Cell Stem Cell
Article

Long-Range Chromatin Contacts in Embryonic Stem Cells Reveal a Role for Pluripotency Factors and Polycomb Proteins in Genome Organization

Matthew Denholtz,[1,4] Giancarlo Bonora,[1,4] Constantinos Chronis,[1] Erik Splinter,[3] Wouter de Laat,[3] Jason Ernst,[1] Matteo Pellegrini,[2,*] and Kathrin Plath[1,*]
[1]Department of Biological Chemistry at the David Geffen School of Medicine
[2]Department of Molecular, Cell, and Developmental Biology
University of California, Los Angeles, CA, 90095, USA
[3]Hubrecht Institute-KNAW & University Medical Center, Utrecht, 3584 CT, The Netherlands
[4]These authors contributed equally to this work
*Correspondence: matteop@mcdb.ucla.edu (M.P.), kplath@mednet.ucla.edu (K.P.)
http://dx.doi.org/10.1016/j.stem.2013.08.013

SUMMARY

The relationship between 3D organization of the genome and gene-regulatory networks is poorly understood. Here, we examined long-range chromatin interactions genome-wide in mouse embryonic stem cells (ESCs), iPSCs, and fibroblasts and uncovered a pluripotency-specific genome organization that is gradually reestablished during reprogramming. Our data confirm that long-range chromatin interactions are primarily associated with the spatial segregation of open and closed chromatin, defining overall chromosome conformation. Additionally, we identified two further levels of genome organization in ESCs characterized by colocalization of regions with high pluripotency factor occupancy and strong enrichment for Polycomb proteins/H3K27me3, respectively. Underlining the independence of these networks and their functional relevance for genome organization, loss of the Polycomb protein Eed diminishes interactions between Polycomb-regulated regions without altering overarching chromosome conformation. Together, our data highlight a pluripotency-specific genome organization in which pluripotency factors such as Nanog and H3K27me3 occupy distinct nuclear spaces and reveal a role for cell-type-specific gene-regulatory networks in genome organization.

INTRODUCTION

Chromosome conformation capture (3C)-based technologies (de Wit and de Laat, 2012; Dekker et al., 2002) have led to a new paradigm wherein gene regulation can be studied in the context of the three-dimensional (3D) organization of the genome (Bickmore and van Steensel, 2013). Recent work has demonstrated an organizational hierarchy to metazoan genome structure (Gibcus and Dekker, 2013). At the smallest scale, up to a few hundred kilobases (kb) of linear DNA, enhancers and promoters come into physical contact to establish cell-type-specific expression programs (Sanyal et al., 2012; Shen et al., 2012; Smallwood and Ren, 2013). These interactions are maintained by the Cohesin complex, which can be recruited to interphase chromatin via the Mediator complex (Kagey et al., 2010) and cell-type-specific transcription factors (TFs) (Denholtz and Plath, 2012; Wei et al., 2013). In mammals, promoter-enhancer interactions are confined to topologically associating domains (TADs), which typically represent ~1 megabase (Mb) cell-type-invariant, self-associating genomic regions whose boundaries are enriched for the insulator protein CTCF (Dixon et al., 2012; Nora et al., 2012).

As a second level of the organizational hierarchy, TADs appear to function as the fundamental modular unit of gene regulation and genome organization, with changes in gene expression and nuclear lamina association during differentiation often occurring in a TAD-wide manner (Dixon et al., 2012; Nora et al., 2012; Shen et al., 2012). A third level in the organizational hierarchy occurs as a result of the preferential colocalization of specific TADs. These interactions can be identified as long-range chromatin contacts between genomic regions many Mb apart on the same chromosome (*cis* or intrachromosomal) or on different chromosomes (*trans* or interchromosomal) (Hakim et al., 2011, 2013; Noordermeer et al., 2011; Osborne et al., 2004; Schoenfelder et al., 2010; Simonis et al., 2006). Although spatially colocalizing distal genomic regions have been reported to be enriched for the sequence motifs of specific TFs (Schoenfelder et al., 2010), coexpressed genes (Osborne et al., 2004), or coregulated genes (Hakim et al., 2013; Noordermeer et al., 2011), the relationship between long-range chromatin interactions and the regulatory features enriched in the colocalizing regions is poorly understood. Superimposed upon the milieu of specific short- and long-range chromatin interactions is a general preference for open, accessible chromatin to colocalize with itself and segregate away from closed, inaccessible chromatin (Lieberman-Aiden et al., 2009).

In this study, we explored long-range (distal) chromatin contacts in mouse embryonic stem cells (ESCs) using 4C-seq to understand the relationship between chromatin contacts and gene-regulatory networks that govern cell identity. We define the distal chromatin interactions made by a variety of genomic "bait" regions representing diverse chromatin and pluripotency

TF binding profiles and extend our findings genome-wide. We further examine how genome organization changes in the absence of a critical chromatin regulator and upon differentiation and transcription-factor-induced reprogramming of somatic cells into induced pluripotent stem cells (iPSCs). Together, our data define a previously unappreciated hierarchy in the organization of long-range chromatin contacts and reveal that distal genomic regions sharing common gene-regulatory features co-localize within the 3D space of the nucleus.

RESULTS

Experimental Approach to Studying Chromatin Contacts

To investigate long-range chromatin interactions between genomic regions Mb away on the same or different chromosomes in ESCs, we performed 3C coupled to high-throughput sequencing (4C-seq) (Splinter et al., 2012) for 16 bait regions (Table S1 available online). 4C-seq allows one to identify any mappable genomic region in close physical proximity to a specific genomic (bait) region within a population of cells at the moment of fixation by means of proximity-based ligation of juxtaposed DNA fragments. The product is a library of chimeric DNA fragments containing the bait region and its interacting DNA partner or partners ligated at a restriction site, in our case HindIII, that can be identified by high-throughput sequencing (Experimental Procedures).

Our 4C-seq data were highly reproducible across biological and technical replicates utilizing distinct primer pairs for amplification of ligated fragments and different HindIII fragments within bait regions as anchor points and passed stringent quality control requirements (Figures S1A, S1D, S1E, and S2A available online, Table S2 and Table S3, and Experimental Procedures). As a result, replicate data sets for each bait locus were pooled for downstream analysis (Experimental Procedures). To obtain a semiquantitative measure of interactions, we calculated an average hit probability, referred to as "hit percentage," within 200 kb windows along the genome (Experimental Procedures). For all of our baits, the hit percentage was higher in *cis* than in *trans*, as exemplified by the 4C-seq interactome of the *Pou5f1*-encoding region (Figure 1Ai). Furthermore, as expected for the nature of chromosome conformation (Lieberman-Aiden et al., 2009), intrachromosomal interactions decay as a function of genomic distance to the bait along the linear DNA (Figure 1Aii).

Significantly interacting regions were identified as those 200 kb windows that showed a markedly higher hit probability than expected based on a binomial test (Figures 1Aii–1Av, 1B, Figure S3A, Table S1, Table S4, Experimental Procedures). To identify distal interactions in *cis*, we empirically modeled the expected background hit probability as a function of distance from the bait locus (Figure 1Aiii). Since *trans* interactions show no such positional biases, the average hit probability across each *trans*-chromosome was used for the expected background level of these interactions. False discovery rates (FDRs) were estimated using simulated data (Table S4, Experimental Procedures).

To test the reliability of our analysis and 4C-seq data, we partitioned our *Pou5f1* 4C-seq replicates into two equal subsets, pooled the libraries within each subset, and ran each pooled subset through our analysis pipeline. We obtained significantly overlapping intrachromosomal and interchromosomal interactomes from the partitioned data sets (Figure S1B), confirming the quality of our data and the robust nature of our analysis.

Several additional approaches were employed to further validate our 4C-seq-defined chromatin interactions and rule out technical biases affecting data generated by 4C-seq. These include 3C confirmation of ligation products (Figure S1F); 3D fluorescence in situ hybridization (FISH) demonstrating that an interacting region 52 Mb away from the *Pou5f1* bait region is significantly closer to the bait in 3D space of the ESC nucleus than a noninteracting region located only 35 Mb away (Figure 1C); reciprocal 4C-seq (Figure 1D); control 4C-seq experiments with unfixed cells and genomic DNA showing no significant interactions (Figures S1A and S1C); the demonstration of comparable mappability inside and outside our interacting regions (Figure S1G); and the comparison of our 4C-seq data with a previously published genome-wide ESC interactome based on Hi-C (Dixon et al., 2012) showing strong correlation (Figure 3H, Table S5, Experimental Procedures).

A Pluripotency-Specific Organization of the Mouse Genome

The *Pou5f1* gene encodes the TF Oct4 that is essential for the establishment and maintenance of mouse ESCs and is specifically expressed in the pluripotent state (Nichols et al., 1998). The initial analysis of our *Pou5f1* 4C-seq data revealed an extensive intrachromosomal interaction network of this bait region (Figure 1A), raising the question of how the spatial interactions of this important locus are regulated. As a first step toward understanding this question, we determined whether the chromatin interactions of this genomic region change upon differentiation. 4C-seq in mouse embryonic fibroblasts (MEFs) demonstrated that the *Pou5f1* interactome in MEFs is distinct from that of ESCs, both in terms of hit probability (Figure 2A and Figure S2A) and interacting domains (Figure 2B) across the *cis* chromosome, indicating that changes in expression and chromatin state that are associated with the silencing of the *Pou5f1* locus during differentiation (Feldman et al., 2006) coincide with changes to that locus' interactions in 3D space. To expand the purview of this analysis, we determined the interactomes of nine additional baits in MEFs for which we already had generated ESC profiles by 4C-seq and found that the interactomes of all these baits differed significantly between ESCs and MEFs (Figures 2E and 2F, Figure S2A, Table S1). We noted that the chromatin contacts of the *Dppa2* bait region are much more different between MEFs and ESCs than those of the other examined bait regions (Figures 2C and 2D, Figure S2A). Consistent with this result, it has been shown that the *Dppa2* region repositions toward the nuclear periphery and switches *cis*-interaction preferences from regions of early DNA replication to regions of late DNA replication during ESC differentiation (Hiratani et al., 2010; Takebayashi et al., 2012). Together, these data highlight an ESC-specific organization of the genome and indicate that changes in 3D chromatin interactions during differentiation are regulated at a locus-specific level.

