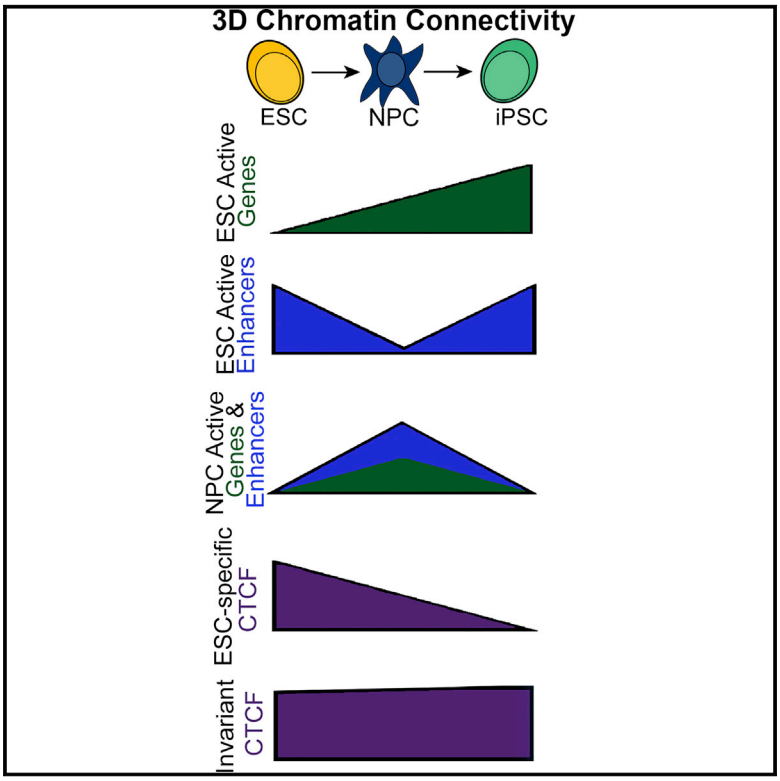
Article

Cell Stem Cell

Local Genome Topology Can Exhibit an Incompletely Rewired 3D-Folding State during Somatic Cell Reprogramming

Graphical Abstract



Authors

Jonathan A. Beagan, Thomas G. Gilgenast, Jesi Kim, ..., Victor G. Corces, Job Dekker, Jennifer E. Phillips-Cremins

Correspondence

jcremins@seas.upenn.edu

In Brief

Phillips-Cremins and colleagues report high-resolution chromatin folding maps in primary NPCs and NPC-derived induced pluripotent stem cells. They find that iPSC genomes can exhibit an imperfectly rewired 3D-folding state linked to poorly reprogrammed, ESC-specific CTCF occupancy and inaccurately reprogrammed gene expression levels. 2i/LIF conditions can fully restore distinct topological hallmarks of pluripotency.

Highlights

- 3D genome architecture is markedly reconfigured during reprogramming
- Some pluripotency genes engage in persistent, NPC-like interactions in iPSCs that break apart in 2i
- ESC-specific interactions that do not reconnect in iPSCs exhibit decreased CTCF binding
- Imperfectly rewired iPSC genome topology is linked to inaccurately reprogrammed expression

Accession Numbers

GSE68582

Beagan et al., 2016, Cell Stem Cell *18*, 611–624
May 5, 2016
http://dx.doi.org/10.1016/j.stem.2016.04.004

CellPress

Cell Stem Cell
Article

Local Genome Topology Can Exhibit an Incompletely Rewired 3D-Folding State during Somatic Cell Reprogramming

Jonathan A. Beagan,[1] Thomas G. Gilgenast,[1] Jesi Kim,[1] Zachary Plona,[1] Heidi K. Norton,[1] Gui Hu,[1] Sarah C. Hsu,[2] Emily J. Shields,[2] Xiaowen Lyu,[3] Effie Apostolou,[5,6] Konrad Hochedlinger,[5] Victor G. Corces,[3] Job Dekker,[4] and Jennifer E. Phillips-Cremins[1,2,*]

[1]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, USA
[2]Epigenetics Program, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[3]Department of Biology, Emory University, Atlanta, GA 30322, USA
[4]Howard Hughes Medical Institute, Program in Systems Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA
[5]Massachusetts General Hospital Cancer Center and Center for Regenerative Medicine, Boston, MA 02114, USA
[6]Present address: Meyer Cancer Center and Department of Medicine, Weill Cornell Medicine, New York, NY 10021, USA
*Correspondence: jcremins@seas.upenn.edu
http://dx.doi.org/10.1016/j.stem.2016.04.004

SUMMARY

Pluripotent genomes are folded in a topological hierarchy that reorganizes during differentiation. The extent to which chromatin architecture is reconfigured during somatic cell reprogramming is poorly understood. Here we integrate fine-resolution architecture maps with epigenetic marks and gene expression in embryonic stem cells (ESCs), neural progenitor cells (NPCs), and NPC-derived induced pluripotent stem cells (iPSCs). We find that most pluripotency genes reconnect to target enhancers during reprogramming. Unexpectedly, some NPC interactions around pluripotency genes persist in our iPSC clone. Pluripotency genes engaged in both "fully-reprogrammed" and "persistent-NPC" interactions exhibit over/undershooting of target expression levels in iPSCs. Additionally, we identify a subset of "poorly reprogrammed" interactions that do not reconnect in iPSCs and display only partially recovered, ESC-specific CTCF occupancy. 2i/LIF can abrogate persistent-NPC interactions, recover poorly reprogrammed interactions, reinstate CTCF occupancy, and restore expression levels. Our results demonstrate that iPSC genomes can exhibit imperfectly rewired 3D-folding linked to inaccurately reprogrammed gene expression.

INTRODUCTION

Mammalian genomes are folded in a hierarchy of architectural configurations that are intricately linked to cellular function. Individual chromosomes are arranged in distinct territories and then further partitioned into a nested series of Megabase (Mb)-sized topologically associating domains (TADs) (Dixon et al., 2012; Nora et al., 2012) and smaller sub-domains (termed subTADs) (Phillips-Cremins et al., 2013; Rao et al., 2014). TADs/subTADs vary widely in size (i.e., 40 kilobase [kb] to 3 Mb) and are characterized by highly interacting chromatin fragments demarcated by boundaries of abruptly decreased contact frequency. Long-range looping interactions connect distal genomic loci within and between TADs/subTADs (Jin et al., 2013; Phillips-Cremins et al., 2013; Rao et al., 2014; Sanyal et al., 2012). Single TADs, or a series of successive TAD/subTADs, in turn congregate into spatially proximal, higher-order clusters termed A/B compartments. Compartments generally fall into two classes: (1) "A" compartments enriched for open chromatin, highly expressed genes, and early replication timing and (2) "B" compartments enriched for closed chromatin, late replication timing, and co-localization with the nuclear periphery (Dixon et al., 2015; Lieberman-Aiden et al., 2009; Pope et al., 2014; Rao et al., 2014). The organizing principles governing genome folding at each length scale remain poorly understood.

Recent high-throughput genomics studies have shed new light on the dynamic nature of chromatin folding during embryonic stem cell (ESC) differentiation. Up to 25% of compartments in human ESCs switch their A/B orientation upon differentiation (Dixon et al., 2015). Compartments that switch between A and B configurations display a modest but correlated alteration in expression of only a small number of genes, suggesting that compartmental switching does not deterministically regulate cell-type-specific gene expression (Dixon et al., 2015). Similarly, lamina associated domains are dynamically altered during ESC differentiation (Peric-Hupkes et al., 2010). For example, the *Oct4*, *Nanog*, and *Klf4* genes relocate to the nuclear periphery in parallel with their loss of transcriptional activity as ESCs differentiate to astrocytes. TADs are largely invariant across cell types and often maintain their boundaries irrespective of the expression of their resident genes (Dixon et al., 2012). By contrast, long-range looping interactions within and between subTADs are highly dynamic during ESC differentiation (Phillips-Cremins et al., 2013; Zhang et al., 2013b). Pluripotency genes connect to their target enhancers through long-range interactions and disruption of these interactions leads to a marked decrease in gene expression (Apostolou et al., 2013; Kagey et al., 2010).

Thus, data are so far consistent with a model in which chromatin interactions at the sub-Mb scale (within TADs) are key effectors in the spatiotemporal regulation of gene expression during development.

In addition to the forward progression of ESCs in development, somatic cells can also be reprogrammed in the reverse direction to induced pluripotent stem cells (iPSCs) via the ectopic expression of key transcription factors (Takahashi and Yamanaka, 2006). Since the initial pioneering discovery, many population-based and single-cell genomics studies have explored the molecular underpinnings of transcription factor-mediated reprogramming (Hanna et al., 2009; Koche et al., 2011; Rais et al., 2013; Soufi et al., 2012). Recent efforts have uncovered changes in transcription, cell surface markers, and classic epigenetic modifications during intermediate stages in the reprogramming process (Buganim et al., 2012; Lujan et al., 2015; Polo et al., 2012). Although there is some evidence of epigenetic traces from the somatic cell of origin (Bock et al., 2011; Kim et al., 2010; Polo et al., 2010), the emerging model is that ESC-like epigenetic and transcriptional states can be generally reset under proper reprogramming conditions (Stadtfeld et al., 2010).

