

UCLA

UCLA Previously Published Works

Title

Oct4:Sox2 binding is essential for establishing but not maintaining active and silent states of dynamically regulated genes in pluripotent cells

Permalink

<https://escholarship.org/uc/item/1195j35n>

Journal

Genes & Development, 36(19-20)

ISSN

0890-9369

Authors

Lo, Jerry Hung-Hao
Edwards, Miguel
Langerman, Justin
[et al.](#)

Publication Date

2022-10-01

DOI

10.1101/gad.350113.122

Peer reviewed

Oct4:Sox2 binding is essential for establishing but not maintaining active and silent states of dynamically regulated genes in pluripotent cells

Jerry Hung-Hao Lo,^{1,2,3} Miguel Edwards,^{1,2,3} Justin Langerman,^{1,2,3} Rupa Sridharan,⁴ Kathrin Plath,^{1,2,5} and Stephen T. Smale^{1,2,3}

¹Molecular Biology Institute, ²Broad Stem Cell Research Center, ³Department of Microbiology, Immunology, and Molecular Genetics, University of California at Los Angeles, Los Angeles, California 90095, USA; ⁴Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, Wisconsin 53715, USA; ⁵Department of Biological Chemistry, University of California at Los Angeles, Los Angeles, California 90095, USA

Much has been learned about the mechanisms of action of pluripotency factors Oct4 and Sox2. However, as with other regulators of cell identity, little is known about the impact of disrupting their binding motifs in a native environment or the characteristics of genes they regulate. By quantitatively examining dynamic ranges of gene expression instead of focusing on conventional measures of differential expression, we found that Oct4 and Sox2 enhancer binding is strongly enriched near genes subject to large dynamic ranges of expression among cell types, with binding sites near these genes usually within superenhancers. Mutagenesis of representative Oct4:Sox2 motifs near such active, dynamically regulated genes revealed critical roles in transcriptional activation during reprogramming, with more limited roles in transcriptional maintenance in the pluripotent state. Furthermore, representative motifs near silent genes were critical for establishing but not maintaining the fully silent state, while genes whose transcript levels varied by smaller magnitudes among cell types were unaffected by nearby Oct4:Sox2 motifs. These results suggest that Oct4 and Sox2 directly establish both active and silent transcriptional states in pluripotent cells at a large number of genes subject to dynamic regulation during mammalian development, but are less important than expected for maintaining transcriptional states.

[*Keywords:* pluripotency; embryonic stem cells; Oct4; Sox2; differentiation; transcription]

Supplemental material is available for this article.

Received September 19, 2022; revised version accepted November 14, 2022.

The groundbreaking experiments of Takahashi and Yamanaka (2006) demonstrating that somatic cells can be reprogrammed to pluripotency by expression of four transcription factors serve as an important example of the powerful role of transcription factors in establishing cell identity. The POU domain transcription factor Oct4 and the Sry-related high-mobility group protein Sox2 are central to reprogramming, with key contributions from other factors, including Klf4, Nanog, and Myc (Wu and Schöler 2014; Schaefer and Lengerke 2020; Deng et al. 2021).

Oct4 and Sox2 are both expressed at the earliest stages of mammalian embryogenesis (Wu and Schöler 2014; Schaefer and Lengerke 2020). Oct4 is critical for establishing and maintaining pluripotency in the inner cell mass

during embryogenesis (Nichols et al. 1998; Niwa et al. 2000; Wu et al. 2013) and possesses unique structural properties that distinguish it from other POU family members (Esch et al. 2013). Sox2 is also critical for the maintenance of pluripotency but is dispensable for its establishment during embryogenesis, possibly due to redundancy with other Sox family members (Ivanova et al. 2006; Masui et al. 2007).

Oct4 and Sox2 bind thousands of sites throughout the genome, with highly prevalent cobinding to consistently spaced and oriented composite motifs (Reményi et al. 2003; Boyer et al. 2005; Zhou et al. 2007; Chen et al. 2008; Kim et al. 2008; Tapia et al. 2015). Binding is biased toward superenhancers (Whyte et al. 2013), with other pluripotency factors often binding in close proximity

Corresponding author: smale@mednet.ucla.edu

Article published online ahead of print. Article and publication date are online at <http://www.genesdev.org/cgi/doi/10.1101/gad.350113.122>. Freely available online through the *Genes & Development* Open Access option.

©2022 Lo et al. This article, published in *Genes & Development*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

(Chen et al. 2014; Chronis et al. 2017). As somatic cells are reprogrammed, they proceed through a continuum of states accompanied by changes in the genomic distribution of Oct4 and Sox2 (Sridharan et al. 2009; Soufi et al. 2012, 2015; Chen et al. 2016; Chronis et al. 2017).

Oct4 and Sox2 possess pioneering activity, with a capacity to engage nucleosomal DNA and promote nucleosome remodeling (Soufi et al. 2012, 2015; Roberts et al. 2021). The two proteins promote remodeling and transcriptional activation in part through the direct or indirect recruitment of coregulatory proteins, including SWI/SNF complexes, Mediator, p300, Trithorax complex, and others (Chen et al. 2008; Pardo et al. 2010; van den Berg et al. 2010; Ang et al. 2011; Esch et al. 2013; Whyte et al. 2013; King and Klose 2017). In addition, Oct4 and Sox2 bind in close proximity to large numbers of silent genes in embryonic stem cells (ESCs) (Boyer et al. 2005) and have been suggested to promote transcriptional repression through genomic approaches (Chronis et al. 2017), studies of promoter-reporter plasmids (Liu et al. 1997), and interactions with corepressor proteins (Liang et al. 2008; Esch et al. 2013).

Genomic studies have revealed that Oct4, Sox2, and other key transcriptional and post-transcriptional regulators participate in intertwined regulatory circuits (Boyer et al. 2005; Zhou et al. 2007; Chen et al. 2008; Kim et al. 2008; Apostolou et al. 2013; Li and Belmonte 2017; Li and Izpisua Belmonte 2018). Genes critical for pluripotency, self-renewal, signaling, and chromatin remodeling are among those thought to be directly regulated by Oct4 and Sox2, with evidence of autoregulatory and feed-forward loops.

Despite extensive progress, much remains to be learned about the functions of Oct4 and Sox2. One incompletely understood issue is the functional roles of the thousands of Oct4 and Sox2 binding events throughout the genome, as binding in a ChIP-seq experiment may or may not indicate a functional role in transcriptional control. Furthermore, in those instances in which binding has functional consequences, it generally is not known whether binding is essential for transcriptional activation or repression or plays a lesser modulatory role in refining expression levels of its target genes, many of which may be broadly expressed.

A second incompletely understood issue is the logic dictating which genes expressed in pluripotent cells are directly regulated by Oct4 and Sox2. Very few if any genes are expressed specifically in pluripotent cells, and large numbers of genes participate in pluripotency, self-renewal, and the survival of pluripotent cells, only a subset of which has been suggested to be direct targets of pluripotency factors. Are there characteristics that determine whether a gene requires direct regulation by Oct4 and Sox2?

One limitation of prior studies has been a focus either on a limited number of genes found to be strongly and preferentially expressed in embryonic stem cells (ESCs) (Mitsui et al. 2003) or on larger sets of differentially expressed genes identified using statistical criteria (e.g., Boyer et al. 2005; Chen et al. 2008; Kim et al. 2008). With this

latter approach, a differentially expressed gene set typically includes all genes whose expression levels consistently differ by approximately twofold or more, despite the fact that the range of differential expression between cell types extends from this low magnitude for some genes to >100-fold for others. We therefore envisioned that greater consideration of the dynamic range of gene expression in ESCs in comparison with other cell types might lead to a greater understanding of the expression properties of Oct4 and Sox2 target genes and allow us to classify potential target genes into defined groups for subsequent functional analysis.

Through this approach, we found that Oct4 and Sox2 binding was selectively enriched near genes that exhibit large dynamic ranges of expression among cell types, with much less frequent binding near genes whose expression levels vary by less than fivefold, even though many of these latter genes would be defined as differentially expressed by common statistical criteria. Mutagenesis of Oct4:Sox2 composite motifs near representative dynamically expressed genes revealed a critical role in establishing a transcriptionally active state, with a lesser role in transcriptional maintenance. Surprisingly, a critical role of Oct4:Sox2 composite binding in establishing but not maintaining a fully silent state in pluripotent cells was also found through mutagenesis of representative motifs near silent genes. Thus, our findings suggest that a large dynamic range of expression is a key characteristic of Oct4:Sox2 target genes and that these factors may rarely regulate genes modulated by smaller magnitudes across cells.

Results

'ESC-specific,' 'dynamic,' and 'broadly expressed' gene classes

We envisioned that a quantitative analysis of the dynamic range of expression of genes differentially expressed between ESCs and somatic cell types could be of value for understanding pluripotency mechanisms. We focused first on a comparison between mouse ESCs and three somatic cell populations: bone marrow-derived macrophages, CD4⁺CD8⁺ thymocytes, and cortical neurons. We chose cell types that could readily be obtained in large quantities, allowing us to perform RNA-seq at considerable depth to increase our ability to accurately determine dynamic ranges of expression; RNA-seq performed at a more conventional depth limits the accuracy of transcript abundance measurements at the low end of the scale. Importantly, we measured chromatin-associated nascent transcripts rather than mRNA to focus the analysis on transcriptional dynamics rather than post-transcriptional events.

We focused on 3030 genes whose nascent transcripts were expressed at a high level (more than five RPKM) in ESCs. The high threshold was used to increase the accuracy of the dynamic range calculations. Using this gene set, we determined the fold difference in nascent transcript level between ESCs and the other three cell types. The

results (Fig. 1A) show that only 52 genes exhibited transcript levels elevated by at least 100-fold in ESCs relative to all three somatic cell types. This small number of 100-fold differentially expressed genes would undoubtedly decline, possibly to zero, if more and more somatic cell types were examined. Moreover, only 91 genes exhibited nascent transcript levels elevated by at least 20-fold in ESCs relative to all three somatic cell types. For the purposes of this initial analysis, we defined these 91 genes as “ESC-specific” (Fig. 1B), but with full recognition that they are not truly ESC-specific, as many if not all of these genes are likely to be expressed in somatic cell types or during developmental stages that were not examined.

We next defined as “dynamic” 248 genes not already defined as “ESC-specific” whose nascent transcript levels were at least 20-fold lower in only one or two somatic cell types but within fivefold of the ESC expression level in the remaining cell type(s) (Fig. 1B). In other words, this set includes genes that require regulatory mechanisms capable of supporting large dynamic ranges of expression, but they are clearly not ESC-specific due to their expression being comparable with the ESC level in one or two of the somatic cell types examined. Finally, for the purpose of comparison, we defined a group of 1931 “broadly expressed” genes, which were highly expressed (more than five RPKM) in ESCs and exhibited expression levels in all three somatic cell types that were no more than fivefold higher or fivefold lower (0.2 of ESC level) than the ESC level. Notably, a large fraction of genes in the “broadly expressed” group would be defined as differentially expressed using conventional criteria for differential expression (data not shown). Figure 1C shows nascent transcript levels for representative genes in each class. Genes that did not meet the criteria for inclusion in any of the three classes were excluded from the analysis to allow us to focus on gene classes with clear distinctions in expression profiles.

To determine whether this classification scheme could be extended to a larger number of cell types and to RNA-seq data obtained with mRNA rather than nascent transcripts, we analyzed data sets from the mouse ENCODE data portal obtained with mouse ESC mRNA and mRNA from 12 different mouse primary somatic cell types or tissues. Focusing on 5921 genes that were highly expressed in ESCs, we found only 40 genes whose transcript levels were at least 20-fold lower in all 12 of the other cell types (Fig. 1D, ESC-specific). Another 369 genes exhibited transcript levels that were at least 20-fold lower than the ESC level in between four and 11 of the cell types but with transcript levels within fivefold of the ESC level in the remaining cell types (Fig. 1D, dynamic). Finally, 3303 genes were identified that were highly expressed in ESCs but with transcript levels that varied by no more than fivefold in any of the 12 other cell types examined (Fig. 1D, broadly expressed). Examples of genes in each of these three groups are shown in Supplemental Figure S1. Thus, although most genes that are highly expressed in ESCs exhibit transcript levels that vary less than fivefold among the 12 somatic cell types examined, hundreds of ESC-expressed genes require regulatory mechanisms to

support a large dynamic range of expression. However, few genes were considered to be ESC-specific in this analysis, and it appears possible if not likely that no genes would be strictly ESC-specific if a larger number of cell types were examined.