We next tested whether the 3D organization of the MEF genome can be reset to an ESC-like state by transcription-factor-mediated reprogramming to iPSCs (Takahashi and Yamanaka, 2006). To this end, we determined the interactomes of eight of our bait regions, including the *Pou5f1* locus, for two

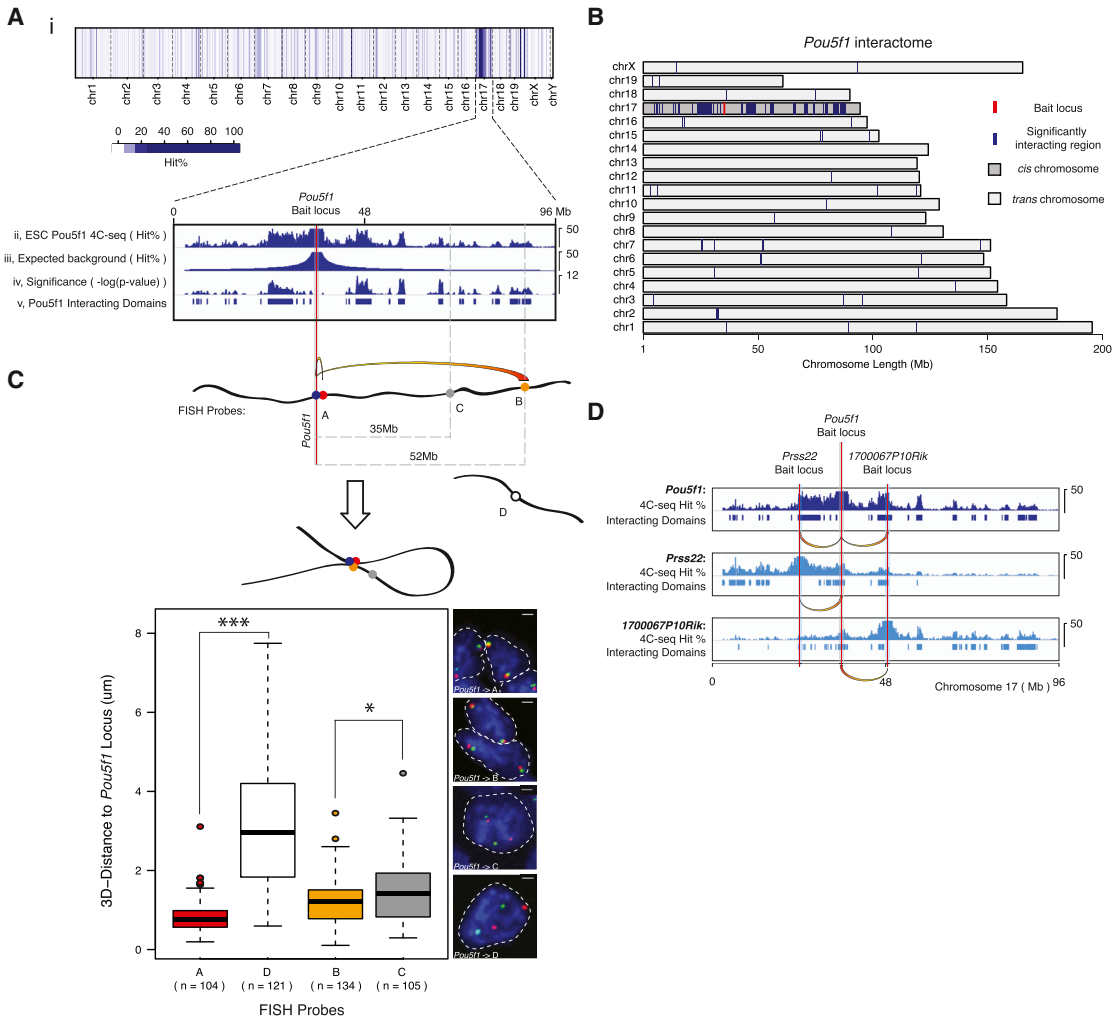


Figure 1. Long-Range Chromatin Contacts of the *Pou5f1* Bait Region in ESCs

(A) Pooled 4C-seq data set, depicting the genome-wide interaction profile of the *Pou5f1* locus, located on chromosome (chr) 17, in mouse ESCs. (Ai) Average hit probability (Hit%) within 200 kb windows tiled across the genome. Blue scale = gradient of hit percentages. (Aii) Hit% in *cis* (thresholded as displayed) was compared to an empirically modeled background Hit% (Aiii) using the binomial test. (Aiv) Binomial test results and (Av) significantly interacting domains of the *Pou5f1* bait region (based on −log(p values) ≥ 1.8). The vertical red line and associated gray bar denote the *Pou5f1* bait locus and the extended 1 Mb bait region, respectively. The latter is always excluded from downstream analysis of interactomes.

(B) Significantly interacting domains (blue) of the *Pou5f1* bait region (red), genome-wide.

(C) DNA FISH-based 3D distance measurements between the *Pou5f1* locus and genomic regions marked by A–C (in *cis*) or by D (in *trans*). Based on 4C-seq data, A and B interact with the *Pou5f1* locus, but C, located in an intervening genomic region closer to *Pou5f1* on linear DNA than B, does not. Boxes demarcate the interquartile range (IQR) with median, and whiskers ±1.5 times the IQR. *p < 0.05, ***p < 0.001, Wilcoxon rank-sum test.

(D) Two genomic regions containing the genes *Prss22* and *1700067P10Rik* were identified as *cis*-interacting partners of the *Pou5f1* locus in (A). 4C-seq experiments using *Prss22* and *1700067P10Rik* as bait regions confirmed their interaction with *Pou5f1*. Bait regions, 4C-seq Hit%, and interacting domains are indicated, and reciprocal interactions highlighted.

See also Figure S1.

additional cell types that represent distinct stages of the reprogramming process (Table S1). First, we performed 4C-seq on faithfully reprogrammed, MEF-derived iPSCs and found that the long-range chromatin contacts in iPSCs are highly similar to those of ESCs, in terms of both hit probability across the *cis* chromosome (Figures 2A, 2C, and 2E, Figure S2A) and interacting domains in *cis* and in *trans* to the bait region (Figures 2B, 2D, and 2F, Figure S2B). Second, we examined chromatin interactions in pre-iPSCs that represent a late reprogramming stage at which the pluripotency expression and chromatin program is not yet fully induced (Sridharan et al., 2009). The long-range chromatin interactomes in these cells are typically distinct from those in ESCs and iPSCs, as well as those in MEFs (Figure 2, Figure S2), indicating that the reorganization of chromatin contacts is not complete at this late stage of reprogramming, in line with previous findings detailing the long-range interactions of the *Nanog* locus (Apostolou et al., 2013). Together, these data show that the large-scale changes in genome organization that

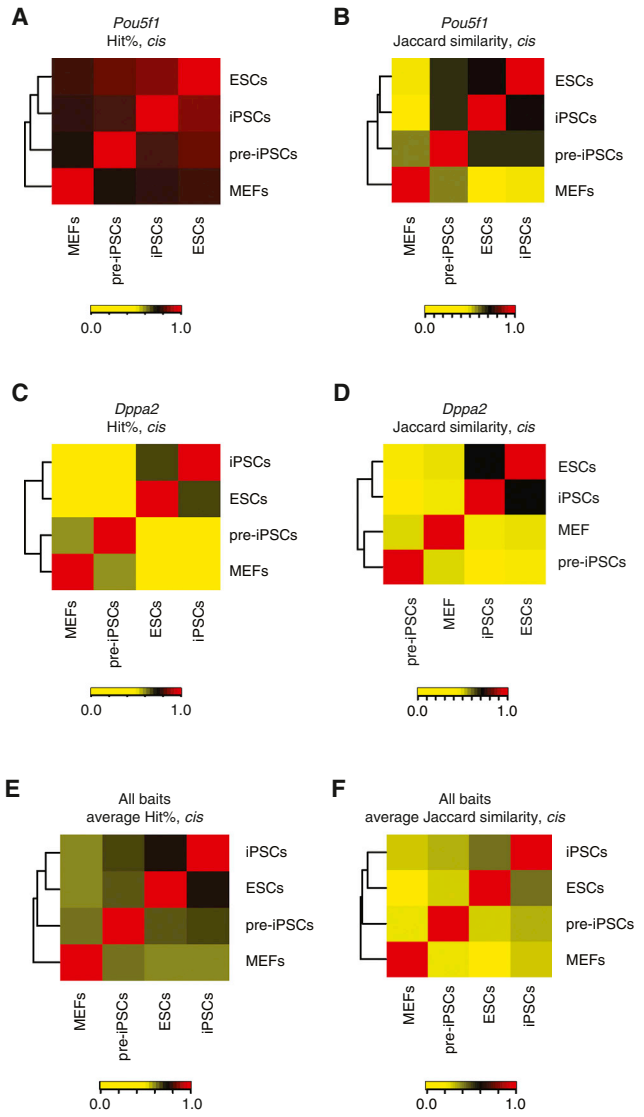


Figure 2. Long-Range Chromatin Contacts Change during Differentiation and Are Reset upon Reprogramming of Somatic Cells to iPSCs

(A) Unsupervised hierarchical clustering of Spearman rank correlation values of the Hit% within 200 kb windows along the *cis* chromosome (chr17) for *Pou5f1* 4C-seq-defined interactomes in ESCs, iPSCs, pre-iPSCs, and MEFs. Color key = Spearman's *rho* values.

(B) Unsupervised hierarchical clustering of Jaccard similarity coefficients for the overlap of interacting domains of the *Pou5f1* bait region in *cis* between ESCs, iPSCs, pre-iPSCs, and MEFs. Color key = Jaccard similarity values.

(C) As in (A), except for the *Dppa2* bait.

(D) As in (B), except for the *Dppa2* bait.

(E) As in (A), except averaged across eight different bait loci (*Pou5f1*, *Stk35*, *1700067P10Rik*, *Nfia*, *Dppa3*, *Rhbdd1*, *Hoxa10,* and *Dppa2*).

(F) As in (B), except averaged for the eight different bait loci named in (E).

See also Figure S2.

occur during differentiation are gradually reset to an ESC-like state during reprogramming to iPSCs, likely in association with progressive changes in chromatin and transcription states (Sridharan et al., 2009). The cell-type specificity of long-range chromatin interactions contrasts with the largely cell-type invariant TAD structure of metazoan genomes (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012), suggesting that the organization of long-range chromatin contacts is a key feature of cell identity.

