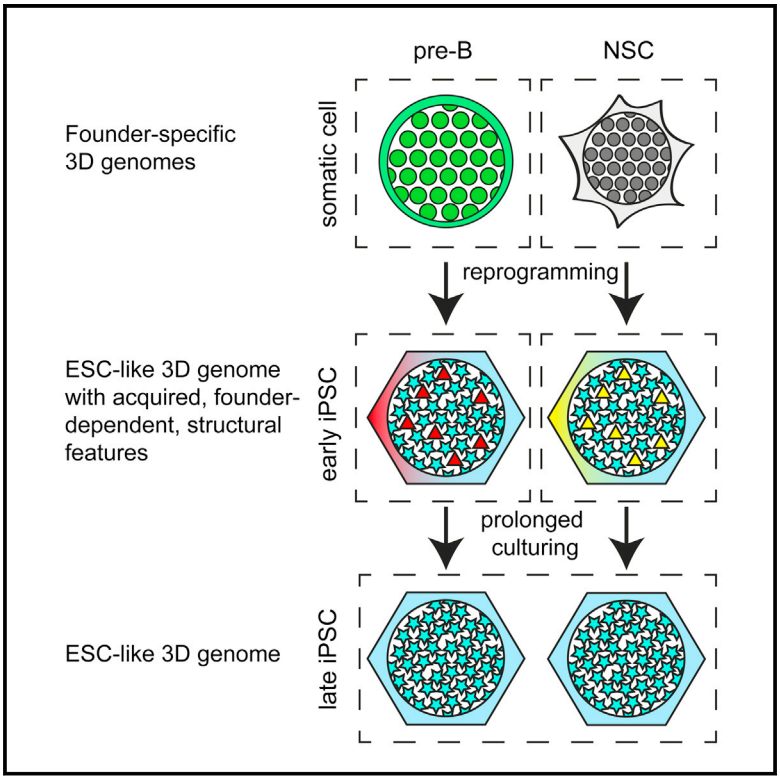
Article

Cell Stem Cell

Cell-of-Origin-Specific 3D Genome Structure Acquired during Somatic Cell Reprogramming

Graphical Abstract



Authors

Peter Hugo Lodewijk Krijger, Bruno Di Stefano, Elzo de Wit, Francesco Limone, Chris van Oevelen, Wouter de Laat, Thomas Graf

Correspondence

w.delaat@hubrecht.eu (W.d.L.), thomas.graf@crg.eu (T.G.)

In Brief

Krijger et al. report that the reprogramming of four somatic cell types with highly distinct 3D genomes results in pluripotent cells with largely identical, ESC-like, genome conformations carrying founder-dependent topological hallmarks. The latter are not remnants of somatic chromosome topologies but are acquired during reprogramming in a cell-of-origin-dependent manner.

Highlights

- The 3D genome topology of four somatic cell types varies greatly and differs from ESCs
- The 3D genomes of iPSCs from different founders and of ESCs are overall highly similar
- Early-passage iPSCs show subtle but reproducible founder-dependent 3D differences
- The distinctive topology features of iPSCs are acquired during reprogramming

Accession Numbers

GSE76481

Krijger et al., 2016, Cell Stem Cell *18*, 597–610
May 5, 2016
http://dx.doi.org/10.1016/j.stem.2016.01.007

Cell Stem Cell
Article

Cell-of-Origin-Specific 3D Genome Structure Acquired during Somatic Cell Reprogramming

Peter Hugo Lodewijk Krijger,[1,3] Bruno Di Stefano,[2,3,5] Elzo de Wit,[1,3,6] Francesco Limone,[2] Chris van Oevelen,[2] Wouter de Laat,[1,4,*] and Thomas Graf[2,4,*]
[1]Hubrecht Institute-KNAW and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands
[2]Gene Regulation, Stem Cells and Cancer Programme, Centre for Genomic Regulation (CRG) and Pompeu Fabra University, Dr Aiguader 88, 08003 Barcelona, Spain
[3]Co-first author
[4]Co-senior author
[5]Present address: Department of Molecular Biology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA
[6]Present address: Division of Gene Regulation, Netherlands Cancer Institute, Plesmanlaan 121, 1066CX Amsterdam, the Netherlands
*Correspondence: w.delaat@hubrecht.eu (W.d.L.), thomas.graf@crg.eu (T.G.)
http://dx.doi.org/10.1016/j.stem.2016.01.007

SUMMARY

Forced expression of reprogramming factors can convert somatic cells into induced pluripotent stem cells (iPSCs). Here we studied genome topology dynamics during reprogramming of different somatic cell types with highly distinct genome conformations. We find large-scale topologically associated domain (TAD) repositioning and alterations of tissue-restricted genomic neighborhoods and chromatin loops, effectively erasing the somatic-cell-specific genome structures while establishing an embryonic stem-cell-like 3D genome. Yet, early passage iPSCs carry topological hallmarks that enable recognition of their cell of origin. These hallmarks are not remnants of somatic chromosome topologies. Instead, the distinguishing topological features are acquired during reprogramming, as we also find for cell-of-origin-dependent gene expression patterns.

INTRODUCTION

Somatic cells can be reprogrammed into induced pluripotent stem cells (iPSCs) by overexpression of the transcription factors OCT4, SOX2, KLF4, and MYC (OSKM) (Takahashi and Yamanaka, 2006). Regardless of tissue origin, IPSCs possess full developmental potential in vitro, form teratomas in vivo, and are even capable of generating "all-iPSC mice" after injection into tetraploid blastocysts (Zhao et al., 2009). Their ability to contribute to all tissues makes iPSCs attractive for disease modeling and for regenerative medicine. Recently, it was reported that the differentiation propensity of iPSCs reflects the tissue of origin, such that neural-derived iPSCs more readily differentiate into neurons, and blood-cell-derived iPSCs are biased toward the hematopoietic lineage (Bar-Nur et al., 2011; Kim et al., 2010; Nishino et al., 2011; Polo et al., 2010). This tissue of origin memory has been shown to be associated with differences in epigenetic features. Residual DNA methylation marks were found at promoters in early iPSCs, presumably stably silencing genes that act in specifying lineages other than the donor cell type (Kim et al., 2010). Early passage iPSCs obtained from different cell types were also found to have distinct gene expression profiles. Some of the distinguishing genes appeared to show residual cell-of-origin-specific transcription, which was interpreted to reflect memory of the transcriptional status in founder cells (Polo et al., 2010). The founder-dependent transcription and DNA methylation profiles were lost upon prolonged passaging of the iPSCs or after treatment with chromatin-modifying drugs (Kim et al., 2010; Polo et al., 2010).

Different cell types also show distinct 3D chromatin structures (Dixon et al., 2015; Rao et al., 2014), and genome topology is increasingly appreciated as an important contributor to genome functioning. Chromosomes can be subdivided into topologically associated domains (TADs), structural units within which sequences preferentially contact each other (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). TADs serve to physically restrain interactions of enhancers with their target gene promoters (Nora et al., 2012). TAD organization is relatively stable during development, but contacts within TADs can dynamically change between cell types (Phillips-Cremins et al., 2013). While some enhancer-promoter contacts seem tissue invariant, others are specifically established during differentiation, contributing to tissue-specific transcription programs (de Laat and Duboule, 2013; Rao et al., 2014). To what degree this is also true for higher levels of structural chromatin organization is not fully understood yet, but some TADs switch between genomic neighborhoods, or compartments, in a cell-type-dependent manner (Dixon et al., 2015; Rao et al., 2014). The genome of embryonic stem cells (ESCs), for example, uniquely brings together distal chromosomal regions that are densely packed with pluripotency factors, which creates a configuration proposed to contribute to maintenance of pluripotency (de Wit et al., 2013). Furthermore, it has been shown that the pluripotency genes *Nanog* and *Oct4* make specific long-range interactions in ESC and iPSCs, which are lost during differentiation (Apostolou et al., 2013; Denholtz et al., 2013; Wei et al., 2013). However, little is known to what extent the overall 3D genome of somatic cells and their iPS derivatives differ, how stable such differences are, and how similar the 3D configurations of iPSC and ESC genomes are.