The role for chromatin topology in the acquisition of pluripotency during reprogramming has not yet been elucidated. Recent studies have suggested that specific long-range interactions between pluripotency genes such as *Nanog* and/or *Oct4* and target enhancers can be reset during reprogramming and precede reactivation of the involved genes (Apostolou et al., 2013; de Wit et al., 2013; Denholtz et al., 2013; Wei et al., 2013; Zhang et al., 2013a). Beyond these initial locus-specific studies, it remains unknown whether the somatic cell genome unfolds/refolds at the sub-Mb scale within TADs and how chromatin topology is linked to gene expression changes during reprogramming. Here we report a detailed analysis of local chromatin folding changes during somatic cell reprogramming. We created ~4–12 kb resolution chromatin architecture maps in primary neural progenitor cells (NPCs), iPSCs derived from primary NPCs, and pluripotent ESCs. We employed Chromosome-Conformation-Capture-Carbon-Copy (5C) to query fine-scale architectural changes in Mb-sized regions around key developmentally regulated genes. We find that chromatin folding is markedly reconfigured within TADs during the transition from primary NPCs to iPSCs. In many cases, pluripotency genes re-engage in fully reprogrammed interactions with their target ESC-specific enhancers. Unexpectedly, we also observe NPC interactions around key pluripotency genes (e.g., *Sox2* and *Klf4*) that remain persistently tethered in our iPSC clone. Pluripotency genes engaged in "persistent NPC-like" interactions can exhibit over/undershooting of gene expression levels in iPSCs, despite the fact that they may have also re-established contact with their target ESC-specific enhancer(s). We also uncover a subset of "poorly reprogrammed" interactions that break apart during differentiation and do not fully reconnect in our iPSC clone. Many poorly reprogrammed interactions exhibit ESC-specific CTCF occupancy that is lost during differentiation and only partially recovered in iPSCs. Importantly, 2i/LIF conditions can (1) abrogate persistent NPC-like interactions, (2) recover poorly reprogrammed interactions, (3) reinstate inadequately reprogrammed CTCF occupancy, and (4) restore precise gene expression levels.

RESULTS

Chromatin Folding Markedly Reconfigures at the Sub-Mb Scale during Reprogramming

To investigate changes in 3D chromatin topology during somatic cell reprogramming, we first generated ~4–12 kb resolution chromatin architecture maps in primary NPCs, iPSCs derived from primary NPCs, and ESCs (Figure 1A). To achieve a comparable genetic background to our pluripotency model (V6.5 ESCs; 129/SvJae × C57BL/6), we selected a previously published iPSC clone derived from primary NPCs isolated from neonatal brains of Sox2-GFP indicator mice (mixed 129/SvJae × C57BL/6 genetic background) (Eminli et al., 2008; Stadtfeld et al., 2008). Hochedlinger and colleagues generated this iPSC clone via the transduction of primary Sox2-GFP NPCs with doxycycline-inducible lentiviral vectors encoding Oct4, Klf4, and c-Myc. Importantly, this iPSC clone was extensively characterized for its pluripotent properties as assessed by (1) expression of endogenous pluripotency markers (Oct4, Sox2, and Nanog), (2) demethylation of *Oct4* and *Nanog* promoters, (3) transgene-independent self-renewal, (4) in vivo teratoma formation of all three germ layers, and (5) generation of chimeric mice (Eminli et al., 2008). Our three cellular states enable a detailed analysis of how chromatin unfolds/refolds between NPCs and iPSCs and also facilitate the comparison of genome topology between ESCs/iPSCs of comparable genetic background.

We employed 5C and high-throughput sequencing to create fine-scale chromatin architecture maps spanning >7 Mb of the mouse genome within a set of TADs (Dostie et al., 2006). 5C combines Chromosome-Conformation-Capture (3C) with a primer-based hybrid capture step to facilitate cost-effective detection of sub-Mb-scale interactions in Mb-sized loci of interest (Dekker et al., 2013). We used a tiled/alternating primer design around *Nanog*, *Sox2*, *Klf4*, *Oct4*, *Nestin*, and *Olig1–Olig2* (described in detail in Phillips-Cremins et al., 2013). Our 5C primer design scheme enabled the creation of ~4–12 kb resolution architecture maps for all loci combined across three cellular states with fewer than 30 million reads per replicate (Table S1). The power in this approach is that it focuses on elucidating fine-scale architecture changes at the sub-Mb scale within TADs (Figure 1B).

We first visualized 5C data with contact frequency heatmaps. To resolve underlying topological features, we developed an analysis pipeline to correct for known biases in 5C data and to normalize samples within and between biological replicates (described in detail in the Supplemental Experimental Procedures). Briefly, raw data (Figure S1A) were quantile normalized to bring the dynamic range of all samples onto equivalent scales and to account for technical differences in sequencing depth and library complexity (Figure S1B). To account for differences in primer efficiency that lead to non-uniformities in coverage across genomic regions, we applied our previously published primer correction algorithm to quantile-normalized data (Figure S1C; Phillips-Cremins et al., 2013). We then applied a blocked binning/smoothing algorithm to attenuate spatial noise in 5C data (Figure S1D). Our "Relative Contact Frequency" heatmaps revealed striking topological patterns that are dynamic across cellular states and are unique to each genomic region (Figure 1C).

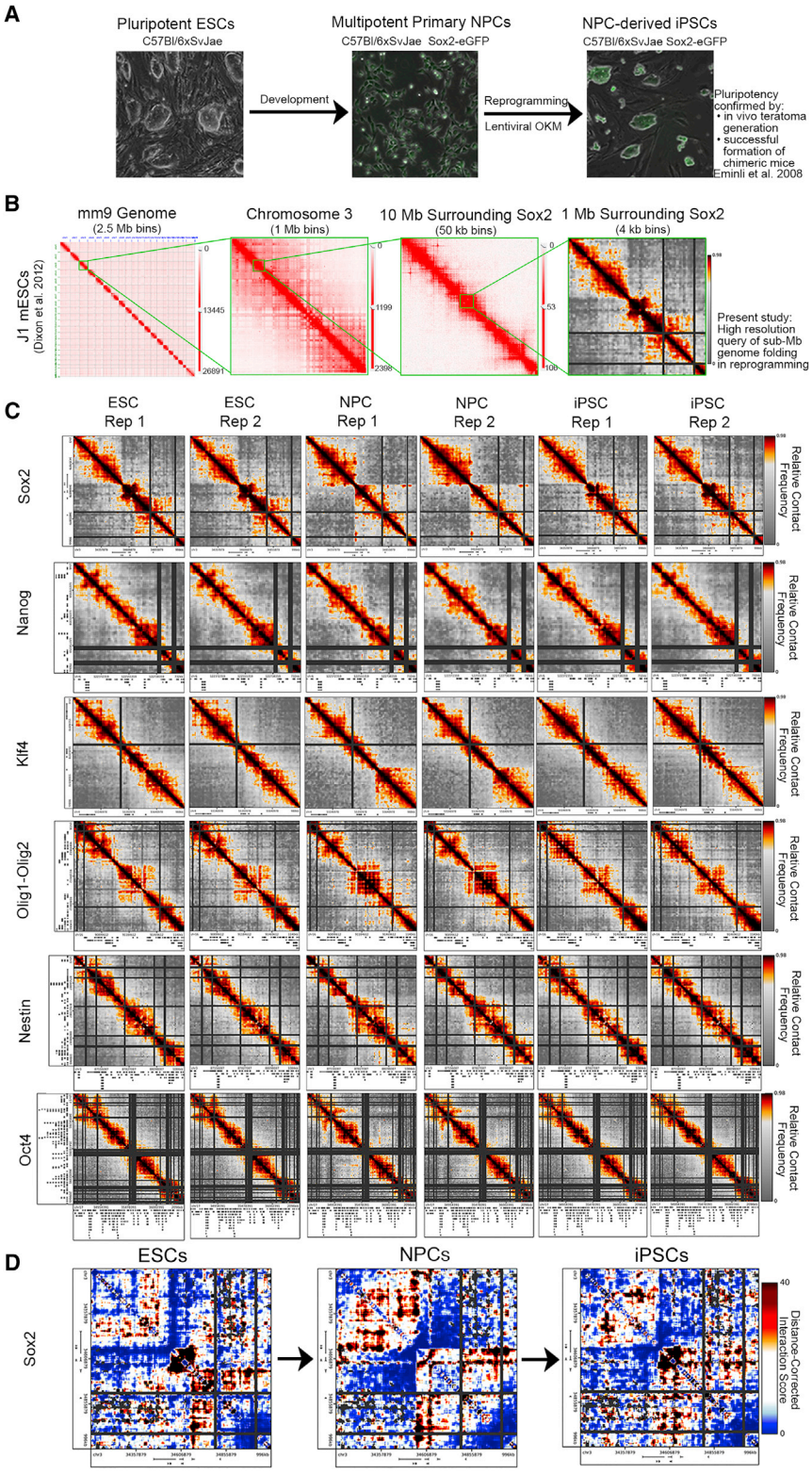


Figure 1. High-Resolution Architecture Maps Reveal Marked Chromatin Reconfiguration during Somatic Cell Reprogramming

(A) Phase contrast images of the reprogramming model system.

(B) Genome-wide ESC Hi-C data (Dixon et al., 2012) at different bin sizes illustrating chromosome territories, A/B compartments, and TADs. Images made with the Juicebox tool (http://www.aidenlab.org/juicebox/). The 4–12 kb resolution heatmaps from the present study query fine-scale genome folding at the sub-Mb scale within TADs.

(C) Relative contact frequency heatmaps are displayed for all biological replicates and regions queried. Color bars range from low (gray) to high (red/black) interaction frequencies.

(D) Distance-corrected interaction score heatmaps for a select region around the *Sox2* gene illustrating the presence of dynamic chromatin architecture among ESCs, NPCs, and iPSCs. Color bars range from low (blue) to high (red/black) interaction scores.

To further resolve the underlying architectural signal, we corrected for the known distance-dependence background in 5C data (Sanyal et al., 2012) (Figures S1E–S1G). Consistent with recent reports (Rao et al., 2014), we found that a local distance-dependence model computed independently for each region would more precisely account for locus-specific differences in chromatin folding that are often over/underestimated by a global background model (Figure S1G). Our "Distance-Corrected Interaction Score" heatmaps showed striking changes in topological features among NPCs, iPSCs, and ESCs (Figure 1D, Figures S1E and S1F), with high consistency between replicates and marked differences among biological conditions (Table S2). A systematic comparative analysis at each stage in the pipeline confirmed that we have reduced known biases in 5C data (Figures S1A–S1I and S2A–S2G).

iPSC Genomes Can Exhibit Imperfectly Rewired Folding Patterns

We next explored fine-scale chromatin folding features within TADs by visually inspecting our heatmaps. Consistent with our previous work (Phillips-Cremins et al., 2013), we observed marked changes in chromatin architecture between ESCs and NPCs. Importantly, we also noticed a striking architectural reconfiguration between NPCs and NPC-derived iPSCs (Figures 1C and 1D). At many loci, iPSC genome folding recapitulated the patterns seen in V6.5 ESCs. However, we also noticed several intriguing cases where iPSC topology retained remnants of the folding patterns from NPCs (Figure 1D).