Open/Closed Chromatin as the Foundation of Genome Organization in ESCs

The finding that changes in distal chromatin contacts occur during differentiation and reprogramming suggested a link between genome organization and genomic features that establish and maintain cell identity. To investigate such an association in the context of mouse ESCs, we compiled a compendium of genomic features available for this cell type (Table S6). Specifically, we considered the binding profiles of the basic transcriptional machinery (TBP, RNA polymerase II), coactivators (Mediator and p300), and architectural proteins (Cohesin and CTCF). We also included transcriptional regulators of three important gene-regulatory networks: the pluripotency TF network (Oct4, Sox2, Nanog, and Klf4), TFs that cooperate with cMyc (cMyc, Max, E2F4), and the repressive Polycomb protein network (Ring1b and Polycomb Repressive Complex 2 [PRC2]) (Young, 2011). Since regulatory genomic regions exhibit extensive co-occupancy by these factors, we grouped them into 11 groups (clusters) based on cobinding at a 1 kb resolution (Figure 3A, Experimental Procedures). We also took into account histone modifications and their combinatorial nature by summarizing the relationship of six histone marks in terms of four functionally distinct chromatin states (ChromHMM states) (Ernst and Kellis, 2012) that are associated with Polycomb repression, transcriptional elongation, enhancers, and promoters. Regions lacking these histone modifications were assigned to a "low signal" state (Figure 3B, Experimental Procedures). In addition, DNA replication timing, DNaseI hypersensitivity, gene density (in terms of transcriptional start sites [TSSs]), RNA-seq- based expression data, and LaminB association were considered.

Upon binning of the resulting 22 linear genomic data sets into 200 kb windows across the genome, principal components analysis (PCA) was used to reduce their high dimensionality (Experimental Procedures). The first three principal components (PCs) were retained for downstream analyses (Figure 3C). To explore the relationship between genomic features and genome organization, we next compared the linear genomic character represented by the PCs to the 4C interactome data for each of the 16 bait regions analyzed in ESCs (Figure S3A, Figure 1B, Table S1).

The first PC (PC1) captures 51% of the variance across all features (Figure 3C) and distinguishes open, accessible chromatin and closed, inaccessible chromatin (Figure 3D). Specifically, regions of the genome with positive PC1 scores are characterized by high gene density, DNaseI hypersensitivity, binding of Cohesin, the basic transcriptional machinery, and TFs, as well as active or Polycomb-repressed chromatin states within 200 kb windows. In contrast, regions of the genome with negative PC1 scores are strongly depleted for these features and instead are LaminB associated and replicate their DNA late in S phase (Figure 3D). Notably, the 1 Mb regions surrounding each bait locus had widely differing open/closed chromatin character as defined by their PC1 scores (Figure 3Ei, top panel). Strikingly, we found that the mean PC1 scores for the 1 Mb bait regions correlated strongly with the mean PC1 scores of the

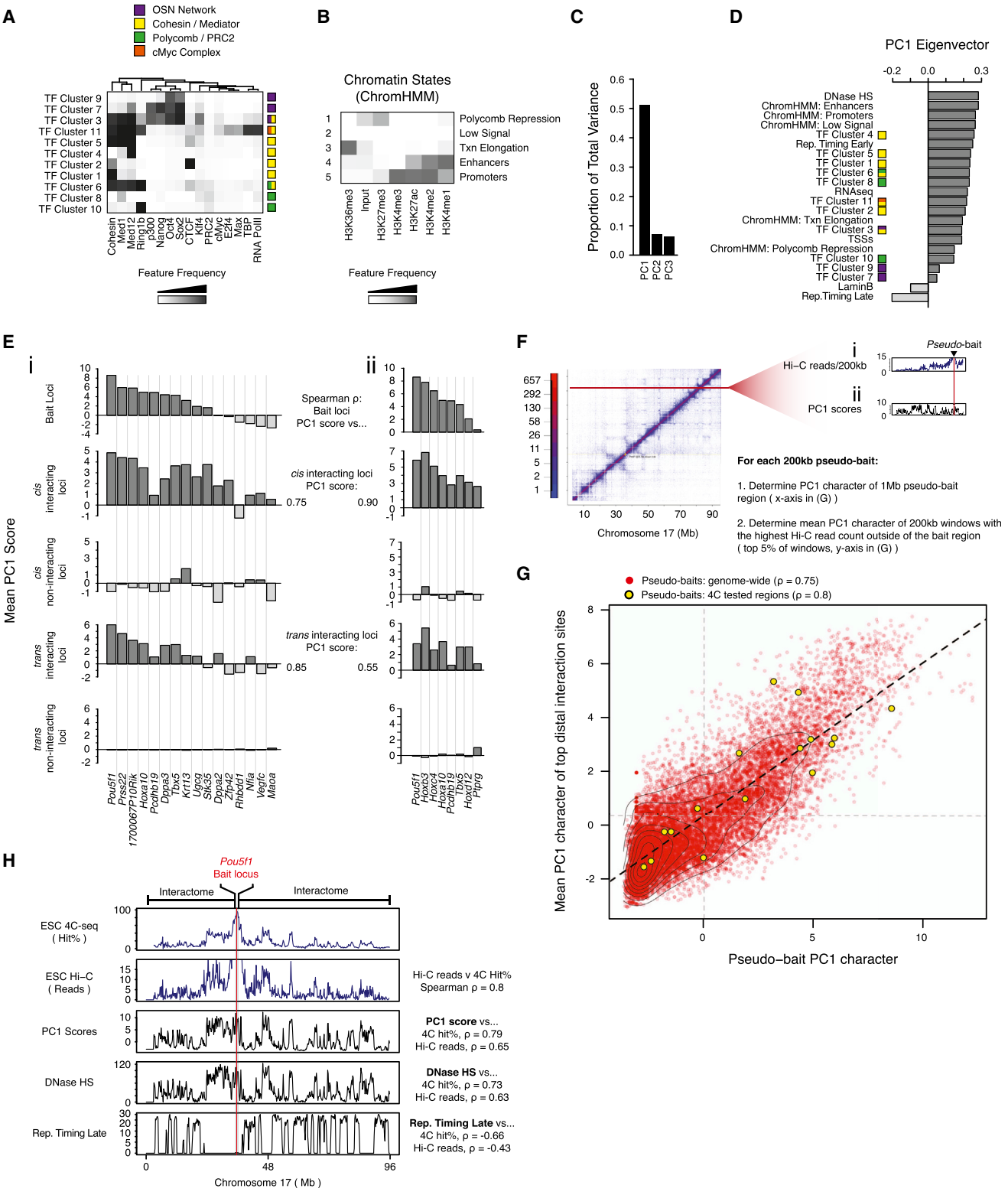


Figure 3. Interactions between Regions with Similar Open/Closed Chromatin Character Are an Intrinsic Aspect of Chromosome Conformation in Mouse ESCs

(A) Transcription factor (TF) clusters defined using k-means clustering at 1 kb resolution for noted mouse ESC data sets, annotated based on feature frequency as represented by the heatmap. Gray scale denotes the frequency with which a given factor is found at genomic positions corresponding to the cluster. The color legend identifies specific gene-regulatory networks.

(legend continued on next page)

corresponding interactomes in both *cis* and *trans* (Figure 3Ei, Spearman's *rho* = 0.75 and 0.85, respectively). That is, baits with highly open and accessible chromatin preferentially interact with regions of the genome with similarly high PC1 scores. By contrast, the interactomes of closed chromatin baits, i.e., those with negative PC1 scores, preferentially colocalize with genomic regions of similarly low PC1 scores in *cis* and *trans*. These findings were confirmed by a 4C-seq analysis for a partially overlapping set of bait regions in a second, independent ESC line (Figure 3Eii).

To test whether this trend holds genome-wide, we analyzed a previously published Hi-C data set for genome-wide chromatin interactions in mouse ESCs (Dixon et al., 2012) in a similar manner (Experimental Procedures). For the Hi-C-based analysis, each 200 kb region of the genome was treated as a "pseudo-bait" and its intrachromosomal interactome was extracted (Figure 3F). Genome-wide, we found a striking positive correlation between the mean PC1 scores of the extended 1 Mb pseudo-bait regions and the mean PC1 scores of their most frequently interacting intrachromosomal sites (Figure 3G), corroborating and extending our 4C-seq-based findings. A comparison of the profiles of the most and least likely interacting regions further demonstrated that genomic regions with very open, high PC1 character, for instance the *Pou5f1* region, interact extensively with most other genomic regions with similarly high PC1 scores on the chromosome in *cis* and do not interact with genomic regions of negative PC1 scores (Figures 3G and 3H, Figure S3Ci). In contrast, genomic regions of closed chromatin character, defined by negative PC1 scores, avoid regions of high PC1 character and appear to interact selectively with only a subset of PC1-negative genomic regions (Figure S3Ci). This may reflect a tethering of closed chromatin regions to the nuclear lamina, thus limiting their sampling of distal interactions. The positive correlation between bait and interactome character in terms of their PC1 scores persisted when the 10 Mb surrounding the bait region were excluded (Figure S3Di), demonstrating that interacting regions far away from the bait in *cis* also show an association with chromatin of similar open/closed character.

Taken together, we conclude that genomic loci with similar PC1 characters preferentially interact (or colocalize) within the 3D space of the ESC nucleus both in *cis* and in *trans*. The Hi-C-based results indicate that these associations are a general, genome-wide feature of long-range chromatin interactions. Our findings suggest that the strong interaction preferences between regions of similar PC1 character are an intrinsic aspect of overall chromosome conformation, in line with previous findings (Imakaev et al., 2012; Lieberman-Aiden et al., 2009; Simonis et al., 2006). The data further demonstrate that interaction preferences in ESCs are not accurately described by a binary model of spatial segregation between open and closed chromatin states into two genome-wide compartments, where genomic regions with open chromatin character may colocalize with any other open chromatin region and vice versa (Lieberman-Aiden et al., 2009). Instead, the ESC interactome follows a more graduated model, where highly open chromatin regions predominantly interact with regions of similarly high PC1 scores, mid PC1 regions predominantly interact with other mid to low PC1 regions, and negative PC1 regions predominantly interact with other closed chromatin regions, supporting findings in other cell types (Imakaev et al., 2012). Notably, PC1 scores across the ESC genome are continuous (Figure S3B) and correlate with the continuum of ESC contact frequencies.