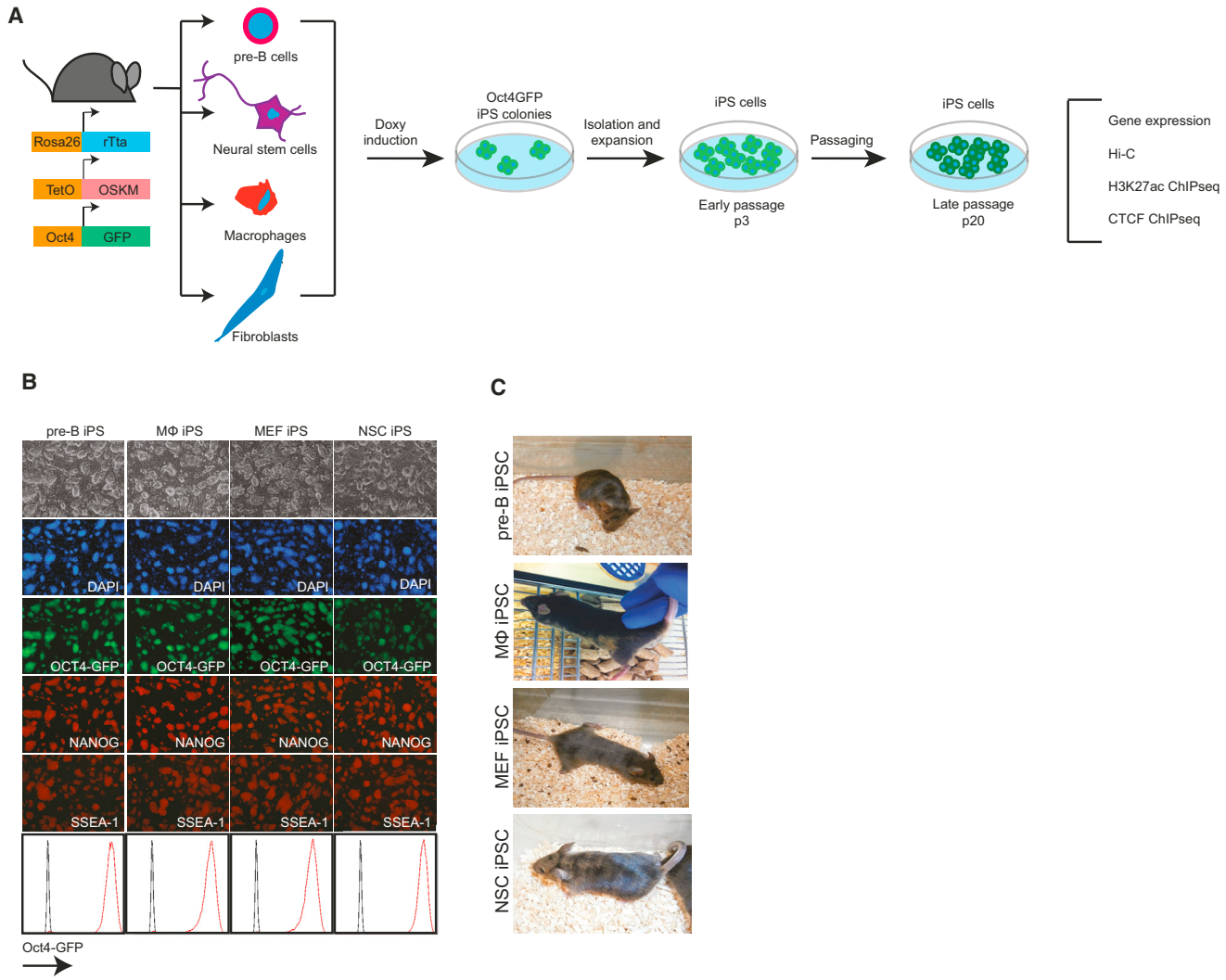


Figure 1. Generation of p3 and p20 iPSCs from Somatic Cells Isolated from Reprogrammable Mice

(A) Schematic representation of the generation and analysis of p3 and p20 iPSCs from pre-B cells, neural stem cells (NSCs), macrophages (MΦ), and mouse embryonic fibroblasts (MEFs) isolated from reprogrammable, OSKM-inducible, Oct4-GFP reporter mice. P3 iPS lines show ESC characteristics. See also Figure S1.

(B) P3 iPSCs derived from pre-B cells, NSC, MΦ and MEF show characteristic ESC-like morphology and express GFP from the Oct4-GFP reporter and NANOG and SSEA-1 proteins.

(C) Chimeras generated by blastocyst injection of p3 iPSCs.

Here we show that somatic cell reprogramming is accompanied by massive changes in genome topology, which, irrespective of the cell type of origin, converge on the 3D structure of the pluripotent genome. Despite this, distinct topological features separate early passage iPSCs according to their cell type of origin, and these differences seem to be acquired during reprogramming in a founder-cell-dependent manner.

RESULTS

To study how reprogramming of somatic cells affects nuclear organization, we used reprogrammable, OSKM-inducible, mice (Carey et al., 2010). We generated three independent iPS cell lines each from four different founder cell types, i.e., pre-B cells, bone-marrow-derived macrophages (MΦ), neural stem cells (NSCs), and mouse embryonic fibroblasts (MEFs) (Figure 1A). iPSCs were established after picking of doxy-independent colonies at day 20 of reprogramming (15 days of reprogramming in the presence of doxy + 5 days without doxy) and were expanded for an additional 3 passages or 20 passages to obtain early (p3) and late (p20) passage iPS lines, respectively. Both p3 and p20 passage iPSC lines showed characteristic ESC-like morphology, expressed markers of pluripotency, as shown by immunofluorescence and fluorescence-activated cell sorting (FACS), and could be maintained in a transgene-independent manner (Figures 1B, 2A, and S1A). Furthermore, p3 iPSCs derived from each cell type gave rise to chimeras upon blastocyst injection (Figure 1C). Importantly, embryoid bodies (EBs)

obtained from the various p3 iPSC lines showed a differentiation bias toward the cell type of origin (Figure S1B). EBs derived from pre-B-iPSCs and MΦ-iPSCs showed higher expression of the hematopoietic-associated genes *Cd45*, *Cd41*, *Itgam* (*Mac-1*), and *Hoxb4*, while the neuronal-associated genes *Nestin* and *Pax6* were more highly expressed in EBs derived from NSC-iPSCs. In contrast, the endoderm associated gene *Sox7* showed no such a bias (Figure S1C). The blood and neural origin bias was lost in p20 iPSCs (Figure S1D). These findings confirm the tissue-of-origin memory of early passage iPSCs described previously (Bar-Nur et al., 2011; Kim et al., 2010; Polo et al., 2010).