Oct4:Sox2 binding is enriched near the ESC-specific and dynamic gene classes

We next examined the prevalence of Oct4 and Sox2 binding in the vicinity of the ESC-specific, dynamic, and broadly expressed genes. By combining two biological replicates of previously published ChIP-seq data sets from mouse ESC line V6.5 (Chronis et al. 2017), we found 15,506 and 11,207 binding peaks genome-wide for Oct4 and Sox2, respectively, with most peaks at intergenic or intronic locations (Fig. 2A). As expected on the basis of prior studies, a large number (8100) of the Oct4 and Sox2 binding sites overlapped (distance between peak summits <100 bp) (Fig. 2B, left). To increase confidence, we limited our analysis to 3092 overlapping binding events with peak scores >20 for both proteins (Fig. 2B, middle). Accurately linking binding sites to their target genes has been a long-standing challenge. In our analysis, we found that 1035 (or roughly one-third) of the 3092 genomic sites cobound by Oct4 and Sox2 (peak score >20) reside within 15 kb of a transcription start site (TSS) (Fig. 2B). We therefore focused on these 1035 cobound sites with the hypothesis that they are more likely to be regulators of the closest gene than if we examined all sites regardless of their distance from the closest gene.

Importantly, an examination of these 1035 cobound sites revealed that they are strongly enriched in the vicinity of ESC-specific and dynamic genes in comparison with the broadly expressed genes (Fig. 2C). Despite the absence of a definitive strategy for assigning binding sites to their target genes, substantial enrichment was observed in both the analysis of nascent transcripts from four cell types (Fig. 2C, left) and the analysis of ENCODE mRNA data from 13 cell types (Fig. 2C, right). In the nascent transcript analysis, Oct4:Sox2 cobinding near ESC-specific genes was enriched by 6.4-fold in comparison with the broadly expressed genes. Oct4:Sox2 cobinding near dynamic genes was enriched by 3.7-fold, with a surprisingly high 20% of the dynamic genes exhibiting a strong Oct4:Sox2-cobound site within 15 kb of their TSSs (Fig. 2C, left). The specific degree of enrichment increased as the peak score threshold for Oct4 and Sox2 increased, which is likely to reflect an increasing reliability of the data (Supplemental Fig. S2A,B). Similar levels of enrichment were found when all Oct4 and Sox2 binding sites were examined, not just cobound sites (data not shown), perhaps related to the fact that very high percentages of Oct4 and Sox2 binding events near the ESC-specific, dynamic, and broadly expressed gene classes that meet our criteria for inclusion represent cobinding (Supplemental Fig. S2C).

A similar enrichment was observed with the ENCODE data (Fig. 2C, right). Strong Oct4:Sox2 cobinding was observed within 15 kb of the TSS for 37.5% of the ESC-specific genes and a surprisingly high 19.8% of the dynamic

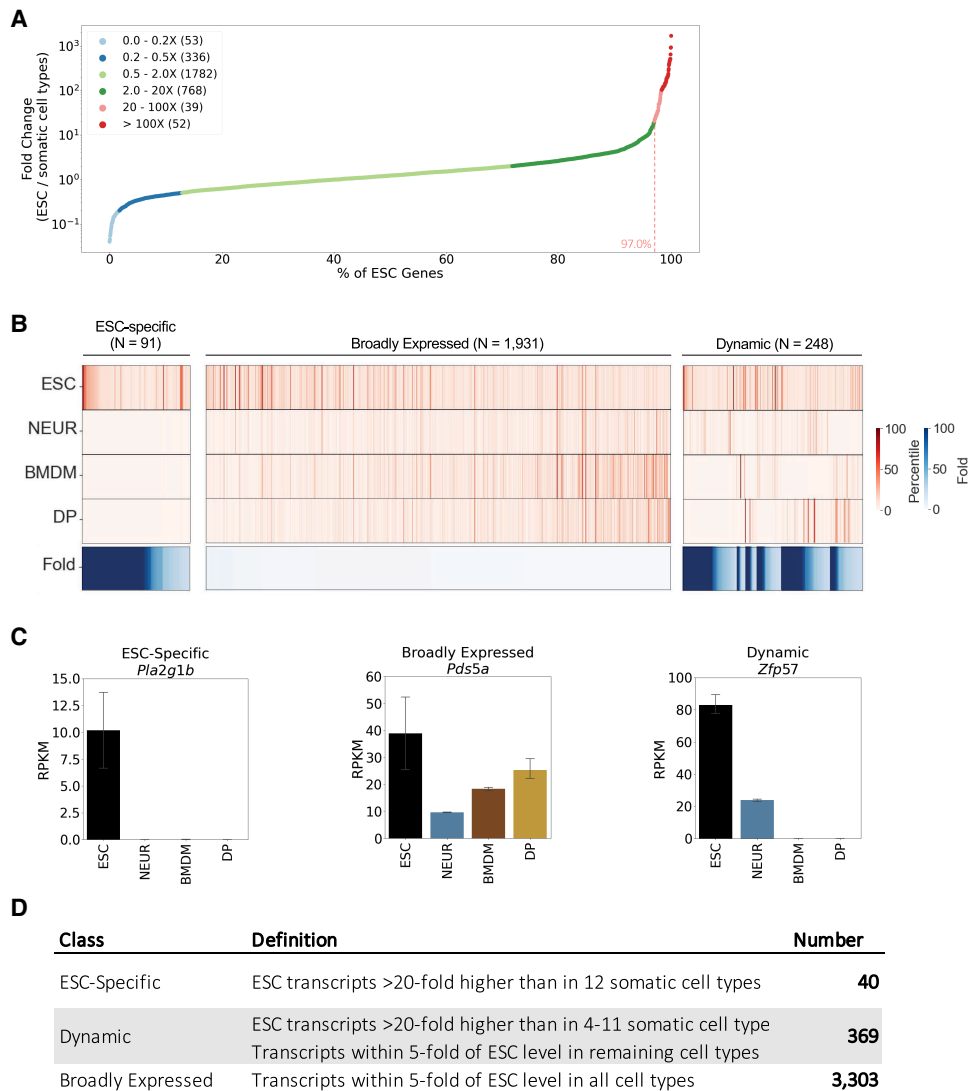


Figure 1. Delineation of ESC-specific, dynamic, and broadly expressed gene classes by deep RNA-seq analysis of nascent transcripts and analysis of ENCODE RNA-seq data sets. (A) Chromatin-associated nascent transcripts from the mouse CCE ESC line (ESC), E14.5 cortical neurons (NEUR), bone marrow-derived macrophages (BMDM), and CD4⁺CD8⁺ double-positive thymocytes (DP) were analyzed by RNA-seq. The smallest fold difference between the ESC RPKM and the RPKMs for the three somatic cells is shown for 3030 ESC-expressed genes (more than five RPKM in ESCs). Six fold difference bins are color-coded, with the number of genes in each bin in parentheses. Only 3% of the genes exhibit a fold difference of >20 for all three somatic cell types. (B) A heat map is shown for genes from the nascent transcript data classified as ESC-specific (91 genes), broadly expressed (1931), and dynamic (248) according to the criteria described in the text. Expression levels are presented as percentiles derived from RPKM, with the highest RPKM among all genes in all cell types defined as 100%. The minimum fold difference in RPKM between ESCs and the three somatic cell types is displayed at the *bottom*. (C) Nascent transcript levels (RPKM) are shown for representative ESC-specific (*Pla2g1b*), broadly expressed (*Pds5a*), and dynamic (*Zfp57*) genes. (D) The numbers of genes in each of the three classes from an analysis of 13 ENCODE data sets are shown, along with the criteria used to assign genes to each class. The analysis was restricted to genes expressed >4.9 RPKM in ESCs.

genes but only 4.9% of the broadly expressed genes, yielding enrichments of 7.6-fold and fourfold, respectively, near ESC-specific and dynamic genes (Fig. 2C, right). The fact that similar findings were obtained with the deep nascent RNA-seq data and the ENCODE mRNA-seq data is noteworthy; this may be because inclusion of a larger number of cell types in the ENCODE analysis offset the benefits of the increased sequencing depth and fo-

cus on nascent transcripts in the nascent transcript analysis.

To summarize, the substantial enrichment of Oct4:Sox2 binding near dynamic, non-ESC-specific genes in comparison with broadly expressed genes and the high prevalence of Oct4:Sox2-cobound sites near dynamic genes suggest a hypothesis in which Oct4:Sox2 cobinding positively regulates a large number of genes that are