Genomic Regions Enriched for Oct4/Sox2/Nanog and Polycomb Proteins Frequently Colocalize in ESCs

Next, we wanted to explore the extent to which long-range chromatin interactions are associated with specific transcriptional networks, beyond their association with the large-scale open/closed chromatin properties demonstrated by the correlation of interactomes with the genomic PC1 character. Since PC1 positive regions reflect enrichment for multiple features with diverse functionalities, we reasoned that an interactome's PC1 character may not reflect any specific mechanistic role of PC1-enriched features, but is most likely a consequence of the overarching chromosome conformation framework (Figures S4A–S4D). Therefore, we considered the second and third principal components (PC2 and PC3), which account for 7% and 6% of

(B) Chromatin states were determined based on the six indicated histone modifications in ESCs by a multivariate hidden Markov model, at 200 bp resolution (Ernst and Kellis, 2012). Gray scale denotes the frequency with which a given histone mark is found at genomic positions corresponding to the chromatin state.
(C) PCA was performed on ESC chromatin states and TF clusters from (A) and (B), RNA-seq expression data, DNaseI hypersensitivity (HS), LaminB binding, early and late DNA replication timing (Rep. timing), and density of transcriptional state sites (TSSs) upon binning of the linear genomic data into 200 kb windows across the genome. Proportion of total variance in genomic features described by each principal component is shown.
(D) PC1 eigenvector ranked by genomic feature contribution.
(E) (Ei) Top to bottom: mean PC1 score within the 1 Mb bait region centered on each listed bait's locus; interacting regions in *cis*; noninteracting regions in *cis*; interacting regions in *trans*; and noninteracting regions in *trans*. Spearman's *rho* values give the rank correlation between the mean PC1 score of the 1 Mb bait regions and their interactomes in both *cis* and *trans*. (Eii) Identical analysis to (Ei) with a partially overlapping set of baits, for an independently derived ESC line, which is discussed in Figure 7 as $Eed^{+/+}$ ESC line.
(F) Schematic of genome-wide, Hi-C-based, pseudo-4C analysis. (Fi) Each extracted row of the Hi-C contact matrix, adapted from (Dixon et al., 2012), represents the interactome of one pseudo-bait, at 200 kb resolution. (Fii) Plot of the PC1 character of the same chromosome. For each 200 kb pseudobait, the mean PC1 score within the extended 1 Mb pseudo-bait region (Pseudo-bait PC1 character) and the mean PC1 score within the pseudo-bait's interactome (top 5% of 200 kb windows ranked by reads and excluding the 1 Mb pseudo-bait region) were obtained and plotted as a red point in the scatterplot shown in (G). Pseudo-bait regions corresponding to genomic regions that we used as baits in our 4C-seq analysis in (Ei) are plotted in yellow (4C-bait loci).
(G) Result of the analysis described in (F). 4C-bait loci show a similar trend when analyzed based on Hi-C data as in (Ei) based on 4C-seq. Correlations between bait and interactome PC1 scores are noted. The Hi-C data are also summarized by the regression line in black, and the mean bait and interactome PC1 scores are demarcated by vertical and horizontal gray lines, respectively. Contour lines represent data density.
(H) Top to bottom: comparison of the *Pou5f1* 4C-seq-based *cis*-interactome (Hit%), the *Pou5f1 cis*-interactome defined by Hi-C read counts, PC1 scores, DNaseI HS, and late DNA replication timing along chr17. Specific correlation values are specified.
See also Figures S3 and S4.

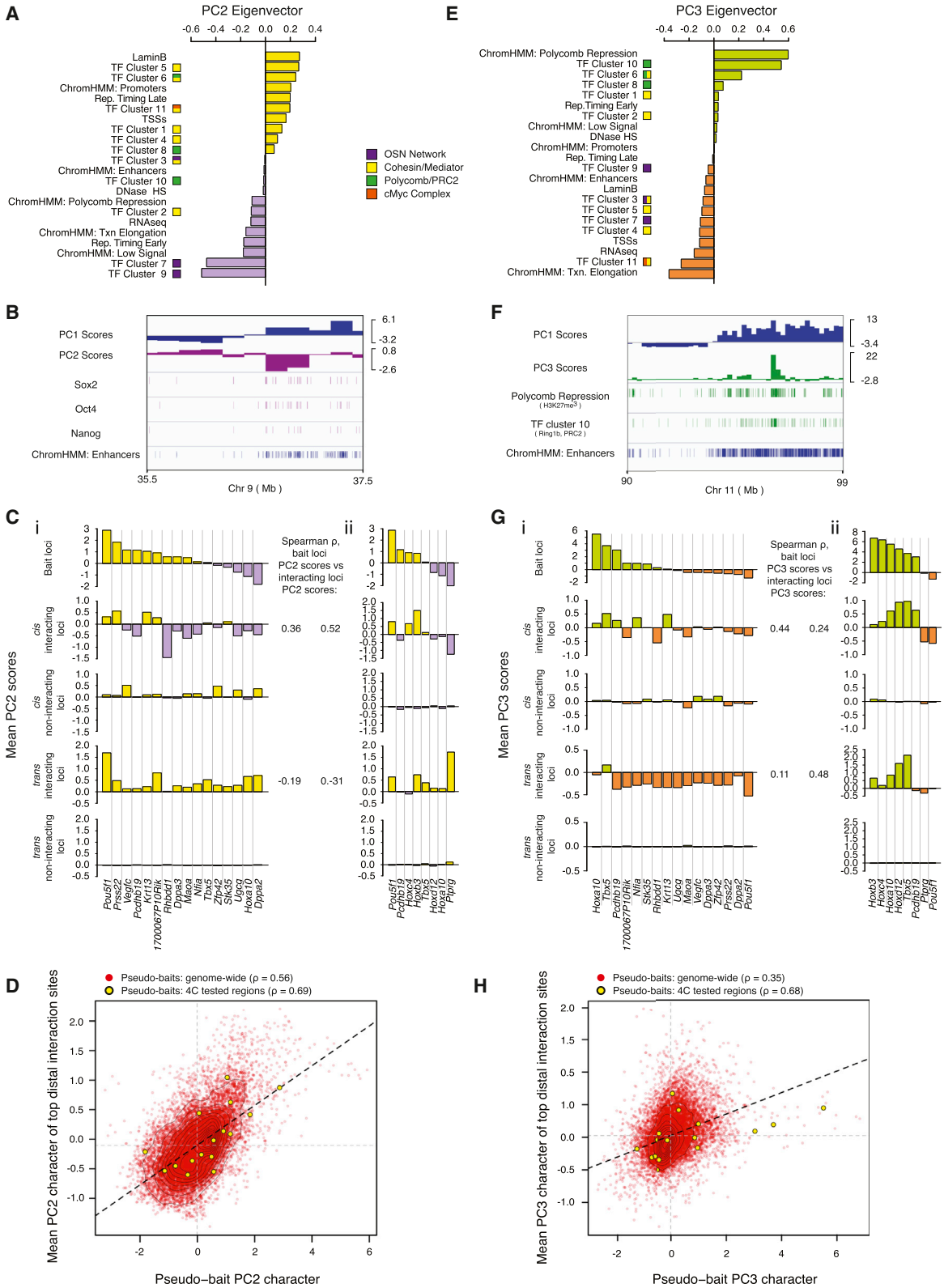


Figure 4. Regions of Shared Transcriptional Network Occupancy Preferentially Interact

(A) PC2 eigenvector with individual feature contributions. TF clusters and chromatin states are as in Figure 3.

(B) Integrative Genomics Viewer tracks of a representative genomic region with (top to bottom) PC1 and PC2 scores; Sox2, Oct4, and Nanog occupancy; and enhancer density.

(legend continued on next page)

the variance in genome features, respectively (Figure 3C). Importantly, these components capture regions of the genome enriched for previously described gene-regulatory networks in ESCs (Figure 4A and 4E). Specifically, genomic regions with positive PC2 scores are exceptionally enriched for Mediator and Cohesin binding, as captured by TF clusters 5, 6, and 11, but also late-replicating and strongly LaminB-bound across 200 kb, while regions with negative PC2 scores are enriched for binding of factors belonging to the pluripotency network, including Nanog, Oct4, and Sox2, as well as p300, which colocalize as TF clusters 7 and 9 (Figure 4A). Figure 4B displays the relationship between PC1 and negative PC2 scores in terms of these features: whereas genomic regions of open chromatin (defined by positive PC1 values) generally have a higher level of pluripotency factor binding than those with negative PC1 scores, genomic regions with negative PC2 scores have an increased density of pluripotency factor co-occupancy over and beyond what is explained by their PC1 state.

Conversely, regions of the genome with positive PC3 values are highly enriched for occupancy by the Polycomb protein complexes PRC1 (Ring1b) and PRC2, as well as histone H3K27me3. Negative PC3 regions, on the other hand, seemingly capture the recently described super-enhancers associated with highly transcribed genes in pluripotent cells (Whyte et al., 2013) and represent a strong enrichment for active transcriptional elongation along with dense occupation by Mediator, Cohesin, the cMyc complex, and the pluripotency TFs Oct4, Sox2, and Nanog (Figure 4E, Figures S4E and S4F). Genomic regions with positive or negative PC3 scores have an elevated density of their characteristic features above what is explained by their PC1 scores (Figure 4F, data not shown). We conclude that PC2 and PC3 describe a finer level of chromatin structure associated with specific TF- and chromatin-regulatory networks that is not captured by the open and closed chromatin character defined by PC1.

To examine whether the genomic states captured by PC2 and PC3 are associated with long-range chromatin contacts, we examined the PC2 and PC3 scores within our 16 ESC bait regions and their respective interactomes in an analogous manner to our analysis of PC1 in Figure 3. Importantly, our bait regions exhibited widely different PC2 characters, and the mean PC2 score of each bait's intrachromosomal interactome showed a strong concordance with the mean PC2 score of the 1 Mb bait region itself (Figure 4Ci). This observation was confirmed for additional bait regions in a second ESC line (Figure 4Cii) and, remarkably, also for *cis* interactions genome-wide based on Hi-C data (Figure 4D, Figures S3Cii and S3Dii). The PC2-based analysis demonstrates that genomic regions with particularly strong Mediator, Cohesin, and Lamin binding preferentially colocalize in *cis* as do regions extremely highly enriched for Oct4, Sox2, and Nanog binding without a concomitant extreme accumulation of Mediator and Cohesin (captured by TF clusters 7 and 9). Notably, however, the colocalization of pluripotency-factor-enriched, PC2-negative genomic regions was not evident for *trans* interactions (Figure 4C).