To further explore the possibility that genome folding might be mis-wired during reprogramming, we conducted principal

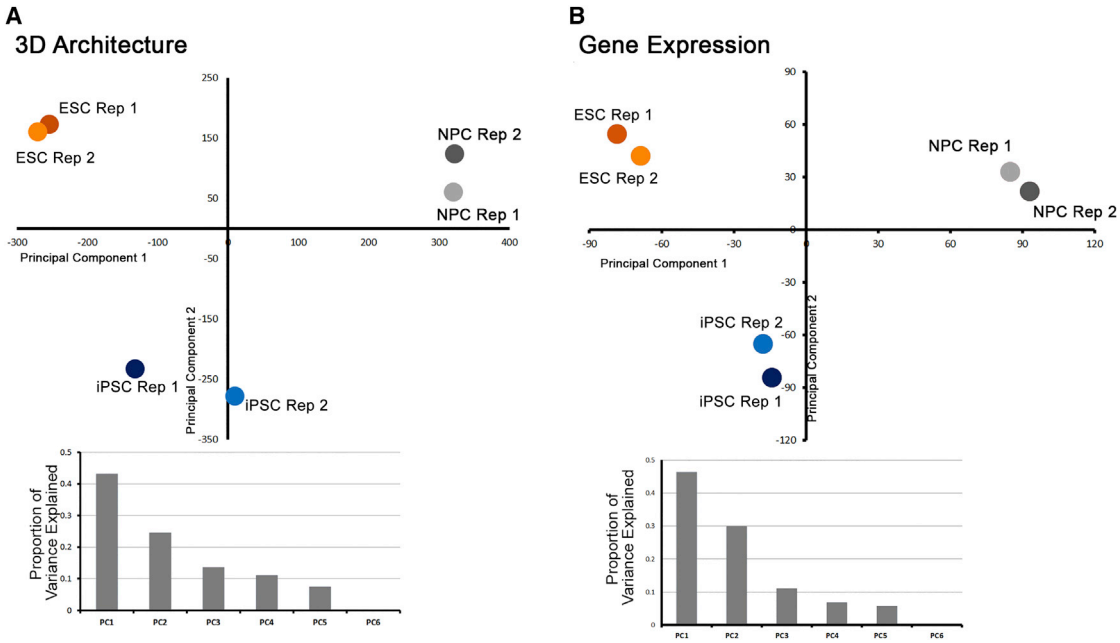


Figure 2. iPSC Genomes Can Exhibit Intermediate Folding and Expression Patterns between Somatic and Pluripotent Stem Cell States

Principal component analysis of (A) Distance-Corrected Interaction Frequency data and (B) normalized RNA-seq data for ESC, NPC, and iPSC replicates. (A and B) Principal components 1 and 2 are scattered and the proportion of variance explained by each principal component is plotted below each scatterplot.

component analysis on our "Distance-Corrected Interaction Frequency" data across all replicates and cellular states. Interestingly, we observed that genome topology in our iPSC clone exhibited folding patterns that were intermediate between NPCs and the pluripotent stem cell state (Figure 2A). To explore the functional significance of potential intermediate iPSC folding patterns, we queried the transcriptome of all three cellular states using RNA-seq. Consistent with our 3D observations, global gene expression profiles in our iPSC clone were also parsed as intermediate between ESCs and NPCs (Figure 2B). Together, these results support the possibility that genome architecture of some iPSC clones might be imperfectly wired within TADs during reprogramming.

Dynamic 3D Interaction Classes during Cell Fate Transitions

To identify high-confidence, long-range interactions across all developmentally regulated loci, we fit our Distance-Corrected Interaction Frequency data with a logistic distribution with location/scale parameters computed independently for each region (Figure S3A, Supplemental Experimental Procedures). We then converted the p value from our fitted models into an interaction score (−10*log2(p values)) that is comparable within and between experiments and allows the robust detection of interactions that are significant above the expected background signal.

We next employed a thresholding strategy to classify 3D interactions by their dynamic contact frequencies across the three cellular states (Figures 3A–3D). To minimize false positives, we required that interaction scores cross the threshold boundaries in both replicates for a given biological condition. Moreover, we iteratively defined thresholds to achieve an empirical False Discovery Rate (eFDR) of <10% when applied to simulated 5C replicates (Figures 3E–3H, Figures S3B and S3C, Supplemental Experimental Procedures). Upon application of our classification scheme, we uncovered several dynamic interaction classes among ESC, NPC, and iPSC cellular states (Figures 3I and 3J), including: (1) 537 interactions present in ESCs, lost in NPCs, and reacquired upon reprogramming (purple class) (Figure 3K); (2) 3,004 interactions present only in ESCs and not reprogrammed (red class) (Figure 3L); (3) 5,043 interactions absent in ESCs, acquired upon differentiation, and lost in iPSCs (green class) (Figure S3D); (4) 1,708 interactions present only in iPSCs (orange class) (Figure S3E); (5) 148 interactions that are high in ESCs and NPCs and not present in iPSCs (gold class) (Figure S3F); and (6) 282 interactions absent in ESCs, acquired in NPCs, and residually connected in iPSCs (blue class) (Figure S3G). Notably, we found that the sensitive detection of these interaction classes, particularly those that distinguish iPSCs from ESCs, was contingent upon the resolution and read depth afforded by the 5C approach (Figures S3H and S3I).

Importantly, we observed that the majority of high-count pixels were spatially adjacent to each other in our Distance-Corrected Interaction Score heatmaps and appear to form larger clusters of enriched 3D contact (Figures 3K–3L and 3N, Figures S3D–S3G). To ensure that our approach was not inflating the number of significant interactions, we clustered adjacent pixels that were similarly classified, resulting in a total of only 1,248 unique interactions across three cellular states in our 5C regions (~7.5 Mb) (Figure 3M). Our clustering approach is similar to the methodology employed by Aiden and colleagues for high-resolution Hi-C data (Rao et al., 2014). We emphasize two important points regarding the 3D interaction classes called in this study: (1) the interactions represent both specific looping contacts and

subTAD boundaries that are dynamic across three cellular states and (2) rather than a traditional peak calling approach in just one cell type, we are reporting seven classes of long-range interactions called across three cellular states with a focus on the regions of the genome that are most likely to undergo dynamic restructuring during the reprogramming process. Overall, these results indicate that chromatin architecture is highly dynamic during cell fate transitions, with unique folding classes emerging during the reprogramming process.

Pluripotency Genes Form Interactions that Can Successfully Reprogram

We next set out to explore the biological relevance of our dynamic interaction classes. We utilized a series of integrative computational approaches to elucidate the underlying relationships among (1) fine-scale chromatin folding, (2) gene expression, (3) histone modifications characteristic of cell-type-specific regulatory elements, and (4) binding profiles of the architectural protein CTCF (Tables S1, S3, and S4).

We first investigated the interactions that were present in ESCs, lost in NPCs, and reconnected during reprogramming (ESC-iPSC; purple class) (Figure 4A). We noticed that the *Sox2* gene formed a strong 3D interaction with a pluripotent enhancer element ~120 kb downstream marked by a large domain of H3K4me1/H3K27ac in ESCs (Figure 4B). Upon differentiation, the *Sox2*-pluripotent enhancer interaction disassembled in parallel with loss of H3K27ac signal and then subsequently reassembled in iPSCs (Figures 4B and 4C). We also identified ESC-iPSC (purple class) interactions between the *Oct4/Pou5f1* gene and a putative enhancer element ~20 kb upstream marked by ESC-specific H3K4me1/H3K27ac (Figure 4D). As expected given the pluripotent properties of our iPSC clone, the *Oct4*-enhancer interaction breaks apart in NPCs and reconnects again in iPSCs (Figures 4D and 4E). We next quantitatively assessed the enrichment of a wide range of genomic elements in the ESC-iPSC class of successfully reprogrammed 3D interactions. Consistent with previous reports (Apostolou et al., 2013) and our qualitative observations, pluripotency genes and putative ESC-specific enhancers were significantly enriched at the base of ESC-iPSC interactions (Figure 4F). Together, these results indicate that pluripotency genes can form long-range connections with ESC-specific enhancer elements and that these interactions can reprogram in iPSCs.