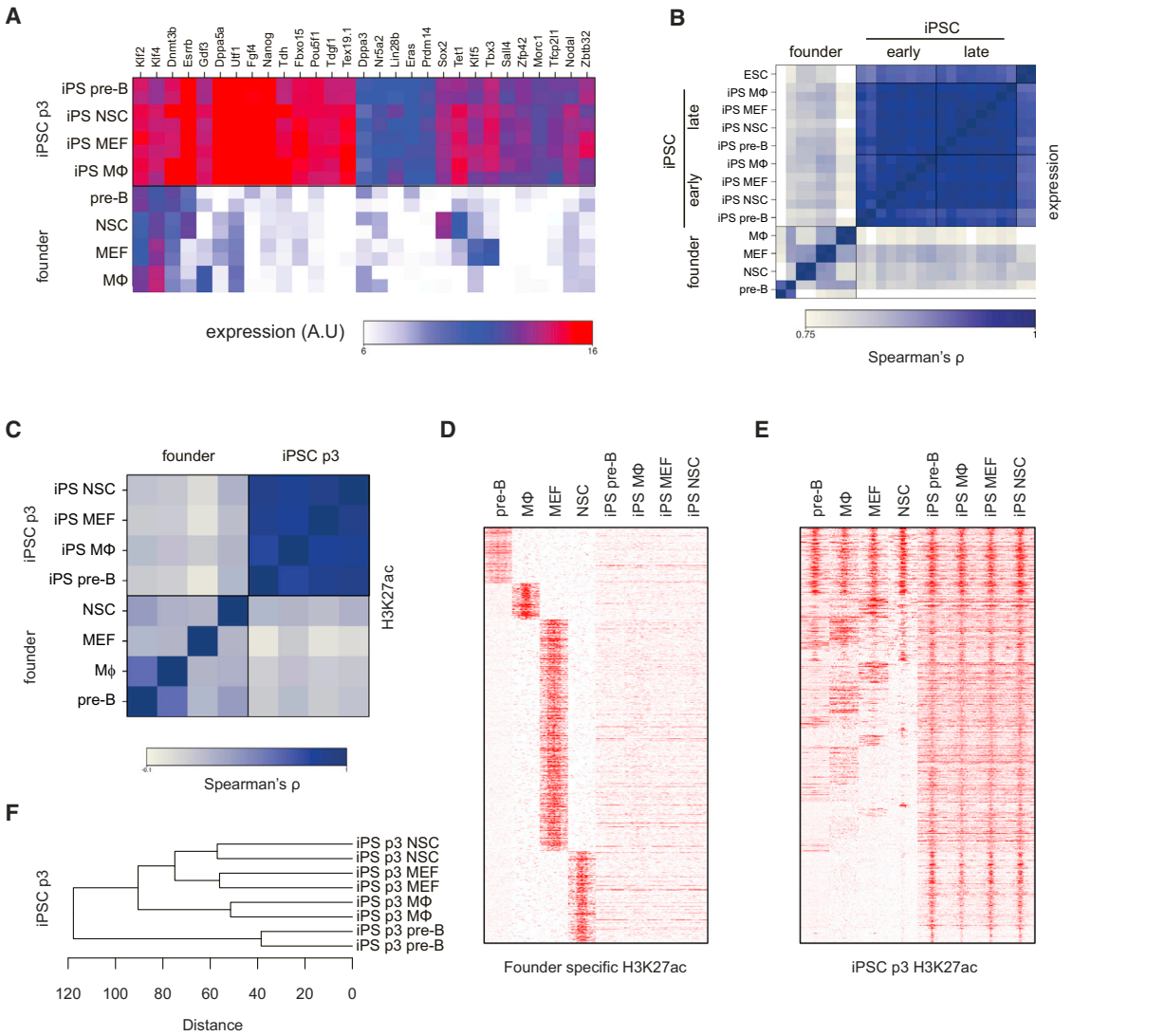
To systematically compare the transcriptomes of the four founder cell types and of their iPSC derivatives, we performed genome-wide expression analysis. Pre-Bs, NSCs, MΦs, and MEFs had highly divergent transcription profiles yet were very similar between biological duplicates (Figure 2B). We identified 13,880 unique genes that were differentially expressed (at an FDR of 0.01) by the four cell types. Reprogramming of the four somatic cells resulted in loss of tissue-specific expression programs and yielded transcriptomes that highly correlated between all iPS lines and strongly corresponded to an ESC-like expression pattern (Figures 2A and 2B). Similarly, active enhancer profiles, as defined by histone H3 lysine 27 acetylation (H3K27ac) were also very different between founder cells (Creyghton et al., 2010; Lane et al., 2014; Yue et al., 2014) (Figure 2D) but became highly similar after reprogramming in all iPSCs (Figures 2C and 2E), with cell-of-origin-specific sites having very little residual enhancer marks (Figure 2D).

Although expression profiles in both p3 and p20 iPSCs were highly correlated, unsupervised hierarchical clustering of the transcription profiles revealed that p3 passage iPSCs derived from the same cell of origin clustered together (Figure 2F). Indeed, 1717 unique genes were differentially expressed (FDR of 0.05) between p3 iPSCs derived from different founders. This indicated that the cell type of origin left a mark on transcription in fully reprogrammed p3 passage iPSCs. In contrast, p20 iPSCs showed clearly reduced gene expression clustering (Figure S2A) and only two genes with reproducibly founder-dependent differential expression (FDR of 0.05). This is consistent with previous reports demonstrating that iPSCs transiently retain cell-type-of-origin differences in gene expression (Kim et al., 2010; Polo et al., 2010). To further understand transcriptional differences and similarities between p3 passage iPSCs derived from different cell types, we performed k-means clustering and identified seven clusters of genes in p3 iPSCs (Figure 2G). To determine whether their differential expression echoed previously established transcription patterns in the cells of origin, we calculated the correlation in expression on a gene-by-gene basis between founders and iPSCs. For p3 iPSCs the expression of genes in most clusters showed little correlation with that of the founders, with the exception of cluster 2 (Figures 2G and S2B) ($p < 0.001$, Wilcoxon rank sum test). In agreement, Gene Ontology analysis using Webgestalt (Zhang et al., 2005) only showed a clear enrichment of functional categories for cluster 2, namely for genes involved in the immune system and collagen (Figure S2C). Consistent with this annotation, genes in this cluster are highly expressed in MΦ-p3 iPSCs and in MEF-p3 iPSCs but not in NSC-derived iPSCs. Vice versa, when we clustered genes according to tissue-specific expression patterns in the founders, we also found no indications for systematic memory of cell-type-specific expression programs in the corresponding p3 and p20 iPSCs (Figure S2D). Furthermore, when we selected tissue-specific genes based on their >4-fold higher expression in one of the founding somatic cell lines compared the other three, we found that none of the genes maintained this difference in transcriptional output in the corresponding iPSCs. Only when we lowered the threshold for differential expression among iPSCs to 1.4-fold, a small number of genes (22) were reproducibly scored across all lineages (16 for MΦ, 5 for pre-B, 1 for MEF, and 0 for NSC) as having a founder-specific expression profile. Collectively, this showed that overall cell-type-specific expression programs were efficiently erased and replaced by ESC-like transcription programs during reprogramming. In addition, reproducible cell-type-of-origin-specific gene expression patterns exist in p3 iPSCs, although only for one gene cluster this is related to a founder-specific gene expression program. The remaining founder-dependent gene expression patterns in p3 iPSCs appear reproducibly acquired during reprogramming, possibly as a consequence of cell-type-specific reprogramming events.

Reprogramming Is Accompanied by Massive Chromosome Topology Changes

To investigate how nuclear organization changes during reprogramming, we used a frequently cutting restriction enzyme (DpnII) to generate genome-wide Hi-C contact maps for each of the four founder cell types and their respective p3 and p20 iPSC derivatives. We prepared Hi-C data from two to three independent clones (with the exception of the NSC founder, from which only one Hi-C library was created), which we combined for each cell type, resulting in Hi-C maps based on 39-72M valid reads (Table S1).

