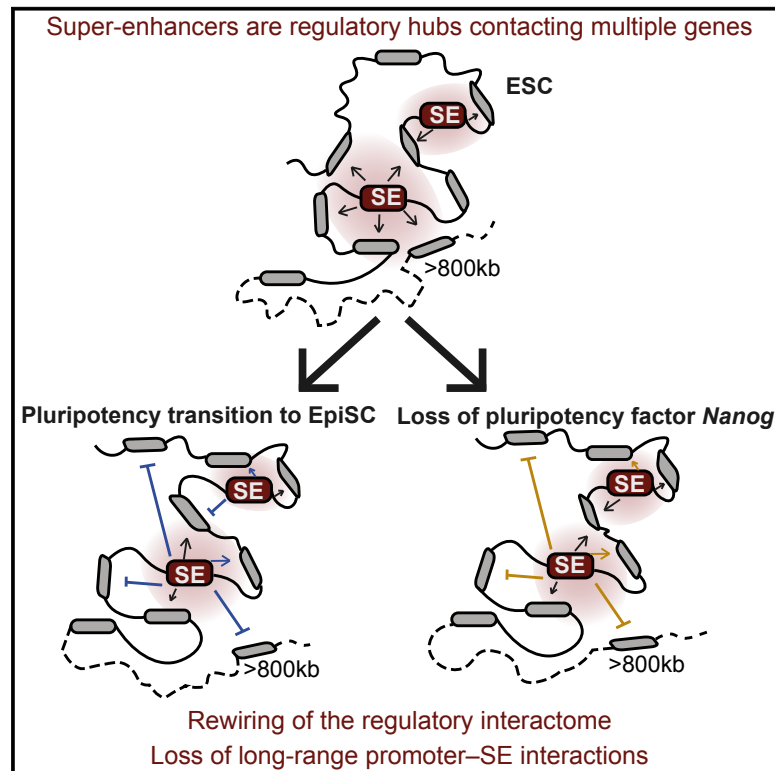


## Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition

### Graphical Abstract



### Authors

Clara Lopes Novo, Biola-Maria Javierre, Jonathan Cairns, ..., Stefan Schoenfelder, Peter Fraser, Peter J. Rugg-Gunn

### Correspondence

peter.rugg-gunn@babraham.ac.uk

### In Brief

Novo et al. use promoter-capture Hi-C to map the target promoters of super-enhancers (SEs) in mouse pluripotent cells. SEs form complex networks, and a subset of promoter-SE interactions was rewired between ESCs and EpiSCs. In ESCs, many SEs form long-range contacts that are not detected in EpiSC or *Nanog*-deficient ESCs.

### Highlights

- Promoter-capture Hi-C identifies 3D interactions in mouse pluripotent cells
- Super-enhancers (SEs) form complex spatial networks contacting multiple promoters
- Rewiring of promoter-SE interactions between ESC and EpiSC pluripotent states
- Long-range SE interactions are a hallmark of mouse ESCs

### Data and Software Availability

GSE103053



# Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition

Clara Lopes Novo,<sup>1</sup> Biola-Maria Javierre,<sup>2,6</sup> Jonathan Cairns,<sup>2,7</sup> Anne Segonds-Pichon,<sup>3</sup> Steven W. Wingett,<sup>3</sup> Paula Freire-Pritchett,<sup>2,8</sup> Mayra Furlan-Magaril,<sup>2,9</sup> Stefan Schoenfelder,<sup>2</sup> Peter Fraser,<sup>2,4</sup> and Peter J. Rugg-Gunn<sup>1,5,10,\*</sup>

<sup>1</sup>Epigenetics Programme, Babraham Institute, Cambridge CB22 3AT, UK

<sup>2</sup>Nuclear Dynamics Programme, Babraham Institute, Cambridge CB22 3AT, UK

<sup>3</sup>Bioinformatics Group, Babraham Institute, Cambridge CB22 3AT, UK

<sup>4</sup>Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA

<sup>5</sup>Wellcome Trust – Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge CB2 1QR, UK

<sup>6</sup>Present address: Josep Carreras Leukemia Research Institute, Campus ICO-Germans Trias i Pujol, Badalona, Spain

<sup>7</sup>Present address: Quantitative Biology Department, Discovery Sciences, IMED Biotech Unit, AstraZeneca, Cambridge, UK

<sup>8</sup>Present address: MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, UK

<sup>9</sup>Present address: Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Mexico City, Mexico

<sup>10</sup>Lead Contact

\*Correspondence: [peter.rugg-gunn@babraham.ac.uk](mailto:peter.rugg-gunn@babraham.ac.uk)

<https://doi.org/10.1016/j.celrep.2018.02.040>

## SUMMARY

Transcriptional enhancers, including super-enhancers (SEs), form physical interactions with promoters to regulate cell-type-specific gene expression. SEs are characterized by high transcription factor occupancy and large domains of active chromatin, and they are commonly assigned to target promoters using computational predictions. How promoter-SE interactions change upon cell state transitions, and whether transcription factors maintain SE interactions, have not been reported. Here, we used promoter-capture Hi-C to identify promoters that interact with SEs in mouse embryonic stem cells (ESCs). We found that SEs form complex, spatial networks in which individual SEs contact multiple promoters, and a rewiring of promoter-SE interactions occurs between pluripotent states. We also show that long-range promoter-SE interactions are more prevalent in ESCs than in epiblast stem cells (EpiSCs) or *Nanog*-deficient ESCs. We conclude that SEs form cell-type-specific interaction networks that are partly dependent on core transcription factors, thereby providing insights into the gene regulatory organization of pluripotent cells.

## INTRODUCTION

Complex, multi-layered compaction and folding enables the eukaryotic genome to undergo functional organization within the 3D nuclear space. Higher-order chromatin architecture forms into topologically associating domains (TADs), which are discrete ~1-Mb structures that compartmentalize and insulate the genome (Nora et al., 2012; Dixon et al., 2012). Within TADs,

DNA loops bring promoters and their distal regulatory elements into close physical proximity. TAD organization is orchestrated by architectural proteins like CTCF and cohesin and is largely cell-type invariant (Rao et al., 2014; Wei et al., 2013; Apostolou et al., 2013; Nora et al., 2017). In contrast, intra-TAD DNA loops are typically cell-type specific and are frequently rewired upon cell state changes (Smith et al., 2016; Denholtz et al., 2013; Schoenfelder et al., 2015b; Freire-Pritchett et al., 2017). This hierarchical nuclear organization permits the coordinated activation and repression of cell-identity genes while restricting the pool of promoters that are able to contact regulatory elements, including transcriptional enhancers (Dixon et al., 2012).

The mechanisms responsible for establishing the interactions between promoters and their regulatory elements include the binding of cell-type-specific transcription factors and the local chromatin landscape (reviewed in Heinz et al., 2015; Vernimmen and Bickmore, 2015). In mouse embryonic stem cells (ESCs), global, low-resolution analyses showed that large regions harboring clusters of NANOG, OCT4, or SOX2 binding sites preferentially interact, and depleting NANOG or OCT4 reduces the frequency of interactions (de Wit et al., 2013; Denholtz et al., 2013). In addition, the recruitment of NANOG to an ectopic site is sufficient to bring different distant regions together, thereby demonstrating a direct role for pluripotency factors in controlling chromatin topology (de Wit et al., 2013). At higher resolution, pluripotent-specific interactomes around the *Nanog* and *Pou5f1* promoters have been mapped (Apostolou et al., 2013), and the maintenance of the chromatin structure at these sites depends on pluripotency factors such as OCT4 (Levasseur et al., 2008). Contacts are also sensitive to differentiation cues that can disrupt promoter-enhancer loops at these loci (Kagey et al., 2010; Gaspar-Maia et al., 2011; Phillips-Cremins et al., 2013). In addition, transcription-factor-induced reprogramming to pluripotency induces a reorganization in the chromatin topology of the donor somatic cells (Krijger et al., 2016; Beagan et al., 2016). These studies collectively provide support for a model



where DNA looping at regulatory elements can be driven and/or regulated by pluripotency-associated transcription factors.

A subset of enhancers, known as super-enhancers (SEs; [Hnisz et al., 2013](#); [Whyte et al., 2013](#)) or stretch enhancers ([Parker et al., 2013](#)), was proposed to be crucial for regulating the expression of cell identity genes. SEs form large domains (typically >3 kb) with high levels of the active enhancer mark, histone 3 lysine 27 acetylation (H3K27ac), and are densely occupied by key transcription factors ([Whyte et al., 2013](#)). In ESCs, for example, most SEs are bound by NANOG, OCT4, and SOX2 ([Whyte et al., 2013](#)). SEs are thought to control gene expression programs by associating with promoters and modulating their transcriptional output. The gene promoter targets of SEs are typically predicted using algorithms that account for linear proximity and high levels of gene expression ([Whyte et al., 2013](#)). This approach in ESCs assigned 231 SEs to 210 genes; the majority of these genes have prominent roles in controlling the pluripotent state. Thus, the binding of cell-specific transcription factors to SEs is predicted to establish self-regulatory feedback that may stabilize cell identity.

The function of several SEs has been examined in ESCs using genetic approaches. Deleting multiple SEs in the vicinity of the *Nanog* and *Sox2* loci leads to variable effects on predicted target gene transcription ([Zhou et al., 2014](#); [Blinka et al., 2016](#)). This phenotypic variability following SE deletion was also observed in another study, which reported a large range in the transcriptional misregulation of predicted target genes ([Moorthy et al., 2017](#)). Interestingly, deleting a SE can also alter the transcription of genes that are not currently assigned to that SE, suggesting that SEs might provide regulatory inputs to multiple promoters and operate within larger networks or interaction hubs ([Blinka et al., 2016](#); [Moorthy et al., 2017](#)). The underlying causes of the observed transcriptional variability are not clear but may be due to the partial redundancy of the individual enhancer elements that comprise a SE and to the failure of current SE predictions to infer additional strong regulatory associations. Mapping the precise regulatory inputs to key cell identity genes will improve the curation of transcriptional networks and lead to a better understanding of the phenotypic variability that is observed upon SE perturbations.

Despite the progress in defining regulatory elements within the genome, inferring promoter-SE associations based on linear proximity has several shortcomings. First, regulatory elements are frequently located at considerable distances away from their target promoters in linear DNA (reviewed in [van Heyningen and Bickmore, 2013](#)). Second, promoters are often contacted by multiple regulatory elements, and individual regulatory elements may interact with multiple target promoters, either sequentially or simultaneously ([Sanyal et al., 2012](#); [Freire-Pritchett et al., 2017](#); [Schoenfelder et al., 2015a](#)). This complexity can easily be overlooked when assigning single promoter-SE pairings. However, methods such as chromosome conformation capture (3C) approaches map the physical interactions between promoters and their regulatory elements, including SEs. Promoter-capture Hi-C (PCHI-C) in ESCs identified interactions between SEs and 503 genes ([Schoenfelder et al., 2015a](#)), which is many more than the predicted number of 210 genes, thereby pointing toward a more complex spatial network than previously thought.

It is important to expand these results by carefully examining this spatial network in more detail, including an assessment of the combinatorial promoter-SE interactions, and in additional cell types. Moreover, it remains unknown how promoter-SE interactions change upon developmental state transitions, and how transcription factors may establish or maintain promoter-SE interactions. These points are important to address in order to assess the accuracy of current SE target gene assignments, understand the complexities of regulatory networks that underpin cell identity, and better interpret current and future functional experiments.