The comparison of bait and interactome PC3 scores demonstrated a preferential colocalization of Polycomb/H3K27me3-enriched, PC3-positive genomic regions (captured by TF clusters 6, 8, and 10 and the Polycomb Repression ChromHMM state), both in *cis* and in *trans* (Figure 4Gi). For example, *Hox* loci, which belong to the most strongly Polycomb-occupied and H3K27me3-enriched (PC3-positive) regions in the ESC genome, were found to interact with other regions characterized by high H3K27me3/Polycomb occupancy in *cis* and *trans* (Figure 4G). The colocalization of distal genomic regions extensively occupied by Polycomb proteins and H3K27me3 may be analogous to the frequent association of Polycomb response elements in *Drosophila* (Bantignies and Cavalli, 2011) and may therefore represent an evolutionarily conserved feature of genome organization linked to gene regulation. Conversely, bait loci with negative PC3 values that are strongly enriched for ESC super enhancers colocalize with genomic regions of similar PC3-negative character in both *cis* and *trans* (Figure 4Gi). These trends were confirmed in a second ESC line (Figure 4Gii) and extended to Hi-C data (Figure 4H, Figures S3Ciii and S3Diii), showing them to be a genome-wide phenomena.

Together, these results argue that regions of the genome enriched for specific gene-regulatory features preferentially colocalize within the 3D space of the nucleus, raising the possibility that specific transcriptional and chromatin-regulatory networks are involved in mediating long-range chromatin contacts in ESCs.

Spatial Segregation of Nanog and H3K27me3 in the ESC Nucleus

The preferential interactions of genomic regions with positive and negative PC3 scores (i.e., Polycomb protein versus super enhancer/Oct4/Sox2/Nanog-enriched regions) suggested a segregation of genomic regions with opposing PC3 character into distinct compartments in the nucleus. To test this hypothesis, we examined the colocalization of Nanog (PC3-negative), RNA polymerase II (PC3-negative), and H3K27me3 (PC3-positive) in the ESC nucleus by immunostaining. Image analysis showed that Nanog and RNA polymerase II have a localization pattern distinct from that of H3K27me3,

(C) (Ci) Top to bottom: mean PC2 score within the 1 Mb region centered on each bait's locus; interacting regions in *cis*; noninteracting regions in *cis*; interacting regions in *trans*; and noninteracting regions in *trans*. Spearman's *rho* values give the correlation between the baits' and the interactomes' PC2 character across all analyzed baits in *cis* and *trans*. (Cii) Identical analysis to (Ci) except for an independently derived ESC line discussed in Figure 7 as $Eed^{+/+}$ ESCs, with a partially overlapping set of bait loci.
(D) Genome-wide pseudo-4C analysis of Hi-C data as described in Figure 3G, except for PC2.
(E) PC3 eigenvector with individual feature contributions.
(F) Integrative Genomics Viewer tracks showing a representative genomic region with (tom to bottom) PC1 and PC3 scores; H3K27me3 and TF cluster 10 (Ring1b/PRC2) enrichment; and enhancer density.
(G) As in (C), except for PC3 scores.
(H) As in (D), except for PC3 scores.
See also Figures S3 and S4.

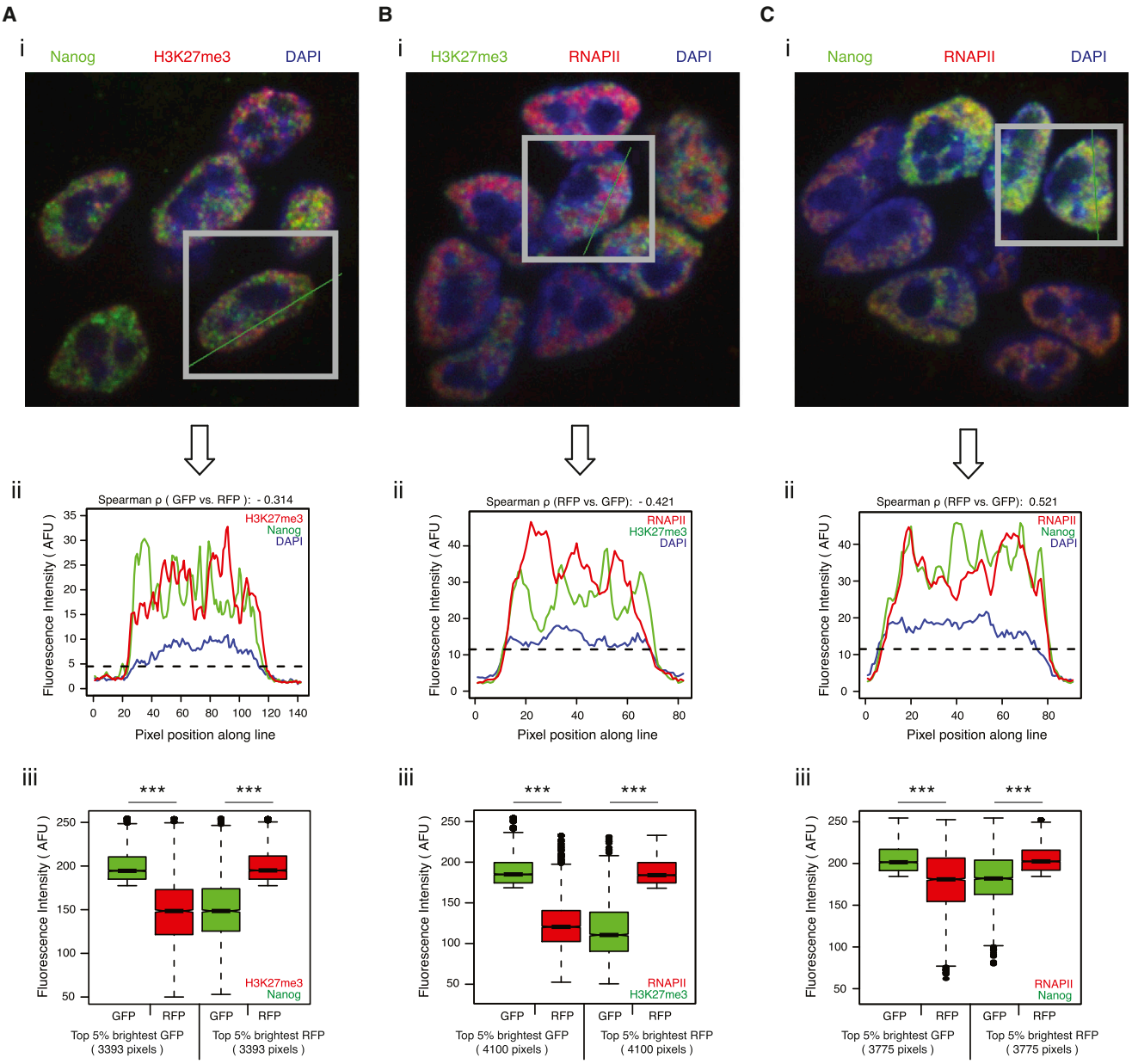


Figure 5. Nanog and H3K27me3 Segregate in the ESC Nucleus

(A) (Ai) Image of ESCs immunostained with antibodies against Nanog (green) and H3K27me3 (red). Nuclei were stained with DAPI (blue). (Aii) Red and green pixel intensities along the line in (Ai) for all pixels whose DAPI signal was above the indicated threshold (dotted line). (Aiii) Quantile normalized fluorescence intensity distribution of the top 5% brightest nuclear (but nonnucleolar) green pixels (GFP) and the normalized red pixel intensity (RFP) at the corresponding position (left), and that of the top 5% brightest nuclear (but nonnucleolar) red pixels and the normalized distribution of green pixel intensity at the corresponding position (right), for all cells in the ESC colony in (Ai). Box and whisker demarcations are as in Figure 1C. ***$p < 2e^{-16}$, Wilcoxon rank-sum test.

(B) As in (A), except for H3K27me3 (green) and RNA polymerase II (red).

(C) As in (A), except for Nanog (green) and RNA polymerase II (red).

with sites strongly enriched for H3K27me3 displaying weak Nanog and RNA polymerase II accumulation and vice versa (Figures 5A and 5B). By contrast, RNA polymerase II and Nanog overlap more extensively, albeit not perfectly (Figure 5C), consistent with their differential contribution to PC2 and PC3 scores. For instance, Nanog occupancy is strongly captured by TF cluster 7 and RNA polymerase binding by TF cluster 11, which have similar contributions to PC3, but opposing contributions to PC2 (Figures 4A and 4E). Overall, the immunofluorescence localization patterns support the spatial segregation of functionally distinct TF- and chromatin-regulatory networks in the ESC nucleus and are consistent with the colocalization of distinct gene-regulatory modules detected by our 4C-seq-based analysis.