We first compared the overall chromosome organization of the different cell types by partitioning the genome into regions of 300kb and plotting all interactions between these regions as a heatmap and a correlation heatmap (Figure 3A; Table S2). As expected, most interactions occurred in *cis* and at close distance, although many long-range contacts beyond the level of TADs can be observed. Closer inspection of the heatmaps revealed clear differences in genome folding between the different somatic cell types. Reprogramming erased many of these tissue-specific configurations and created a 3D genome that was highly similar between all iPS lines. Previous Hi-C studies have shown that chromosomal regions can be segregated into two main nuclear compartments (Lieberman-Aiden et al., 2009). Regions within the same compartment preferentially interacted with each other and were highly enriched for, respectively, active (compartment A) or inactive chromatin (compartment B) (Lieberman-Aiden et al., 2009). We found that the distribution of genomic regions between these compartments differed strongly between the somatic lines (Figures 3A–3C), with 28% of the genome located in a different compartment in at least one of the founders. This percentage is not very different from the 36% of the genome that was found to change compartments during in vitro differentiation (Dixon et al., 2015). Genes tissue specifically residing in the A compartment showed increased expression levels in the corresponding cell type, while those tissue specifically located in the B compartment showed reduced



(legend on next page)

transcriptional output as compared with that in the other tissues (Figure S3). An example of such a gene (*Ly6d*) with tissue-specific activity and corresponding nuclear location in pre-B cells is shown in Figure 3D.

For every cell type of origin, reprogramming efficiently erased the tissue-specific division of genomic regions over the A and B compartments and induced a compartment structure that is very similar to what is found in ESCs (Figures 3A–3C). Already in p3 passage iPSCs, 99.9% of the genome resided in identical compartments, which increased to 99.95% in p20 iPSCs. For example, *Sox2* was found to be expressed and located in the active compartment in NSCs, but inactive and located in the B compartment in the other somatic cell types. Reprogramming induced the expression of endogenous *Sox2* and relocated the gene to the active compartment in iPSCs derived from pre-B, MΦ, and MEF (Figure 3E). Genes that switched compartments during reprogramming changed their expression levels more often than genes that did not switch compartments (Figure 3F), with genes relocating from B to A showing an overall increase in expression, and vice versa, genes switching from A to B showing an overall reduction in expression in iPSCs. A similar, albeit more subtle, correlation between expression changes and compartment switching was observed before during in vitro differentiation (Dixon et al., 2015).

These data demonstrate that somatic founder cells generally structure their chromosomes very differently, but reprogramming induces these differences to disappear and genomes to adopt an ESC-like higher order structure (Figure 3C). Thus, independently of the somatic founder cell type, reprogramming into iPSCs leads to a convergence of the 3D genomes to an ESC-like topology.

Reprogramming Establishes Topological Hallmarks of the Pluripotent Genome

The pluripotent genome was previously found to have some unique topological features (de Wit et al., 2013; Denholtz et al., 2013). One such hallmark is the preferential long-range contacts between regions with high-density binding sites of the pluripotency factors NANOG, OCT4, and SOX2 (de Wit et al., 2013). To understand whether reprogramming reshapes the genome to establish this ESC-specific configuration, we performed paired-end spatial chromatin analysis (PE-SCAn), an algorithm combining ChIP-seq data with Hi-C data (de Wit et al., 2013). PE-SCAn applied to founder cells showed, as expected, no spatial clustering of NANOG-, OCT4-, and SOX2-rich regions. P3 passage iPSCs, however, did show strong preferential clustering of these regions (Figures 4A and 4B). This indicates that during reprogramming chromosomes are refolded such that clusters of binding sites of pluripotency factors contact each other in nuclear space. We found that 3D configurations specific for somatic cells were simultaneously lost during reprogramming. For example, regions dense in binding sites of PU.1, an essential factor for the development of immune cells, showed preferential contacts in pre-B cells but not in the other somatic cell types or in p3 iPSCs, including the pre-B-cell-derived lines (Figure 4A). Another interesting example is SOX2, which is expressed both in NSCs and iPSCs but binds to a completely different repertoire of sites in the two cell types (Lodato et al., 2013). PE-SCAn analysis revealed that the NSC-specific SOX2-associated regions exclusively clustered in NSCs, whereas the ES-specific SOX2-bound regions only did so in iPSCs (Figure 4B). Altogether, the results demonstrate that reprogramming reshapes the genome to erase somatic cell-type-specific topological features and establishes a conformation unique to the pluripotent cell genome.

Reprogramming Alters Domain Organization

Our data so far established that the relative positioning of chromosomal regions is strongly reorganized during reprogramming, raising the question whether this is also true for more local topological features such as TADs. To assess how domain organization is affected by reprogramming, we annotated TADs in the Hi-C map from p3 passage iPSCs (Table S3, with a size range of 20.4 kb to 7.6 Mb, median size 526 kb) and compared chromatin interactions in these domains between all cell types.

Recently various reports introduced a domain score, a measure for the degree of connectivity within a TAD (Chandra et al., 2015; Dixon et al., 2015). For each TAD we calculated this domain score as the fraction of intradomain contacts over its total number of *cis* contacts (Figure 4C). As seen before (Dixon et al., 2015), genes in domains with a high domain score showed higher expression than genes in domains with a lower score (Figure S4A). A high domain score therefore does not reflect compaction, but rather compartmentalization of the domain. In agreement, lamina-associated domains (LADs), which have been linked with compacted chromatin (Peric-Hupkes et al., 2010), overlapped mostly with TADs that had a low domain score (Figure S4B).

Comparing the domain score between the different Hi-C maps revealed a strong correlation between biological replicates, but not between different somatic cell types (Figure 4D). Tissue-specific genes often resided in TADs with an above average,

Figure 2. Cell of Origin Influences Gene Expression in p3 iPSCs

(A) Heatmap showing the expression of pluripotency genes in founder cells and p3 iPSCs.

(B) Correlation matrix (Spearman's ρ) of expression data of founder cells, p3 iPSCs, p20 iPSCs, and E14 ESC (n = 2 for all).

(C) Correlation matrix (Spearman's ρ) of H3K27ac ChIP-seq of pro-B, MΦ, MEFs, NSCs, and p3 iPSCs.

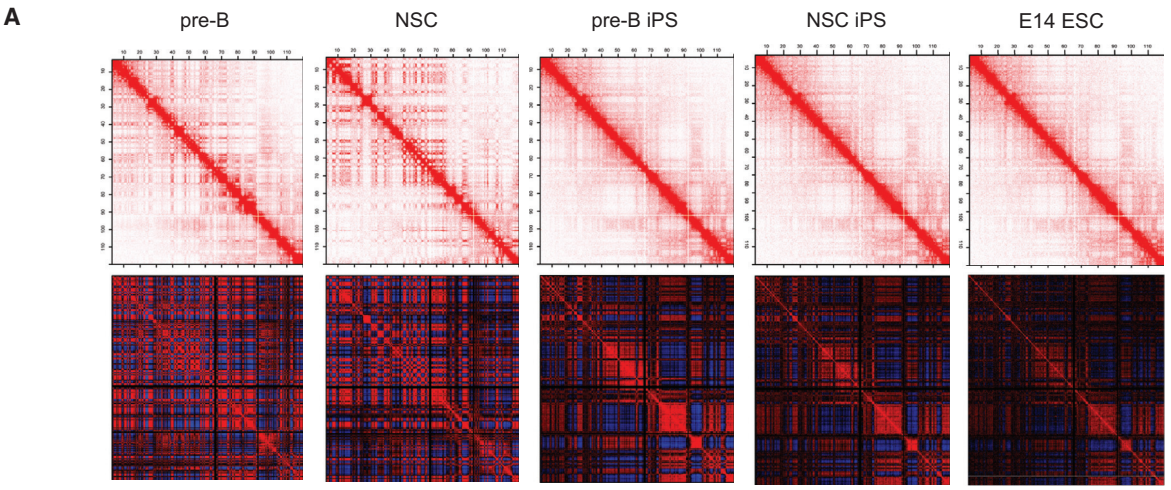
(D) Heatmap representing H3K27ac enrichment in pro-B, MΦ, MEFs, NSCs, and p3 iPSC. The coverage within 3-kb upstream and downstream of the summit of each H3K27ac peak was calculated and shown for the indicated cell types.