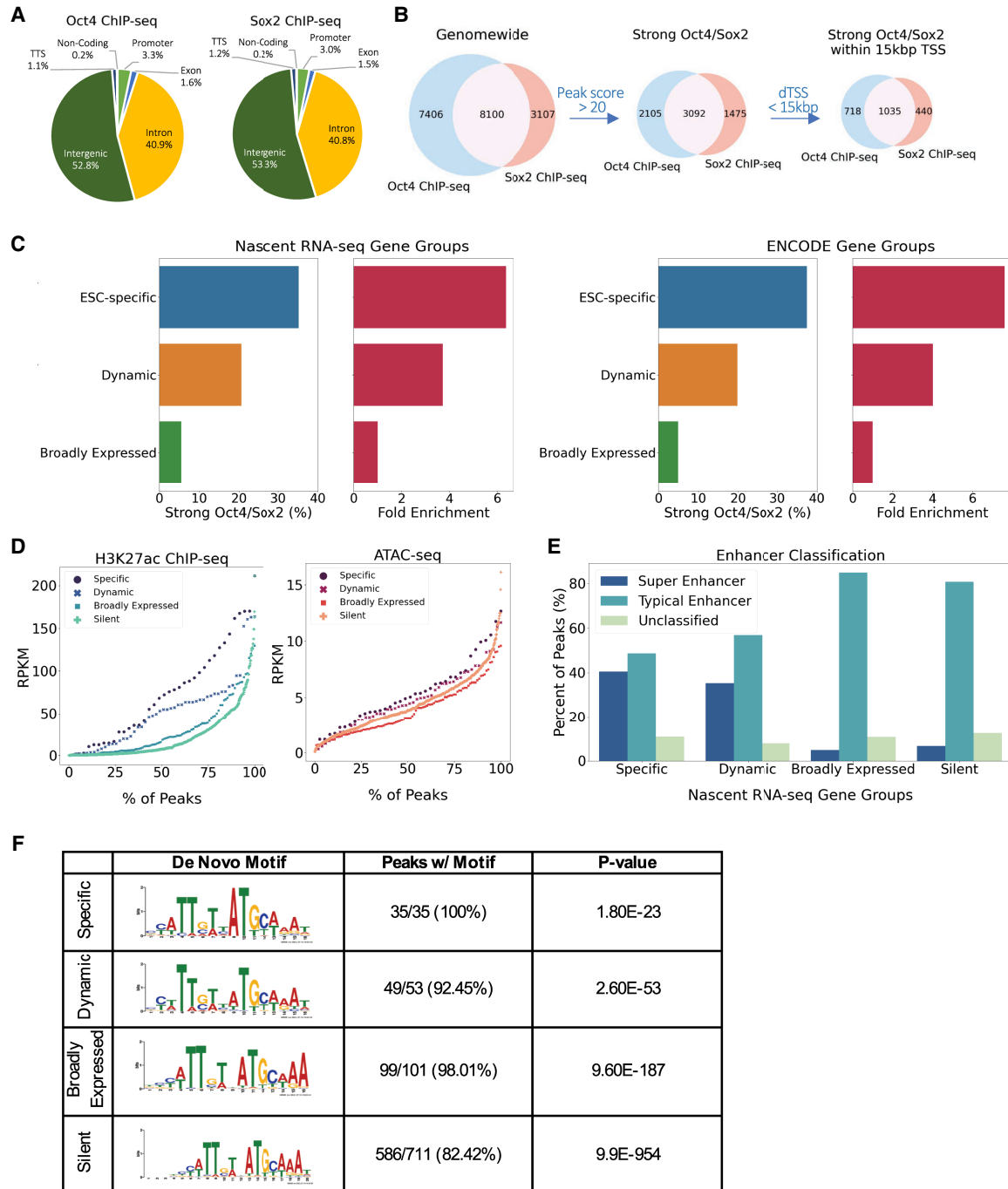


Figure 2. Oct4 and Sox2 binding is enriched near ESC-specific and dynamic gene classes in comparison with broadly expressed genes. (A) The genomic distribution of Oct4 and Sox2 binding peaks is shown, based on ChIP-seq data sets from the mouse ESC line V6.5 (Chronis et al. 2017). (B) The overlap of Oct4 and Sox2 peaks is shown, with a peak summit distance <100 bp required for inclusion as a cobound site. Venn diagrams display overlap when all Oct4 and Sox2 called peaks are analyzed (*left*), when the analysis is restricted to peaks with peak scores >20 (*middle*), and when the analysis was further restricted to peaks within 15 kb of a TSS (*right*). (C) The degree of enrichment is shown for Oct4:Sox2-cobound peaks near ESC-specific and dynamic genes in comparison with broadly expressed genes. The percentage of genes in each class that exhibit nearby (<15 kb from TSS) Oct4:Sox2-cobound peaks (peak scores >20) is also shown. (D) De novo motif discovery at Oct4:Sox2-cobound peaks (peak score >20) near ESC-specific, dynamic, broadly expressed, and silent genes shows strong enrichment of Oct4:Sox2 composite motifs but no large differences at the different gene classes. (E) The distribution of histone H3K27ac ChIP-seq and ATAC-seq signals (RPKM) coinciding with Oct4:Sox2-cobound sites near genes in the ESC-specific, dynamic, broadly expressed, and silent gene classes (from the nascent transcript analysis) is shown. (F) The percentage of Oct4:Sox2-cobound sites that fit the criteria of superenhancers and typical enhancers (Whyte et al. 2013) is shown for cobound sites near each of the four classes of genes identified in the nascent transcript analysis.

highly expressed in ESCs and have a strong preference toward genes that exhibit large dynamic ranges of expression among cell types. Thus, among the hundreds or thousands of ESC-expressed genes that participate in pluripotency, self-renewal, and pluripotent cell survival, direct regulation by Oct4:Sox2 may occur preferentially at those genes requiring dynamic regulatory mechanisms during development.

Properties of Oct4:Sox2-cobound sites at different gene classes

To examine more carefully the characteristics of Oct4:Sox2-cobound sites near different gene classes, we first performed de novo motif analysis. For this analysis, we focused on the nascent transcript-derived gene classes and included a fourth class that comprised 711 Oct4:Sox2 cobinding events located within 15 kb of the TSS of a gene that is poorly expressed in ESCs ("silent"). The results revealed that a previously described composite motif containing juxtaposed and strictly spaced Oct4 and Sox2 motifs (Tapia et al. 2015) is highly prevalent at Oct4:Sox2-cobound sites near all four gene classes examined (Fig. 2D). This result confirms the high prevalence of this composite motif within the genome but suggests that the binding sequence itself cannot distinguish the four gene classes.

Next, we examined ESC ChIP-seq data sets for Nanog, which contributes to gene regulation with Oct4 and Sox2 and is often cobound with the two proteins (Boyer et al. 2005; Jauch et al. 2008). Nanog bound a high percentage of Oct4:Sox2-cobound sites near genes in all four classes, with only a small preference for ESC-specific and dynamic genes in comparison with broadly expressed and silent genes (Supplemental Fig. S2D). Interestingly, Oct4:Sox2-cobound sites were generally closer to the TSSs of nearby ESC-specific genes, but with no significant difference among the dynamic, broadly expressed, and silent genes (Supplemental Fig. S2E).

To further characterize Oct4:Sox2-cobound sites, we analyzed histone H3K27ac ChIP-seq and ATAC-seq data sets, as well as superenhancer properties (Whyte et al. 2013; Buecker et al. 2014; Ji et al. 2015; Di Stefano et al. 2016). H3K27ac was substantially enriched in ESCs at Oct4:Sox2-cobound sites near both ESC-specific and dynamic genes in comparison with broadly expressed and silent genes (Fig. 2E, left). The magnitude of the difference was further revealed by separating H3K27ac peak signals associated with the 1035 Oct4:Sox2-cobound sites into multiple bins (Supplemental Fig. S2F). In contrast, chromatin accessibility as measured by ATAC-seq was similar among the gene classes (Fig. 2E, right). Finally, using the criteria of Whyte et al. (2013), superenhancer characteristics were found to be highly prevalent at Oct4:Sox2-cobound sites associated with ESC-specific and dynamic genes but were rarely observed at cobound sites near broadly expressed or silent genes (Fig. 2F).

To summarize, the substantial enrichment of Oct4:Sox2-cobound sites near both ESC-specific and dynamic genes coincides with a dramatic enrichment of superen-

hancer characteristics and an enrichment of histone H3K27ac. However, clear distinctions between gene classes were not observed in composite motif sequences, accessibility of the composite motif, or Nanog association.

*Limited role of an Oct4:Sox2 motif for transcription of the ESC-specific *Pla2g1b* gene in ESCs*

The results presented above support the hypothesis that Oct4:Sox2 cobinding may positively and preferentially regulate a large number of genes that require dynamic regulation across cell types. To perform a functional test of this hypothesis, we used CRISPR/Cas9 mutagenesis with homology-directed repair (CRISPR-HDR) to disrupt Oct4:Sox2 composite motifs in the mouse CCE ESC line in the vicinity of representative genes in each class. We first focused on *Pla2g1b* as an example of an ESC-specific gene as defined by our criteria. *Pla2g1b* encodes a secreted phospholipase A2 family member known to be highly expressed in ESCs. Notably, however, it has not been well-studied as a key contributor to a pluripotency network (Liu et al. 2008). Expression of this gene was >20-fold higher in mouse ESCs than in any of the other cell types examined in either our nascent transcript analysis or the ENCODE analysis (Figs. 1C, 3A), although a broader examination using public data sets revealed expression in the pancreas and stomach (data not shown). The strongest Oct4 and Sox2 binding sites within 20 kb of this gene display cobinding and are located in a superenhancer 1.3 kb upstream of the *Pla2g1b* TSS. This cobound site contains an Oct4:Sox2 composite motif and coincides with open chromatin as assessed by ATAC-seq and active chromatin as assessed by H3K27ac enrichment (Fig. 3B).

CRISPR-HDR mutagenesis was used to introduce homozygous substitution mutations into 11 of 12 bp in the Oct4:Sox2 composite motif (Fig. 3B), with two independent mutant clones confirmed by DNA sequencing and selected for further analysis. Notably, a large decrease in both Oct4 and Sox2 ChIP signals was observed in both mutant clones (Fig. 3C,D). *Pla2g1b* transcription as measured by qRT-PCR was diminished in both mutant clones in comparison with the wild-type ESC control (Fig. 3E). Surprisingly, however, the transcript levels were reduced by only 60% (Fig. 3E). Although this reduction provides support for a role of Oct4:Sox2 cobinding in *Pla2g1b* transcription in ESCs, a much larger reduction in transcription was expected, given the dominant role of Oct4 and Sox2 in pluripotency and given that the *Pla2g1b* expression profile revealed greater ESC specificity than observed at almost any other gene in the mouse genome.

Establishment of a secondary reprogramming assay to distinguish initial transcriptional activation from transcriptional maintenance

Redundancy offers one possible explanation for the marginal importance of the Oct4:Sox2 motif in *Pla2g1b* transcription. However, no other Oct4 or Sox2 peaks of comparable strength and associated with features of active chromatin were identified near *Pla2g1b*. We therefore

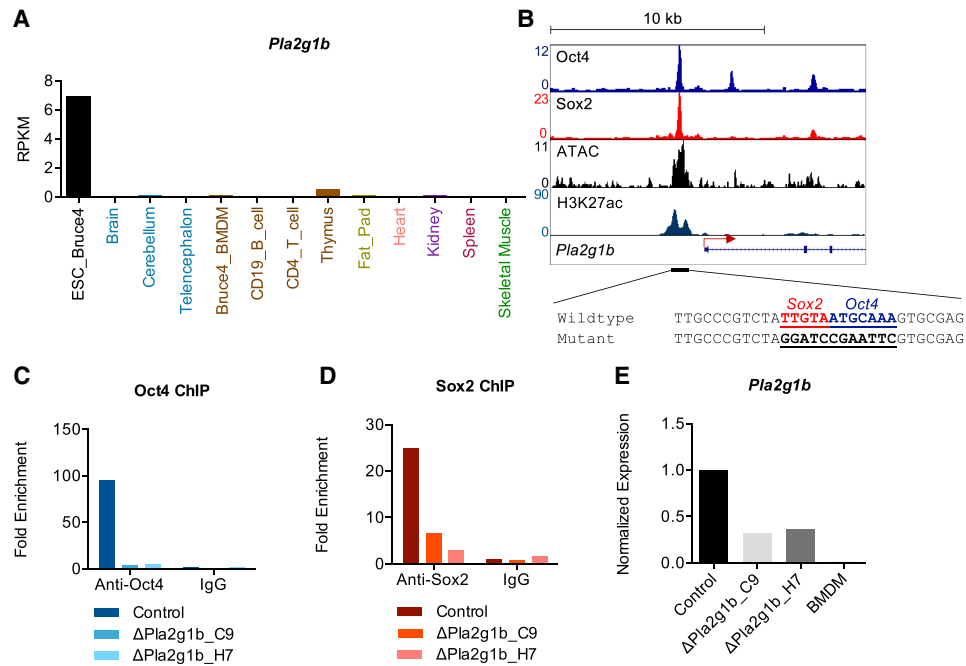


Figure 3. CRISPR-HDR mutagenesis reveals only a moderate role for an Oct4:Sox2 composite motif near the ESC-specific *Pla2g1b* gene. (A) The bar graph shows the *Pla2g1b* mRNA profile (RPKM) derived from an analysis of 13 mouse ENCODE RNA-seq data sets. (B) Genome browser snapshots display Oct4 and Sox2 binding, as well as ATAC-seq and histone H3K27ac ChIP-seq peaks, upstream of *Pla2g1b*. The Oct4:Sox2 composite motif at this location and the mutant sequence introduced by CRISPR-HDR are also shown. (C,D) Oct4 and Sox2 binding at the *Pla2g1b* upstream region was examined by ChIP-qPCR in wild-type ESCs and in two independent mutant clones. Fold enrichment was calculated as the fold change of percentage of input between the *Pla2g1b* ChIP-qPCR signal and a negative control region signal (*Hbb-b2*). IgG was used as a negative control. (E) The normalized expression levels for *Pla2g1b* mRNA were determined in the two independent clones by qRT-PCR, with BMDM mRNA analyzed as a negative control.

tested an alternative explanation, which is that Oct4:Sox2 binding might be critical for the initial activation of *Pla2g1b* transcription during reprogramming to pluripotency but less important for the maintenance of *Pla2g1b* transcription in established ESC lines.

To address this possibility, we established an assay for secondary reprogramming to pluripotency (Fig. 4A), taking advantage of a primary mouse embryonic fibroblast (MEF) line, tetO-OSKM, which harbors a single doxycycline (DOX)-inducible polycistronic cassette coding for Oct4, Sox2, Klf4, and Myc (OSKM) (Sridharan et al. 2013). This line can be efficiently converted to iPSCs upon addition of DOX. Moreover, we envisioned that iPSCs derived from the tetO-OSKM line (referred to here as Tet-on iPSCs) could be differentiated and then subject to secondary reprogramming by readdition of DOX. A similar DOX-inducible system for secondary reprogramming has previously been described for human iPSCs (Hockemeyer et al. 2008).