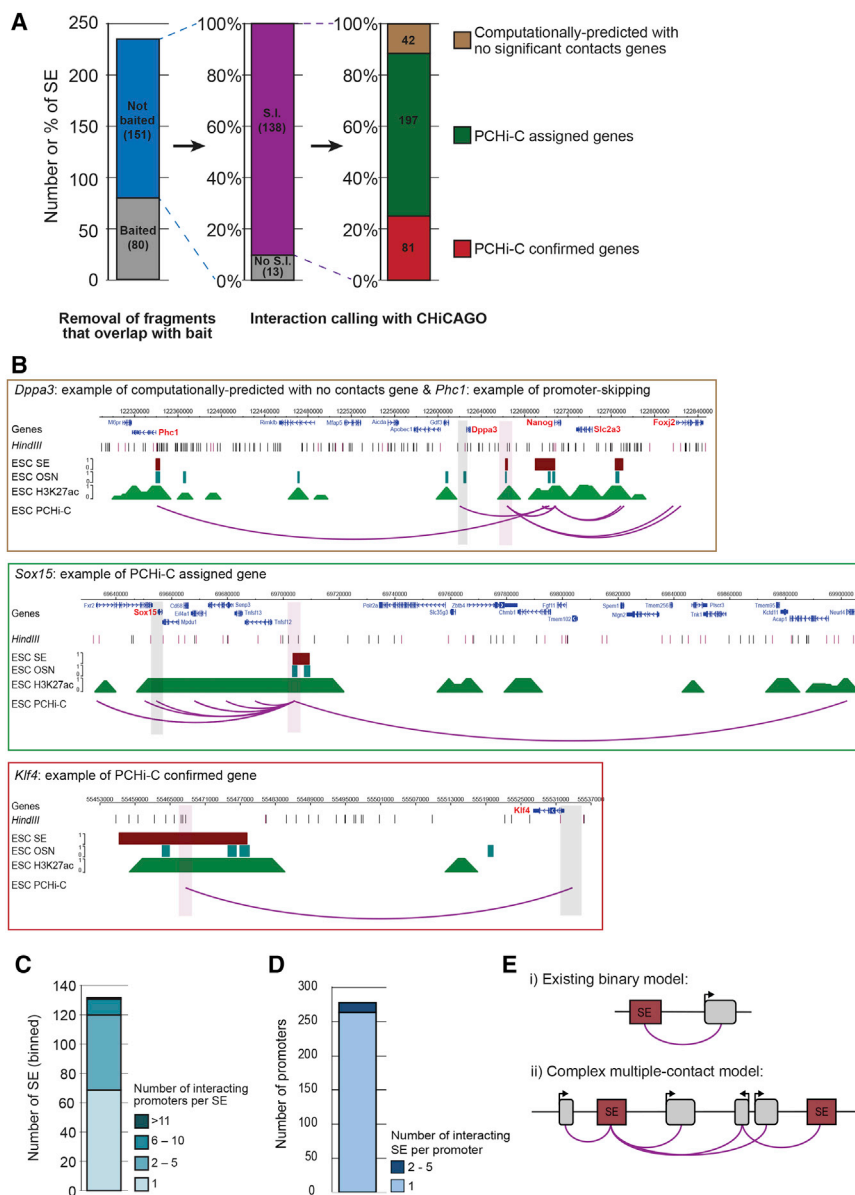
ESCs require precise regulation of transcriptional programs to enable the balance between self-renewal and effective differentiation and provide an informative system to study how gene regulatory interactions are altered on cell state change. Here, we used PCHI-C to identify promoter-based interactions in ESCs, and we focused our analysis on investigating the interactions with SEs. We assigned promoter targets to previously defined SEs, which both confirmed and extended the number of known regulatory contacts. We also mapped promoter-SE interactions in a more developmentally advanced pluripotent cell type, epiblast stem cells (EpiSCs), and found that a subset of interactions was rewired between ESCs and EpiSCs. Interestingly, promoter-SE interactions frequently spanned large distances (>800 kb) in ESCs, but not in EpiSCs or in *Nanog*-deficient ESCs. Together, these results provide insights into the organization of chromatin topology in ESCs and lead to a better understanding of how gene regulatory networks can switch between pluripotent states.

## RESULTS

### SEs Are Highly Interactive Hubs in ESCs

To map the gene promoters that are in close physical proximity to SEs in mouse ESCs, we performed PCHI-C, a high-throughput 3C-based technique with a capture step to enrich for interactions at >22,000 promoters at single-restriction-fragment resolution ([Mifsud et al., 2015](#); [Schoenfelder et al., 2015a](#)). We identified significant interactions using Capture Hi-C Analysis of Genomic Organization (CHICAGO) ([Cairns et al., 2016](#)), including several previously reported contacts between promoters and their regulatory elements in ESCs for *Pax2*, *Tbx5*, and *Wnt6* ([Figure S1](#); [Schoenfelder et al., 2015a, 2015b](#)).

We focused on the 901 HindIII fragments overlapping the 231 SEs that were defined previously in ESCs ([Whyte et al., 2013](#)). We assigned each SE to one or more gene promoters based on their interactions ([Table S1](#)). We restricted our analysis to the 151 SEs (~70%) that did not overlap with a promoter to ensure that detected interactions were mediated by promoter-SE, and not by promoter-promoter, contacts ([Figure 1A](#)). Of these, we detected significant promoter interactions with 138 SEs (91%). The original description of SEs in ESCs assigned 210 genes to SEs by linear proximity ([Whyte et al., 2013](#)). Our PCHI-C data provide direct evidence for promoter-SE interactions at 81 of those genes, including pluripotency-associated factors such as *Klf4*, *Sox2*, *Nanog*, and *Fgf4* ([Figures 1A and 1B](#); [Table S1](#)). No significant SE contacts were detected for 42 genes assigned to a SE by computational predictions,



**Figure 1. Promoter-Capture Hi-C Reveals that SEs Form Spatial Networks in Mouse ESCs**

(A) From the 231 previously defined SEs in ESCs (Whyte et al., 2013), 80 overlapped baited Hi-C fragments and were not considered further in this analysis. Significant promoter-SE interactions were identified at 138 of the remaining SEs, and a comparison of SE target genes between PCHI-C and computational assignments is shown.

(B) Example of a computationally predicted SE-target gene (*Dppa3*) without significant PCHI-C interactions (top). Examples of a newly identified SE-target gene (*Sox15*; middle) and a computationally predicted SE-target gene that was confirmed by PCHI-C (*Klf4*; bottom). Tracks show HindIII sites (black lines), baited HindIII sites (purple lines), ESC SEs (maroon), and ESC OSN (OCT4, SOX2, and NANOG) binding sites (teal). Arcs show significant interactions. Red and gray vertical bars highlight the location of SEs and promoters, respectively. The upper genome browser screenshot also exemplifies SE promoter skipping, where the SE close to *Phc1* contacts the *Nanog* promoter, skipping over six genes.

(C) Stacked bar chart showing the number of SEs interacting with 1, 2–5, 6–10, or >11 gene promoters.

(D) Stacked bar chart showing the number of gene promoters interacting with 1 or 2–5 SEs.

(E) Schematic proposing an updated model of SE interactions from a binary promoter-SE interaction (i) to a more complex network where SEs can contact multiple promoters, and promoters can interact with more than one SE (ii). Red and gray rectangles represent SEs and genes, respectively. See also Figure S1 and Table S1.

confirming that linear distance does not accurately infer regulatory contacts (Figures 1A and 1B; Table S1). Excitingly, we identified an additional 197 target promoters that have not been previously associated with a SE in ESCs (Figures 1A and 1B; Table S1). Notably, several newly identified genes encode for components of the glycolysis pathway, which is highly active in serum-grown mouse ESCs (Zhang et al., 2012, 2016). Mapping the SE-interactome in ESCs, therefore, has provided direct evidence for predicted promoter-SE contacts, and expanded the gene regulatory networks to reveal an unanticipated connection with the control of metabolic activity.

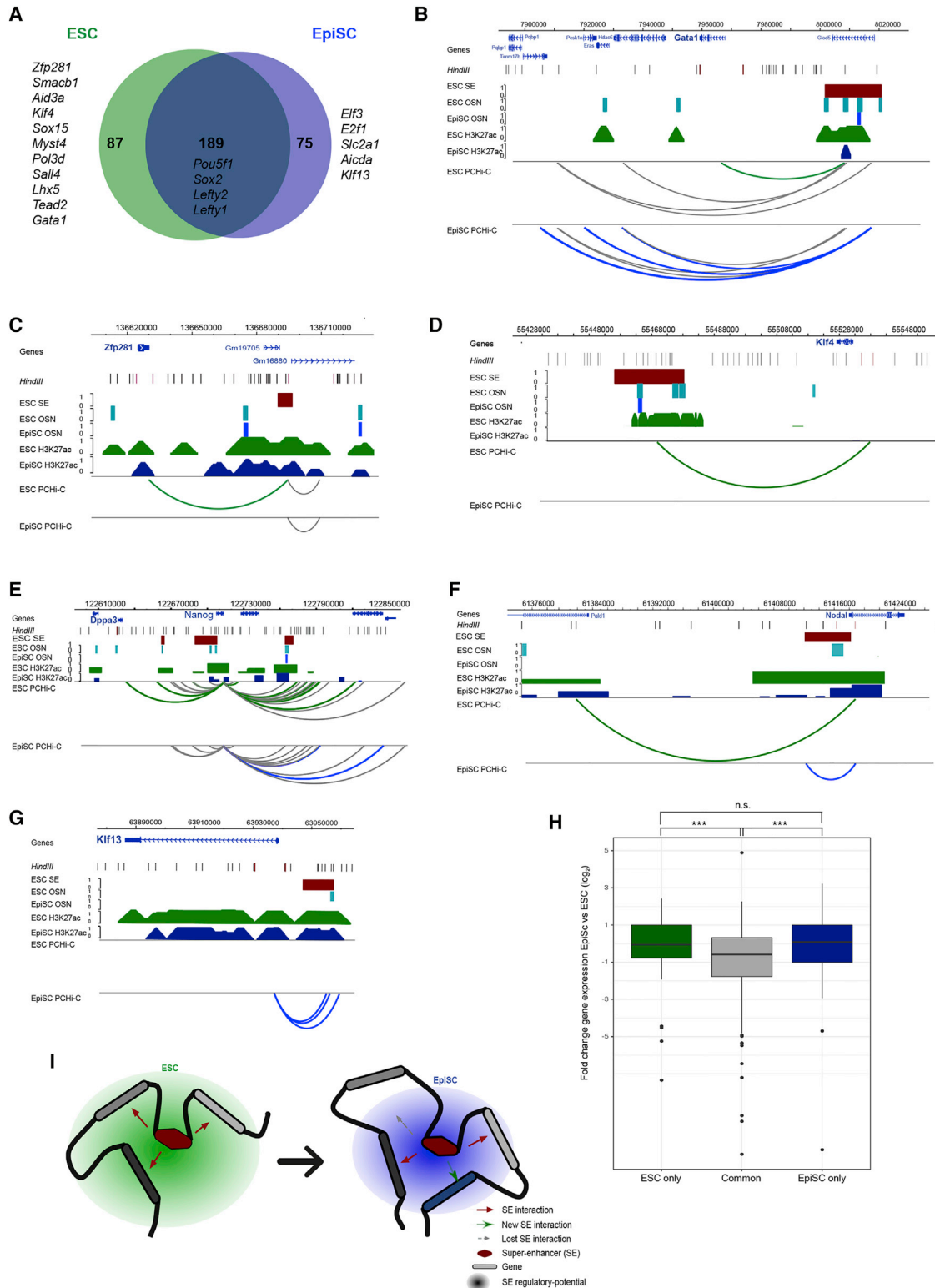
Further analysis of the PCHI-C data revealed frequent promoter skipping by SEs, as exemplified by the SE proximal to *Phc1* that contacts the more distal *Nanog* promoter, skipping over six other promoters (Figure 1B, upper). This observation is

similar to the skipping of active promoters by “regular” enhancers (Schoenfelder et al., 2015a). We also observed that half of the interacting SEs contacted two or more promoters, with ~5% interacting with >6 different promoters (Figure 1C). Most promoters contacted one individual SE, and only 14 promoters contacted 2–5 SEs (Figure 1D). Together, this high-resolution contact map shows that SEs are highly interactive and complex hubs in ESCs, where several promoters can be in close physical contact to an individual SE (Figure 1E).