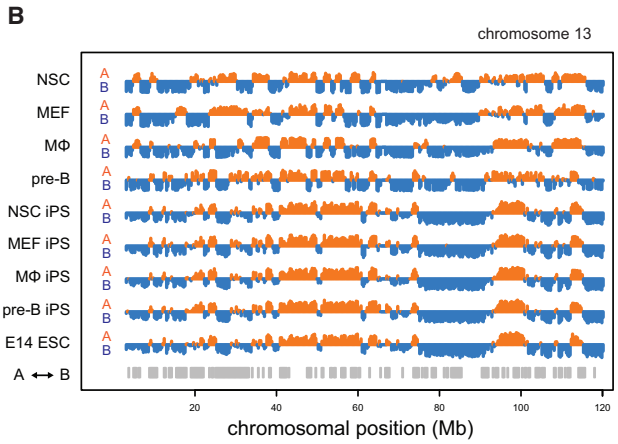
(E) Heatmap representing H3K27ac enrichment in pro-B, MΦ, MEFs, NSCs, and p3 iPSCs (similar to D). Genomic regions are the combined ChIP-seq peaks identified in p3 iPSCs.

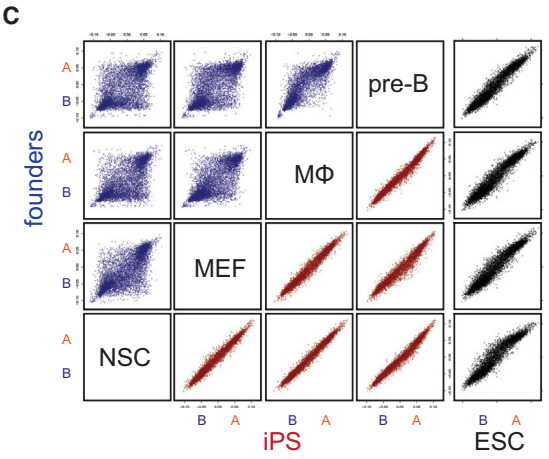
(F) Unsupervised hierarchical clustering of the transcription profiles of p3 iPSCs.

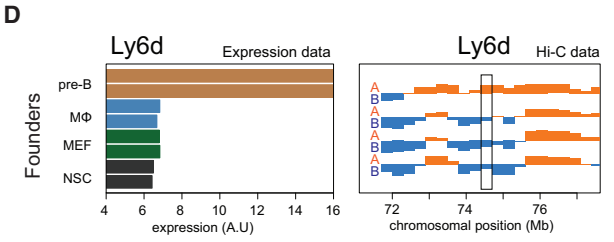
(G) k-means clustering of 1,717 differentially expressed genes between the p3 iPSCs. Relative expression change of each differentially expressed gene (as compared with their median expression level across all experiments) is indicated for the founder cells, p3 iPSC and p20 iPSC (n = 2, for every cell type). Clusters and expression difference range (color gradient) are shown.

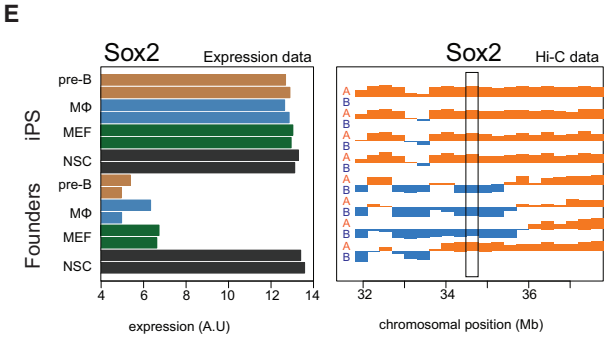
See also Figure S2.

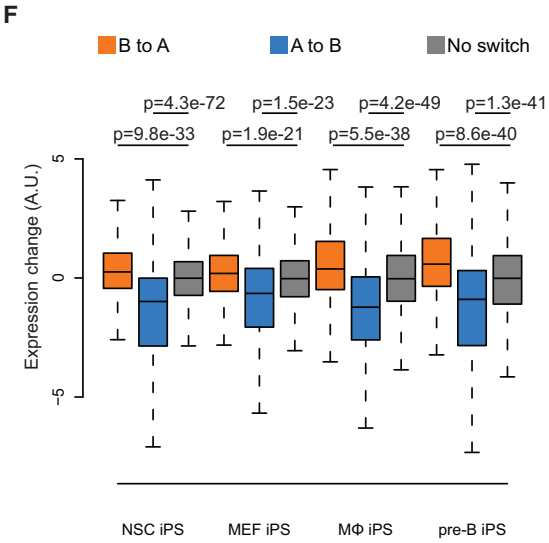












(legend on next page)

cell-type-specific, domain score (Figure S4C). Reprogramming of the four cell types resulted in highly correlated domain scores across all iPSCs and similar to the scores measured in ESCs (Figure 4D). Thus, also at the intra-TAD level reprogramming was accompanied by large-scale topological changes, resulting in domains that overall are fairly uniformly organized between all iPSC lines.

Chromatin Looping Changes at Tissue-Specific Genes

Although our datasets are relatively sparse compared with published multibillion read datasets (Rao et al., 2014), we could clearly see chromatin loops at multiple sites in the genome (exemplified for two genes in Figure 5A). To identify chromatin loops in a systematic manner, we developed an algorithm that deals with relatively sparse data (see Supplemental Experimental Procedures). We combined all the p3 iPSC line data into a dataset that contained over 220 million unique contacts. Based on these data, we identified 5,168 chromatin loops, 97% of which were found within a TAD (Table S4). The median loop size was 249 kb; 5% spanned less than 99 kb, and 5% spanned more than 590 kb. Intersecting these loops with CTCF binding sites observed by ChIP-seq in the iPSCs (see below) revealed that 67% of the chromatin loops were formed between pairs of CTCF sites. Consistent with previous reports (Rao et al., 2014), the participating CTCF sites were found overwhelmingly oriented in a convergent manner (Figure S5A). Analysis of the Hi-C datasets of the founder tissues revealed 2,741, 2,666, 846, 536 loops for pre-B cells, NSCs, MEFs, and MΦ, respectively (Table S4), which is in line with the number of informative read pairs sequenced per library.

To study the structural dynamics during reprogramming at the level of chromatin loops, we developed a Hi-C meta-loop analysis (Figure S5B). In this analysis, we aligned one of our Hi-C datasets on the loop calling data from another Hi-C dataset, scaled the data between the beginning and end of each loop, and projected all contact maps on top of each other. The resulting plots visualize whether loops scored in one tissue are overall conserved in other tissues. When we performed a meta-loop analysis of the p3 iPS loops in E14 ESC Hi-C data, we found a strong enrichment of contacts at the site of the loops (Figure 5B), showing that most loops called in iPSCs were conserved in ESCs. When we intersected the same iPS loops with Hi-C data from the founder tissues, we also found enrichment of contacts at the site of the loops (shown for pre-B cells in Figure 5B). Thus, the majority of chromatin loops scored in iPSCs is tissue invariable.