To evaluate the feasibility of this approach, we differentiated the Tet-on iPSC line into embryoid bodies (EBs) and subsequently cultured the cells for multiple passages under neural progenitor cell (NPC) growth conditions in the presence of 0.5 μ M retinoic acid (Fig. 4A; Lee et al. 2000; Sagner et al. 2018). The EBs displayed classic spheroid-like colonies in suspension culture, and the NPCs acquired short spindle-shaped morphology (Fig. 4B). For secondary reprogramming to a pluripotent state (secondary

iPSCs), we plated the cells in the presence of 2 μ g/mL DOX and LIF under mouse ESC culture conditions. After DOX treatment for 14–20 d, colonies with typical ESC-like morphology emerged (Fig. 4B). Selected single-cell colonies were further expanded into secondary iPSCs and maintained in the absence of DOX. Both primary and secondary iPSCs expressed several pluripotency markers that were not expressed in the NPCs (Fig. 4C). In contrast, several neural lineage-specific markers were highly expressed only in the NPCs (Fig. 4D).

Critical role for the Oct4:Sox2 composite motif in *Pla2g1b* transcriptional activation following secondary reprogramming

We next created mutant Tet-on iPSC lines by CRISPR-HDR with the same *Pla2g1b* Oct4:Sox2 motif mutations that we previously introduced into ESCs (see above). As expected, Oct4 and Sox2 binding was greatly reduced in two independent mutant lines (Fig. 5A). Importantly, *Pla2g1b* transcript levels were only moderately reduced in these lines, similar to the ESC results (Fig. 5B). *Pla2g1b* transcript levels declined dramatically upon differentiation into NPCs in both the wild-type and mutant lines. Strikingly, although the *Pla2g1b* transcript level recovered upon secondary reprogramming of the wild-type Tet-on iPSC line, it remained extremely low in the

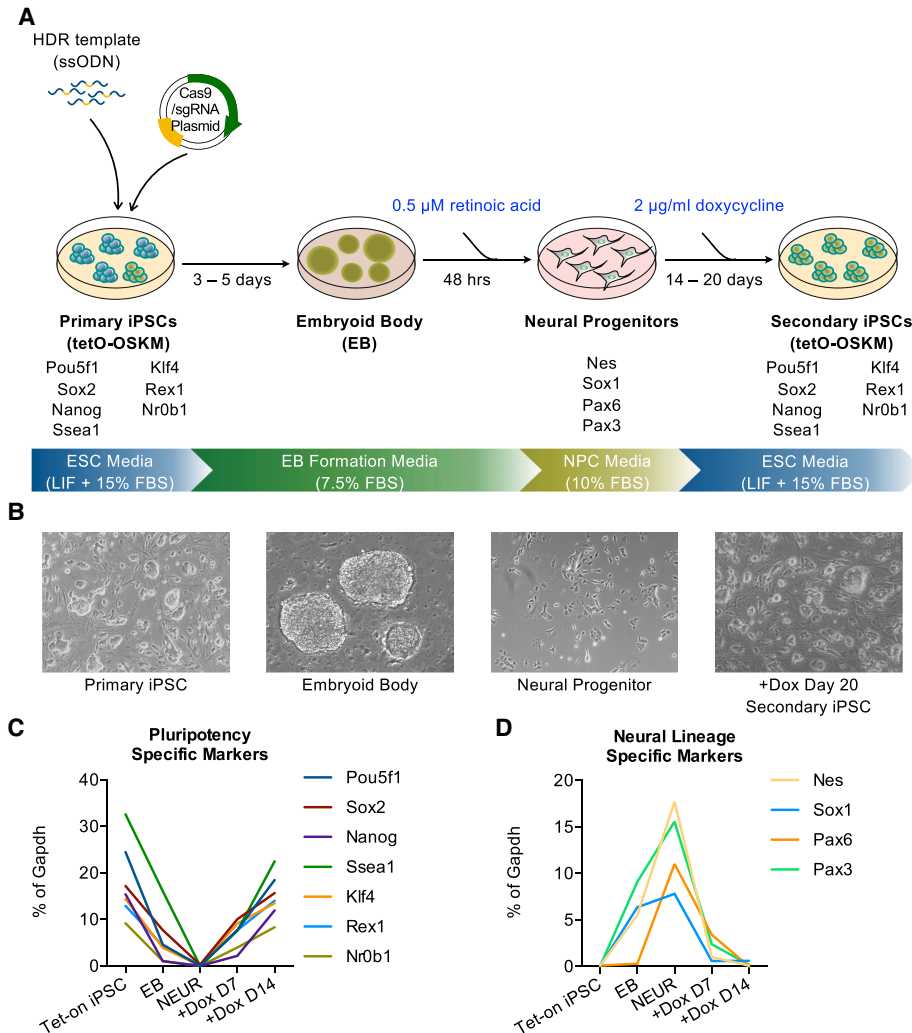


Figure 4. A tetO-OSKM iPSC line can be used to study gene expression changes during differentiation and secondary reprogramming. (A) The schematic diagram displays the experimental design used to edit the tetO-OSKM iPSC line (by cotransfection with HDR template and Cas9/sgrRNA plasmid, as well as selection/genotyping of single-cell colonies). (B) Representative cell culture morphologies are shown for primary iPSCs, EBs, neural progenitors, and secondary iPSCs. (C) The line graph shows mRNA levels monitored by qRT-PCR for seven genes often used to define a pluripotent state. mRNA levels are displayed as a percentage of *Gapdh* levels. (D) The line graph shows mRNA levels monitored by qRT-PCR for four genes known to be selectively expressed in neural lineage cells.

mutant lines (Fig. 5B). Notably, transcript levels following secondary reprogramming were far lower than those observed in the mutant lines prior to differentiation (>20-fold difference between wild type and mutant after secondary reprogramming in comparison with 2.5-fold difference prior to differentiation). This effect was not due to a broad impact of the Oct4:Sox2 mutation on reprogramming, as pluripotency genes were regulated normally in the mutant cells (Supplemental Fig. S3). These results provide evidence that the Oct4:Sox2 composite motif upstream of the *Pla2g1b* TSS is critical for the initial activation of *Pla2g1b* transcription, consistent with the established role of these proteins as pioneer factors. However, after transcription is established, it can be maintained with only a moderate reduction in transcription upon mutagenesis of the Oct4:Sox2 motif.

We considered the possibility that heterogeneity within the ESC and iPSC cell populations might be responsible for the moderate impact of the Oct4:Sox2 motif mutation prior to differentiation. However, isolation of single cells by limiting dilution, followed by expansion of several subcloned colonies, revealed that each subclone generally exhibited moderately reduced *Pla2g1b* transcript levels, similar to the original mutant lines (Supplemental Fig. S4A). Moreover, the moderately reduced levels were maintained for many passages (Supplemental Fig. S4B).

To determine whether Oct4:Sox2 binding near *Pla2g1b* contributes to the deposition of active histone marks, we measured H3K27ac enrichment in the wild-type and mutant lines. Prior to differentiation, the Oct4:Sox2 motif mutant iPSC lines exhibited moderately reduced H3K27ac levels on both sides of the Oct4:Sox2 binding

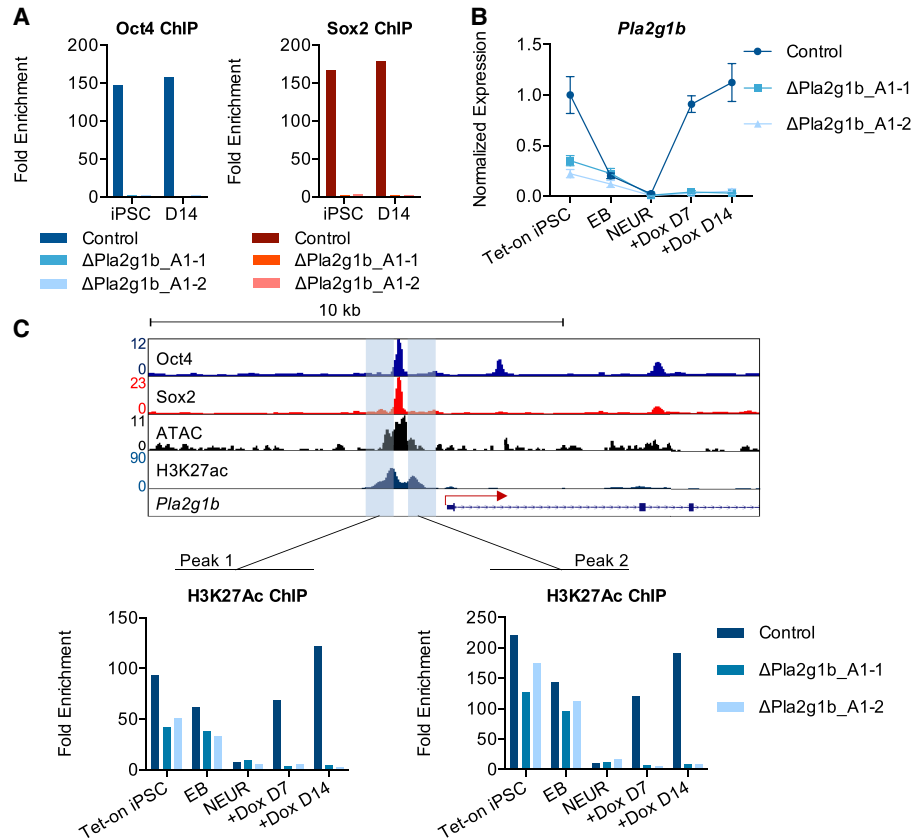


Figure 5. Critical role for the *Pla2g1b* Oct4:Sox2 composite motif in gene activation during secondary reprogramming but not for transcriptional maintenance. (A) Bar graphs show Oct4 and Sox2 binding monitored by ChIP-qPCR at the *Pla2g1b* enhancer in wild-type and in two independent mutant tetO-OSKM lines in primary iPSCs and day 14 secondary iPSCs. Data are displayed as in Figure 3C. (B) The line graph shows normalized *Pla2g1b* mRNA levels (by qRT-PCR) in wild-type and two independent mutant tetO-OSKM lines at each stage of differentiation and secondary reprogramming. Values represent means of three independent samples along with standard errors. (C) A genome browser snapshot displays the Oct4 and Sox2 ChIP-seq peaks, ATAC-seq peak, and H3K27ac peak at the *Pla2g1b* enhancer in ESCs. Blue shades highlight two regions with enriched H3K27ac. At the bottom, bar graphs show H3K27ac ChIP-qPCR levels at the two highlighted regions in wild-type and two independent mutant tetO-OSKM lines at each stage of differentiation and reprogramming. Comparable results were obtained in three independent experiments.