### Rewiring of the SE Interactome upon Pluripotent State Transition

ESCs and EpiSCs represent functionally distinct pluripotent states with differing transcriptional and epigenetic programs that reflect their similarities to pre- and post-implantation epiblast cells, respectively. Many enhancers undergo changes in activity between the two pluripotent cell types (Buecker et al., 2014; Factor et al., 2014), and transcriptional differences include the reduced expression of genes such as *Nanog* and *Klf4* (Brons et al., 2007; Osorno et al., 2012; Tesar et al., 2007;





**Figure 2. Pluripotent State-Specific Wiring of SE Interactions**

(A) The promoters in close physical proximity to ESC SEs were largely maintained in EpiSCs, but subsets of significant SE interactions were detected uniquely in each cell type.

(legend continued on next page)

Guo et al., 2009; Novo et al., 2016). To investigate how the ESC SE interactome is reorganized upon this developmental transition, we performed PCHI-C in EpiSCs. We found that the majority (70%) of genes interacting with SEs in ESCs were the same in EpiSCs (Figure 2A). These genes encode key pluripotency factors such as *Pou5f1* and *Sox2*, plus signaling regulators like *Lefty1* and *Smad1* (Figure 2A).

The remaining 30% of SE-interacting genes were specific to each pluripotent state (Figure 2A). There were 87 genes that significantly interacted with SEs only in ESCs, including *Sall4*, *Tead2*, *Zfp281*, *Klf4*, *Sox15*, and *Gata1* (Figures 2A–2E; Table S1). We also detected in EpiSCs a set of promoters that gained significant interactions with ESC SEs, including *Aicda* and *Klf13* (Figures 2A, 2F, and 2G; Table S1). The majority (92%) of these SEs remained either a SE or an active enhancer in EpiSCs. Interestingly, two-thirds of the EpiSC-specific gene promoters are interacting with a SE that is in contact with a different gene promoter in ESCs. This suggests that the same SE can interact with different promoters in different cell types. Although there was no global association between cell-type-specific SE interaction and target gene expression (Figure 2H), the rewiring of a set of ESC SE interactions was consistent with transcriptional changes at individual loci. For example, loss of contacts between the *Klf4* promoter and a SE, and between the *Dppa3* promoter and a SE, is consistent with their transcriptional downregulation in EpiSCs compared to ESCs (Figures 2D and 2E). In addition, the gain of a significant interaction in EpiSCs between the *Nodal* promoter and its pluripotent-specific enhancer (Papanayotou et al., 2014) is in line with the transcriptional upregulation of *Nodal* in EpiSCs (Figure 2F). Taken together, these results show that the majority of promoter-SE interactions are conserved between ESCs and EpiSCs, but a subset of the ESC SE interactome is reorganized upon the transition to an EpiSC state (Figure 2I).

We examined ESC chromatin immunoprecipitation sequencing (ChIP-seq) data (Sánchez-Castillo et al., 2015) to identify proteins that were enriched at the gene promoters in contact with SEs only in ESCs (ESC only). This revealed that ESC-only promoters are commonly bound by pluripotency factors, including TFCEP2L1, KLF4, PRDM14, and SOX2 (Figure S2A). We also examined the gene promoters that are not in contact with SEs in ESCs but acquire SE interactions in EpiSCs (EpiSC only). This category of promoters was enriched for a different set of proteins that included JARID2, EZH2, TBX3, TET1, and KDM2A (Figure S2A). Core factors, such as OCT4 and MYC, were present at similar levels for ESC-only and EpiSC-only gene promoters. The analysis indicates that different categories of SE-interacting promoters are associated with distinct sets of transcription factor occupancy. This configura-

tion might help to stabilize active SE interactions in ESCs and also promote the “priming” of new SE interactions upon transition to EpiSCs.

### Long-Range SE Interactions Are Detected in ESCs, but Not in EpiSCs

To investigate whether the rewiring of interactions between pluripotent states was a consequence of altered SE status, we defined SEs in EpiSCs by running H3K27ac ChIP-seq data (Factor et al., 2014) through the SE-calling ranking of super-enhancer (ROSE) pipeline (Whyte et al., 2013). H3K27ac levels are one of the most informative features to identify SEs (Lovén et al., 2013; Khan and Zhang, 2016) and have been used to operationally define SEs in a range of tissues and species (Whyte et al., 2013; Chapuy et al., 2013; Ding et al., 2015; Pérez-Rico et al., 2017; Cao et al., 2017). This approach identified 896 SEs in EpiSCs (EpiSC-SE). To generate a comparable list in ESCs, we re-called SEs using the same pipeline, resulting in 927 SEs (ESC-SE), which incorporated the vast majority (85%) of the 231 ESCs SEs considered thus far. As expected, the number of ROSE-called SEs is higher than the original 231 ESCs SEs, as MED1 occupancy was initially used as additional criteria (Whyte et al., 2013). Approximately one-third of the ROSE-called SEs were the same in ESCs and EpiSCs. In both cell types, SEs were highly interactive when compared to “regular” enhancers or control regions that were randomly selected to size match each individual SE (Figures 3A and 3B).

To facilitate the comparison of promoter-SE interactions between ESCs and EpiSCs, we modified the CHICAGO pipeline to concatenate all the individual HindIII fragments that overlapped a single ROSE-called SE and mapped significant interactions between promoters and each SE region. We also selected additional categories of regions with medium (normal enhancers [NEs]), low (low enhancers [LEs]) or undetectable (control [CTRL]) levels of H3K27ac and extended their sizes to match those of SE regions. As before, there were many more significant interactions with the concatenated SE regions as compared to the size-matched regular enhancers or control regions, thereby demonstrating that SEs are contacted by a particularly high number of promoter interactions (Figure 3C). The transcriptional output of genes interacting with SEs was higher than from the set of genes contacting NE, LE, or control regions (Figure 3D).

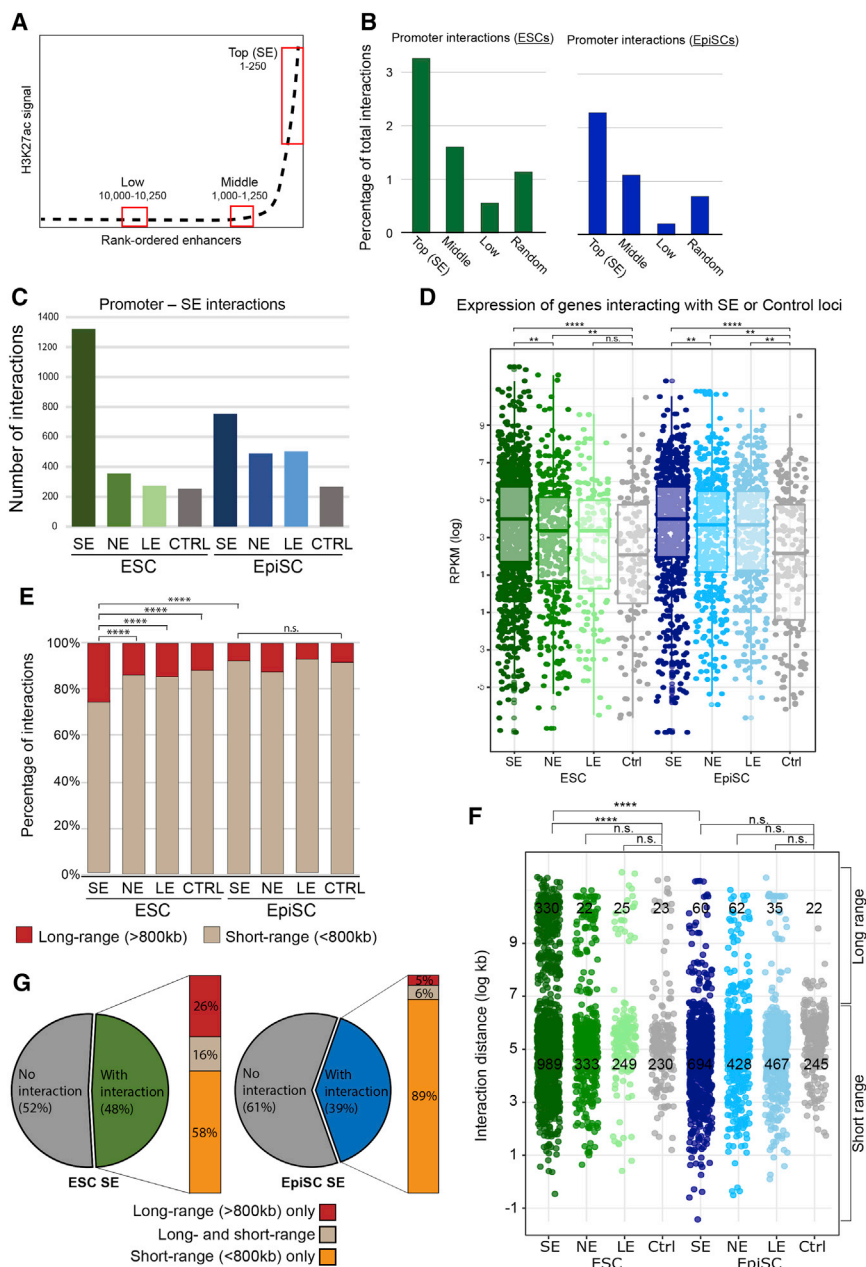
Interestingly, the proportion of long-range interactions (LRIs; defined as >800 kb) between promoters and SEs was significantly higher in ESCs than in EpiSCs. Overall, 25% of promoter-SE interactions spanned >800 kb in ESCs, which is more than for regular enhancers and control regions ( $p < 0.0001$ ) and also when compared to SEs in EpiSCs ( $p < 0.0001$ ; Figures 3E and 3F). Indeed, while 58% of the ESC

(B–G) Differences in the SE interactome between ESCs and EpiSCs are shown for the *Gata1* (B), *Zfp281* (C), *Klf4* (D), *Nanog* (E), *Nodal* (F), and *Klf13* (G) loci. Arcs indicate significant interactions colored by cell-type specificity (ESC only, green; EpiSC only, blue; shared, gray).

(H) Boxplot showing the log<sub>2</sub> fold change in gene expression between EpiSCs and ESCs, for genes interacting with SEs only in ESCs, in ESCs and EpiSCs, and only in EpiSCs. \*\* $p < 0.01$  (Mann-Whitney *U* test).

(I) Model illustrating the rewiring of interactions between pluripotent states at an individual SE. The red arrows represent the physical contacts between promoters detected by PCHI-C, creating a spatial hub for potential SE regulation (green/blue circles). Upon transitioning to EpiSCs, some contacts are lost (gray arrow), while new ones are established (green arrow).

See also Figure S2 and Table S1.



**Figure 3. Long-Range Promoter-SE Interactions Are Prevalent in ESCs**

(A) After ROSE analysis, we selected three enhancer categories: the top 250 H3K27ac-ranked regions (comprising a subset of SEs), the first 250 enhancers immediately below the SE-threshold (middle), and 250 enhancers with low levels of H3K27ac. Red boxes highlight the selected enhancers within each category.

