintains self-renewal by its interaction with Oct3/4 in mouse embryonic stem cells. *Stem Cells* 2009; **27**:3001-3011.

27 Ang YS, Tsai S-Y, Lee D-F, *et al*. Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* 2011; **145**:1-15.

28 Liang J, Wan M, Zhang Y, *et al*. Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nat Cell Biol* 2008; **10**:731-739.

29 Zhang L, Rayner S, Katoku-Kikyo N, Romanova L, Kikyo N. Successful co-immunoprecipitation of Oct4 and Nanog using cross-linking. *Biochem Biophys Res Commun* 2007; **361**:611-614.

30 Tesar PJ, Chenoweth JG, Brook FA, *et al*. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* 2007; **448**:196-199.

31 Ivanova N, Dobrin R, Lu R, *et al*. Dissecting self-renewal in stem cells with RNA interference. *Nature* 2006; **442**:533-538.

32 Hailesellasse Sene K, Porter CJ, Palidwor G, *et al*. Gene function in early mouse embryonic stem cell differentiation. *BMC Genomics* 2007; **8**:85.

33 Huang J, Chen T, Liu X, *et al*. More synergetic cooperation of Yamanaka factors in induced pluripotent stem cells than in embryonic stem cells. *Cell Res* 2009; **19**:1127-1138.

34 Berger SI, Posner JM, Ma'ayan A. Genes2Networks: connecting lists of gene symbols using mammalian protein interactions databases. *BMC Bioinformatics* 2007; **8**:372.

35 Ho L, Jothi R, Ronan JL, Cui K, Zhao K, Crabtree GR. An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc Natl Acad Sci USA* 2009; **106**:5187-5191.

36 Ho L, Ronan JL, Wu J, *et al*. An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc Natl Acad Sci USA* 2009; **106**:5181-5186.

37 Singhal N, Graumann J, Wu G, *et al*. Chromatin-remodeling components of the BAF complex facilitate reprogramming. *Cell* 2010; **141**:943-955.

38 Dejosez M, Krumenacker JS, Zitur LJ, *et al*. Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. *Cell* 2008; **133**:1162-1174.

39 Cao S, Bendall H, Hicks GG, *et al*. The high-mobility-group box protein SSRP1/T160 is essential for cell viability in day 3.5 mouse embryos. *Mol Cell Biol* 2003; **23**:5301-5307.

40 Gaspar-Maia A, Alajem A, Polesso F, *et al*. Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* 2009; **460**:863-868.

41 Fazzio TG, Huff JT, Panning B. An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* 2008; **134**:162-174.

42 Foster CT, Dovey OM, Lezina L, *et al*. Lysine-specific de-methylase 1 regulates the embryonic transcriptome and Co-REST stability. *Mol Cell Biol* 2010; **30**:4851-4863.

43 Wang J, Hevi S, Kurash JK, *et al*. The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat Genet* 2009; **41**:125-129.

44 Yang P, Wang Y, Chen J, *et al*. Rcor2 is a subunit of the LSD1 complex that regulates ES cell property and substitutes for Sox2 in reprogramming somatic cells to pluripotency. *Stem Cells* 2011; **29**:791-801.

45 Stopka T, Skoultchi AI. The ISWI ATPase Snf2h is required for early mouse development. *Proc Natl Acad Sci USA* 2003; **100**:14097-14102.

46 Landry J, Sharov AA, Piao Y, *et al*. Essential role of chromatin remodeling protein Bptf in early mouse embryos and embryonic stem cells. *PLoS Genet* 2008; **4**:e1000241.

47 Kim D, Probst L, Das C, Tucker PW. REKLES is an ARID3-restricted multifunctional domain. *J Biol Chem* 2007; **282**:15768-15777.

48 Hu G, Kim J, Xu Q, Leng Y, Orkin SH, Elledge SJ. A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev* 2009; **23**:837-848.

49 Ding L, Buchholz F. RNAi in embryonic stem cells. *Stem Cell Rev* 2006; **2**:11-18.

50 Chia NY, Chan YS, Feng B, *et al*. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* 2010; **468**:316-320.

51 Wong CC, Gaspar-Maia A, Ramalho-Santos M, Reijo Pera RA. High-efficiency stem cell fusion-mediated assay reveals Sall4 as an enhancer of reprogramming. *PLoS One* 2008; **3**:e1955.

52 Tsubooka N, Ichisaka T, Okita K, Takahashi K, Nakagawa M, Yamanaka S. Roles of Sall4 in the generation of pluripotent stem cells from blastocysts and fibroblasts. *Genes Cells* 2009; **14**:683-694.

53 de Boer E, Rodriguez P, Bonte E, *et al.* Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci USA* 2003; **100**:7480-7485.

(**Supplementary information** is linked to the online version of the paper on the *Cell Research* website.)