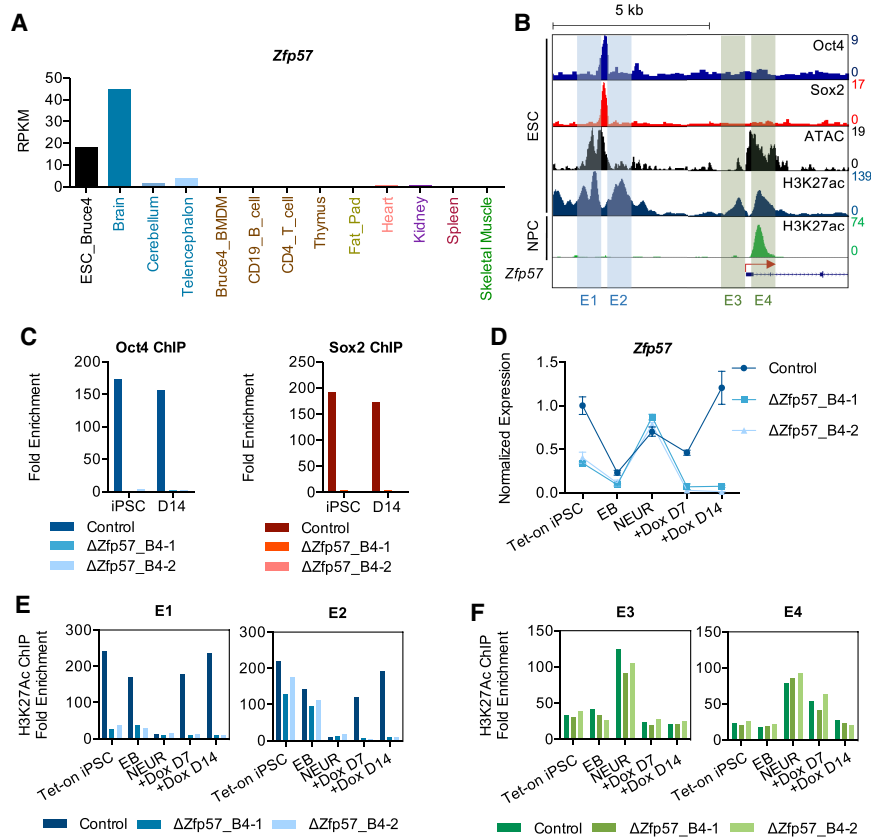
peak, consistent with the moderate impact of the mutation on transcription (Fig. 5C). When differentiated into EBs and then into NPCs, H3K27ac declined to a low level in both the wild-type and mutant lines (Fig. 5C). Notably, upon secondary reprogramming, H3K27ac levels were restored in the wild-type line but remained low in the mutant lines (Fig. 5C). These results suggest that Oct4:Sox2 binding is critical for both transcription and for the deposition of H3K27ac during the initial activation of *Pla2g1b* upon reprogramming, but both transcription and H3K27ac can be maintained at substantial levels in the pluripotent state in the absence of the Oct4:Sox2 motif.

Critical role of an Oct4:Sox2 composite motif for transcriptional activation of the dynamic *Zfp57* gene during secondary reprogramming

We hypothesized above that Oct4:Sox2 binding may be critical not only for the transcription of genes that exhibit considerable ESC specificity, but also for the transcription in pluripotent cells of a large number of genes that are ex-

pressed more broadly but require a large dynamic range of transcription among cell types. To test this hypothesis, we used CRISPR-HDR to introduce substitution mutations into the Oct4:Sox2 composite motif underlying an Oct4:Sox2-cobound site upstream of the dynamic *Zfp57* gene. This gene, encoding a KRAB domain zinc finger protein that binds imprinting control regions in ESCs (Quenneville et al. 2011), was defined as dynamic in our classification scheme; it is highly expressed in ESCs and is expressed at a much lower level (>20-fold) in many other cell types, but it also exhibits high expression in the brain (Figs. 1C, 6A). Notably, the Oct4:Sox2-cobound site coincides with a superenhancer and with ATAC-seq sensitivity and high H3K27ac in ESCs (Fig. 6B). However, the H3K27ac mark is absent at this location in *Zfp57*-expressing NPCs (Fig. 6B). As expected, mutagenesis of the Oct4:Sox2 composite motif eliminated Oct4 and Sox2 binding in two independent mutant lines (Fig. 6C).

Similar to the impact of the Oct4:Sox2 mutation at *Pla2g1b*, the Oct4:Sox2 motif substitution mutations near *Zfp57* reduced *Zfp57* transcription only moderately



at each stage of differentiation and reprogramming. Comparable results were obtained in three independent experiments. (F) Bar graphs show H3K27ac ChIP-qPCR levels at the *Zfp57* E3 and E4 regions in wild-type and mutant tetO-OSKM lines and reprogramming.

in the primary iPSCs (Fig. 6D). Upon differentiation to NPCs, *Zfp57* transcripts rose in the mutant lines to levels comparable with that observed in the wild-type line, consistent with the neuronal expression of endogenous *Zfp57* (Fig. 6D): This result suggests that the Oct4:Sox2 motif plays no role in neuronal *Zfp57* transcription. Strikingly, however, upon secondary reprogramming, *Zfp57* transcript levels declined to near background in the mutant lines while remaining high in the wild-type line (Fig. 6D). As in the *Plag1b* experiment, several pluripotency genes were regulated normally in the mutant cells (Supplemental Fig. S5).

An examination of histone modifications revealed that H3K27ac levels on both sides of the Oct4:Sox2 motif (E1 and E2 in Fig. 6B) were greatly diminished following secondary reprogramming in the mutant lines in comparison with the wild-type line (Fig. 6E). In the primary iPSCs, the mutant lines exhibited slightly reduced H3K27ac at E2, but more substantially reduced H3K27ac at E1 (Fig. 6E). Notably, H3K27ac levels at both E1 and E2 were low in NPCs. Interestingly, H3K27ac levels were selectively elevated in NPCs at regions immediately upstream of and downstream from the *Zfp57* TSS, with no significant change in the Oct4:Sox2 motif mutants (Fig. 6F); although these regions are in close proximity to the promoter, which is active in both

Figure 6. Critical role for an Oct4:Sox2 composite motif near the dynamic *Zfp57* gene in gene activation during secondary reprogramming but not in transcriptional maintenance. (A) The bar graph shows *Zfp57* mRNA levels (RPKM from mouse ENCODE RNA-seq data sets) in 13 cell populations. (B) Genome browser snapshots display the Oct4 and Sox2 ChIP-seq, ATAC-seq, and H3K27ac ChIP-seq peaks in ESCs, as well as H3K27ac ChIP-seq peaks in NPCs. Blue shading highlights two genomic regions (E1 and E2) with H3K27ac ChIP-seq peaks flanking the Oct4:Sox2 composite motif. Green shading highlights two additional genomic regions (E3 and E4) with H3K27ac peaks, one of which exhibits a strong H3K27ac peak in NPCs. (C) Bar graphs show Oct4 and Sox2 binding monitored by ChIP-qPCR from the *Zfp57* enhancer in wild-type and two independent mutant tetO-OSKM lines in primary iPSCs and day 14 secondary iPSCs. Data are displayed as in Figure 3C. (D) The line graph shows normalized *Zfp57* mRNA levels (by qRT-PCR) in wild-type and two independent mutant tetO-OSKM lines at each stage of differentiation and secondary reprogramming. Values represent means of three independent samples along with standard errors. (E) Bar graphs show H3K27ac ChIP-qPCR levels at the *Zfp57* E1 and E2 regions in wild-type and mutant tetO-OSKM lines

iPSCs and NPCs, they may possess additional regulatory regions that support transcription only in NPCs.

Together, these results demonstrate that Oct4:Sox2 binding to an Oct4:Sox2 composite motif is critical for the transcriptional activation of a gene classified as dynamic rather than ESC-specific. As with *Plag1b*, the *Zfp57* Oct4:Sox2 motif was critical for transcriptional activation during reprogramming but was less important for transcriptional maintenance. Importantly, although *Zfp57* is highly expressed in NPCs via a mechanism that is apparently independent of the Oct4:Sox2 composite motif, the Oct4:Sox2 motif remained critical for *Zfp57* transcriptional activation upon secondary reprogramming.

Critical role of an Oct4:Sox2 composite motif at the dynamic *Epb4.115* gene

We next extended this analysis to the *Epb4.115* gene, which is also classified as dynamic in our scheme. This gene is expressed much more broadly than *Zfp57* but with very low expression (>20-fold below the ESC level) in several cell types examined (Supplemental Fig. S6A). *Epb4.115* encodes a FERM domain protein that, to our knowledge, has not been implicated in a pluripotency network, although it plays multiple developmental roles in

mice (Lee et al. 2010). Oct4:Sox2 cobinding was observed in an *Epb4.115* intronic region with characteristics of a typical enhancer (Whyte et al. 2013), coinciding with an Oct4:Sox2 composite motif and with strong ATAC-seq and H3K27ac peaks (Supplemental Fig. S6B). Despite the broader expression of this dynamic gene, the results obtained following Oct4:Sox2 composite motif mutagenesis by CRISPR-HDR in the Tet-on iPSCs were very similar to those obtained with *Zfp57* (Supplemental Fig. S6C–F), with a modest effect of the mutation on transcription in the primary iPSCs and no effect of the mutation on the high-level transcription observed in NPCs, but a dramatic impact on transcription after secondary reprogramming. Moreover, the Oct4:Sox2 composite motif mutation influenced H3K27ac levels near the composite motif but not at regions closer to the TSS that were selectively elevated in NPCs (Supplemental Fig. S6E,F). Thus, *Zfp57* and *Epb4.115* represent two independent examples of dynamic genes that have not been well studied as members of pluripotency networks yet are directly regulated by Oct4:Sox2 during reprogramming to pluripotency, but with only modest roles of Oct4:Sox2 binding in transcriptional maintenance. Given that >20% of genes defined as dynamic exhibit strong Oct4:Sox2 cobinding within 15 kb of their TSSs, which represents a much higher percentage than observed in the broadly expressed gene set, the results support the hypothesis that a large dynamic range of expression is a critical and common characteristic of genes that are direct targets of Oct4 and Sox2.

Oct4:Sox2 composite motifs do not contribute to the transcription of two broadly expressed genes

Although Oct4:Sox2 binding is enriched near ESC-specific and dynamic genes, many broadly expressed genes also exhibit cobinding within 15 kb of their TSSs. Oct4:Sox2 binding could be important for supporting a high expression level for these genes in pluripotent cells. To examine the functional significance of the less prevalent Oct4:Sox2 cobinding near broadly expressed genes, we first used CRISPR-HDR to mutate composite motifs in the Tet-on iPSCs in regions near the *Pds5a* and *Hnrnp1* genes, both defined as broadly expressed using our criteria. The *Pds5a* Oct4:Sox2 motif is located 2.4 kb upstream of the TSS (Supplemental Fig. S7A, Pds5a_E1), and the *Hnrnp1* motif is located in an intron 6.9 kb downstream from the TSS (Supplemental Fig. S7D, Pds5a_E2). Both of these Oct4:Sox2-cobound sites possess characteristics of typical enhancers (Whyte et al. 2013) and coincide with ATAC-seq peaks but not with H3K27ac peaks (Supplemental Fig. S7A,D).

Importantly, in both primary Tet-on iPSCs and following secondary reprogramming, transcription of *Pds5a* and *Hnrnp1* was largely unaffected following CRISPR-HDR mutagenesis of the Oct4:Sox2 composite motifs (Supplemental Fig. S7C,F) despite loss of Oct4 and Sox2 binding (Supplemental Fig. S7B,E). H3K27ac remained undetectable near the mutant sites and was unaffected at other locations in the loci (Supplemental Fig. S7G,H).

To extend this analysis, we disrupted Oct4:Sox2 composite motifs near broadly expressed genes at which Oct4:Sox2 cobinding coincided with H3K27ac peaks in ESCs (Supplemental Fig. S8). (Note that Oct4:Sox2 binding accompanied by H3K27ac is very rare near broadly expressed genes [see Fig. 2D].) One motif mutated is located in a region characterized as a typical enhancer 7.1 kb upstream of the TSS for the broadly expressed *Dido1* gene, with three other genes residing within 200 kb (Supplemental Fig. S8A). Importantly, despite the presence of an H3K27ac peak coinciding with the Oct4:Sox2-cobound site, mutation of the underlying composite motif had no significant impact on transcription of any of the four genes in the secondary reprogramming assay (Supplemental Fig. S8C). This mutation also had no impact on the H3K27ac signal (data not shown).

Next, we mutated an Oct4:Sox2 composite motif that coincided with Oct4:Sox2 cobinding, an H3K27ac peak, and superenhancer characteristics in an intron of the broadly expressed *Ift52* gene (14.4 kb from its TSS) (Supplemental Fig. S8D). In two independent mutant lines, this mutation had no impact on transcription of the *Ift52* gene in the secondary reprogramming assay (Supplemental Fig. S8F). It also had no impact on the transcription of two nearby genes: *Sgk2* and *Gtsf11* (Supplemental Fig. S8D,F). Interestingly, however, a fourth nearby gene, *Mybl2*, which was classified as dynamic according to our classification criteria, maintained a low expression level following secondary reprogramming of the mutant cells, despite activation of *Mybl2* transcription in the wild-type line (Supplemental Fig. S8F). *Mybl2*, encoding the transcription factor B-Myb, is important for inner mass formation and contributes to the control of cell proliferation but has been studied only minimally for its role in pluripotent cells (Zhan et al. 2012). These results provide an interesting example of an Oct4:Sox2 composite site in an intron of a broadly expressed gene that appears to directly regulate a nearby dynamic gene. The result strengthens the relationship between Oct4:Sox2 and dynamically regulated genes by suggesting that some Oct4:Sox2-cobound sites near broadly expressed genes may actually be regulators of more distant dynamic genes.