(B) Proportion of PCHI-C interactions detected for the three enhancer categories and control regions in ESCs (left) and in EpiSCs (right). Control regions were randomly selected and size-matched to SEs.

(C) Number of interactions after concatenation of the fragments spanning an individual SE, normal enhancer (NE), low enhancer (LE), or randomly selected control regions (CTRL) in ESCs and EpiSCs.

(D) Expression (log<sub>2</sub> reads per kilobase per million mapped reads [RPKM]) of genes interacting with the three classes of enhancers and control regions in ESCs and EpiSCs. \*\*\*\*p < 0.0001; \*\*p < 0.01 (Mann-Whitney U test).

(E) ESCs have a high proportion (25%) of promoter-SE interactions spanning >800 kb (red) when compared to interactions in EpiSCs or to size-matched NEs and control regions (CTRL; Benjamini-Hochberg adjusted chi-square p < 0.0001).

(F) A significant proportion of LRIs between promoters and SEs in ESCs (green, 25%) compared to EpiSCs (blue, 7%) or to size- and distance-matched random control regions (light green, 13%; Benjamini-Hochberg adjusted p < 0.0001). NE and LE are not significantly different from controls. See Table S2 for a list of LRIs.

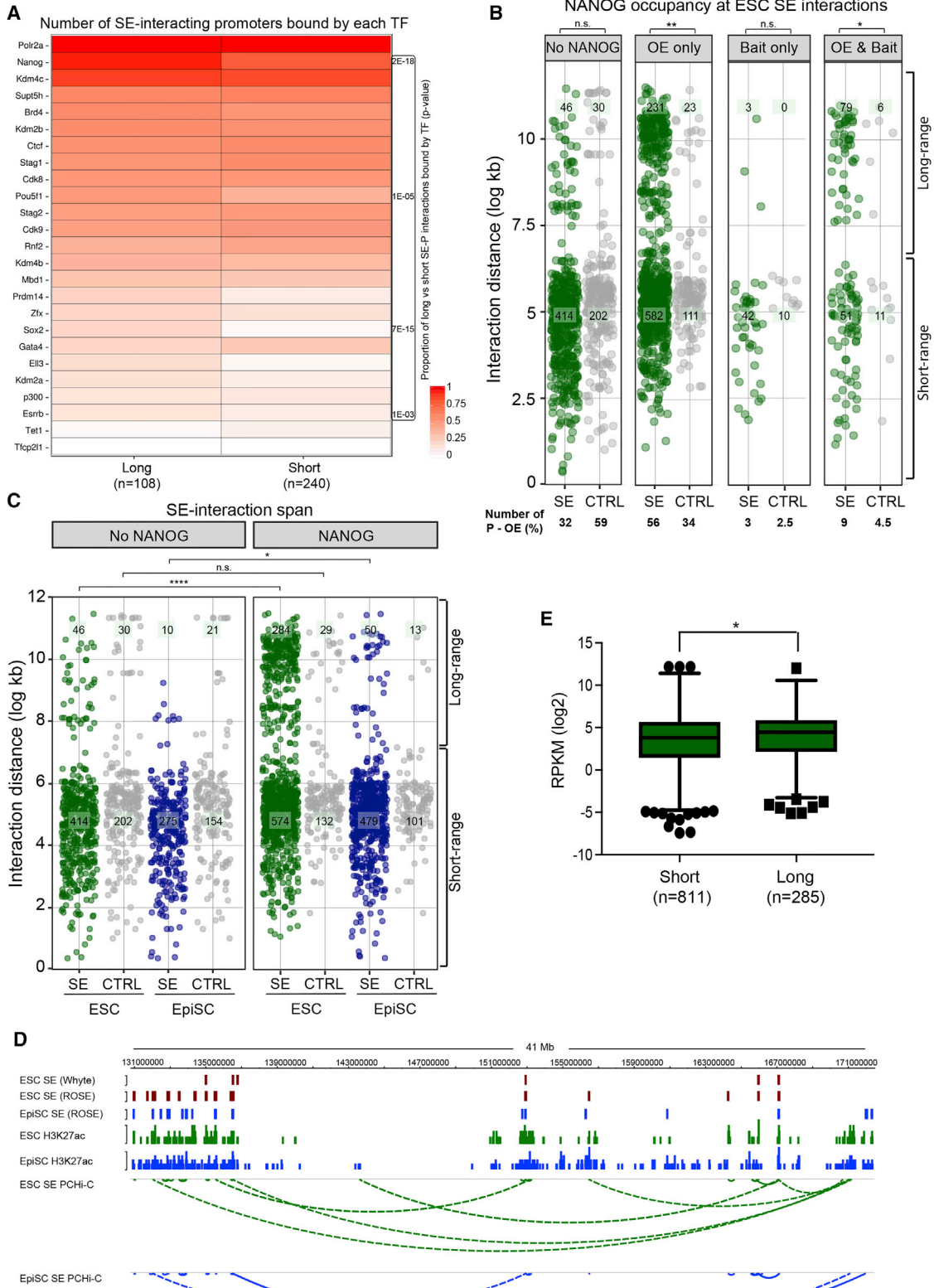
(G) Classification of concatenated SE fragments according to the genomic distance of their target promoters. In ESCs, 58% of interacting SE regions only engage in short-range (<800 kb), 26% in long-range only (>800 kb), and 16% in both short- and long-range promoter interactions. In EpiSCs, most interacting SEs (89%) engage in short-range interactions only. Note that the proportion of interactive SE regions is likely to be underestimated due to the stringent statistical analysis applied to the concatenated regions, and also due to the inability to detect interactions when the baited promoter and SE are within neighboring HindIII fragments. See also Table S2.

SEs were involved in short-range interactions, 26% of the interacting SEs were engaged exclusively in LRIs and 16% in both (Figure 3G). Importantly, when considering all significant promoter-genome interactions, there was no difference in the interaction distance when comparing between ESCs and EpiSCs (Figure S2B). In addition, interaction distances between promoters and non-SE regions, such as the Polycomb-mediated *Hox* network (Schoenfelder et al., 2015b; Joshi et al., 2015), were similar in ESCs and EpiSCs (Figure S2C), suggesting that the enrichment for LRI is a specific feature of ESC SEs. Furthermore, the CHiCAGO scores for long-range and short-range SE interactions were comparable (Figure S2D). Taken together,

our results reveal a prevalence for SEs to establish long-range promoter interactions in ESCs, but not EpiSCs.

### Long-Range Regulatory Contacts in ESCs Are Enriched for NANOG Occupancy

As a first step toward understanding the determinants of LRIs, we examined the CODEX ChIP-seq database (Sánchez-Castillo et al., 2015) for transcription factors that are bound at gene promoters engaged in long-range versus short-range SE interactions in ESCs. Interestingly, the binding of several pluripotency factors, including NANOG, OCT4, SOX2, and KLF4, were significantly enriched at long-range SE interactions compared to



(legend on next page)



short-range ones (Figures 4A and S3A). In particular, the pluripotency factor NANOG was the second most prevalent factor at promoters that interact with SEs (present at 12% of all promoters contacting ESC SEs) and with a skew toward those promoters connected by LRIs (Figure 4A). Moreover, when considering both ends of an interaction, NANOG binds the majority (68%) of sites that overlap with promoter-SE contacts (Figure 4B), with a strong preference for binding at the SE end of the interaction only (56%). This extends the observations that NANOG is a core protein that occupies many SEs in ESCs (Whyte et al., 2013) and that large regions harboring clusters of NANOG preferentially interact (de Wit et al., 2013).

Examining this in further detail, we found that long-range promoter-SE interactions were bound by NANOG; 284 are NANOG-bound versus 46 without NANOG, whereas in contrast, the short-range SE-interactions did not show this enrichment (574 bound versus 414 not bound; Figures 4B–4D and S3A). In contrast, there was no significant difference in NANOG occupancy at interactions between promoters and randomly selected control regions that were matched for fragment size and distance from the bait ( $p = 0.39$ ; Figures 4C and S3A). Although there were much fewer long-range SE interactions in EpiSCs than ESCs (total 60), we detected a modest but significant association with NANOG occupancy at long-range SE interactions compared to short-range ones ( $p = 0.008$ ; Figures 4C and 4D). We found no enrichment for other genomic features like Cohesin or POLII occupancy (Figure S3B). Finally, the expression of genes involved in long-range SE interactions was slightly higher than that of genes engaged in short-range SE interactions ( $p = 0.01$ , Mann-Whitney test; Figure 4E). Taken together, long-range SE contacts in ESCs are associated with the presence of pluripotency transcription factors, in particular NANOG, and have increased expression levels compared to genes contacting SEs over a shorter distance.

### Long-Range SE Contacts Are Depleted in *Nanog*-Deficient ESCs

Although the occupancy of core transcription factors is a defining feature of SEs, it is not known whether the transcription factors themselves have a role in establishing SE contacts. To investigate a potential role for NANOG in coordinating the SE interactome, we generated PChi-C libraries from *Nanog*<sup>-/-</sup> ESCs. Although depleted in a core transcription factor, *Nanog*-deficient cells retain key properties of ESCs, including the ability to self-

renew and to undergo multi-lineage differentiation (Chambers et al., 2007). We identified the gene promoters that interacted with SEs in *Nanog*<sup>-/-</sup> ESCs and compared this list to those that we assigned in wild-type (WT) ESCs. Overall, ~80% (187) of the genes interacting with SEs in *Nanog*<sup>-/-</sup> ESCs are common to WT ESCs, demonstrating that NANOG is not essential to maintain most promoter-SE interactions in ESCs. Notably, we did identify a set of gene promoters that interacted with SEs only in *Nanog*<sup>-/-</sup> ESCs, such as *Elf3* and *Tet1* (47; 20%), and another set that interacted with SEs only in WT ESCs, including *Zfp281* and *Lefty1* (87; Figures 5A and 5B).

Interestingly, some of the changes in SE interactions upon *Nanog* deletion resembled the differences observed between ESCs and EpiSCs. For example, 24% of the newly interacting promoters in *Nanog*<sup>-/-</sup> ESCs were also present in EpiSCs (Figure S3C). Moreover, LRIs were particularly susceptible to the deletion of *Nanog*, as they were 6-fold reduced in *Nanog*<sup>-/-</sup> ESCs compared to WT ESCs (4% of all promoter-SE interactions span  $\geq 800$  kb, compared to the 25% in WT ESCs,  $p < 0.0001$ ; Figures 5C–5E). Deleting *Nanog* in ESCs, therefore, triggers an interaction rewiring that partially recapitulates the differences between ESCs and EpiSCs. These findings lead us to propose that the presence of NANOG contributes to the stability of a subset of ESC-specific SE contacts, and particularly those interactions that span large distances.

## DISCUSSION

In recent years, a subset of regulatory elements termed SEs have gained attention across a wide range of research fields (Hnisz et al., 2013; Whyte et al., 2013; Parker et al., 2013; Lovén et al., 2013). The interest in SEs stems from their potential to act as responsive, regulatory hubs that can control cell identity. Equally, questions have been raised as to whether SEs should be considered as separate entities from clusters of canonical enhancers or locus control regions (Moorthy et al., 2017; Bojcsuk et al., 2017; Hay et al., 2016; Pott and Lieb, 2015), and thus, it is fundamental to investigate the properties of SEs.