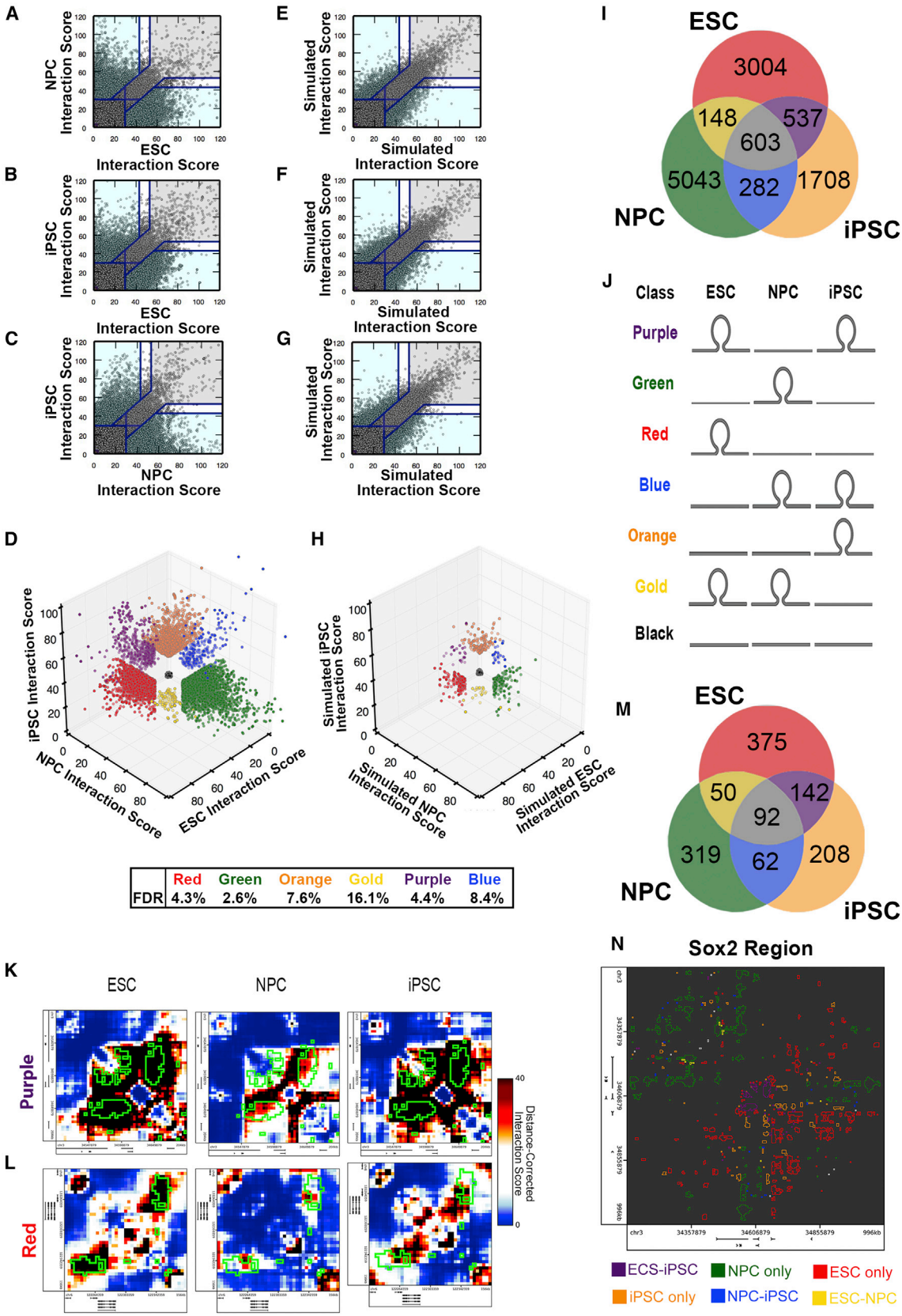
To explore the functional significance of fully reprogrammed interactions, we next conducted genome-wide RNA-seq analysis in ESCs, NPCs, and iPSCs. We examined *Oct4* and *Sox2* gene expression after normalization among libraries to account for any potential batch effects and differences in sequencing depth (Figures S4A–S4D; Table S3, Table S5, and Table S6). Unexpectedly, despite reconnection with target pluripotent enhancers, *Sox2* expression was markedly lower than target ESC expression levels (Figure 4G), whereas *Oct4* expression was more than 2-fold higher than target ESC expression levels (Figure 4H). Our observations highlight the importance of further understanding the relationship between genome folding and expression and led us to question if more global architectural connections around these pluripotent enhancer-promoter interactions could be linked to inaccurately reprogrammed gene expression levels in iPSCs.

Some Pluripotency Genes Reconfigure into New NPC Interactions that Remain Persistent in iPSCs

We next sought to understand larger-scale chromatin folding patterns around *Sox2* (Figure 5A). We hypothesized that chromatin architecture dynamics surrounding the short-range enhancer-promoter interaction might impact the incompletely reprogrammed *Sox2* expression in our iPSC clone. Unexpectedly, we observed that *Sox2* is also engaged in NPC-iPSC (blue class) interactions, which are classified by (1) absence of connection in ESCs, (2) acquisition of connection in NPCs, and (3) residual tethering in iPSCs (Figures 5A and 5B). In NPCs, the *Sox2*-pluripotent enhancer interaction breaks apart and the gene forms long-range contacts with two distal NPC-specific enhancers marked by NPC-specific H3K27ac/H3K4me1. Intriguingly, although the *Sox2*-pluripotent enhancer interaction is reassembled (purple box), the gene also remains partially tethered to the NPC-specific enhancer in iPSCs (blue box) (Figure 5A). We observed a similar phenomenon at the *Klf4* locus, where the *Klf4* gene is highly expressed in ESCs and interacts with a putative ESC-specific enhancer element marked by ESC-specific H3K4me1/H3K27ac ~75 kb upstream of the gene (Figures S5A–S5D). In NPCs, *Klf4* disconnects from its pluripotent enhancer and engages with a downstream NPC-specific enhancer (Figures S5E and S5F). In iPSCs, *Klf4* retains its interaction with the NPC-specific enhancer (blue box) while also partially re-tethering to its target pluripotent enhancer (purple box) (Figure S5F).

We hypothesized that the dual tethering of *Sox2/Klf4* genes to their target ESC-specific pluripotent enhancers and their decommissioned NPC-specific enhancers might lead to inaccurate reprogramming of proper expression levels in our iPSC clone. As a first step toward testing this hypothesis, we cultured our iPSC clone under 2i/LIF conditions to promote a naive, ground state of pluripotency and ensure morphological/phenotypic uniformity across the population (Marks et al., 2012; Ying et al., 2008). Strikingly, we noticed that 2i/LIF culture of iPSCs resulted in (1) loss of the *Sox2*- or *Klf4*-NPC enhancer (blue class) interactions, (2) a further amplification in strength of the *Sox2*- or *Klf4*-pluripotent enhancer (purple class) interactions, and (3) a fine-tuning of *Sox2* or *Klf4* expression to ESC levels (Figures 5A and 5C–5D, Figures S5E and S5F). These results indicate that 2i/LIF conditions are capable of untethering persistent somatic cell chromatin architecture in a population of iPSCs and restoring inaccurately reprogrammed gene expression to levels equivalent to those found in V6.5 ESCs. Future causative studies will be necessary to further dissect the link among architectural persistence, naive versus primed pluripotency, and precise gene expression levels during reprogramming.

We then set out to further understand the mechanistic basis of NPC-iPSC (blue class) interactions. Quantitative enrichment analysis revealed three key genomic annotations enriched at the base of NPC-iPSC contacts: (1) ESC-specific genes, (2) NPC-specific CTCF, and (3) constitutive CTCF (Figure 5E). We then computed "sided" enrichments by accounting for the presence/absence of genomic annotations in both anchoring loci at the base of the NPC-iPSC interactions (see schematic, Figure 5F). Consistent with our qualitative observations, ESC-specific genes most significantly contact NPC-specific enhancers when located at the base of NPC-iPSC interactions (Figure 5F).



	Red	Green	Orange	Gold	Purple	Blue
FDR	4.3%	2.6%	7.6%	16.1%	4.4%	8.4%

(legend on next page)

We note that *Sox2* and *Klf4* are classified as ESC-specific genes in our study due to their markedly increased expression in ESCs versus NPCs. However, both genes are still expressed at levels at least 8-fold higher than background in NPCs. Together, these results led us to hypothesize that genes with developmental roles in both ESCs and NPCs, but regulated by different enhancers in the two cellular states, might be particularly susceptible to inappropriate tethering to off-lineage enhancers in iPSCs.

Our quantitative enrichment analyses also indicated that ESC-specific genes formed significant 3D connections with NPC-specific and constitutively bound CTCF sites (Figures 5E and 5F). Consistent with this quantitative result, we noticed a constitutively bound CTCF site at the base of the *Sox2* NPC-specific enhancer (Figure 5A) and an NPC-specific CTCF site at the base of the *Klf4* NPC-specific enhancer (Figure S5F), suggesting that CTCF might work together with enhancers to facilitate 3D connections to the correct target gene(s). To understand how CTCF binding might be altered during reprogramming, we performed CTCF ChIP-qPCR across all five of our cellular states. We queried CTCF occupancy levels in the NPC-specific and ESC-specific enhancers (Figure 5A, blue and red stars, respectively) at the *Sox2* locus. We found that the NPC-specific enhancer remains constitutively bound by CTCF in ESC, NPC, iPCS, ESC+2i, and iPSC+2i conditions (Figure 5G, left). By contrast, the ESC-specific enhancer exhibited high CTCF in ESCs, loss of binding in NPCs, sustained low CTCF occupancy in iPSCs, and subsequent restoration of occupancy in 2i/LIF (Figure 5G, right).

Intriguingly, CTCF binding patterns correlate with the changes in chromatin architecture around *Sox2*. In ESCs, the constitutive CTCF site interacts with the ESC-specific CTCF site, resulting in spatial co-localization of the ESC- and NPC-specific enhancers (Figure 5A, red box). Loss of CTCF binding at the ESC-specific enhancer correlates with disconnection of the enhancer-enhancer interaction in NPCs. In parallel, the constitutive CTCF site at the NPC-specific enhancer forms a strong NPC-iPSC (blue class) interaction with the *Sox2* gene (Figure 5A, blue box). We posit that the *Sox2*-NPC-enhancer interaction remains tethered in iPSCs because CTCF does not fully rebind to the ESC-specific enhancer (Figure 5G, right). In support of this idea, 2i/LIF leads to (1) reacquisition of CTCF binding at the ESC-specific enhancer, (2) reconnection of the interaction between both ESC-specific and NPC-specific enhancers, and (3) abrogation of the *Sox2*-NPC-specific enhancer interaction. These observations are consistent with a working model in which "persistent-NPC" interactions can remain in iPSCs when some developmentally regulated genes are tethered to NPC-specific enhancers, possibly at constitutive or NPC-specific CTCF sites.

We highlight that somatic cell-specific elements were not specifically enriched in NPC-iPSC interactions (Figures S6A–S6C). For example, NPC-specific genes and enhancers were primarily enriched in NPC (green class) interactions only, supporting our finding that it is ESC-specific genes, particularly those that remain somewhat active in NPCs, that are redirected into NPC-iPSC contacts. An example illustrating this idea can be found at the *Olig1* and *Olig2* genes that are expressed in an NPC-specific manner and equivalently form NPC (green class) interactions only with a downstream NPC-specific enhancer (Figures S6D and S6E). Expression of *Olig1* and *Olig2* is lost in parallel with loss of the green class 3D interaction. Together, these results support the intriguing possibility that ESC-specific genes that remain partially active in NPCs form new interactions with somatic cell-specific enhancers during differentiation and that these contacts can remain tethered as a form of architectural persistence in iPSCs. Finally, we note that 5C is performed on a population of millions of cells, we cannot distinguish between the possibilities that (1) pluripotency genes simultaneously form both ESC-iPSC and NPC-iPSC contacts in individual cells and (2) pluripotency genes form two different sets of interactions in distinct ESC-like subpopulations.