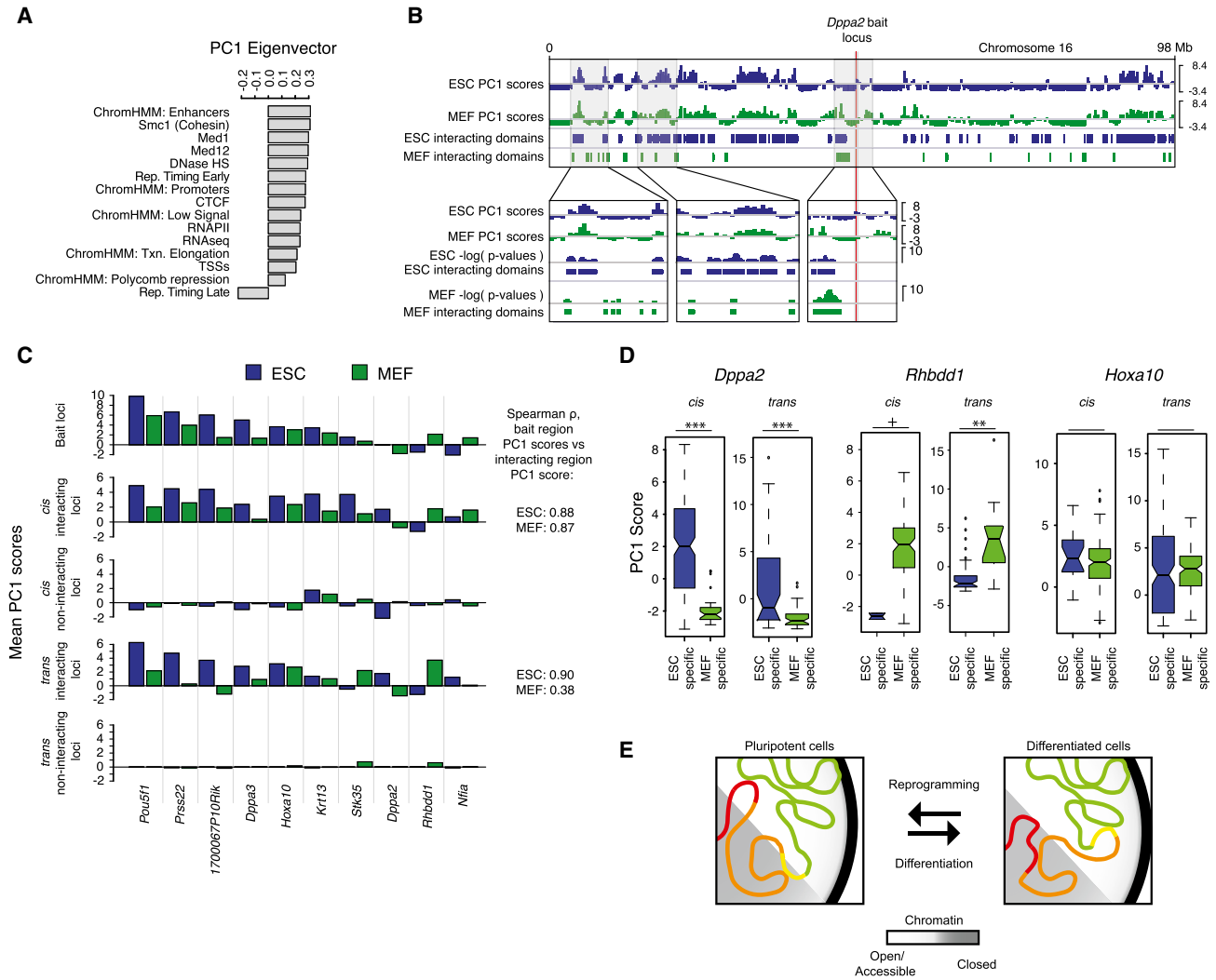


Figure 6. Changes in Open/Closed Chromatin Character between ESCs and MEFs Correspond to Changes in Interaction Preferences

(A) PCA was performed for concatenated ESC+MEF data and included the indicated features. The resulting PC1 eigenvector is depicted.

(B) Top to bottom: Integrative Genomics Viewer tracks showing the PC1 scores along chr16 in ESCs (blue) and MEFs (green) and the 4C-seq-defined *cis*-interacting domains of the *Dppa2* locus in ESCs and MEFs. Zoom-ins highlight the switch in PC1 bait character of the *Dppa2* locus between MEFs and ESCs (right box) and corresponding changes in interaction preferences (left two boxes).

(C) Top to bottom: mean PC1 score within the 1 Mb region centered on each listed bait's locus; interacting regions in *cis*; noninteracting regions in *cis*; interacting regions in *trans*; and noninteracting regions in *trans*, for ESC (blue) and MEF data (green). Spearman's *rho* values give the rank correlation between the PC1 bait character and interactome character per cell type, across all analyzed baits in *cis* and *trans*.

(D) PC1 score distributions of ESC- and MEF-specific, significantly interacting domains of *Dppa2*, *Rhbdd1*, and *Hoxa10*. Box and whisker demarcations are as in Figure 1C, with notches ≈95% confidence interval around medians. $^{+}p = 0.099$; $^{**}p < 0.01$; $^{***}p < 0.001$; Wilcoxon rank-sum test.

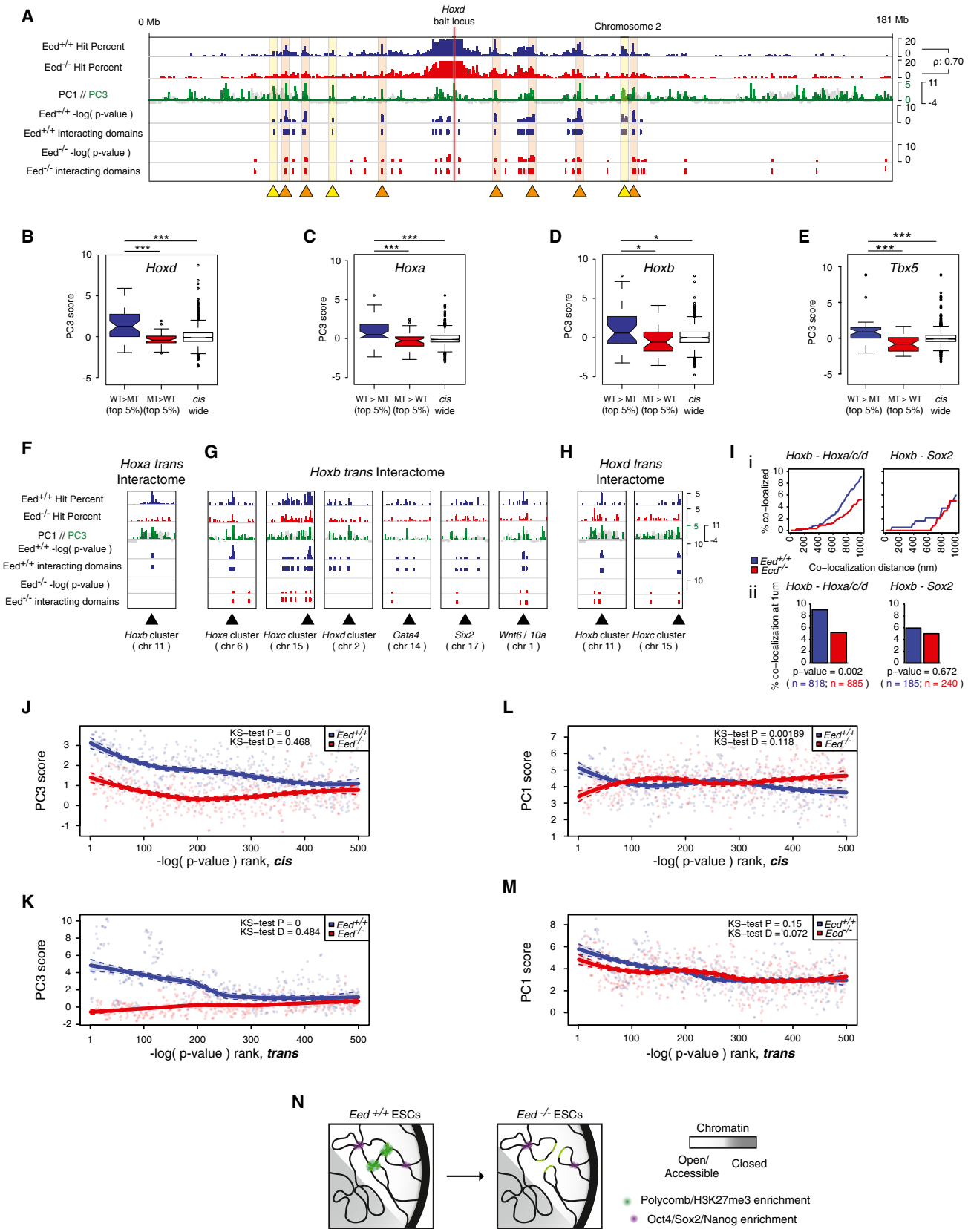
(E) Chromatin interaction model, wherein large-scale changes in chromatin interactions mirror changes in open/closed chromatin (PC1) character upon ESC differentiation or reprogramming to pluripotency. Gray scale reflects the continuum between open/accessible chromatin (light) and closed chromatin (dark). Note, for instance, the different interactions of the genomic regions marked in red and yellow, switching between open and closed chromatin.

See also Figure S5.

Changes in Open/Closed Chromatin Character Mirror Changes in Genome Organization during Differentiation

Our analysis revealed a close relationship between chromatin character and spatial interactions, with the combination of linear genomic features summarized by PC1 (open/closed chromatin) showing the strongest association with the fundamental organization of chromatin interactions. Based on these results, we predicted that dramatic changes in open/closed chromatin character that occur during differentiation should coincide with strong changes in interactome character.

To test this, we examined differences in chromatin interactions between ESCs and MEFs with respect to changes in their open/closed chromatin character. To this end, we performed PCA on concatenated ESC and MEF genomes using linear genomic feature data sets that were available for both cell types but did not include cell-type-specific TFs (Figure 6A, Table S6,



(legend on next page)

Experimental Procedures). This resulted in a new PC1 eigenvector that allowed the comparison of PC1 scores for ten bait regions and their interactomes across both cell types (Figures S5A–S5D). We found that, as in ESCs, regions with similar PC1 character in MEFs preferentially colocalize in both *cis* and *trans*. (Figure 6C). Furthermore, changes in bait character between MEFs and ESCs are generally associated with similar changes in their respective interactomes (Figures 6B–6D). For instance, the *Dppa2* locus participates in extensive interactions with genomic regions of positive PC1 character in ESCs, in accordance with its early replicating and highly transcribed state in pluripotent cells (Takebayashi et al., 2012). In MEFs, the *Dppa2* bait region displays a PC1-negative, repressed, late replicating state, and the MEF-specific interactions likewise exhibit negative PC1 scores, both in *cis* and in *trans* (Figures 6B–6D, Figure S5E). In contrast, the *Rhbdd1* bait region transitions from negative to positive PC1 scores from ESCs to MEFs, and its MEF-specific interactions have significantly higher PC1 scores than the ESC-specific interactions (Figures 6C and 6D). In addition, bait regions without a change in PC1 character do not change interaction preferences with regards to PC1 scores (Figures 6C and 6D, *Hoxa10*).

Together, these data support a model where open/closed chromatin character is the strongest predictor of interaction preferences between distal genomic loci, and long-range chromatin interaction preferences are subject to change during differentiation in concordance with changes to the PC1 nature of the regions in question (Figure 6E). Of note, although the *Pou5f1* gene itself becomes repressed during differentiation (Feldman et al., 2006), the extended 1 Mb *Pou5f1* bait region is strongly positive in PC1 character even in MEFs (Figure 6C), likely explaining the more limited difference in spatial interactions between ESCs and MEFs for this bait region compared to the *Dppa2* bait (Figure 2), which shows a more dramatic change in PC1 character upon differentiation.