In this study, we generated high-resolution, global maps of promoter interactions in two distinct pluripotent cell types, ESCs and EpiSCs. We found that SEs form complex spatial networks in which individual SEs have the potential to regulate several genes. This observation expands on the one promoter to one SE assignments that have been proposed using Hi-C

### Figure 4. Long-Range Promoter-SE Interactions Are Enriched for NANOG Binding

(A) CODEX database was integrated to screen for the binding of transcription factors (TFs) at the promoter of genes involved in long- and short-range SE interactions in ESCs. The heatmap shows the top 25 TFs bound at promoters engaged in LRIs with SEs, normalized by column. The side panel shows the Benjamini-Hochberg adjusted chi-square  $p$  values for the proportion of long- versus short-range interactions bound/not bound by NANOG, POU5F1, SOX2, and ESRRB.

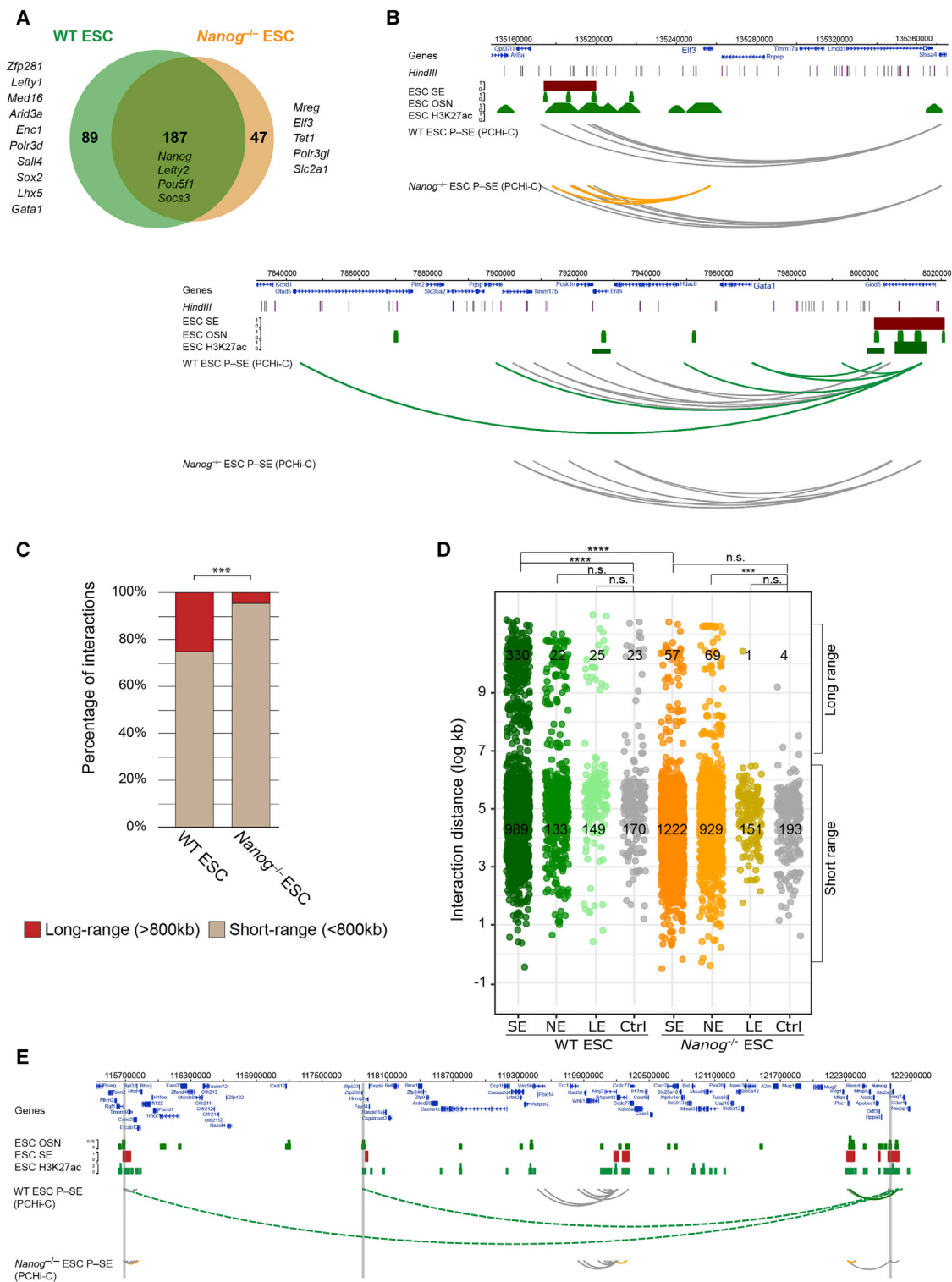
(B) Either end of the promoter-SE/CTRL contacts was classified according to the binding of NANOG. The majority (68%) of promoter-SE interactions in ESCs were bound by NANOG, particularly at the SE end (56%). The percentage of interactions within each category is shown. OE, other end of interaction (either SE or matched control). \*\* $p < 0.01$ ; \* $p < 0.05$  (Mann-Whitney  $U$  test).

(C) LRIs were significantly enriched for NANOG occupancy when compared to short-range (<800 kb) interactions in ESCs (green dots, Benjamini-Hochberg adjusted chi-square  $p = 2E-18$ ; Figure S3A). This enrichment was moderately significant in EpiSCs (blue dots, Benjamini-Hochberg adjusted chi-square  $p = 8E-03$ ) and not significant at control regions (gray dots).

(D) Genome browser illustrating an example of LRIs between promoters and SEs in ESCs (green dashed arcs) and in EpiSCs (blue dashed arcs).

(E) Gene expression levels (log<sub>2</sub> RPKM) of genes engaged in long- and short-range SE interactions in ESCs (Mann-Whitney  $U$  test;  $p = 0.01$ ).

See also Figure S3 and Tables S2 and S3.



**Figure 5. Long-Range Promoter-SE Interactions Are Depleted in *Nanog*-Deficient ESCs**

(A) Comparison of the SE-interactome between WT and *Nanog*<sup>-/-</sup> ESCs. The Venn diagram shows the overlap of genes that significantly interact with the 231 SEs (Whyte et al., 2013) in wild-type (WT) ESCs and *Nanog*<sup>-/-</sup> ESCs. A similar overlap was detected for ROSE-called WT SEs (not shown).

(B) Examples of significant promoter-SE interactions in *Nanog*<sup>-/-</sup> ESCs (top, *Eif3*) and in WT ESCs (bottom, *Gata1*). Significant interactions common to WT and *Nanog*<sup>-/-</sup> ESCs are shown as gray arcs, while orange or green arcs represent interactions detected only in the knockout or in the WT ESCs.

(legend continued on next page)

data and computational predictions (Whyte et al., 2013; Ing-Simmons et al., 2015). Whether this multi-modular form of gene regulation ensures robustness of transcriptional control or opportunities for co-regulation remains to be tested experimentally. The latter hypothesis is supported by a previous study where deleting an individual SE triggered the dysregulation of several nearby genes (Moorthy et al., 2017). Extending these observations to additional SEs, particularly in association with our promoter-SE interactome, will provide an exciting line of future studies.

Genes assigned to SEs have been proposed to confer unique cellular identities (Parker et al., 2013; Hnisz et al., 2013). By comparing the SE interactome of ESCs and EpiSCs, we detected the gain and loss of significant SE interactions at a subset of potentially important loci. For example, we detected a significant interaction between the *Klf4* promoter and a SE in ESCs only and between *Klf13* and a SE only in EpiSCs. Thus, the rewiring of SE interactions at genes encoding transcription factors could underlie the regulatory signals required for pluripotent state transitions. This is further supported by the differential enrichment for the binding of distinct sets of transcription factors at genes interacting with SEs in ESCs and EpiSCs, which might stabilize active SE interactions in ESCs while promoting the “priming” of new SE interactions upon transition to EpiSCs. However, we also found many genes interacting with SEs in both cell types that are not currently associated with pluripotency. For example, we uncovered a subset of genes engaged in SE interactions in ESCs with functions related to the synthesis of acetyl-coenzyme A (acetyl-CoA). The glycolytic production of acetyl-CoA promotes histone acetylation and helps to maintain the “open” chromatin organization of ESCs (Moussaieff et al., 2015). By connecting genes to highly active regulatory elements, our PCHI-C resource could be used to identify additional and previously unexplored mediators of pluripotent states.

Although the occupancy of core transcription factors is a hallmark of SEs, whether such factors contribute to SE function has not been investigated previously. We found that the deletion of *Nanog*, which is a SE-associated transcription factor in ESCs, had a surprisingly modest impact on the majority of promoter-SE contacts. However, we did observe that particular subsets of SEs were disrupted in several important ways. First, we detected a reduction in the span of promoter-SE interactions, such that significant LRIs were less frequent in *Nanog*-deficient ESCs. Second, interaction rewiring resulted in a different set of promoters in contact with SEs. We are unable to distinguish between whether the interaction differences were driven predominantly by the loss of SE activity in *Nanog*-deficient ESCs or by the loss of the interaction itself. We also cannot rule out that

some of these changes could be due to the increased prevalence of partially differentiated cells within the *Nanog*-deficient cultures. Nevertheless, it is interesting to note that both of these genomic changes resembled the SE interactome in EpiSCs. Given that NANOG levels are lower in EpiSCs than ESCs, differences between the pluripotent states seem to be accompanied by a reorganization of the network of promoters contacting regulatory regions. In particular, this involves the decommissioning of regulatory regions associated with promoters that are involved in pluripotency/stem cell maintenance (like *Zfp281* or *Sall4*) and the rewiring to promoters that are involved in chromatin organization and differentiation priming (such as *Tet1*, *Etf3*, and *Slc2a1*).

Another interesting feature that emerged from our analysis was the prevalence of LRIs between promoters and SEs in ESCs, with 25% of such interactions spanning 800 kb to 90 Mb. Notably, these LRIs were rarely detected in EpiSCs. This finding suggests that the ESC genome has a permissive organization, which enables a more highly folded spatial arrangement with DNA loops that can span greater distances. Interestingly, previous studies reported that ESCs form long-range promoter-promoter interactions, mainly involving the *Hox* clusters and other Polycomb-associated sites (Denholtz et al., 2013; Joshi et al., 2015; Schoenfelder et al., 2015b). In addition, a recently developed method called genome architecture mapping identified chromatin contacts from thin nuclear sections and found that TADs containing SEs are highly interactive in ESCs and their contacts can span up to 116 Mb (Beagrie et al., 2017). Together with our findings, this leads to a model in which ESCs are particularly permissive to establishing physical contacts with very distant sites, both at transcriptionally repressed promoters and at highly active promoter-SE regions. Importantly, very few significant long-range SE contacts were detected in EpiSCs or in *Nanog*<sup>-/-</sup> ESCs, suggesting that it might be a feature resultant from the distinct nuclear organization of WT ESCs. Indeed, we and others have previously shown that ESCs have a more open and dynamic chromatin architecture than EpiSCs or *Nanog*<sup>-/-</sup> ESCs (Meshorer and Misteli, 2006; Fussner et al., 2011; Efroni et al., 2008; Novo et al., 2016). One interesting exception to this is that LRIs at the *Hox* network seem to be unaffected by the global changes in chromatin organization, as these interactions remained in EpiSCs. Further experiments are needed to define how Polycomb-group proteins are able to control the large spatial *Hox* networks within this environment. As long-range promoter-SE interactions also seem to be largely associated with NANOG, we hypothesize that the increased chromatin compaction that is triggered by reduced NANOG levels in EpiSCs or deletion of *Nanog* in mutant ESCs could restrict the connectivity between SEs and their distal target promoters. Indeed, although there

(C) The proportion of long-range promoter-SE interactions was significantly reduced in *Nanog*<sup>-/-</sup> ESCs compared to WT ESCs (57 versus 330 LRIs, respectively; Benjamini-Hochberg adjusted chi-square  $p < 0.0001$ ).