Pluripotent Interactions that Do Not Reprogram Display Dynamic CTCF Occupancy

Finally, we explored the interactions that are present in ESCs and lost in NPCs but do not reconnect in iPSCs (red group, Figures 6A–6B, Figures S7A and S7B). A noteworthy illustration of these poorly reprogrammed interactions is found at the *Zfp462* gene (highlighted in green, Figure 6A), which interacts with a downstream putative ESC-specific enhancer element in ESCs. *Zfp462* expression is reduced in NPCs in parallel with loss of H3K27ac at the putative downstream enhancer and loss of the interaction. By contrast to the previously discussed ESC-iPSC (purple) group, this gene-enhancer interaction is not reassembled in iPSCs. Similarly, the genes *Mis18a* and *Urb1* form interactions in ESCs that are not reprogrammed (highlighted in yellow and green, respectively; Figure S7A). Together, these

Figure 3. Genome Architecture Can Be Classified into Several Distinct Dynamic Groups during Cell Fate Transitions

(A–C) Scatterplot comparison of distance-corrected interaction scores between (A) ESCs and NPCs, (B) ESCs and iPSCs, and (C) NPCs and iPSCs. Thresholds are displayed as blue lines. For pairwise plots, cell-type-specific, invariant, and background interactions are represented by blue, gray, and brown colored shading, respectively.

(D) 3D scatterplot of distance-corrected interaction scores for cellular states in which both replicates cross the thresholds displayed in (A)–(C). Interaction classes are indicated by color (red, ESC only; green, NPC only; orange, iPSC only; gold, ESC-NPC; purple, ESC-iPSC; blue, NPC-iPSC; black, background). Empirical false discovery rates computed from simulated data in (E)–(G) are reported for each classification.

(E–G) Scatterplots of Distance-Corrected Interaction Scores from simulated replicates. Empirical false discovery rates were computed based on the number of interactions that cross pre-established thresholds in the simulated data versus the real data.

(H) 3D scatterplot of distance-corrected interaction scores for simulated libraries that cross the thresholds displayed in (A)–(C) and (E)–(G).

(I) Number of interactions called significant in each cell-type-specific interaction class.

(J) Schematic illustrating the 3D interaction behavior for each interaction class.

(K and L) Zoomed-in heatmaps of distance-corrected interaction scores for specific (K) ESC-iPSC (purple class) and (L) ESC only (red class) interactions. Classified interaction pixels are outlined in green.

(M) Number of interactions called significant for each 3D classification after clustering directly adjacent 4 kb bins.

(N) Depiction of all interactions called significant in the Sox2 region. Each interaction is outlined by the corresponding classification color.

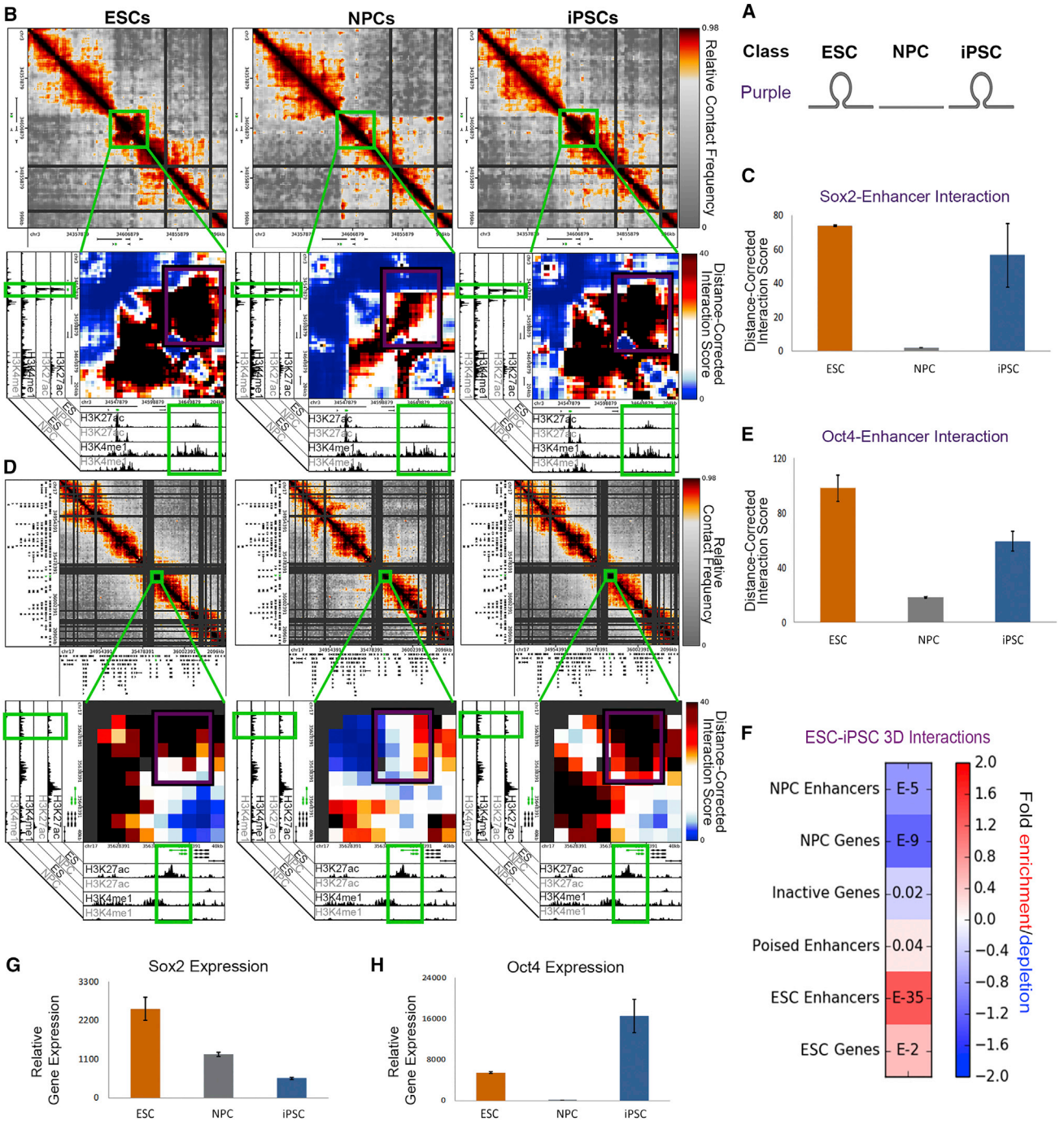


Figure 4. Pluripotency Gene-Enhancer Interactions Can Be Re-established in iPSCs

(A) Schematic illustrating the ESC-iPSC (purple) interaction class.

(B and D) Relative contact frequency heatmaps (top) and zoomed-in distance-corrected interaction score heatmaps (bottom) highlighting key ESC-iPSC interactions (purple class) between (B) *Sox2* and (D) *Oct4* genes and their target enhancers. Heatmaps are overlaid on ChIP-seq tracks of H3K27ac and H3K4me1 in ESCs and NPCs.

(C and E) Distance-corrected interaction score changes at (C) the *Sox2*-enhancer interaction and (E) *Oct4*-enhancer interaction among ESCs, NPCs, and iPSCs. Error bars represent the standard deviation across two 5C replicates.

(F) Fold enrichment of cell-type-specific regulatory elements in ESC-iPSC (purple class) interactions compared to the enrichment expected by chance across the genome. Color bar represents fold change enrichment over background (blue, depletion; red, enrichment). p values are computed with Fisher's Exact test and listed in each bin.

(G and H) Normalized gene expression is plotted for (G) *Sox2* and (H) *Oct4* genes. Error bars represent standard deviation across two RNA-seq replicates.