Increased transcription of silent genes following Oct4:Sox2 composite motif mutagenesis and secondary reprogramming

As noted above, a large number of Oct4:Sox2-cobound sites are located within 15 kb of genes that are silent or poorly expressed in ESCs. To test the role of Oct4:Sox2 cobinding at these locations, we disrupted Oct4:Sox2 composite motifs that coincide with Oct4:Sox2 cobinding in close proximity to two different silent genes. One motif is located in an intron 8.7 kb downstream from the silent *Oxgr1* gene (Fig. 7A, region E1; Supplemental Fig. S9), and the second is located in an intron 7.5 kb downstream from the *Gnrhr* gene (Fig. 7E, region E1; Supplemental Fig. S9). The *Oxgr1* motif is in a region characterized as a typical enhancer, and the *Gnrhr* motif does not meet the criteria of either a superenhancer or typical enhancer. Both of

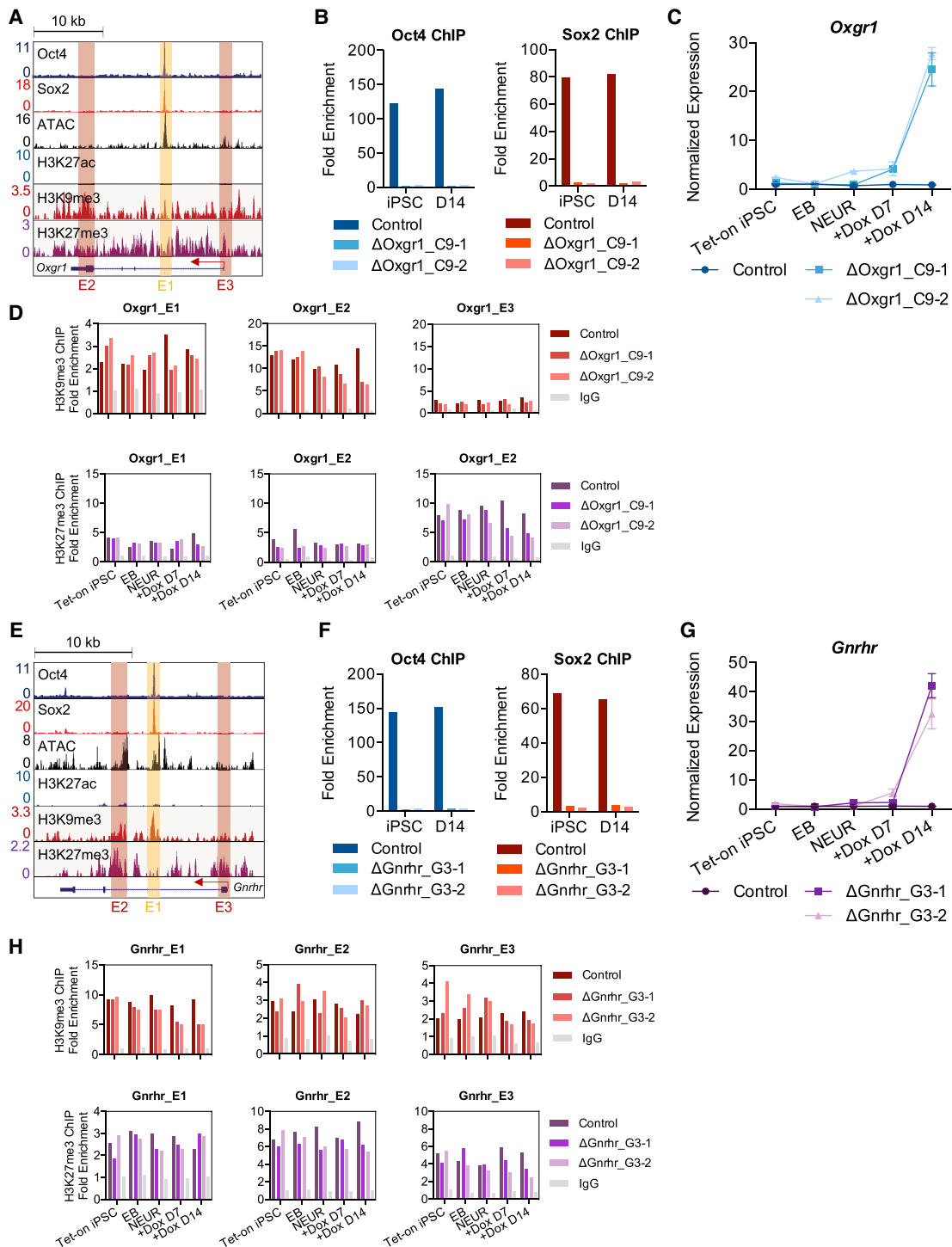


Figure 7. Roles for Oct4:Sox2 composite motifs near the silent *Oxgr1* and *Gnhr* genes in establishing a silent state during secondary reprogramming but not for maintaining silencing in tetO-OSKM iPSCs. (A) Genome browser snapshots display Oct4 and Sox2 ChIP-seq, ATAC-seq, H3K27ac ChIP-seq, H3K9me3 ChIP-seq, and H3K27me3 ChIP-seq tracks at the silent *Oxgr1* locus in mouse ESCs. Yellow shading highlights the region containing the Oct4:Sox2 composite motif. Red shading highlights two other regions (E2 and E3) analyzed here. (B) Bar graphs show Oct4 and Sox2 binding monitored by ChIP-qPCR at the *Oxgr1* enhancer in wild-type and two independent mutant tetO-OSKM lines in primary iPSCs and day 14 secondary iPSCs. Data are displayed as in Figure 3C. (C) The line graph shows normalized *Oxgr1* mRNA levels (by qRT-PCR) in wild-type and two independent mutant tetO-OSKM lines at each stage of differentiation and secondary reprogramming. Values represent means of three independent samples along with standard errors. (D) Bar graphs show H3K9me3 and H3K27me3 ChIP-qPCR levels at the *Oxgr1* E1, E2, and E3 regions in wild-type and mutant tetO-OSKM lines at each stage of differentiation. (E–H) Data analogous to those in A–D are shown for the silent *Gnhr* gene.

these genes encode G protein-coupled receptors and both are expressed in somatic cell types but are silent in ESCs (Supplemental Fig. S9).

The *Oxgr1* and *Gnrhr* Oct4:Sox2-cobound sites coincide with ATAC-seq peaks but not with histone H3K27ac (Fig. 7A,E). The repressive histone modifications—H3K27me3 and H3K9me3—span the two loci at a low level, but peaks for these modifications did not coincide with the Oct4:Sox2-cobound sites (Fig. 7A,E).

Mutagenesis of the Oct4:Sox2 composite motifs had no significant impact on transcription of the genes in the primary iPSCs (Fig. 7C,G). Surprisingly, however, following secondary reprogramming, greatly enhanced transcript levels for each gene emerged in two independent lines harboring each of the motif mutations (Fig. 7C,G). In all mutant lines, transcript levels increased by a large magnitude of 30-fold to 40-fold (Fig. 7C,G), although the transcript levels appeared to remain below those observed in somatic cells expressing the genes (Supplemental Fig. S9). Notably, the mutations led to only small, variable decreases in the H3K9me3 and H3K27me3 modifications that did not reach statistical significance (Fig. 7D,H; data not shown). These results suggest that Oct4:Sox2 cobinding at these genes is not needed to maintain the silent state in pluripotent cells. However, cobinding is needed for the genes to enter into a fully silent state following reprogramming.

Discussion

Transcription factor loss-of-function experiments and genomic studies have provided many important insights into the roles of Oct4 and Sox2 in the establishment and maintenance of pluripotency and in the regulation of pluripotency networks (see above). We set out to increase our understanding of pluripotency by first quantitatively examining the dynamic range of expression of genes expressed in ESCs. We then used CRISPR-HDR to examine the roles of representative Oct4:Sox2 composite motifs in gene transcription. Our results suggest that, during reprogramming to pluripotency, Oct4 and Sox2 preferentially activate genes in pluripotent cells that are subject to large dynamic ranges of expression during mammalian development, but with lesser roles in transcriptional maintenance. Oct4 and Sox2 also appear to contribute to the establishment of a silent state at somatic cell-specific genes, but again with a lesser role in maintenance of the silent state.

Prior analyses of ChIP-seq data sets have revealed networks of genes that appear to be directly regulated by Oct4 and Sox2, with the target genes involved in signaling pathways, chromatin structure, self-renewal, and survival, all of which contribute to the pluripotent state (Boyer et al. 2005; Zhou et al. 2007; Chen et al. 2008; Kim et al. 2008). However, the characteristics that dictate whether a gene that contributes to pluripotency will be directly regulated by Oct4 and Sox2 remained undefined. Our results suggest that a key characteristic of Oct4:Sox2 direct targets is the dynamic range of expres-

sion required for each gene. In other words, genes that are highly expressed in ESCs and involved in pluripotency are much less likely to be directly regulated by Oct4:Sox2 if their expression level is modulated to only a limited extent during development. In addition, ESC-expressed genes that are subject to large dynamic ranges of expression may be directly regulated by Oct4:Sox2 regardless of their previously established role in pluripotency. Notably, our criteria for defining a gene as dynamic requires a large dynamic range of expression (>20-fold) that greatly exceeds the statistical criteria typically used to define differential expression (often 1.5-fold to twofold). In fact, a substantial fraction of the genes included in our “broadly expressed” category are differentially expressed from the perspective of statistical significance but by small magnitudes (less than fivefold) in comparison with the 20-fold differential expression used to define the “dynamic” gene class.

The results suggest that dynamic and broadly expressed genes may require distinct regulatory strategies, with the former category benefiting from direct regulation by Oct4 and Sox2 for reasons that remain to be defined. Our findings suggest that, in other cell types, these dynamic genes are activated by other transcription factors, often acting through distinct regulatory regions. An example is *Zfp57*, which is highly expressed in NPCs despite the absence of a role for the Oct4:Sox2 composite motif and despite the absence of ATAC-seq accessibility and H3K27ac in NPCs at the enhancer bound by Oct4 and Sox2 in ESCs. Many examples of genes regulated by different enhancers in different cell types have been reported.

Recent studies have begun to uncover different regulatory features for classes of genes with different expression profiles. For example, Zabidi et al. (2015) found that enhancers for developmental and housekeeping genes, which may correspond to our dynamic and broadly expressed genes, respectively, display distinct core promoter preferences, suggesting fundamental regulatory differences. More recently, Bergman et al. (2022) found that promoters for housekeeping genes possess features that reduce their responsiveness to enhancers in comparison with the promoters of variably expressed genes (which may be analogous to the developmental genes of Zabidi et al. [2015] and our dynamic genes). The unique regulatory requirements of these genes, which remain poorly understood, may contribute to their requirement for direct regulation by Oct4:Sox2.

The finding that Oct4:Sox2 motifs near dynamic genes are less important for transcriptional maintenance than for transcriptional activation may at face value appear to contradict extensive evidence from loss-of-function studies that Oct4 and Sox2 are both required for the maintenance of pluripotency (Nichols et al. 1998; Niwa et al. 2000; Ivanova et al. 2006; Masui et al. 2007). However, a key distinction is that motif mutagenesis provides an assessment of the impact of Oct4:Sox2 binding on expression of an individual gene, whereas the Oct4 and Sox2 loss-of-function studies assessed the collective impact of eliminating or reducing binding throughout the genome. It therefore may not be surprising that, despite a limited

role of Oct4:Sox2 in transcriptional maintenance at individual genes, modest reductions in transcription genome-wide would eliminate pluripotency.