(D) The proportion of long-range versus short-range promoter-SE interactions in WT ESCs is significantly higher compared to *Nanog*<sup>-/-</sup> ESCs (Benjamini-Hochberg adjusted chi-square  $p < 1E-05$ ). No significance in the proportion of long- versus short-range was detected between promoter-SE and promoter-CTRL in *Nanog*<sup>-/-</sup> ESCs.

(E) Promoter-SE interactions at a large region encompassing the *Nanog* locus. Grey arcs show promoter-SE interactions common to both WT and *Nanog*<sup>-/-</sup> ESCs, and orange and green arcs represent interactions detected in each, respectively. Dashed arcs indicate LRIs. Vertical gray lines denote SE interactions at sites bound by the pluripotency transcription factors OCT4, SOX2, and NANOG (OSN).

See also Figure S3 and Tables S1 and S3.

were few long-range contacts involving control regions, this number was further reduced in *Nanog*-depleted cells, which points toward a global constriction in interaction distances. Alongside these global effects, it is possible that NANOG also has a direct role in stabilizing long-range SE interactions in ESCs, and future work could examine this prediction through acute gain and loss of NANOG binding.

A regulatory network where several gene promoters are in close proximity to strong regulatory elements like SEs confers the potential for a rapid and dynamic response to altered stimuli. For example, variations in the recruitment of transcription factors or signaling pathway members to SEs could quickly impact the regulation of the genes in contact with the SEs. Interestingly, the joint engagement of multiple promoters and enhancers in “chromatin hubs” has been observed in other contexts, and it is possible that SEs could be involved in a similar mode of gene regulation (Tolhuis et al., 2002; Patrinos et al., 2004; Jiang et al., 2016). We did not find a correlation between global transcription changes and SE interactions, suggesting that the spatial association between a promoter and a SE is not predictive of the cell-type-specific transcriptional status of that gene and that additional features (e.g., other regulatory inputs, binding of transcription factors, chromatin accessibility) mediate the differential gene expression changes. Contacts between multiple active regions are common in individual cells (Beagrie et al., 2017), and single-cell approaches will be particularly valuable for interrogating the potential of individual SEs for multiple gene regulation.

Taken together, our work provides an annotated view of interactions between putative regulatory elements, including SEs, and their target promoters across multiple pluripotent cell types. These results lead us to conclude that SEs form spatial networks in pluripotent cells that are partly dependent on core transcription factor occupancy. The data can be used to better define the gene regulatory architecture in pluripotency and understand the transcriptional variability that is observed upon perturbation of regulatory elements. This work will also be informative for understanding gene regulation in other systems including adult stem cells (Adam et al., 2015) and alternative developmental models, and it has valuable applications for cell therapies that are currently focused on targeting SEs as drivers of disease (Hnisz et al., 2013).

## EXPERIMENTAL PROCEDURES

### PChI-C Quantification and Statistical Analysis

Raw sequencing reads were processed using Hi-C User Pipeline (HiCUP) (Wingett et al., 2015). Interactions were called using CHICAGO (Cairns et al., 2016), and the resulting p values were adjusted with a weighted false discovery control procedure. Interactions were called at the level of individual HindIII fragments based on two biological replicates for each cell type that were normalized and combined. Interactions with a CHICAGO score  $\geq 5$  were classified as high-confidence interactions. Additionally, interactions with scores  $\geq 4$  were included in the analysis if they scored  $>5$  in one biological replicate and  $\leq 5$  in both replicates of the other cell types. Table S1 lists significant interactions.

### SE Calling

H3K27ac ChIP-seq data for ESCs (Creyghton et al., 2010) and EpiSCs (Factor et al., 2014) were mapped to GRCm38/mm10. SEs were identified by ROSE

([https://bitbucket.org/young\\_computation/rose](https://bitbucket.org/young_computation/rose)) as described previously (Lovén et al., 2013; Whyte et al., 2013). Briefly, after model-based analysis of ChIP-seq (MACS) peaks identification, peaks within 12.5 kb distance of each other were stitched with a TSS exclusion zone size of 2 kb. The signal of stitched enhancers was determined by the total normalized number of reads minus the number of normalized reads in the input. Stitched and normalized peaks were then ranked by H3K27ac density and those higher than the inflection point on the density curve were defined as SEs (genome locations in Table S2).

### SE Interaction Calling

We improved the power of promoter-SE interaction calling by “stitching” the fragments together. Only *cis*-chromosomal interactions were considered. For each promoter-SE interaction test, we stitched together all HindIII fragments overlapping the appropriate SE, except for the bait and its adjacent fragments. We then assigned an observed and expected read count to each promoter-SE pair as follows. The observed count is the total number of reads between the baited fragment and any of the HindIII fragments overlapping the SE. Similarly, the expected count for the composite interaction is the sum of the expected Brownian counts retrieved for each bait-HindIII-fragment pair. Finally, we tested if the observed count was greater than expected by performing a one-tailed hypothesis test - the null distribution was Negative Binomial, with mean equal to the expected count and with dispersion parameter retrieved from the original CHICAGO analysis, thereby obtaining a p value. The p values were weighted and transformed into scores using the default methods of CHICAGO (Cairns et al., 2016), and scores  $\geq 5$  were classified as high-confidence interactions. Randomized control regions were defined by shuffling SEs within each chromosome until no two control regions overlapped. NE and LE regions were defined by H3K27ac signals. Each region was assigned randomly to an SE, and the regions were size extended from the midpoint of the fragment until their size matched the assigned SE. Extended regions that shared a fragment with an SE were discarded and replaced with an alternative size-matched region. A list of the interactions can be found in Table S2.

### LRI

The midpoints of each interacting genomic feature were subtracted to define the distances of individual interactions:  $\text{distance} = |(a + b)/2 - (c + d)/2|$ , where (a,b) and (c,d) are the (start,end) coordinates of each interacting genomic feature. 25% of SE-interactions in ESCs span  $>800$  kb, and this value was used as the threshold to define LRIs. To calculate the LRI enrichment, we compared the proportion of SE interactions spanning  $\geq 800$  kb in each cell type. The proportion of LRI versus short-range ( $<800$  kb) SE interactions was calculated and represented as percentages of all SE interactions for a cell type. Any proportion differences of LRI versus short-range SE-interactions between cell types was tested using the chi-square test of independence.

## DATA AND SOFTWARE AVAILABILITY

The accession number for the raw and processed sequencing data reported in this paper is GEO: GSE103053.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and three tables and can be found with this article online at <https://doi.org/10.1016/j.celrep.2018.02.040>.

## ACKNOWLEDGMENTS

We thank Kristina Tabbada and Clare Murnane at the Babraham Institute Sequencing Facility for sequencing support, Laura Biggins at the Babraham Bioinformatics Group for help with CODEX analysis, and members of the P.J.R.-G. group for helpful discussions and feedback. P.J.R.-G. is supported by the Wellcome Trust (WT093736), the Biotechnology and Biological Sciences Research Council (BB/M022285/1 and BB/P013406/1), and the



European Commission Network of Excellence EpiGeneSys (HEALTH-F4-2010-257082). This work was also supported by grants to P.F. from the Medical Research Council (MR/L007150/1, MC\_UP\_1302/1, MC\_UP\_1302/3, and MC\_UP\_1302/5), and the Biotechnology and Biological Sciences Research Council (BB/J004480/1).

## AUTHOR CONTRIBUTIONS

Conceptualization, C.L.N., P.J.R.-G.; Methodology, C.L.N., J.C., S.W.W.; Investigation, C.L.N., B.-M.J., M.F.-M., S.S.; Formal Analysis, C.L.N., J.C., A.S.-P., S.W.W., P.F.-P.; Writing, C.L.N., P.J.R.-G.; Visualization, C.L.N.; P.J.R.-G.; Resources, P.F., P.J.R.-G.; Supervision, P.F., P.J.R.-G.; Funding Acquisition, P.F., P.J.R.-G.; Project Administration, P.J.R.-G.

## DECLARATION OF INTEREST

The authors declare no competing interests.

Received: August 21, 2017

Revised: January 26, 2018

Accepted: February 9, 2018

Published: March 6, 2018

## REFERENCES

- Adam, R.C., Yang, H., Rockowitz, S., Larsen, S.B., Nikolova, M., Oristian, D.S., Polak, L., Kadaja, M., Asare, A., Zheng, D., and Fuchs, E. (2015). Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature* **521**, 366–370.
- 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.
- Beagan, J.A., Gilgenast, T.G., Kim, J., Plona, Z., Norton, H.K., Hu, G., Hsu, S.C., Shields, E.J., Lyu, X., Apostolou, E., et al. (2016). Local genome topology can exhibit an incompletely rewired 3D-folding state during somatic cell reprogramming. *Cell Stem Cell* **18**, 611–624.
- Beagrie, R.A., Scialdone, A., Schueler, M., Kraemer, D.C., Chotalia, M., Xie, S.Q., Barbieri, M., de Santiago, I., Lavitas, L.M., Branco, M.R., et al. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524.
- Blinka, S., Reimer, M.H., Jr., Pulakanti, K., and Rao, S. (2016). Super-enhancers at the Nanog locus differentially regulate neighboring pluripotency-associated genes. *Cell Rep.* **17**, 19–28.
- Bojcsuk, D., Nagy, G., and Balint, B.L. (2017). Inducible super-enhancers are organized based on canonical signal-specific transcription factor binding elements. *Nucleic Acids Res.* **45**, 3693–3706.
- Brons, I.G., Smithers, L.E., Trotter, M.W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., and Vallier, L. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191–195.
- Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838–853.
- Cairns, J., Freire-Pritchett, P., Wingett, S.W., Várnai, C., Dimond, A., Plagnol, V., Zerbino, D., Schoenfelder, S., Javierre, B.M., Osborne, C., et al. (2016). CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* **17**, 127.
- Cao, F., Fang, Y., Tan, H.K., Goh, Y., Choy, J.Y.H., Koh, B.T.H., Hao Tan, J., Bertin, N., Ramadass, A., Hunter, E., et al. (2017). Super-enhancers and broad H3K4me3 domains form complex gene regulatory circuits involving chromatin interactions. *Sci. Rep.* **7**, 2186.
- Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–1234.
- Chapuy, B., McKeown, M.R., Lin, C.Y., Monti, S., Roemer, M.G., Qi, J., Rahl, P.B., Sun, H.H., Yeda, K.T., Doench, J.G., et al. (2013). Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**, 777–790.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* **107**, 21931–21936.
- de Wit, E., Bouwman, B.A., Zhu, Y., Klous, P., Splinter, E., Verstegen, M.J., Krijger, P.H., Festuccia, N., Nora, E.P., Welling, M., et al. (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature* **501**, 227–231.
- Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., Pellegrini, M., and Plath, K. (2013). Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell* **13**, 602–616.
- Ding, J., Huang, X., Shao, N., Zhou, H., Lee, D.F., Faiola, F., Fidalgo, M., Guallar, D., Saunders, A., Shliaha, P.V., et al. (2015). Tex10 coordinates epigenetic control of super-enhancer activity in pluripotency and reprogramming. *Cell Stem Cell* **16**, 653–668.
- 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.
- Efroni, S., Dutttagupta, R., Cheng, J., Dehghani, H., Hoepfner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D., Buetow, K.H., et al. (2008). Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* **2**, 437–447.
- Factor, D.C., Corradin, O., Zentner, G.E., Saiakhova, A., Song, L., Chenoweth, J.G., McKay, R.D., Crawford, G.E., Scacheri, P.C., and Tesar, P.J. (2014). Epigenomic comparison reveals activation of “seed” enhancers during transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 854–863.
- Freire-Pritchett, P., Schoenfelder, S., Várnai, C., Wingett, S.W., Cairns, J., Collier, A.J., García-Vílchez, R., Furlan-Magaril, M., Osborne, C.S., Fraser, P., et al. (2017). Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells. *eLife* **6**, e21926.
- Fussner, E., Djuric, U., Strauss, M., Hotta, A., Perez-Iratxeta, C., Lanner, F., Dilworth, F.J., Ellis, J., and Bazett-Jones, D.P. (2011). Constitutive heterochromatin reorganization during somatic cell reprogramming. *EMBO J.* **30**, 1778–1789.
- Gaspar-Maia, A., Alajem, A., Meshorer, E., and Ramalho-Santos, M. (2011). Open chromatin in pluripotency and reprogramming. *Nat. Rev. Mol. Cell Biol.* **12**, 36–47.
- Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* **136**, 1063–1069.
- Hay, D., Hughes, J.R., Babbs, C., Davies, J.O.J., Graham, B.J., Hanssen, L., Kassouf, M.T., Marieke Oudelaar, A.M., Sharpe, J.A., Suci, M.C., et al. (2016). Genetic dissection of the  $\alpha$ -globin super-enhancer in vivo. *Nat. Genet.* **48**, 895–903.
- Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* **16**, 144–154.
- Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947.
- Ing-Simmons, E., Seitan, V.C., Faure, A.J., Flicek, P., Carroll, T., Dekker, J., Fisher, A.G., Lenhard, B., and Merckenschlager, M. (2015). Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. *Genome Res.* **25**, 504–513.
- Jiang, T., Raviram, R., Snetkova, V., Rocha, P.P., Proudhon, C., Badri, S., Bonneau, R., Skok, J.A., and Kluger, Y. (2016). Identification of multi-loci hubs from 4C-seq demonstrates the functional importance of simultaneous interactions. *Nucleic Acids Res.* **44**, 8714–8725.
- Joshi, O., Wang, S.Y., Kuznetsova, T., Atlasi, Y., Peng, T., Fabre, P.J., Habibi, E., Shaik, J., Saeed, S., Handoko, L., et al. (2015). Dynamic reorganization of