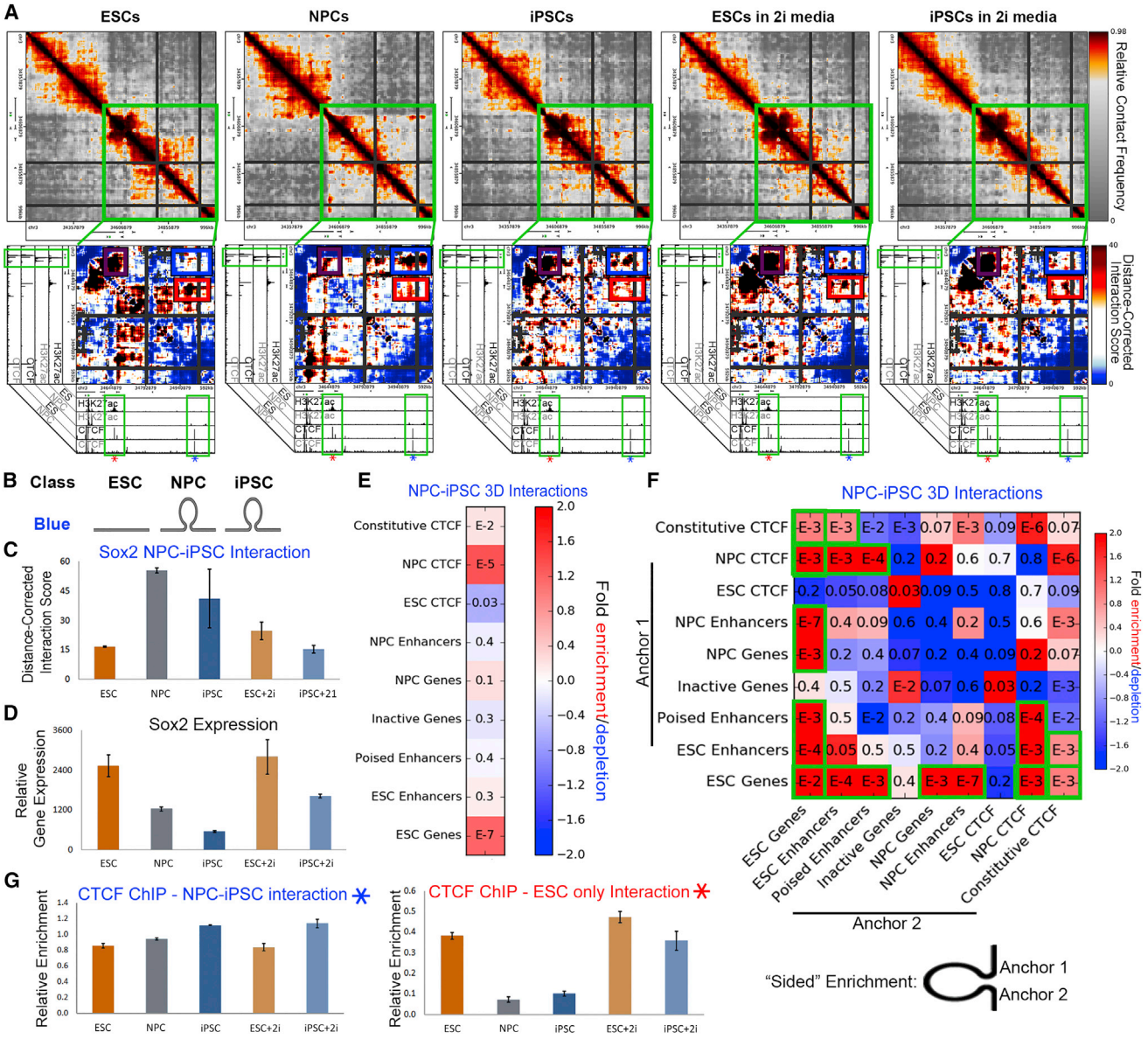


Figure 5. Pluripotency Genes Can Exhibit "Persistent NPC-like" Folding Patterns in iPSCs

(A) Relative contact frequency heatmaps (top) and zoomed-in distance-corrected interaction score heatmaps (bottom) highlighting an NPC-iPSC interaction (blue class) around the *Sox2* gene. Heatmaps are overlaid on ChIP-seq tracks of H3K27ac and CTCF in ESCs and NPCs.

(B) Schematic illustrating the NPC-iPSC (blue) interaction class.

(C) Distance-corrected interaction score changes at an NPC-iPSC interaction around the *Sox2* gene among ESC, NPC, iPSC, ESC+2i, and iPSC+2i conditions. Error bars represent standard deviation across two 5C replicates.

(D) Normalized expression for the *Sox2* gene. Error bars represent standard deviation across two RNA-seq replicates.

(E and F) Fold enrichment of cell type-specific regulatory elements in NPC-iPSC (blue class) interactions compared to the enrichment expected by chance across the genome. p values are computed with Fisher's Exact test and listed in each bin. (E) Enrichment for any given genomic annotation at the base of NPC-iPSC interactions. (F) Enrichment for any given pairwise combination of genomic annotations in the two anchoring bins at the base of NPC-iPSC interactions.

(G) Relative ChIP-qPCR enrichment of CTCF binding at the NPC-iPSC interaction (left, denoted by blue star in A) and ESC only interaction (right, denoted by red star in A). Error bars represent SD across three technical replicates.

genomic loci reveal a class of interactions that are refractory to reprogramming in iPSCs.

To investigate the mechanistic basis for poorly reprogrammed (red class) interactions, we again looked for possible dynamic CTCF binding. We noticed that genomic loci where CTCF is bound in ESCs, but severely depleted in NPCs, were preferentially located at the base of poorly reprogrammed interactions (green boxes; Figures 6A and S7A). Consistent with this observation, ESC-specific CTCF sites were significantly enriched in ESC only (red class) interactions (Figures 6C and 6D). ChIP-qPCR

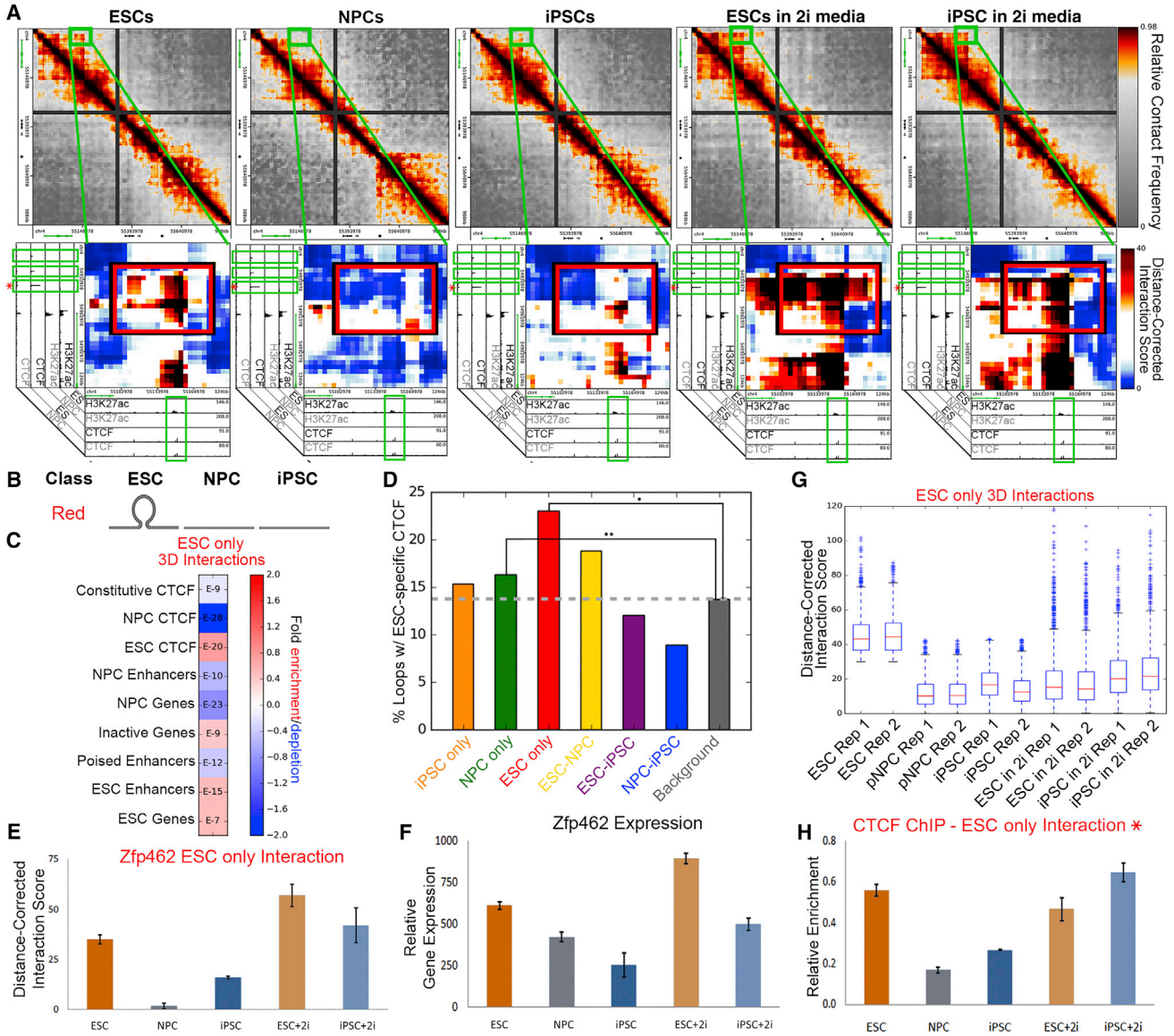


Figure 6. Interactions that Do Not Reprogram Display Poorly Reprogrammed CTCF Occupancy

(A) Relative contact frequency heatmaps (top) and zoomed-in distance-corrected interaction score heatmaps (bottom) highlighting an ESC only (red class) interaction at ESC-specific CTCF binding sites at the *Zfp462* gene (indicated in green). Heatmaps are overlaid on ChIP-seq tracks of H3K27ac and CTCF in ESCs and NPCs.

(B) Schematic illustrating the ESC only (red class) interactions.

(C) Fraction of ESC only (red class) interactions enriched with distinct cell type-specific regulatory elements compared to the expected enrichment in background. p values are computed with Fisher's Exact test and listed in each bin.

(D) Bar plot displaying the fraction of each interaction class containing ESC-specific CTCF binding sites compared to the expected background fraction. Fisher's Exact test: *p = 2.06016e−21; **p = 0.000541696.

(E) Distance-corrected interaction score changes at an ESC only interaction around the *Zfp462* gene among ESC, NPC, iPSC, ESC+2i, and iPSC+2i conditions. Error bars represent standard deviation across two 5C replicates.

(F) *Zfp462* gene expression among ESC, NPC, iPSC, ESC+2i, and iPSC+2i conditions. Error bars represent standard deviation across two RNA-seq replicates.

(G) Aggregate distance-corrected interaction score changes among ESC, NPC, iPSC, ESC+2i, and iPSC+2i conditions for loci anchoring red class.