We then defined founder-specific loops as those being present in a given cell type but absent in all other somatic cell types. We could identify tissue-specific loops in pre-B, NSC, and iPSCs (346, 27, and 246, respectively). Genes associated with pre-B cell-specific loops were expressed significantly higher in pre-B cells than in the other cells ($p < 0.0001$, for all comparisons, Wilcoxon rank sum test; Figures 5C and S5C), supporting the idea that lineage-restricted chromatin loops contribute to tissue-specific transcriptional activity (Rao et al., 2014). We next asked whether such somatic cell-type-specific loops are retained in the corresponding iPSCs. Figure 5D shows that in general this was not the case. When we took the pre-B cell-specific chromatin loops, the meta-loop plot of pre-B cell derived iPSCs no longer showed preferential contacts between the loop anchors: in fact, the plot looked identical to that of another iPS line or of NSCs (Figure 5D). In all tissues, however, the intervening sequences appeared to form contact domains even in the absence of an encompassing chromatin loop. The expression bias seen in somatic cells (Figure 5C) disappeared with the loss of tissue-specific loops during reprogramming. Thus, the genes involved in pre-B-cell-specific chromatin loops, which are more highly expressed in pre-B cells than in other somatic cells, showed no residual increased expression in pre-B-derived iPSCs as compared with other iPSCs (Figure S5D). In addition, when inspecting individual gene loci, we could clearly see the disappearance of tissue-specific chromatin loops. For example, *Ikaros*, *Ctgf*, *Anks1b*, and *F7* loci all contained strong chromatin loops in pre-B, MEF, NSC, and MΦ, respectively, of which most disappeared during reprogramming (Figure 5E). While the Hi-C meta-loop analysis showed that the majority of founder-specific chromatin loops dissolve during reprogramming, it still remained possible that rare but possibly important tissue-specific loops are transmitted to their iPSC derivatives. To investigate this in more detail, we visually inspected the 346 pre-B-specific loops in the Hi-C datasets of pre-B cells and its reprogrammed derivatives. Also, upon visual inspection, the great majority (326) of these loops was found to dissolve upon reprogramming (for example, see Figure 5E). Twenty pre-B cell loops also showed an (often weak) looping signal in the pre-B derived p3 iPSCs, but 17 of these also had a looping signal in one or more of the unrelated iPSCs (for examples, see Figure S5E). The three pre-B-cell-specific loops that were transmitted to and exclusively detectable in pre-B-derived iPSCs located near *Mcl1*, *Ccdc69*, and *Amph* genes not known to be important for B cell identity. We therefore find no evidence for looping memory and conclude that tissue-specific chromatin loops near cell identity genes effectively dissolve during reprogramming.

Figure 3. Reprogramming Is Accompanied by Massive Chromosome Topology Changes

(A) Normalized Hi-C contact frequency heatmap (top) and correlation heatmap (bottom) at 300-kb resolution for chr13.

(B) The first eigenvector of the Hi-C correlation matrix is plotted along the linear sequence of chr13 and used as the A (orange) and B (blue) compartment segmentation. Grey bars indicate regions located in differential compartments in one of the cell lines.

(C) Scatterplot of the eigenvector for all 300-kb regions in the genome comparing the founders (blue), the p3 iPSC (red), and the p3 iPSC compared with E14 ESC (black).

(D) Expression and A/B compartment location of the pre-B-specific Ly6d gene. Box indicates chromosomal position of Ly6d. See also Figure S3.

(E) Expression and A/B compartment location of the *Sox2* gene. Box indicates chromosomal position of *Sox2*.

(F) Distribution of gene expression change for genes that switched from the B to A compartment (orange), from the A to the B (blue) or that did not switch compartments (gray). Genes that relocated from the B compartment to the A compartment during reprogramming showed an increase in expression compared with genes that did not switch compartments, while genes that switched from the A to the B compartment showed reduced expression in iPSCs.

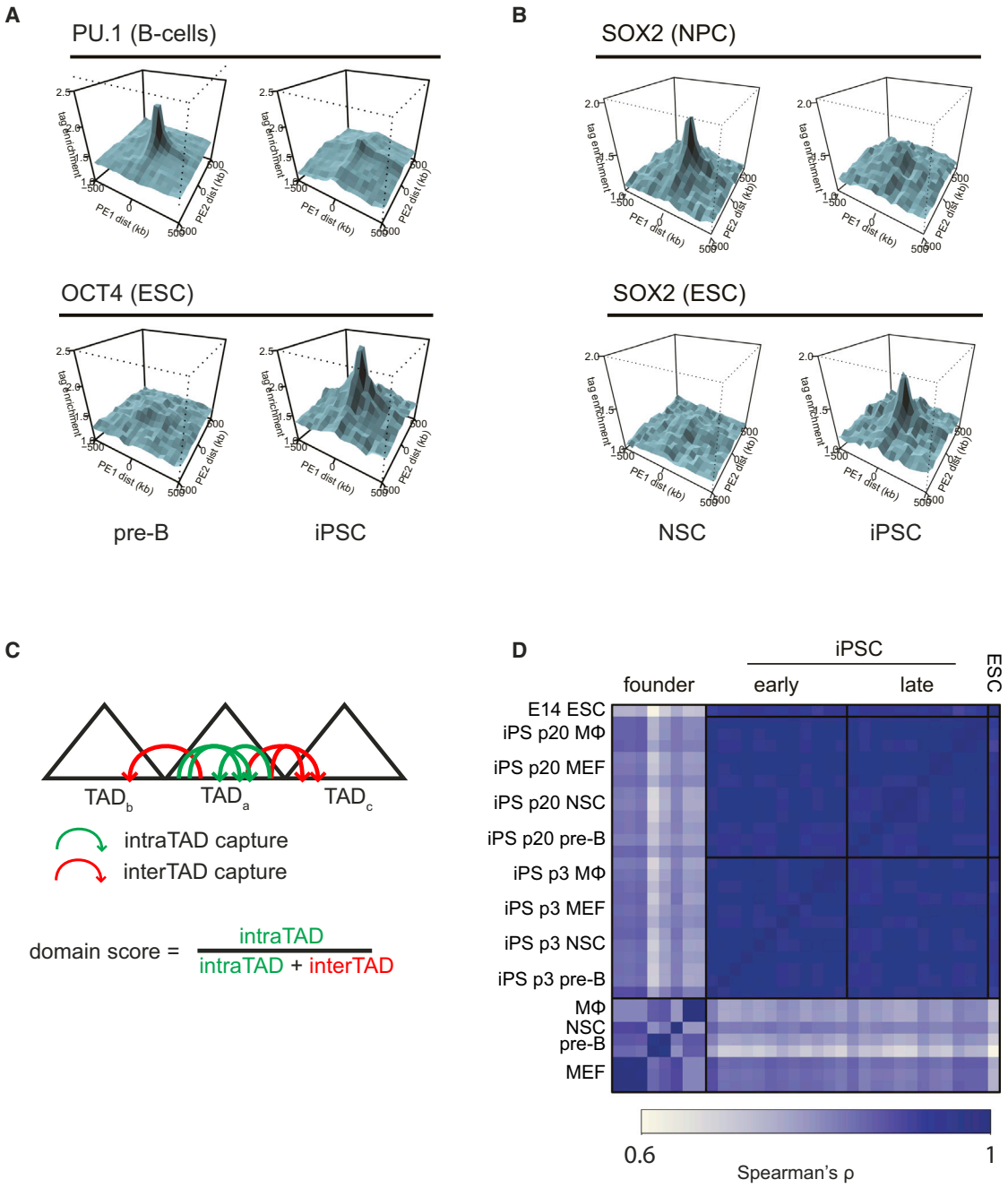


Figure 4. Reprogramming Establishes Topological Hallmarks of the Pluripotent Genome

(A) PE-SCAn analysis plots in which long-range intrachromosomal Hi-C contacts (>5 Mb) in pre-B and corresponding p3 iPSC are aligned to ChIP-seq binding site clusters (more than five sites in 50 kb) of PU.1 in B cells (top) and OCT4 in ESCs (bottom).

(B) Same as (A), but using Hi-C data from NSC and corresponding p3 iPSC and SOX2 binding site clusters in NPCs (top) and ESCs (bottom).

(C) Schematic explaining the domain score.