The Preferential Colocalization of Polycomb-Enriched Genomic Regions Is Eed Dependent

Our data demonstrated that genomic regions are more likely to contact each other when they share strong enrichment of similar regulatory proteins (Figure 4). Therefore, we considered testing the functional importance of a specific gene-regulatory network for long-range chromatin interactions. To this end, we determined the long-range chromatin contacts in ESCs lacking the protein Eed, a subunit of Polycomb complex PRC2 that is required for all genomic H3K27me3 (Montgomery et al., 2005). We examined particularly the interactions of bait regions with positive PC3 scores, which capture high occupancy by PRC2 and H3K27me3 (Table S1). Importantly, despite the complete loss of H3K27me3 (Figure S6A), $Eed^{-/-}$ ESCs continue to express pluripotency-specific TFs including Oct4, Sox2, and Nanog at normal levels, maintain their ability to self-renew, and do not spontaneously differentiate when cultured appropriately (Chamberlain et al., 2008), allowing us to test the role of PRC2 in genome organization without a change in cell identity.

Notably, 4C-seq analysis showed that the intrachromosomal long-range interactions of the *Hoxd* cluster, a Polycomb-targeted genomic region with highly positive PC3 scores, correlated strongly between the $Eed^{+/+}$ and the $Eed^{-/-}$ ESC lines (Figure 7A). However, despite this overall similarity, numerous intrachromosomal interactions present in $Eed^{+/+}$ ESCs are lost or have less significant interactions by p value, indicative of a reduced interaction frequency, in $Eed^{-/-}$ ESCs (Figure 7A, yellow and orange highlights, respectively). A visual inspection of chromatin contacts indicated that these losses and reductions appear to occur at regions of high Polycomb enrichment in wild-type ESCs, as defined by high PC3 scores (Figure 7A). In

Figure 7. *Eed* Is Required for the Colocalization of Polycomb-Occupied Genomic Regions

(A) Integrative Genomics Viewer tracks showing the *Hoxd12* interactome in *cis* in terms of Hit% for $Eed^{+/+}$ (blue) and $Eed^{-/-}$ (red) ESCs, chromosome-wide PC1 scores (black) overlaid with PC3 scores (green, positive values shown only), binomial test −log(p values), and interacting domains. Regions that lose significant interactions with the *Hoxd12* bait upon *Eed* ablation are marked with yellow triangles and shading; those that do not lose interactions with the *Hoxd12* locus upon *Eed* ablation but show a decrease in interaction strength are marked with orange triangles and shading. The Spearman's *rho* value shows the rank correlation of the Hit% between $Eed^{+/+}$ and $Eed^{-/-}$ ESCs.

(B) *Hoxd12* 4C-seq Hit% tracks in $Eed^{+/+}$ and $Eed^{-/-}$ ESCs from (A) were subtracted and the 200 kb windows with the top and bottom 5% of resulting values were used to define regions of the *cis* chromosome that showed stronger interactions in $Eed^{+/+}$ (WT > MT) and $Eed^{-/-}$ (MT > WT) ESCs, respectively. The PC3 score distributions of these genomic regions and of the entire chromosome are shown. Box and whisker demarcation are as in Figure 1C, with notches ≈95% confidence interval around medians. ***p < 0.001, Wilcoxon rank-sum test.

(C–E) As in (B), but for the *Hoxa10*, *Hoxb3*, and *Tbx5* 4C-seq *cis* interactomes. *p < 0.05.

(F–H) *Trans* interactions between the indicated (PC3-positive) *Hox* loci in $Eed^{+/+}$ and $Eed^{-/-}$ ESCs, displayed as in (A). *Hoxa* refers to results of the *Hoxa10* bait locus; *Hoxb*, to *Hoxb3*; *Hoxc*, to *Hoxc4*; and *Hoxd*, to *Hoxd12*.

(I) DNA FISH analysis of the *trans* interactions between *Hox* clusters. (Ii) Cumulative frequency distribution plots of colocalization frequencies between *Hoxb3* (chr11) and the other three *Hox* loci (*Hoxa10-chr6*, *Hoxc4-chr15*, *Hoxd12-chr2*) (left), as well as between *Hoxb3* and the *Sox2* (chr3) locus (right), with colocalization distances noted on the x axis, measured in $Eed^{+/+}$ (blue) and $Eed^{-/-}$ (red) ESCs. (Iii) Colocalization frequencies at 1 μm for *Hoxb3* and the other *Hox* loci (left), as well as for *Hoxb3* and *Sox2* (right), derived from (Ii). n = FISH signal pairs analyzed in both (Ii) and (Iii); p value from two-tailed Fisher exact test.

(J) The *cis* interactomes of the six PC3-positive (Polycomb/H3K27me3 enriched) bait loci (*Hoxa10*, *Hoxb3*, *Hoxc4*, *Hoxd12*, *Pcdhb19*, and *Tbx5*; see Figure 4Gii) were ranked by −log(p value) for both $Eed^{+/+}$ (blue) and $Eed^{-/-}$ (red) ESCs, and the 500 top genomic sites were plotted against their average PC3 scores in wild-type ESCs. Loess regression was used for curve generation. KS, Kolmogorov-Smirnov test to determine the probability that the two underlying probability distributions differ (D = KS − test D statistic).

(K) As in (J), but for the *trans* interactomes.

(L) As in (J), except for PC1 scores.

(M) As in (K), except for PC1 scores.

(N) Chromatin interaction model wherein in the absence of Eed, the frequency of interactions between regions with high PC3 scores is reduced, but large-scale chromosome conformation is largely conserved. Gray scale reflects the regions of the genome that are more open/accessible (light) versus more closed (dark). See also Figure S6.

agreement with this, long-range chromatin interactions of the *Hoxd* cluster that specifically occur in $Eed^{+/+}$ ESCs have a significantly more positive PC3 character than those that are specific for $Eed^{-/-}$ ESCs (Figure 7B). This result extends to other Polycomb-regulated regions such as the *Hoxa* and *Hoxb* clusters and the *Tbx5* locus (Figures 7C–7E, Figure S6D).

A comparison of chromatin interactions in *trans* between $Eed^{+/+}$ and $Eed^{-/-}$ ESCs revealed that the *Hox* clusters interact with each other as well as with other regions of high Polycomb/H3K27me3 enrichment that encode developmental regulators in wild-type ESCs (Figures 7F–7H). Importantly, many of these interchromosomal contacts are also lost or reduced in the absence of Eed (Figures 7F–7H). For instance, interactions of the *Hoxb* cluster with the *Hoxa*, *Hoxc*, and *Hoxd* clusters are observed in wild-type ESCs and are diminished in knockout cells (Figures 7F–7H). We also found that the colocalization frequency between the *Hoxb* cluster and the *Hoxa*, *Hoxc*, and *Hoxd* clusters in wild-type ESCs was significantly higher than in knockout ESCs when measured by FISH (Figure 7I, Figures S6B and SBC), which is consistent with our 4C-seq results.

To explore whether the absence of Eed specifically affects chromatin contacts that occur between genomic regions characterized by positive PC3 scores, we examined the chromatin character of the most significantly interacting distal regions across six PC3-positive baits (Table S1). We found that in $Eed^{+/+}$ ESCs, the highest-ranking (and likely most frequent) interactions fall within genomic regions with highly positive PC3 and PC1 scores and with less strongly positive PC2 scores both in *cis* and in *trans* (Figures 7J–7M, Figure S6E). In the absence of Eed, the strongest chromatin contacts no longer occur with regions that are highly positive for PC3 in wild-type ESCs (Figures 7J and 7K), although they still take place between distal genomic regions of similarly positive PC1 and PC2 scores both in *cis* and in *trans* (Figures 7L and 7M, Figure S6E). The corollary to this finding is demonstrated by two bait regions that are not enriched for Polycomb binding (*Pou5f1* and *Ptprg*): they do not show a similarly dramatic difference in interaction preferences between $Eed^{+/+}$ and $Eed^{-/-}$ ESCs with regards to any of the three principal components (Figures S6F–S6I). These data indicate that loss of PRC2 and H3K27me3 specifically alters the coassociation of PC3-positive genomic regions but does not dramatically affect spatial interactions associated with PC1 and PC2 character.

Based on these data we conclude that Eed is required for the establishment and/or stable maintenance of interchromosomal and intrachromosomal chromatin interactions between Polycomb-occupied, PC3-positive regions in ESCs. Our data also suggest that the overall chromosome topology does not dramatically change upon loss of Eed (Figure 7A). The results indicate that regions that are spatially interacting and Polycomb protein-bound in wild-type ESCs remain confined by a similar chromosome topology in the absence of Eed, but their interaction frequency, i.e. their proximity, is dramatically reduced (Figures 7F–7H, model in Figure 7N).

DISCUSSION

Our work describes a pluripotency-specific organization of the mouse genome and suggests that distal regions of the genome bound by similar regulatory proteins colocalize within the 3D space of the ESC nucleus. Based on our data, we propose a model with two layers of regulation for long-range chromatin contacts in ESCs. (1) We posit that, at the largest scale, the open/closed chromatin character (described by the PC1 character of the genome) defines the regions of the genome that have the potential to come into close spatial proximity with one another, both intrachromosomally and interchromosomally, which is intricately linked to the overall folding of the chromosome and in agreement with other recent findings (Imakaev et al., 2012; Lieberman-Aiden et al., 2009). (2) Our data also suggest that on a finer scale, and within the constraints established by the open/closed chromatin architecture, genomic regions are more likely to contact each other when they share strong enrichment of similar regulatory proteins, such as binding by the pluripotency TFs Oct4, Sox2, and Nanog with Mediator and Cohesin (represented by negative PC3 scores) or binding of the repressive Polycomb complex (captured by positive PC3 scores).

Our results demonstrate that the depletion of a single gene-regulatory network in ESCs specifically affects long-range interactions of genomic regions particularly strongly enriched for occupancy by this network (in our case PRC2 and H3K27me3) without altering the global interaction network associated with open/closed chromatin character. A potential explanation for the limited effect on overall chromatin interactions may be that the chromosomal conformation chassis is maintained by many combinatorially acting regulatory factors that probably involve numerous interactions mediated by Cohesin (Apostolou et al., 2013; Phillips-Cremins et al., 2013). Notably, in addition to our description of PRC2 as a critical regulator of specific long-range chromatin interactions (Figure 7), a functional requirement for the TF Klf4 in the maintenance of long-range chromatin contacts in ESCs has recently been reported (Wei et al., 2013), extending previous findings that demonstrated a requirement for Oct4 in the organization of short-range chromatin interactions within the *Nanog* locus (Levasseur et al., 2008).