(H) Relative ChIP-qPCR enrichment of CTCF binding at the ESC only interaction (denoted by blue star in A). Error bars represent SD across three technical replicates.

analysis of CTCF occupancy revealed consistent depletion of CTCF in our iPSC clone compared to ESCs (Figures 5G and 6H, Figure S7G). Importantly, culture of our iPSC clone in 2i/LIF media resulted in (1) reacquisition of the red group interactions, (2) re-establishment of CTCF occupancy, and (3) restoration of gene expression levels in iPSCs (Figures 6E–6H, Figures S7C–S7G). Corroborating locus-specific observations, a global analysis of red class interactions demonstrated a marked

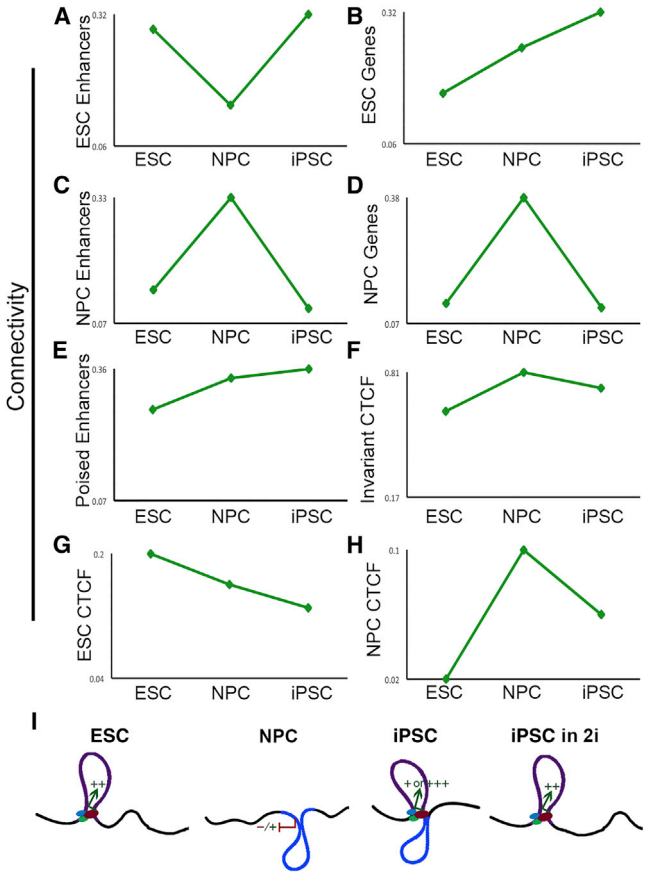


Figure 7. Pluripotency Genes Can Be Hyperconnected in iPSCs

Connectivity of distinct regulatory elements in ESCs, NPCs, and NPC-derived iPSCs. (A) ESC-specific enhancers; (B) ESC-specific genes; (C) NPC-specific enhancers; (D) NPC-specific genes; (E) poised enhancers; (F) invariant CTCF; (G) ESC-specific CTCF; (H) NPC-specific CTCF. (I) Schematic illustrating a model of the "hyperconnectivity" of certain pluripotency genes in our NPC-derived iPSC clone. Key ESC-specific genes (denoted by colored arrows) display the ability to reprogram their connections with ESC-specific enhancers (denoted by green/blue "transcription factor" binding sites) and retain remnants of their somatic connections. This intermediate architectural state correlates with inaccurate reprogramming of gene expression levels (represented by colored "±") and can be fully restored upon culture in 2i/LIF media.

increase in interaction score upon the addition of 2i/LIF media to iPSCs (Figure 6G). On the basis of these results, we posit that the loss of CTCF binding at critical developmentally regulated loci can be inefficiently restored during a cell fate transition like somatic cell reprogramming.

Somatic Elements Are Disconnected and Pluripotent Genes Are Hyperconnected in our iPSC Clone

We hypothesized that distinct types of regulatory elements exhibit differential connectivity patterns as ESCs transition to NPCs and back to iPSCs. To address this hypothesis, we computed a "connectivity" metric for each class of genomic element in each of the three cellular states. ESC-specific enhancers lose their connectivity in NPCs and then reconnect in iPSCs (Figure 7A). Intriguingly, ESC-specific genes become increasingly more connected upon differentiation and subsequent reprogramming (Figure 7B). By contrast, NPC-specific genes/enhancers increase connectivity in NPCs, but then resume ground state ESC-like connectivity in iPSCs (Figures 7C and 7D). Poised enhancers and invariant CTCF sites display minor differences in connectivity across the three cellular states (Figures 7E and 7F), whereas ESC-specific CTCF sites lose their interactions upon differentiation and only partially gain back connectivity in iPSCs (Figure 7G). NPC-specific CTCF sites increase in connectivity in NPCs and then partially resume their disconnected state in iPSCs (Figure 7H).

Overall, our results support a model in which somatic cell regulatory elements reconfigure to a ground connectivity state during reprogramming, whereas pluripotency genes (particularly those that retain a low level of activity in NPCs) can be "hyperconnected" in our iPSC clone due to persistent cell-of-origin interactions (Figure 7I). We hypothesize that persistent-NPC and poorly reprogrammed interactions contribute to inaccurate reprogramming of gene expression levels. Consistent with this idea, 2i/LIF can erase persistent-NPC interactions, restore poorly reprogrammed interactions, and re-establish precise ESC-like expression levels in our iPSC clone.

DISCUSSION

Understanding the molecular mechanisms governing somatic cell reprogramming is of paramount importance to our knowledge of cell fate commitment and the use of iPSCs for regenerative medicine applications. Mechanistic studies have primarily focused on profiling gene expression and classic epigenetic modifications at intermediate stages in the reprogramming process (Koche et al., 2011; Polo et al., 2012; Soufi et al., 2012; Stadtfeld et al., 2008). However, the molecular roadblocks that impede the efficiency and timing of epigenome resetting in iPSCs are just beginning to emerge. Here we examine a unique aspect of reprogramming: the higher-order folding of chromatin in the 3D nucleus. We demonstrate that iPS genome architecture at the sub-Mb scale within TADs can be imperfectly rewired during transcription factor-mediated reprogramming.

Recent studies focusing on individual loci (e.g., *Nanog* or *Oct4*) reported that pluripotency genes can re-establish long-range connections with their target enhancers in iPSCs (Apostolou et al., 2013; de Wit et al., 2013; Denholtz et al., 2013; Wei et al., 2013; Zhang et al., 2013a). Motivated by the need to understand how chromatin unfolds/refolds more generally in iPSCs, we created high-resolution maps of chromatin architecture in Mb-sized regions around developmentally regulated genes. Consistent with previous reports, we observe that many pluripotency genes interact with ESC-specific enhancers in ESCs; these interactions break apart in NPCs and then reassemble in iPSCs. Additionally, we find that somatic cell interactions between NPC-specific genes and NPC-specific enhancers generally disconnect in iPSCs. Thus, our data confirm and extend several known locus-specific principles of genome folding during reprogramming.

We also uncover new classes of chromatin interactions that do not behave in the expected manner. We identified a small subset of NPC-iPSC (blue class) interactions representing persistent chromatin folding patterns from the somatic cell of origin in

iPSCs. Unexpectedly, we find that some key pluripotency genes can form new 3D connections in NPCs that remain tethered in our iPSC clone. For example, *Klf4* and *Sox2* are dually tethered to their target ESC-specific enhancers and their decommissioned NPC-specific enhancers in iPSCs. We posit that this rare but intriguing form of ''architectural persistence'' might be causally linked to inaccurate reprogramming of target gene expression levels in certain iPSC clones. In support of this working model, we find that 2i/LIF conditions are capable of untethering persistent somatic cell chromatin architecture and restoring the inaccurately reprogrammed expression to levels equivalent to those found in a genetically comparable ESC line. Notably, we highlight that NPC-specific genes/enhancers form contacts in NPCs that subsequently disassemble in iPSCs, suggesting that somatic genes are not driving the architectural persistence in iPSCs. These results agree with previous studies suggesting that somatic cell gene expression is downregulated during the initiation phase of reprogramming and precedes the reactivation of the pluripotency network (Polo et al., 2012). We favor a model in which reconfiguration of higher-order chromatin topology could be a potential rate-limiting step in the reprogramming process as a result of architectural persistence or incomplete architectural reprogramming (discussed below) blocking the formation of fully reprogrammed iPSCs (Buganim et al., 2012; Tanabe et al., 2013).

CTCF is a key player in the organization of the 3D genome and anchors the base of a large number of long-range interactions in ESCs (Dixon et al., 2012; Rao et al., 2014; Handoko et al., 2011; Phillips-Cremins et al., 2013). Here we provide a new link between CTCF and reprogramming. We identify a new class of chromatin interactions that are high in ESCs, break apart in NPCs, and are not fully reconfigured in iPSCs. Importantly, we find that these poorly reprogrammed interactions often contain ESC-specific CTCF binding sites that lose occupancy in NPCs and do not reacquire full binding in our iPSC clone. CTCF has largely stable occupancy patterns during development, with 60%–90% of sites remaining bound to the genome between cell types (Kim et al., 2007). Thus, we speculate a model in which CTCF binding is difficult to lose during differentiation, but once occupancy is abolished it is inefficiently re-established during reprogramming. Importantly, DNA methylation is refractory to CTCF binding (Bell and Felsenfeld, 2000), suggesting a possible link between poorly reprogrammed chromatin contacts and previously reported sources of cell-of-origin epigenetic persistence (Kim et al., 2010; Polo et al., 2010). Indeed, because ESCs cultured in 2i/LIF display global hypomethylation (Ficz et al., 2013; Habibi et al., 2013), we speculate that the interplay between CTCF and dynamic DNA methylation might serve as a mechanism underlying our observation that 2i/LIF media can fully restore CTCF occupancy and poorly reprogrammed interactions.

Epigenetic and transcriptional signatures are generally reset in fully reprogrammed iPSCs cultured under optimal conditions (Cahan et al., 2014; Stadtfeld et al., 2010). However, variations in epigenetic profiles among iPSC clones have been attributed to reprogramming method, passage number, genetic background, or laboratory-to-laboratory procedural discrepancies (Bock et al., 2011; Polo et al., 2010). Therefore, we sought to confirm that our observations were truly linked to inefficiencies in the reprogramming of our iPSC clone, and not experimental artifacts due to (1) residual somatic cells in our iPSC population or (2) laboratory-specific culture conditions. Importantly, Hochedlinger and colleagues have extensively characterized the iPSC clone used in this manuscript for its pluripotent properties (Eminli et al., 2008). Additionally, our iPSC clone was cultured to >15 passages in serum+LIF-containing growth conditions not amenable to NPC proliferation/survival. Finally, known NPC markers are not upregulated in our iPSC population versus ESCs (Figures S6E–S6G). Thus, we see no evidence of contaminating NPCs in our iPSCs. Although somatic cells are absent, we cannot rule out the possibility that there could be a gradient of pluripotent properties (e.g., a continuum between naive and primed pluripotency) across single cells within our fully reprogrammed iPSC clonal population. Because we are conducting population-based assays, we would detect all interactions that exist across the different pluripotent states. Consistent with this possibility, we see that conversion of the population to a uniform, naive pluripotent state with 2i/LIF media abrogates architectural persistence interactions and reinstates poorly reprogrammed interactions. Additionally, although we subjected our iPSCs with or without 2i/LIF to the same number of passages ($p > 15$), we cannot rule out the possibility that further long-term passaging might also resolve any mis-wired chromatin interactions. Noteworthily, these results raise the interesting possibility that an iPSC clone capable of creating transgenic mice might still exhibit some level of architectural heterogeneity that can be fully resolved with 2i/LIF media. Exciting lines of future inquiry will query genome folding in higher passages, alternative reprogramming conditions, tetraploid-complementation verified iPSCs, and a range of iPSC clones derived from multiple somatic cell lineages.