(D) Correlation matrix (Spearman's ρ) of the domain score for pre-B (n = 2), MΦ (n = 2), MEFs (n = 3), NSCs (n = 1) and p3 and p20 iPSCs derived from pre-B (n = 3), MΦ (n = 3), MEFs (n = 3), and NSCs (n = 3).

See also Figure S4.

Next, we analyzed the iPSC-specific loops in more detail. Intersections with other Hi-C data showed that these loops were also present in ESCs but absent in all founder Hi-C datasets (Figure 5F). This confirmed that they were formed de novo during reprogramming and were specific for pluripotent cells. When we checked the individual loci involved, we noticed that they often contained pluripotency genes, including *Sox2*, *Prdm14*, *Nanog*, *Fgf4*, and *Sox1* (Figure 5G). Thus, while many chromatin loops,

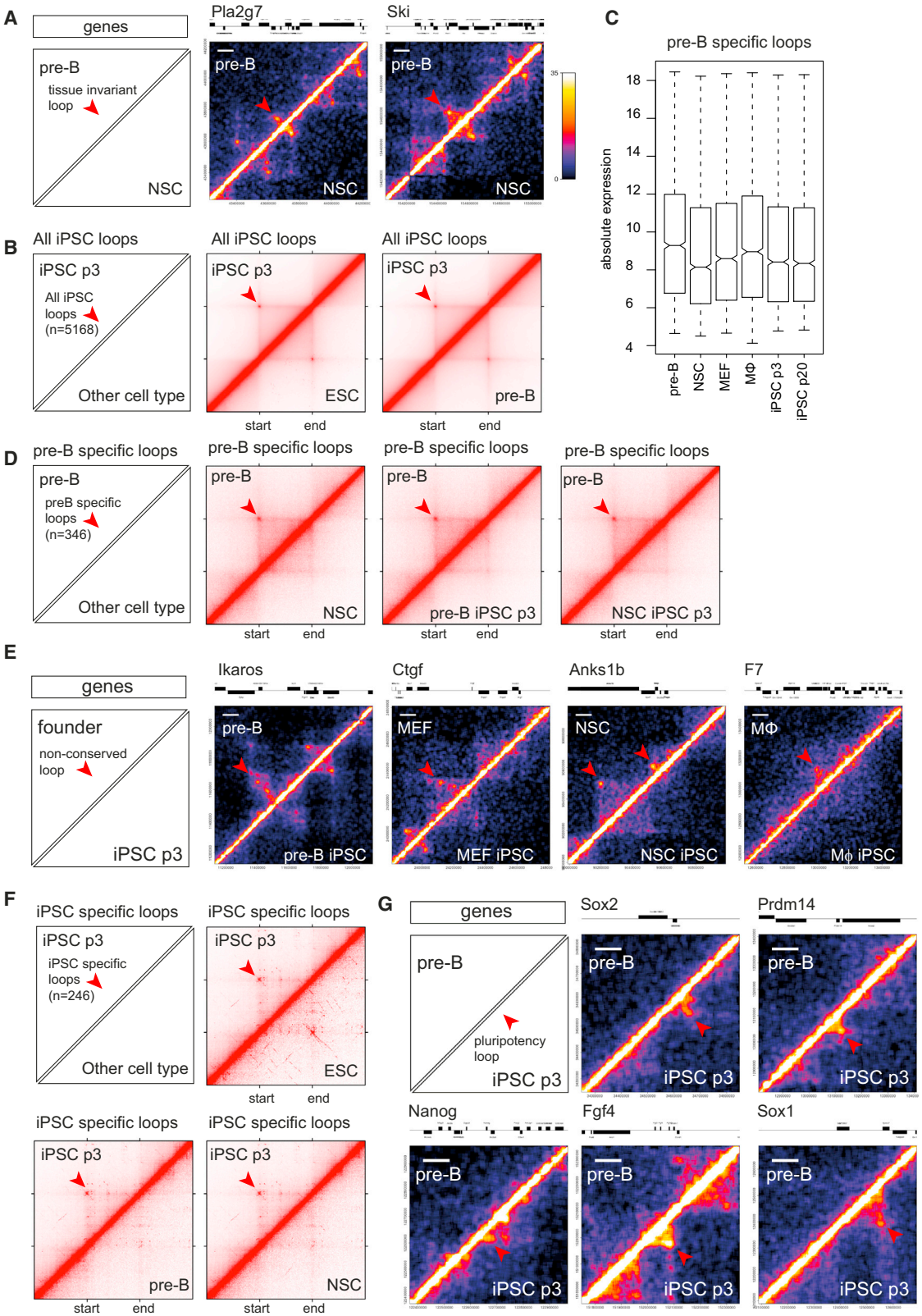


Figure 5. Chromatin Looping Changes at Tissue-Specific Genes

(A) Hi-C interaction heatmap showing chromatin loops in pre-B cells (upper triangle) and NSC (lower triangle) at the Pla2g7 (left) and Ski (right) locus. Tissue-invariant loops are indicated with a red arrowhead. Color scale indicates normalized Hi-C score (see Experimental Procedures). White scale indicates 100 kb.

(legend continued on next page)

particularly those between domain boundaries, were conserved between ESCs, iPSCs and somatic cells types, there were also tissue-specific chromatin loops that contributed to cell-specific expression programs. During reprogramming, most somatic cell-specific loops were lost while new specific regulatory contacts, often involving pluripotency genes, were established.

We considered the possibility that not all important pluripotent loops are properly established during reprogramming. Our Hi-C datasets were of insufficient resolution to conclusively analyze this, but to get an impression, we focused on 207 loci previously established to contain super-enhancers in ESCs (Whyte et al., 2013), reasoning that the topologies of these loci are likely important for pluripotency. We could appreciate chromatin loops at 84 of these loci in ESCs, 37 of which were not appreciable in our pre-B Hi-C dataset. All but two of these loops were also detectable in the iPSC Hi-C data (for examples, see Figures S5F and S5G). The exceptions included the genes *Tsc22d1* and, interestingly, *Sall4*, a known key pluripotency gene, where ES-specific loops were less appreciable in iPSCs. *Sall4* expression was nevertheless induced in all iPSCs (Figure 2A), making it difficult to judge the functional significance of these differential DNA contacts. We conclude that reprogramming is accompanied by efficient loop formation at nearly all analyzed key pluripotent regulatory sequences, demonstrating that it robustly reshapes the chromatin not only at the TAD but also at the sub-TAD level to adopt a pluripotent conformation.

Acquisition of a Cell-of-Origin-Dependent 3D Genome during Reprogramming

To further search for a possible topological memory during reprogramming, we asked whether structural domains exist which keep a tissue-restricted positioning in either the A or B compartment during reprogramming. As shown before, nearly all domains adopted the same nuclear environment in p3 passage iPSCs irrespective of their cell of origin (Figure 3). Only three domains in pre-B cells and none in the other founder lines were found to maintain a cell-of-origin-specific location during reprogramming. The fact that only in pre-B cells a few domains memorized their founder-specific location during reprogramming suggests that topological memory during reprogramming, if it exists, is not detectable at this level of genome organization.