The role of cell-type-specific gene-regulatory networks in defining specific long-range chromatin interactions potentially allows TADs enriched for specific gene-regulatory features to colocalize in the 3D space of the nucleus. Interestingly, while our results demonstrate that Polycomb complexes are important for long-range chromatin contacts between Polycomb targets in mammals, it has recently been shown that the TAD structure within a specific locus, the X chromosome inactivation center, is not affected by the *Eed* knockout (Nora et al., 2012), indicating different regulatory mechanisms at the different hierarchies of genome organization. It is also interesting to speculate that the interaction between the *Hoxd* and *Hoxc* clusters may provide a mechanism for how the noncoding RNA HOTAIR, encoded within the *Hoxc* cluster, finds its target genes within the *Hoxd* cluster located on a different chromosome (Rinn et al., 2007): by exploiting 3D conformation of the genome in a manner similar to that employed by the long-noncoding RNA *Xist* (Engreitz et al., 2013).

Based on the data presented here and on other reports (Hakim et al., 2011; Noordermeer et al., 2011), we suggest that cell-type-specific gene-regulatory networks generally guide specific spatial interactions within the context of a cell-type-specific chromosome folding pattern that relates to the open/closed

chromatin state. A preferential colocalization of distal genomic regions with similar regulatory networks tens of Mb apart on the same chromosome and in *trans* suggests a previously unappreciated role for transcriptional networks in influencing the 3D positional preferences of chromatin in mammalian cells, which may represent an evolutionarily conserved interaction between eukaryotic genome organization and gene regulation (Sexton et al., 2012; Tanizawa et al., 2010). We speculate that this organizational hierarchy facilitates the recruitment of regulatory proteins and potentially noncoding RNAs to their genomic target sites and the establishment of chromatin environments, which are both critical for the efficient regulation of gene expression.

EXPERIMENTAL PROCEDURES

4C-seq and 3C

4C-seq libraries listed in Table S1 were prepared as described (Splinter et al., 2012) using the primers given in Table S2 and sequenced on Illumina machines. Reads that aligned to unique HindIII sites in the genome (build mm9) with at most two mismatches were retained. Read distribution statistics for all libraries are given in Table S3. 3C libraries were prepared in an analogous manner to 4C libraries, ending with the first ligation step and amplified with primers listed in Table S2.

Data Analysis

For each 4C-seq library, read counts at each unique HindIII site were collapsed to a hit to reduce effects of clonal amplification. Replicate libraries that passed stringent quality control were pooled for downstream analysis by calculating the probability of a hit at each HindIII site across all replicates. Next, we determined the average hit probability within 200 kb windows tiled along each chromosome for each pooled data set, referring to this as the hit percentage. A binomial test with different background models for *cis* and *trans* interactions was used to identify significantly interacting regions of each bait locus.

For PCA, 31 linear genomic features (summarized in Table S6) were transformed into five chromatin states and 11 TF clusters (Figures 3A and 3B), which were in turn aggregated within 200 kb windows across the genome, together with six unclustered features, to obtain a feature density matrix. To describe the linear genomic feature state of ESCs and MEFs, vectors containing feature density within 200 kb windows for each cell type were concatenated, allowing PCA to be conducted on the combined feature matrix (Table S6). To obtain a PC score enrichment value for each 4C bait, the mean PC score within the 200 kb bait window and the four flanking windows was calculated (i.e., five 200 kb bait windows = 1 Mb bait region). The PC score enrichment within the bait's interactome was calculated as the mean PC score within 200 kb windows that overlapped 4C positive domains (as determined by the binomial test) by at least 25%, excluding the five bait windows. The rankings of the bait and interactome PC score enrichment values were correlated using Spearman's *rho* statistics.

For the Hi-C data comparison, normalized, mouse ESC Hi-C interaction matrices based on 40 kb bins were downloaded from the Ren Lab website (Dixon et al., 2012) and rebinned into 200 kb bins to match the resolution of our 4C and feature data. Each 200 kb window of rebinned Hi-C data was treated as pseudo-bait and its intrachromosomal interactome was extracted from the chromosome-wide contact matrix. Bait and interactome PC scores were calculated as described above for 4C-seq interactomes, except that the interactome was defined as those 5% of 200 kb windows that had the highest read count, excluding the 1 Mb bait region.

FISH and Immunostaining Analysis

3D FISH and immunostaining were done following standard methods.

ACCESSION NUMBERS

Our 4C-seq data are available at GEO under GSE50029.

SUPPLEMENTAL INFORMATION

Supplemental Information for this article includes Supplemental Experimental Procedures, six figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2013.08.013.

ACKNOWLEDGMENTS

We thank Shawn Cokus for advice, Sanjeet Patel for RNA-seq data analysis, and Terry Magnuson for *Eed*$^{-/-}$ ESCs. K.P. is supported by the UCLA Broad Center of Regenerative Medicine and Stem Cell Research, NIH (DP2OD001686; P01GM099134), and CIRM (RN1-00564); M.P., by NIH (P01GM099134), NSF (HYS-1066293), and the hospitality of the Aspen Center for Physics; G.B., by the Whitcome Pre-doctoral Training Program; M.D., by fellowships from the UCLA Broad Center of Regenerative Medicine and Stem Cell Research, the UCLA Graduate Division, and CIRM; C.C., by a Leukemia and Lymphoma Research Fellowship 10040; J.E., by funds from the UCLA David Geffen School of Medicine; and W.d.L., by an ERC Starting Grant (209700). M.D. and G.B. designed research, performed experiments, analyzed data, and wrote the paper; C.C. performed experiments; E.S. and W.d.L. trained M.D. in 4C-seq library preparation in 2008; J.E. provided analytic tools; M.P. offered supervision and data analysis; and K.P. designed research, analyzed data, wrote the paper, and supervised the overall project.

Received: August 1, 2013
Revised: August 26, 2013
Accepted: August 28, 2013
Published: September 12, 2013

REFERENCES

Apostolou, E., Ferrari, F., Walsh, R.M., Bar-Nur, O., Stadtfeld, M., Cheloufi, S., Stuart, H.T., Polo, J.M., Ohsumi, T.K., Borowsky, M.L., et al. (2013). Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. Cell Stem Cell *12*, 699–712.

Bantignies, F., and Cavalli, G. (2011). Polycomb group proteins: repression in 3D. Trends Genet. *27*, 454–464.

Bickmore, W.A., and van Steensel, B. (2013). Genome architecture: domain organization of interphase chromosomes. Cell *152*, 1270–1284.

Chamberlain, S.J., Yee, D., and Magnuson, T. (2008). Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. Stem Cells *26*, 1496–1505.

de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. Genes Dev. *26*, 11–24.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.

Denholtz, M., and Plath, K. (2012). Pluripotency in 3D: genome organization in pluripotent cells. Curr. Opin. Cell Biol. *24*, 793–801.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., et al. (2013). The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. Science *341*, 1237973.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods *9*, 215–216.

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat. Cell Biol. *8*, 188–194.

Gibcus, J.H., and Dekker, J. (2013). The hierarchy of the 3D genome. Mol. Cell *49*, 773–782.

Hakim, O., Sung, M.H., Voss, T.C., Splinter, E., John, S., Sabo, P.J., Thurman, R.E., Stamatoyannopoulos, J.A., de Laat, W., and Hager, G.L. (2011). Diverse

gene reprogramming events occur in the same spatial clusters of distal regulatory elements. Genome Res. *21*, 697–706.

Hakim, O., Sung, M.H., Nakayamada, S., Voss, T.C., Baek, S., and Hager, G.L. (2013). Spatial congregation of STAT binding directs selective nuclear architecture during T-cell functional differentiation. Genome Res. *23*, 462–472.

Hiratani, I., Ryba, T., Itoh, M., Rathjen, J., Kulik, M., Papp, B., Fussner, E., Bazett-Jones, D.P., Plath, K., Dalton, S., et al. (2010). Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. Genome Res. *20*, 155–169.

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat. Methods *9*, 999–1003.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

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

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Montgomery, N.D., Yee, D., Chen, A., Kalantry, S., Chamberlain, S.J., Otte, A.P., and Magnuson, T. (2005). The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. Curr. Biol. *15*, 942–947.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379–391.

Noordermeer, D., de Wit, E., Klous, P., van de Werken, H., Simonis, M., Lopez-Jones, M., Eussen, B., de Klein, A., Singer, R.H., and de Laat, W. (2011). Variegated gene expression caused by cell-specific long-range DNA interactions. Nat. Cell Biol. *13*, 944–951.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W., and Fraser, P. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. Nat. Genet. *36*, 1065–1071.

Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281–1295.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., and Chang, H.Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311–1323.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109–113.

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N.F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J.A., Umlauf, D., Dimitrova, D.S., et al. (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat. Genet. *42*, 53–61.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., and Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116–120.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat. Genet. *38*, 1348–1354.

Smallwood, A., and Ren, B. (2013). Genome organization and long-range regulation of gene expression by enhancers. Curr. Opin. Cell Biol. *25*, 387–394.

Splinter, E., de Wit, E., van de Werken, H.J., Klous, P., and de Laat, W. (2012). Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: From fixation to computation. Methods *58*, 221–230.

Sridharan, R., Tchieu, J., Mason, M.J., Yachechko, R., Kuoy, E., Horvath, S., Zhou, Q., and Plath, K. (2009). Role of the murine reprogramming factors in the induction of pluripotency. Cell *136*, 364–377.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Takebayashi, S., Dileep, V., Ryba, T., Dennis, J.H., and Gilbert, D.M. (2012). Chromatin-interaction compartment switch at developmentally regulated chromosomal domains reveals an unusual principle of chromatin folding. Proc. Natl. Acad. Sci. USA *109*, 12574–12579.

Tanizawa, H., Iwasaki, O., Tanaka, A., Capizzi, J.R., Wickramasinghe, P., Lee, M., Fu, Z., and Noma, K. (2010). Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. Nucleic Acids Res. *38*, 8164–8177.

Wei, Z., Gao, F., Kim, S., Yang, H., Lyu, J., An, W., Wang, K., and Lu, W. (2013). Klf4 organizes long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency. Cell Stem Cell *13*, 36–47.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Young, R.A. (2011). Control of the embryonic stem cell state. Cell *144*, 940–954.