- extremely long-range promoter-promoter interactions between two states of pluripotency. *Cell Stem Cell* 17, 748–757.
- 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.
- Khan, A., and Zhang, X. (2016). dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res.* 44, 164–171.
- Krijger, P.H., Di Stefano, B., de Wit, E., Limone, F., van Oevelen, C., de Laat, W., and Graf, T. (2016). Cell-of-origin-specific 3D genome structure acquired during somatic cell reprogramming. *Cell Stem Cell* 18, 597–610.
- 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.
- Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153, 320–334.
- Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. *Nat. Rev. Mol. Cell Biol.* 7, 540–546.
- Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., et al. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* 47, 598–606.
- Moorthy, S.D., Davidson, S., Shchuka, V.M., Singh, G., Malek-Gilani, N., Langroudi, L., Martchenko, A., So, V., Macpherson, N.N., and Mitchell, J.A. (2017). Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. *Genome Res.* 27, 246–258.
- Moussaieff, A., Rouleau, M., Kitsberg, D., Cohen, M., Levy, G., Barasch, D., Nemirovski, A., Shen-Orr, S., Laevsky, I., Amit, M., et al. (2015). Glycolysis-mediated changes in acetyl-CoA and histone acetylation control the early differentiation of embryonic stem cells. *Cell Metab.* 21, 392–402.
- 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.
- Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted degradation of ctcf decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* 169, 930–944.e22.
- Novo, C.L., Tang, C., Ahmed, K., Djuric, U., Fussner, E., Mullin, N.P., Morgan, N.P., Hayre, J., Sienerth, A.R., Elderkin, S., et al. (2016). The pluripotency factor Nanog regulates pericentromeric heterochromatin organization in mouse embryonic stem cells. *Genes Dev.* 30, 1101–1115.
- Osorno, R., Tsakiridis, A., Wong, F., Cambray, N., Economou, C., Wilkie, R., Blin, G., Scotting, P.J., Chambers, I., and Wilson, V. (2012). The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. *Development* 139, 2288–2298.
- Papanayotou, C., Benhaddou, A., Camus, A., Perea-Gomez, A., Jouneau, A., Mezger, V., Langa, F., Ott, S., Sabéran-Djoneidi, D., and Collignon, J. (2014). A novel nodal enhancer dependent on pluripotency factors and smad2/3 signaling conditions a regulatory switch during epiblast maturation. *PLoS Biol.* 12, e1001890.
- Parker, S.C., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., Black, B.L., et al.; NISC Comparative Sequencing Program; National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program Authors; NISC Comparative Sequencing Program Authors (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. USA* 110, 17921–17926.
- Patrinos, G.P., de Krom, M., de Boer, E., Langeveld, A., Imam, A.M., Strouboulis, J., de Laat, W., and Grosveld, F.G. (2004). Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes Dev.* 18, 1495–1509.
- Pérez-Rico, Y.A., Boeva, V., Mallory, A.C., Bitetti, A., Majello, S., Barillot, E., and Shkumatava, A. (2017). Comparative analyses of super-enhancers reveal conserved elements in vertebrate genomes. *Genome Res.* 27, 259–268.
- 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.
- Pott, S., and Lieb, J.D. (2015). What are super-enhancers? *Nat. Genet.* 47, 8–12.
- Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680.
- Sánchez-Castillo, M., Ruau, D., Wilkinson, A.C., Ng, F.S., Hannah, R., Diamanti, E., Lombard, P., Wilson, N.K., and Gottgens, B. (2015). CODEX: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities. *Nucleic Acids Res.* 43, D1117–D1123.
- 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., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B.M., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S.W., et al. (2015a). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* 25, 582–597.
- Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B.M., Armstrong, H., Mifsud, B., Dimitrova, E., Matheson, L., Tavares-Cadete, F., Furlan-Magaril, M., et al. (2015b). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat. Genet.* 47, 1179–1186.
- Smith, E.M., Lajoie, B.R., Jain, G., and Dekker, J. (2016). Invariant TAD boundaries constrain cell-type-specific looping interactions between promoters and distal elements around the Cfr locus. *Am. J. Hum. Genet.* 98, 185–201.
- Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* 448, 196–199.
- Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* 10, 1453–1465.
- van Heyningen, V., and Bickmore, W. (2013). Regulation from a distance: long-range control of gene expression in development and disease. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368, 20120372.
- Vernimmen, D., and Bickmore, W.A. (2015). The hierarchy of transcriptional activation: from enhancer to promoter. *Trends Genet.* 31, 696–708.
- 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.
- Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., and Andrews, S. (2015). HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res.* 4, 1310.
- Zhang, J., Nuebel, E., Daley, G.Q., Koehler, C.M., and Teitell, M.A. (2012). Metabolic regulation in pluripotent stem cells during reprogramming and self-renewal. *Cell Stem Cell* 11, 589–595.
- Zhang, H., Badur, M.G., Divakaruni, A.S., Parker, S.J., Jäger, C., Hiller, K., Murphy, A.N., and Metallo, C.M. (2016). Distinct metabolic states can support self-renewal and lipogenesis in human pluripotent stem cells under different culture conditions. *Cell Rep.* 16, 1536–1547.
- Zhou, H.Y., Katsman, Y., Dhaliwal, N.K., Davidson, S., Macpherson, N.N., Sakthidevi, M., Collura, F., and Mitchell, J.A. (2014). A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes Dev.* 28, 2699–2711.

Cell Reports, Volume 22

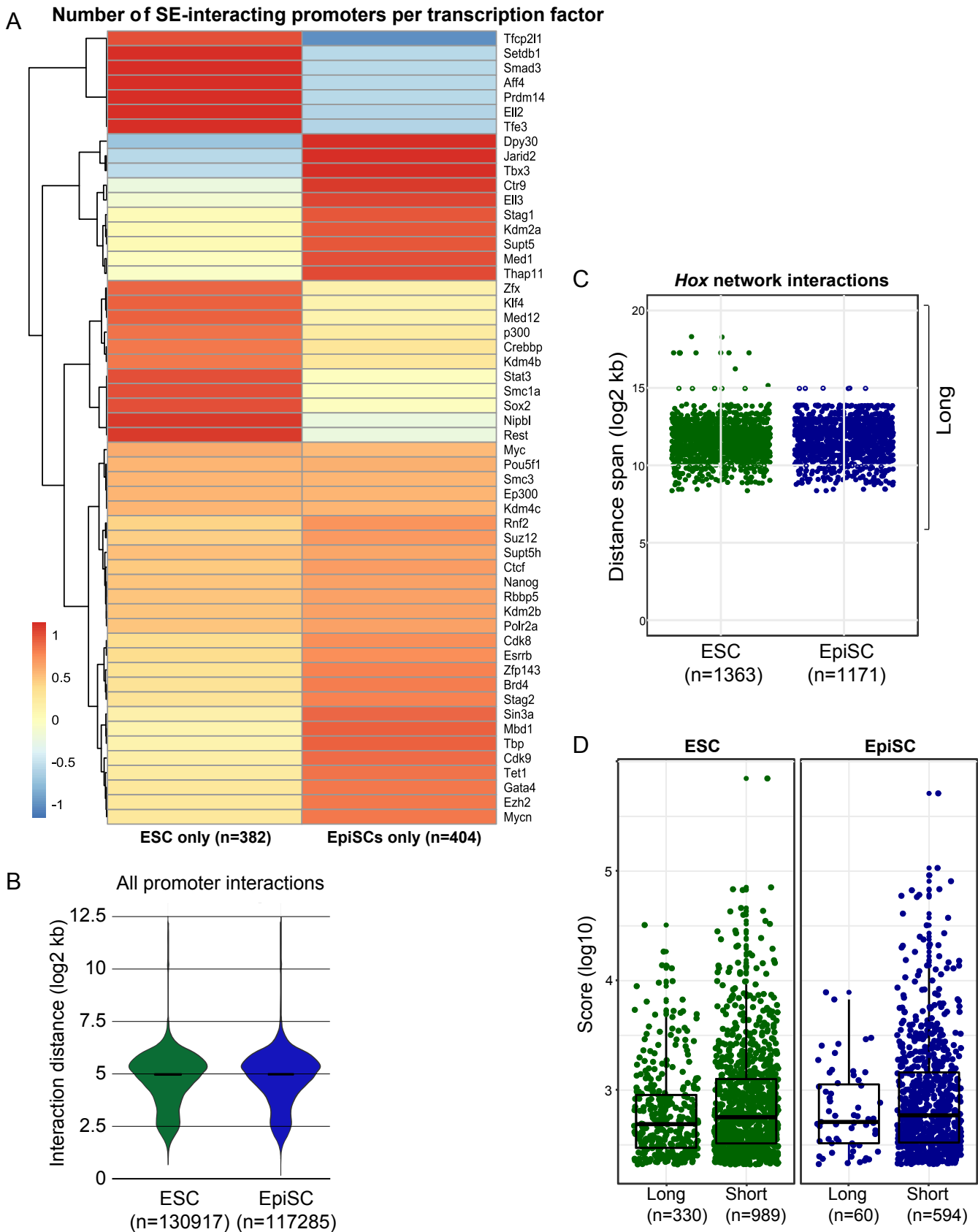
## Supplemental Information

### **Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition**

**Clara Lopes Novo, Biola-Maria Javierre, Jonathan Cairns, Anne Segonds-Pichon, Steven W. Wingett, Paula Freire-Pritchett, Mayra Furlan-Magaril, Stefan Schoenfelder, Peter Fraser, and Peter J. Rugg-Gunn**