The critical roles of Oct4:Sox2 during reprogramming are consistent with extensive evidence that these factors possess pioneering activity, allowing them to access chromatin at inactive genes and promote transcriptional activation during the transition to a pluripotent state (Soufi et al. 2012, 2015). The lesser impact of Oct4:Sox2 motifs on transcriptional maintenance suggests that other transcriptional activators and coactivators associated with the Oct4:Sox2 target genes can support transcription at a substantial level in pluripotent cells after the Oct4:Sox2 motif is eliminated. Although we cannot rule out the possibility of redundancy with additional Oct4:Sox2 binding sites near the dynamic genes examined, these sites would need to be far from the target gene, as the genes studied generally lacked additional identifiable Oct4 and Sox2 binding events accompanied by H3K27ac. The modest reduction of histone H3K27ac that was generally observed in the Oct4:Sox2 motif mutants suggests that these enhancers remain partially active after Oct4 and Sox2 binding is eliminated. Notably, although the Oct4:Sox2 composite motifs near dynamic genes were greatly enriched for super-enhancers, one that was examined functionally (near *Epb4.115*) was in a region classified as a typical enhancer. One possibility is that the characteristics used to define superenhancers are imperfect.

The absence of a functional impact of Oct4:Sox2 composite motif mutations near broadly expressed genes will require further examination. One possibility is that these composite motifs entirely lack functional roles, perhaps due to the absence of additional factors bound in close proximity that may be needed to support enhancer activity. An alternative possibility is that these motifs regulate the activation of more distant dynamic genes that were not examined. One example of this scenario was obtained upon mutagenesis of the Oct4:Sox2 motif near the broadly expressed *Ift52* gene; disruption of this motif did not alter *Ift52* transcription, but transcription of the adjacent dynamic gene, *Mybl2*, was eliminated following secondary reprogramming.

The most surprising results were arguably obtained upon mutagenesis of Oct4:Sox2 motifs near genes that are silent in pluripotent cells. These composite motifs did not contribute to the maintenance of the pre-established silent state in pluripotent cells, but they contributed to the initial establishment of silencing during secondary reprogramming. Our results suggest that pluripotent cells express transcription factors capable of supporting the transcription of these genes in the absence of an Oct4:Sox2-dependent mechanism for establishing a silent state. Alternatively, the relatively open chromatin structure characteristic of pluripotent cells (Lim and Meshorer 2021) may allow promiscuous transcription of these genes if an Oct4:Sox2-dependent silencing mechanism is absent. The repressive histone modifications H3K9me3 and H3K27me3 were implicated in Oct4:Sox2-mediated silencing, but the changes in modification levels were small and will require further study. Never-

theless, our findings are consistent with a number of prior studies that implicated Oct4 and Sox2 in transcriptional repression and in the silencing of somatic cell-specific genes during reprogramming (Liu et al. 1997; Liang et al. 2008; Esch et al. 2013; Chronis et al. 2017).

Materials and methods

Cell culture, iPSC differentiation, and secondary reprogramming

Mouse CCE ESCs were grown as described (Langerman et al. 2021). TetO iPSCs (Sridharan et al. 2013) were differentiated into NPCs as described (Sagner et al. 2018). In brief, an established TetO iPSC line was grown in the absence of doxycycline for 3–5 d in Corning Ultra-low attachment culture dishes to form EBs in EB formation medium (DMEM, 7.5% FBS, 2 mM L-glutamine, 1% penicillin/streptomycin, 0.1mM nonessential amino acid, 100 μ M β -mercaptoethanol). EBs were subsequently plated onto adherent culture dishes with 0.5 μ M retinoic acid (Millipore Sigma) to induce NPC differentiation. Neural-committed EBs were collected 48 h later and grown in NPC culture medium (DMEM plus 10% FBS) for 5–7 d. Differentiation into NPCs was monitored by measuring *Sox1*, *Nes*, *Pax6*, and *Pax3* transcript levels by qRT-PCR. To induce secondary reprogramming, NPCs were grown in DMEM plus 10% FBS at densities of 1.0×10^4 cells per 35 mm on gelatin-coated dishes with monolayer mouse feeders for 48 h. Culture medium was then removed and replaced by ESC medium supplemented with 2 μ g/mL doxycycline and cultured as described. Cell identity was monitored by qRT-PCR analysis of pluripotency genes (*Oct4*, *Sox2*, *Nanog*, *Ssea1*, *Klf4*, *Rex1*, and *Nr0b1*) and NPC genes (*Sox1*, *Nes*, *Pax6*, and *Pax3*).

CRISPR/Cas9 mutagenesis

Single guide RNAs (sgRNAs) and homology-directed repair (HDR) templates were designed using Massachusetts Institute of Technology CRISPR Designer (<http://crispr.mit.edu>) and Benchling CRISPR Guide Design (<https://www.benchling.com/crispr>). HDR templates containing EcoRI/BamHI sequences were designed to introduce substitution mutations into Oct4:Sox2 motifs. pSpCas9(BB)-2A-Puro (PX459) V2.0 (Addgene 62988) expressing both Cas9 and the sgRNAs (Cong and Zhang 2015) was cotransfected with the HDR template into CCE ESCs or iPSCs. Puromycin-resistant cells were collected, diluted, and plated in 96-well plates to obtain single-cell colonies for genotyping. Genomic regions flanking the Oct4:Sox2 composite motifs were amplified by PCR and sequenced to confirm the presence of the substitution mutations.

ChIP-qPCR

Nuclei from 4×10^7 cells were collected as described (Ramirez-Carrozzi et al. 2006). Nuclear pellets were sonicated with a Misonix 3000 sonicator and subsequently incubated overnight with ChIP-grade antibodies for Oct4 (R&D AF1759), Sox2 (R&D AF2018), H3K27ac (Active Motif 39133), H3K27me3 (Active Motif 39155), or H3K9me3 (Abcam ab8898). Antibody-bound complexes were collected using protein G Dynabeads (Invitrogen 10004D) and then reverse-cross-linked with proteinase K (Thermo Fisher Scientific EO0491) overnight at 60°C. Immunoprecipitated DNA was purified using phenol-chloroform extraction (Sigma P3803) and quantified by Qubit (Thermo Fisher Q32854). The enrichment of chromatin fragments was measured by qPCR with primer pairs designed to generate 100- to 125-bp amplified products within ± 200 bp of the genomic sites of

interest. Quantification of fold enrichment was calculated based on the fold change of the percentage of input between target genomic loci and a negative control region (*Hbb-b2*).

RNA-seq

Chromatin-associated RNA was fractionated and isolated as described (Bhatt et al. 2012). Ribosomal RNA (rRNA) was depleted using the RiboMinus transcriptome isolation kit (Thermo Fisher Scientific). Strand-specific cDNA libraries were prepared using the Illumina TruSeq RNA sample preparation kit v2 (Illumina) with the dUTP second strand method (Levin et al. 2010). All cDNA libraries were single-end-sequenced (50 bp) on an Illumina HiSeq 2000 at the University of California at Los Angeles Broad Stem Cell Research Center High-Throughput Sequencing Core. Reads were mapped to the mouse NCBI37/mm9 reference genome by HISAT2 v2.1.0, and only those uniquely mapping reads with no more than two mismatches were retained (Kim et al. 2015). Chromatin RNA RPKM was calculated as previously described (Tong et al. 2016). mRNA RPKM was calculated by counting all mapped exonic reads and dividing by the length of the spliced product (Mortazavi et al. 2008). All RPKMs represent an average from two or three biological replicates in each tissue/cell type. Published mRNA sequencing data sets were obtained from the Mouse ENCODE Project (<http://www.mouseencode.org>) and are listed in Supplemental Figure S11.

ChIP-seq and ATAC-seq

Public ChIP-seq data sets obtained from GEO are listed in Supplemental Figure S11 (Buecker et al. 2014; Chronis et al. 2017; Mollitor et al. 2017). ChIP-seq data processing was performed as previously described (Tong et al. 2016). Peak calling and gene annotation were done by HOMER (<http://homer.ucsd.edu/homer>) with false discovery rate (FDR) <0.01 and enrichment over input. Only reproducible peaks from replicates were retained for downstream analyses. Called peaks were annotated to the nearest TSSs of genes. Oct4 and Sox2-cobound sites were determined by the distance of peak summit with <100 bp. Histone modification enrichments were analyzed by calculating the RPKM values of 1.5-kb windows centered on the center of Oct:Sox2-cobound sites. ChIP-seq-discovered peaks were analyzed by MEME-ChIP (Bailey et al. 2009; Machanick and Bailey 2011) for motif discovery. For the de novo motif analyses at Oct4:Sox2-cobound sites, the peak center was defined as the midpoint of two peak summits. Motif width between 10 and 30 bp and *P*-value threshold of <0.05 were used to identify composite motifs.

Public ATAC-seq data sets obtained from GEO are listed in Supplemental Figure S11 (Di Stefano et al. 2016; Xu et al. 2017). Analysis of ATAC-seq data was conducted as described (Tong et al. 2016). Chromatin accessibility (ATAC sensitivity) was analyzed by calculating the RPKM values of 1.0-kb windows centered on the center of the Oct4:Sox2-cobound sites.

RNA extraction and qRT-PCR

Cells grown in six-well plates at a confluency of 80%–90% were lysed in TRI reagent (Molecular Research Center TR118). RNA was extracted by Qiagen RNeasy kit following the manufacturer's instructions and reverse-transcribed into cDNA by SuperScript III reverse transcriptase (Thermo Fisher Scientific). cDNA levels were quantified by RT-PCR using PowerUp SYBR Green master mix (Thermo Fisher Scientific) on a Bio-Rad CFX384 real-time PCR system. Gene expression levels were calculated relative to a standard curve and normalized to the house-

keeping gene *Gapdh* in triplicate. Primers used for qRT-PCR are listed in Supplemental Figure S10.

Data availability

All data have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE61540.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

This work was funded by National Institutes of Health grants P01 GM099134 (to S.T.S. and K.P.), R01 GM086372 (to S.T.S.), and T32 GM007185 (to M.E. and J.L.). S.T.S. is a Senior Scientific Officer with the Howard Hughes Medical Institute.

Author contributions: J.H.-H.L. designed the experiments; acquired, analyzed, and interpreted the data; and prepared the manuscript. M.E. and J.L. designed the experiments and acquired, analyzed, and interpreted the data. R.S. and K.P. designed the tools and assisted with data interpretation and critical manuscript revisions. S.T.S. designed the experiments, interpreted the data, and prepared the manuscript.