In parallel with this manuscript, de Laat, Graf, and colleagues published a genome-wide analysis of chromatin architecture in iPSCs derived from four independent somatic cell lineages (Krijger et al., 2016). The authors take a top-down approach in which they generate genome-wide, albeit low-resolution, Hi-C maps suited to query higher-order levels of genome organization (i.e., A/B compartments, TADs, and nuclear positioning of TADs). Importantly, they demonstrate that A/B compartments are largely reset during reprogramming. Moreover, consistent with the leading idea that TADs are largely invariant among cell types (Dixon et al., 2012), TAD boundaries remained for the most part consistent among iPSC clones and ESCs. At the level of sub-Mb-scale genome folding, however, the design of the two studies is such that different findings arise. Here we take a bottom-up approach in which we create high-resolution, high-complexity maps focused on fine-scale chromatin folding dynamics within TADs around developmentally regulated genes. Given the sensitivity and statistical power afforded by the 5C assay, it is not surprising that we detect a larger number of dynamic subTAD boundaries and looping interactions than reported in Krijger et al. during the transition among ESC, iPSC, and NPC cellular states. It is noteworthy that when we increase our bin size from 4 kb up to 300 kb (Figure S3H), we can recapitulate the author's high level of correlation between the ESCs and iPSCs (Figure S3I). Krijger et al. and our manuscript offer complementary viewpoints into genome architecture dynamics across a wide range of length scales and resolutions during

reprogramming. Together, the findings from these studies are consistent with our working hypothesis that architectural changes causally linked to developmentally relevant alterations in gene expression occur within TADs at the sub-Mb scale.

Overall, we present high-coverage, fine-scale maps of chromatin folding within TADs in iPSCs and use our maps to uncover several new organizing principles for genome folding during reprogramming. We find that different cell type-specific regulatory elements exhibit contrasting 3D connectivity patterns as cells switch fates in forward and reverse directions. A deeper understanding of the role for chromatin folding at each step in the reprogramming process is of critical importance toward the use of iPSCs for disease modeling and regenerative medicine purposes. Future work combining high- and low-resolution mapping approaches will provide a comprehensive view of genome folding across length scales and cellular states to create a catalog of "hotspots" of incomplete architectural reprogramming to address whether specific somatic cell types are more or less resistant to topological changes.

EXPERIMENTAL PROCEDURES

Cell Culture, Differentiation, and Reprogramming

V6.5 ESCs, primary NPCs, and NPC-derived iPSCs were cultured as described in the Supplemental Experimental Procedures. Briefly, ESCs were expanded on Mitomycin-C inactivated MEFs under standard pluripotent conditions and passaged onto feeder-free gelatin-coated plates before fixation. Primary NPCs were isolated from whole brains of P1 129SvJae × C57BL/6, Sox2-eGFP mice and cultured as neurospheres for two passages before adherent culture and fixation. The iPSC clone used in this paper was derived and characterized in Eminli et al. (2008) and expanded/cultured for use in this study to >15 passages with or without 2i/LIF media as described in the Supplemental Experimental Procedures.

Generation and Analysis of 5C Libraries

5C libraries were generated according to standard procedures described in the Supplemental Experimental Procedures. Paired-end reads were aligned to a pseudo-genome consisting of all 5C primers using Bowtie. Interactions were counted when both paired-end reads were uniquely mapped to the 5C primer pseudo-genome. Counts were converted to contact matrices for each genomic region queried, processed, normalized, and modeled as described in the Supplemental Experimental Procedures. Customized algorithms for classification of 5C interactions and the downstream integration of interaction classes with ChIP-seq and RNA-seq data are detailed in the Supplemental Experimental Procedures.

RNA-Seq Library Preparation and Analysis

Cells were lysed with Trizol and total RNA was extracted as detailed in the Supplemental Experimental Procedures. Samples were prepared for sequencing using the Illumina TruSeq Stranded Total RNA Library Prep kit with RiboZero (Illumina RS-122-2202) following the supplier's protocol and sequenced on the Illumina NextSeq500. Libraries were analyzed and corrected for any sequencing depth or batch effect differences with methods described in the Supplemental Experimental Procedures.

ChIP-Seq Analysis and ChIP-qPCR

A summary of published ChIP-seq libraries re-analyzed in this study is provided in Table S4. Reads were aligned to mm9 with Bowtie using default parameters. Only uniquely mapped reads were used for downstream analyses. ChIP-seq peak calling and ChIP-qPCR experiments are detailed in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The accession number for the data reported in this paper is GEO: GSE68582.

SUPPLEMENTAL INFORMATION

Supplemental Information for this article includes seven figures, seven tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2016.04.004.

AUTHOR CONTRIBUTIONS

T.G.G., J.K., and Z.P. contributed equally to this work. J.E.P.C, V.G.C., and J.D. conceived the project. J.A.B, G.H., and J.E.P.C designed and performed 3C/5C experiments. K.H. and E.A. provided iPSCs and Sox2-GFP neonates. H.K.N. performed RNA-seq experiments. J.A.B., S.C.H., and X.L. performed ChIP experiments. T.G.G., J.K., Z.P., E.J.S., J.A.B., and J.E.P.C. developed code and analyzed data. J.A.B. and J.E.P.C wrote the paper with input from E.A., K.H., V.G.C., and J.D.

ACKNOWLEDGMENTS

We thank members of the Cremins laboratory for helpful discussions and Michael Duong for cell culture assistance. J.E.P.C. is a New York Stem Cell Foundation Robertson Investigator and an Alfred P. Sloan Foundation Fellow. This work was funded by The New York Stem Cell Foundation (J.E.P.C), the Alfred P. Sloan Foundation (J.E.P.C), the NIH Director's New Innovator Award from the National Institute of Mental Health (1DP2MH11024701; J.E.P.C), 4D Nucleome Common Fund grants (1U01HL12999801 to J.E.P.C; U54DK107980 to J.D.), and an R01 from the National Human Genome Research Institute (R01 HG003143; J.D.). J.D. is an investigator of the Howard Hughes Medical Institute.

Received: June 3, 2015
Revised: December 31, 2015
Accepted: April 15, 2016
Published: May 5, 2016

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.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482–485.

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H., et al. (2011). Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell *144*, 439–452.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*, 1209–1222.

Cahan, P., Li, H., Morris, S.A., Lummertz da Rocha, E., Daley, G.Q., and Collins, J.J. (2014). CellNet: network biology applied to stem cell engineering. Cell *158*, 903–915.

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.

Dekker, J., Marti-Renom, M.A., and Mirny, L.A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat. Rev. Genet. *14*, 390–403.

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.

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.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331–336.

Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res. *16*, 1299–1309.

Eminli, S., Utikal, J., Arnold, K., Jaenisch, R., and Hochedlinger, K. (2008). Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression. Stem Cells *26*, 2467–2474.

Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.L., Walter, J., et al. (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. Cell Stem Cell *13*, 351–359.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., et al. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell *13*, 360–369.

Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W., Ye, C., Ping, J.L., Mulawadi, F., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. Nat. Genet. *43*, 630–638.

Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C.J., Creyghton, M.P., van Oudenaarden, A., and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature *462*, 595–601.

Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature *503*, 290–294.

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.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell *128*, 1231–1245.

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I., et al. (2010). Epigenetic memory in induced pluripotent stem cells. Nature *467*, 285–290.

Koche, R.P., Smith, Z.D., Adli, M., Gu, H., Ku, M., Gnirke, A., Bernstein, B.E., and Meissner, A. (2011). Reprogramming factor expression initiates widespread targeted chromatin remodeling. Cell Stem Cell *8*, 96–105.

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*, this issue, 597–610.

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.

Lujan, E., Zunder, E.R., Ng, Y.H., Goronzy, I.N., Nolan, G.P., and Wernig, M. (2015). Early reprogramming regulators identified by prospective isolation and mass cytometry. Nature *521*, 352–356.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., and Stunnenberg, H.G. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell *149*, 590–604.

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.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., et al. (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol. Cell *38*, 603–613.

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.

Polo, J.M., Liu, S., Figueroa, M.E., Kulalert, W., Eminli, S., Tan, K.Y., Apostolou, E., Stadtfeld, M., Li, Y., Shioda, T., et al. (2010). Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat. Biotechnol. *28*, 848–855.

Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., et al. (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. Cell *151*, 1617–1632.

Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., et al. (2014). Topologically associating domains are stable units of replication-timing regulation. Nature *515*, 402–405.

Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A.A., Caspi, I., Krupalnik, V., Zerbib, M., et al. (2013). Deterministic direct reprogramming of somatic cells to pluripotency. Nature *502*, 65–70.

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.

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

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994–1004.

Stadtfeld, M., Maherali, N., Breault, D.T., and Hochedlinger, K. (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell Stem Cell *2*, 230–240.

Stadtfeld, M., Apostolou, E., Akutsu, H., Fukuda, A., Follett, P., Natesan, S., Kono, T., Shioda, T., and Hochedlinger, K. (2010). Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature *465*, 175–181.

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.

Tanabe, K., Nakamura, M., Narita, M., Takahashi, K., and Yamanaka, S. (2013). Maturation, not initiation, is the major roadblock during reprogramming toward pluripotency from human fibroblasts. Proc. Natl. Acad. Sci. USA *110*, 12172–12179.

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.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

Zhang, H., Jiao, W., Sun, L., Fan, J., Chen, M., Wang, H., Xu, X., Shen, A., Li, T., Niu, B., et al. (2013a). Intrachromosomal looping is required for activation of endogenous pluripotency genes during reprogramming. Cell Stem Cell *13*, 30–35.

Zhang, Y., Wong, C.H., Birnbaum, R.Y., Li, G., Favaro, R., Ngan, C.Y., Lim, J., Tai, E., Poh, H.M., Wong, E., et al. (2013b). Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature *504*, 306–310.