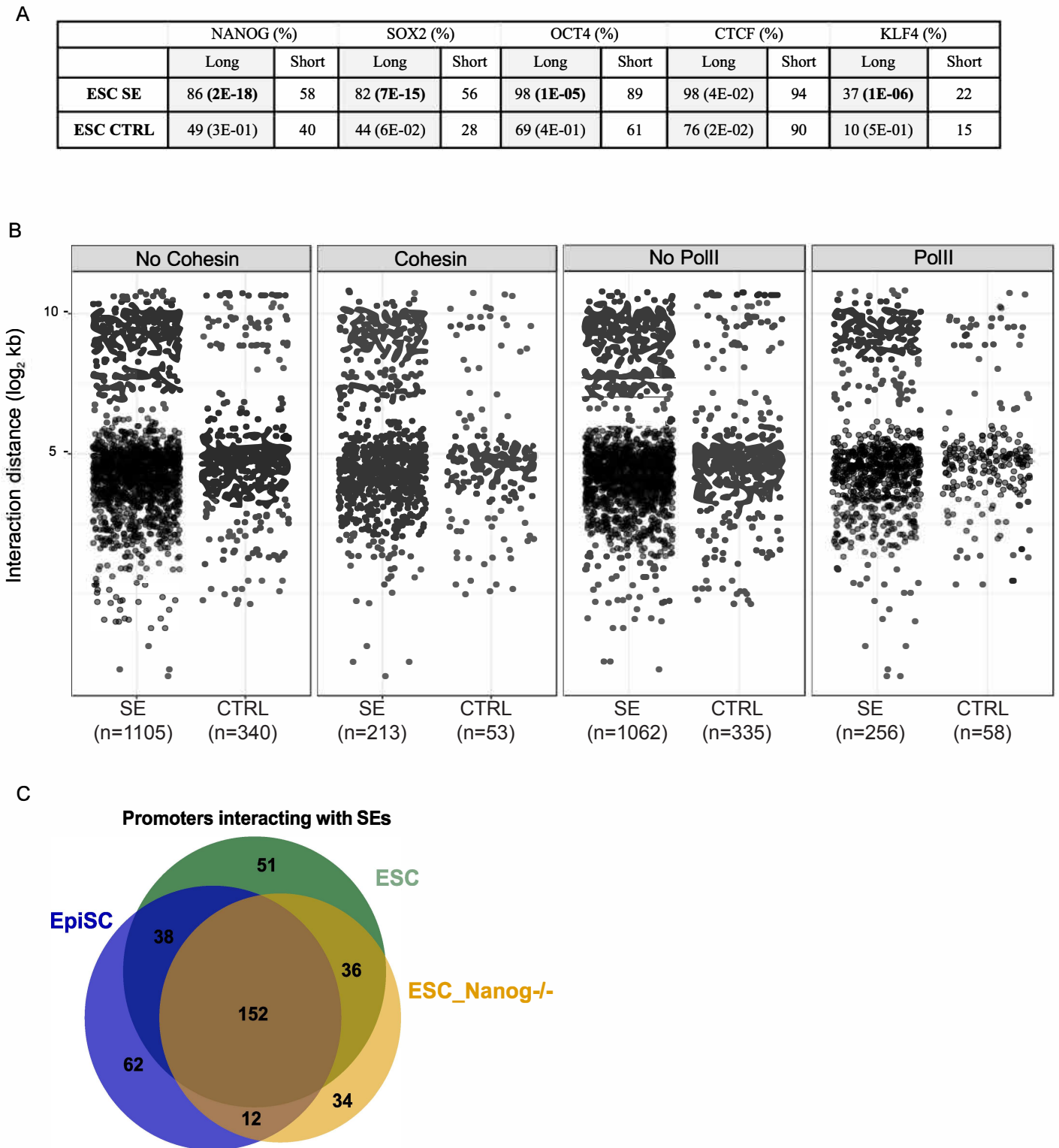






**Figure S2, related to Figure 2 - Characterization of SE interactome**

- a)** The number of SE-interacting promoters bound by each transcription factor in the CODEX ChIPseq database was calculated. The heatmap shows these values, mean centred and scaled by row, for ESC-only and EpiSC-only sets. **b)** Violin plot of the genomic distances (log<sub>2</sub> kb) of all significant interactions in ESCs and EpiSCs. **c)** Genomic distances (log<sub>2</sub> kb) of significant interactions within the *Hox* network in both pluripotent states. **d)** Dotplot of CHIcAGO scores of long- and short-range interactions between SE and promoters, in both cell-lines.



**Figure S3, related to Figure 4 - Screening for different genomic features at SE-P in all cell-types analysed.**

a) Percentage of long-range (>800kb) or short-range (<800kb) interactions bound by different proteins in ESCs for SEs and for control regions (CTRL). In brackets, the Benjamini-Hochberg adjusted Chi-squared p-values for the pairwise comparison between the proportion of bound/not bound in long-range versus short-range interactions are shown. Significant differences are highlighted in bold.

b) Each SE or Control - P interaction was annotated for the binding of Cohesin or POUi and plotted according to the genomic distance.

c) Venn-diagram with the promoters interacting with SE (Whyte et al 2013) in the three pluripotent cell-lines analysed.

## Supplemental Experimental Procedures

### Cell culture

E14Tg2a (129P2/OlaHsd (Hooper et al., 1987) and RCN $\beta$ H-B(t) (E14Tg2a-derived Nanog<sup>-/-</sup>, passages 20-30) (Chambers et al., 2007) ESCs were cultured on gelatin-coated surfaces in standard ESC media (DMEM supplemented with 15% FBS, 1mM sodium pyruvate, 0.1mM 2-mercaptoethanol, 0.1mM non-essential amino acids, 2mM glutamax and 1000 U/ml LIF), and 25 $\mu$ g/ml hygromycin was added during RCN $\beta$ H-B(t) ESC expansion to select for undifferentiated Nanog<sup>-/-</sup> cells. Embryo-derived 129S2 (Brons et al., 2007) EpiSC were cultured on fibronectin (10 $\mu$ g/ml) in N2B27 media supplemented with 20ng/ml Activin A and 12ng/ml bFGF.

### Hi-C and PCHi-C preparation

Promoter Capture Hi-C was carried out with SureSelect target enrichment, using a custom-designed biotinylated RNA bait library for mouse promoters and custom paired-end blockers according to the manufacturer's instructions (Agilent Technologies) (Schoenfelder et al., 2015a) and following the in-nucleus Hi-C protocol detailed in (Nagano et al., 2015). Briefly, HindIII-digested chromatin was end-filled with biotinylated d-ATP and ligated in preserved nuclei. After de-crosslinking, DNA was purified and sheared to an average size of 400bp (Covaris). Sheared ends were repaired and adenine-tailed prior to immobilization of the biotinylated ligation fragments using MyOne Streptavidin C1 DynaBeads (Invitrogen). Promoter capture Hi-C was performed with 500ng of Hi-C library using the custom-made RNA bait library mentioned above. After paired-end adapter ligation, libraries were amplified by PCR for four cycles.

### ChIP-seq analysis and data processing

ChIP-seq in EpiSCs was performed according to (Schoenfelder et al., 2015b), with 200  $\mu$ g of chromatin and 5  $\mu$ g of antibody for each IP (NANOG, ab80892 Abcam; OCT4 sc-5279, Santa Cruz and SOX2, ab97959 Abcam). Libraries were prepared using the NEBNext DNA Library Preparation kit and sequenced on an Illumina HiSeq2500.

The publicly available ChIP-seq data sets used are listed in Supplementary Table 2. Raw ChIP-seq data was processed by re-alignment of reads to GRCm38/mm10 using Bowtie with default parameters (Langmead et al., 2009), followed by MACS peak-calling (Zhang et al., 2008) to generate a map of protein binding sites (using p-value of 1E-9). When raw data was not available, binding sites (peak location) and, when required, genome coordinates were converted to mm10 using the UCSC Genome Browser liftOver command line tool (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>). Interactions were considered to be occupied by a protein factor if they overlapped at least one ChIP-seq peak (bedtools).

### RNA-seq libraries and analysis

Total RNA was purified using an RNeasy kit (Qiagen). Indexed mRNA-Seq libraries were constructed from 500ng total RNA using the Tru-Seq RNA Library Prep Kit v2 (Illumina). Library fragment size and concentration was determined using an Agilent Bioanalyzer 2100 and KAPA Library Quantification Kit (KAPA Biosystems). Samples were sequenced on an Illumina HiSeq as 50bp or 100bp single-end libraries at the Babraham Institute Sequencing Facility.

Reads were trimmed using trim galore ([www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)) using default parameters to remove the Illumina adapter sequence. Reads were mapped to the mouse NCBI37 genome assembly using TopHat guided by the gene models from Ensembl. BAM files were imported to Seqmonk ([www.bioinformatics.babraham.ac.uk/projects/seqmonk/](http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/)). Read counts per transcript were calculated using the RNA-Seq quantitation pipeline using non-directional counts. Raw read counts per transcript were calculated using the RNA-seq quantitation pipeline on the Ensembl v70 gene set using non-directional counts. Differential analysis of gene expression was

performed using the default settings in DESeq2 (Love et al., 2014) without independent filtering of the results. Differentially expressed genes were called at  $\text{padj} < 0.05$  and  $\log_2$  fold change above 1.5 or below  $-1.5$ .

#### Statistical analysis

Proportions of long versus short-range and bound versus not bound interactions were compared using Chi-Square tests followed by Benjamini-Hochberg corrections to account for multiple comparisons.

#### CODEX analysis

The number of genes bound by each transcription factor in the CODEX ChIPseq database was calculated for both sets of genes interacting with SE in only ESC or EpiSC (ESC- and EpiSC-only, respectively) and for a set common to EpiSC and ESC. Only the CODEX ChIP-seq peak sites that overlapped TSSs were included in the analysis. The pheatmap package in R was used and values were scaled by row to enable the relative number of genes per transcription factor to be compared between the EpiSConly, ESConly and common datasets. The default scaling method in R was used so, for each set of 3 values, the data was mean centred and then divided by the standard deviation of the 3 values. For long- versus short-range interactions in ESC, the analysis performed was similar except for the scaling, which was performed by column using the dplyr package in R.



## Supplemental References

BRONS, I. G., SMITHERS, L. E., TROTTER, M. W., RUGG-GUNN, P., SUN, B., CHUVA DE SOUSA LOPES, S. M., HOWLETT, S. K., CLARKSON, A., AHRLUND-RICHTER, L., et al. 2007. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*, 448, 191-5.

CHAMBERS, I., SILVA, J., COLBY, D., NICHOLS, J., NIJMEIJER, B., ROBERTSON, M., VRANA, J., JONES, K., GROTEWOLD, L. & SMITH, A. 2007. Nanog safeguards pluripotency and mediates germline development. *Nature*, 450, 1230-4.

HOOPER, M., HARDY, K., HANDYSIDE, A., HUNTER, S. & MONK, M. 1987. HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature*, 326, 292-5.

LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10, R25.

NAGANO, T., VARNAI, C., SCHOENFELDER, S., JAVIERRE, B. M., WINGETT, S. W. & FRASER, P. 2015. Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome Biol*, 16, 175.

SCHOENFELDER, S., FURLAN-MAGARIL, M., MIFSUD, B., TAVARES-CADETE, F., SUGAR, R., JAVIERRE, B. M., NAGANO, T., KATSMAN, Y., et al 2015a. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res*, 25, 582-97.

SCHOENFELDER, S., SUGAR, R., DIMOND, A., JAVIERRE, B. M., ARMSTRONG, H., MIFSUD, B., DIMITROVA, E., MATHESON, L., TAVARES-CADETE, F., et al. 2015b. Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat Genet*, 47, 1179-86.

ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSBAUM, C., MYERS, R. M., et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, 9, R137.