References

- Ang YS, Tsai SY, Lee DF, Monk J, Su J, Ratnakumar K, Ding J, Ge Y, Darr H, Chang B, et al. 2011. Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145**: 183–197. doi:10.1016/j.cell.2011.03.003
- Apostolou E, Ferrari F, Walsh RM, Bar-Nur O, Stadtfeld M, Che-loufi S, Stuart HT, Polo JM, Ohsumi TK, Borowsky ML, et al. 2013. Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. *Cell Stem Cell* **12**: 699–712. doi:10.1016/j.stem.2013.04.013
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME suite: tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202–W208. doi:10.1093/nar/gkp335
- Bergman DT, Jones TR, Liu V, Ray J, Jagoda E, Siraj L, Kang HY, Nasser J, Kane M, Rios A, et al. 2022. Compatibility rules of human enhancer and promoter sequences. *Nature* **607**: 176–184. doi:10.1038/s41586-022-04877-w
- Bhatt D, Pandya-Jones A, Tong AJ, Barozzi I, Lissner MM, Natoli G, Black DL, Smale ST. 2012. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**: 279–290. doi:10.1016/j.cell.2012.05.043
- Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, et al. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**: 947–956. doi:10.1016/j.cell.2005.08.020
- Buecker C, Srinivasan R, Wu Z, Calo E, Acampora D, Faial T, Simeone A, Tan M, Swigut T, Wysocka J. 2014. Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**: 838–853. doi:10.1016/j.stem.2014.04.003
- Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, et al. 2008. Integration of external

- signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**: 1106–1117. doi:10.1016/j.cell.2008.04.043
- Chen J, Zhang Z, Li L, Chen B-C, Revyakin A, Bassam H, Legant W, Dahan M, Lionnet T, Betzig E, et al. 2014. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**: 1274–1285. doi:10.1016/j.cell.2014.01.062
- Chen J, Chen X, Li M, Liu X, Gao Y, Kou X, Zhao Y, Zheng W, Zhang X, Huo Y, et al. 2016. Hierarchical Oct4 binding in concert with primed epigenetic rearrangements during somatic cell reprogramming. *Cell Rep* **14**: 1540–1554. doi:10.1016/j.celrep.2016.01.013
- Chronis C, Fizev P, Papp B, Butz S, Bonora G, Sabri S, Ernst J, Plath K. 2017. Cooperative binding of transcription factors orchestrates reprogramming. *Cell* **168**: 442–459.e20. doi:10.1016/j.cell.2016.12.016
- Cong L, Zhang F. 2015. Genome engineering using CRISPR–Cas9 system. *Methods Mol Biol* **1239**: 197–217. doi:10.1007/978-1-4939-1862-1_10
- Deng W, Jacobson EC, Collier AJ, Plath K. 2021. The transcription factor code in iPSC reprogramming. *Curr Opin Genet Dev* **70**: 89–96. doi:10.1016/j.gde.2021.06.003
- Di Stefano B, Collombet S, Jakobsen JS, Wierer M, Sardina JL, Lackner A, Stadhouders R, Segura-Morales C, Francesconi M, Limone F, et al. 2016. C/EBP α creates elite cells for iPSC reprogramming by upregulating Klf4 and increasing the levels of Lsd1 and Brd4. *Nat Cell Biol* **18**: 371–381. doi:10.1038/ncb3326
- Esch D, Vahokoski J, Groves MR, Pogenberg V, Cojocaru V, Vom Bruch H, Han D, Drexler HCA, Araúzo-Bravo MJ, Ng CKL, et al. 2013. A unique Oct4 interface is crucial for reprogramming to pluripotency. *Nat Cell Biol* **15**: 295–301. doi:10.1038/ncb2680
- Hockemeyer D, Soldner F, Cook EG, Gao Q, Mitalipova M, Jaenisch R. 2008. A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. *Cell Stem Cell* **3**: 346–353. doi:10.1016/j.stem.2008.08.014
- Ivanova N, Dobrin R, Lu R, Kotenko I, Levorse J, DeCoste C, Schafer X, Lun Y, Lemischka IR. 2006. Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**: 533–538. doi:10.1038/nature04915
- Jauch R, Ng CKL, Saikatendu KS, Stevens RC, Kolatkar PR. 2008. Crystal structure and DNA binding of the homeodomain of the stem cell transcription factor Nanog. *J Mol Biol* **376**: 758–770. doi:10.1016/j.jmb.2007.11.091
- Ji X, Dadon DB, Abraham BJ, Lee TI, Jaenisch R, Bradner JE, Young RA. 2015. Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proc Natl Acad Sci* **112**: 3841–3846. doi:10.1073/pnas.1502971112
- Kim J, Chu J, Shen X, Wang J, Orkin SH. 2008. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**: 1049–1061. doi:10.1016/j.cell.2008.02.039
- Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**: 357–360. doi:10.1038/nmeth.3317
- King HW, Klose RJ. 2017. The pioneer factor OCT4 requires the chromatin remodeler BRG1 to support gene regulatory element function in mouse embryonic stem cells. *Elife* **6**: e22631. doi:10.7554/eLife.22631
- Langerman J, Lopez D, Pellegrini M, Smale ST. 2021. Species-specific relationships between DNA and chromatin properties of CpG islands in embryonic stem cells and differentiated cells. *Stem Cell Reports* **16**: 899–912. doi:10.1016/j.stemcr.2021.02.016
- Lee SH, Lumelsky N, Studer L, Auerbach JM, McKay RD. 2000. Efficient generation of midbrain and hindbrain neurons from mouse embryonic stem cells. *Nat Biotechnol* **18**: 675–679. doi:10.1038/76536
- Lee JD, Migeotte I, Anderson KV. 2010. Left-right patterning in the mouse requires Epb4.115-dependent morphogenesis of the node and midline. *Dev Biol* **346**: 237–246. doi:10.1016/j.ydbio.2010.07.029
- Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, Gnirke A, Regev A. 2010. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods* **7**: 709–715. doi:10.1038/nmeth.1491
- Li M, Belmonte JCI. 2017. Ground rules of the pluripotency gene regulatory network. *Nat Rev Genet* **18**: 180–191. doi:10.1038/nrg.2016.156
- Li M, Izpisua Belmonte JC. 2018. Deconstructing the pluripotency gene regulatory network. *Nat Cell Biol* **20**: 382–392. doi:10.1038/s41556-018-0067-6
- Liang J, Wan M, Zhang Y, Gu P, Xin H, Jung SY, Qin J, Wong J, Cooney AJ, Liu D, et al. 2008. Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nat Cell Biol* **10**: 731–739. doi:10.1038/ncb1736
- Lim PSL, Meshorer E. 2021. Organization of the pluripotent genome. *Cold Spring Harb Perspect Biol* **13**: a040204. doi:10.1101/cshperspect.a040204
- Liu L, Leaman D, Villalta M, Roberts RM. 1997. Silencing of the gene for the α -subunit of human chorionic gonadotropin by the embryonic transcription factor Oct-3/4. *Mol Endocrinol* **11**: 1651–1658.
- Liu X, Huang J, Chen T, Wang Y, Xin S, Li J, Pei G, Kang J. 2008. Yamanaka factors critically regulate the developmental signaling network in mouse embryonic stem cells. *Cell Res* **18**: 1177–1189. doi:10.1038/cr.2008.309
- Machanic P, Bailey TL. 2011. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**: 1696–1697. doi:10.1093/bioinformatics/btr189
- Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, Takahashi K, Okochi H, Okuda A, Matoba R, Sharov AA, et al. 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* **9**: 625–635. doi:10.1038/ncb1589
- Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S. 2003. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**: 631–642. doi:10.1016/S0092-8674(03)00393-3
- Molitor J, Mallm JP, Rippe K, Erdel F. 2017. Retrieving chromatin patterns from deep sequencing data using correlation functions. *Biophys J* **112**: 473–490. doi:10.1016/j.bpj.2017.01.001
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**: 621–628. doi:10.1038/nmeth.1226
- Nichols J, Zevnik B, Anastasiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Schöler H, Smith A. 1998. Formation of pluripotent stem cells in the mammalian embryo depend on the POU transcription factor Oct4. *Cell* **95**: 379–391. doi:10.1016/S0092-8674(00)81769-9
- Niwa H, Miyazaki J, Smith AG. 2000. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* **24**: 372–376. doi:10.1038/74199
- Pardo M, Lang B, Yu L, Prosser H, Bradley A, Babu MM, Choudhary J. 2010. An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell* **6**: 382–395. doi:10.1016/j.stem.2010.03.004

- Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Balivo I, Pedone PV, Grimaldi G, Riccio A, et al. 2011. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol Cell* **44**: 361–372. doi:10.1016/j.molcel.2011.08.032
- Ramirez-Carrozzi VR, Nazarian AA, Li CC, Gore SL, Sridharan R, Imbalzano AN, Smale ST. 2006. Selective and antagonistic functions of SWI/SNF and Mi-2 β nucleosome remodeling complexes during an inflammatory response. *Genes Dev* **20**: 282–296. doi:10.1101/gad.1383206
- Reményi A, Lins K, Nissen LJ, Reinbold R, Schöler HR, Wilmanns M. 2003. Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev* **17**: 2048–2059. doi:10.1101/gad.269303
- Roberts GA, Ozkan B, Gachulincová I, O'Dwyer MR, Hall-Ponsole E, Saxena M, Robinson PJ, Soufi A. 2021. Dissecting OCT4 defines the role of nucleosome binding in pluripotency. *Nat Cell Biol* **23**: 834–845. doi:10.1038/s41556-021-00727-5
- Sagner A, Gaber ZB, Delile J, Kong JH, Rousso DL, Pearson CA, Weicksel SE, Melchionda M, Mousavy Gharavy SN, Briscoe J, et al. 2018. Olig2 and Hes regulatory dynamics during motor neuron differentiation revealed by single cell transcriptomics. *PLoS Biol* **16**: e2003127. doi:10.1371/journal.pbio.2003127
- Schaefer T, Lengerke C. 2020. Sox2 protein biochemistry in stemness, reprogramming, and cancer: the PI3K/AKT/SOX2 axis and beyond. *Oncogene* **39**: 278–292. doi:10.1038/s41388-019-0997-x
- Soufi A, Donahue G, Zaret K. 2012. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**: 994–1004. doi:10.1016/j.cell.2012.09.045
- Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, Zaret KS. 2015. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**: 555–568. doi:10.1016/j.cell.2015.03.017
- Sridharan R, Tchieu J, Mason MJ, Yachechko R, Kuoy E, Horvath S, Zhou Q, Plath K. 2009. Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136**: 364–377. doi:10.1016/j.cell.2009.01.001
- Sridharan R, Gonzales-Cope M, Chronis C, Bonora G, McKee R, Huang C, Patel S, Lopez D, Mishra N, Pellegrini M, et al. 2013. Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1 γ in reprogramming to pluripotency. *Nat Cell Biol* **15**: 872–882. doi:10.1038/ncb2768
- Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**: 663–676. doi:10.1016/j.cell.2006.07.024
- Tapia N, MacCarthy C, Esch D, Gabriele Marthaler A, Tiemann U, Araúzo-Bravo MJ, Jauch R, Cojocaru V, Schöler HR. 2015. Dissecting the role of distinct OCT4–SOX2 heterodimer configurations in pluripotency. *Sci Rep* **5**: 13533. doi:10.1038/srep13533
- Tong AJ, Liu X, Thomas BJ, Lissner MM, Baker MR, Senagolage MD, Allred AL, Barish GD, Smale ST. 2016. A stringent systems approach uncovers gene-specific mechanisms regulating inflammation. *Cell* **165**: 165–179. doi:10.1016/j.cell.2016.01.020
- van den Berg DLC, Snoek T, Mullin NP, Yates A, Bezstarosti K, Demmers J, Chambers I, Poot RA. 2010. An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* **6**: 369–381. doi:10.1016/j.stem.2010.02.014
- Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**: 307–319. doi:10.1016/j.cell.2013.03.035
- Wu G, Schöler HR. 2014. Role of Oct4 in the early embryo development. *Cell Regen* **3**: 7.
- Wu G, Han D, Gong Y, Sebastiano V, Gentile L, Singhal N, Adachi K, Fishedick G, Ortmeier C, Sinn M, et al. 2013. Establishment of totipotency does not depend on Oct4A. *Nat Cell Biol* **15**: 1089–1097. doi:10.1038/ncb2816
- Xu J, Carter AC, Gendrel AV, Attia M, Loftus J, Greenleaf WJ, Tibshirani R, Heard E, Chang HY. 2017. Landscape of monoallelic DNA accessibility in mouse embryonic stem cells and neural progenitor cells. *Nat Genet* **49**: 377–386. doi:10.1038/ng.3769
- Zabidi MA, Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O, Stark A. 2015. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**: 556–559. doi:10.1038/nature13994
- Zhan M, Riordon DR, Yan B, Tarasova YS, Bruweleit S, Tarasov KV, Li RA, Wersto RP, Boheler KR. 2012. The B-MYB transcriptional network guides cell cycle progression and fate decisions to sustain self-renewal and the identity of pluripotent stem cells. *PLoS One* **7**: e42350. doi:10.1371/journal.pone.0042350
- Zhou Q, Chipperfield H, Melton DA, Wong WH. 2007. A gene regulatory network in mouse embryonic stem cells. *Proc Natl Acad Sci* **104**: 16438–16443. doi:10.1073/pnas.0701014104