We then looked at the domain score; as explained before, this gives a measure for intra-TAD connectivity. Unsupervised hierarchical clustering based on the domain score revealed that, with three exceptions, all p3 iPSC lines clustered based on their cell of origin (Figure 6A). Intriguingly, the three exceptions (one Mϕ, one MEF, and one NSC-derived iPSC line) were the only three lines showing aneuploidy, carrying an extra copy of a chromosome (Figure S6A). Such genomic instability during reprogramming has been reported before (Weissbein et al., 2014). To identify the domains with a cell-of-origin-specific topology, we carried out k-means clustering on domain scores from p3 iPSCs lines with a normal karyotype. Figure 6B shows clusters of domains grouped according to conserved differences in domain score across the four types of p3 iPSCs. Unexpectedly, the same domain clusters in the founders showed uncorrelated scores within and between tissues, indicating that the cell-of-origin-dependent domain structures are not reflective of residual founder-specific structures and therefore not a consequence of 3D memory. To further understand, this we again performed k-means clustering but now based on the domain scores measured in the founder lines (Figure 6C). Seven distinct clusters of domains with unique founder-dependent structures could be delineated, but their tissue-restricted architectural features were not appreciable in either p3 or p20 iPSCs and were therefore efficiently erased during reprogramming. Indeed, there was no significant overlap between the founder and p3 iPSC defining domains ($p = 0.3$, hypergeometric test). Collectively, this shows that the 3D genome of iPSCs carries features that allow linking the cells to their cell of origin. These topological features, however, are not remnants of founder-specific 3D structures and therefore do not reflect "memory." Rather, we propose that these distinguishing topological domain features are acquired during reprogramming in a reproducible and cell-of-origin-dependent manner.

Exploring the Cell-of-Origin-Dependent Topological Features

To explore the origin of founder-dependent domain structures, we first investigated whether they are associated with distinct expression patterns in p3 iPSCs. For this, we intersected the clusters of structural domains with the clusters of genes that showed cell-of-origin-dependent expression patterns in p3 iPSCs. No correlation was observed (data not shown), not even when considering all expressed genes in these domains and asking whether the domains showed any distinguishable transcriptional output in p3 iPSCs (Figure 6D). Thus, there seems no (causal) relationship between domains having a cell-of-origin-specific topology in p3 iPSCs and the expression of genes contained within these domains.

One could speculate that the erasure of founder-specific transcription programs is accompanied by tissue-restricted deposition of epigenetic marks at the corresponding genes, which, if maintained during reprogramming, causes cell-of-origin-specific structural features in p3 iPSCs. To investigate this, we performed an identical analysis as above, but now intersecting the Hi-C data from p3 iPSCs with founder-specific transcriptome data. Two iPSC clusters with cell-of-origin-specific domain

(B) Meta-loop analysis of all loops identified in p3 iPSCs. Hi-C data from all p3 iPSCs combined (upper triangle) are compared with Hi-C data from E14 ESC and pre-B cells (lower triangle). Loops are scaled to have the same size. In addition to the loop itself also 50% of the genome upstream of the loop and 50% of the genome downstream of the loop are shown. See also Figure S5.

(C) Absolute gene expression for genes located in pre-B-cell-specific loops.

(D) Meta-loop analysis of pre-B-specific loops.

(E) Hi-C interaction heatmap at the *Ikaros*, *Ctgf*, *Anks1b*, and *F7* loci.

(F) Meta-loop analysis of iPSC-specific loops.

(G) Hi-C interaction heatmap for pre-B cells (upper triangle) and pre-B p3 iPSC (lower triangle) at the *Sox2*, *Prdm14*, *Nanog*, *Fgf4*, and *Sox1* loci. Arrowheads at the *Sox2* and *Nanog* loci correspond to previously identified loops in ESC and iPSC using 3C and 5C (Apostolou et al., 2013; Kagey et al., 2010; Phillips-Cremins et al., 2013).

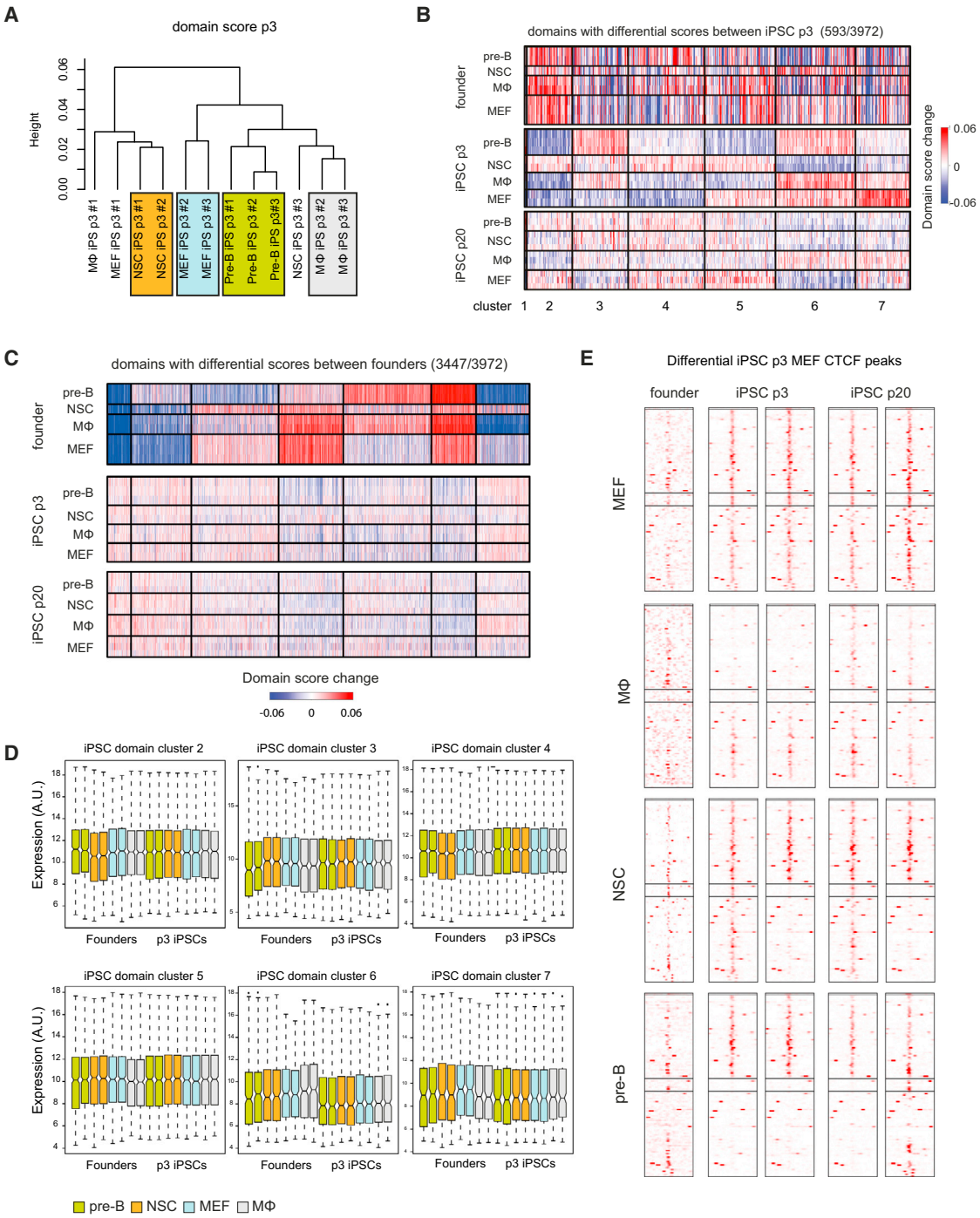


Figure 6. Cell-of-Origin-Specific 3D Genome Features in p3 iPSCs

(A) Unsupervised hierarchical clustering of the domain scores of p3 iPSCs.

(B) k-means clustering of TADs with differential domain scores between the p3 iPSCs. The relative domain score of each differential TAD is indicated.

(C) k-means clustering of TADs with differential domain scores between the founders. The relative domain score of each differential TAD is indicated.

(D) Absolute expression in the founders and in p3 iPSCs, of genes located in the clusters of TADs with differential domain scores at the p3 iPSC stage (see Figure 6B).

(E) Heatmap representing differential CTCF peaks in MEF p3 iPSCs, showing their presence in all founders, p3 iPSCs, and p20 iPSCs.

See also Figure S6.

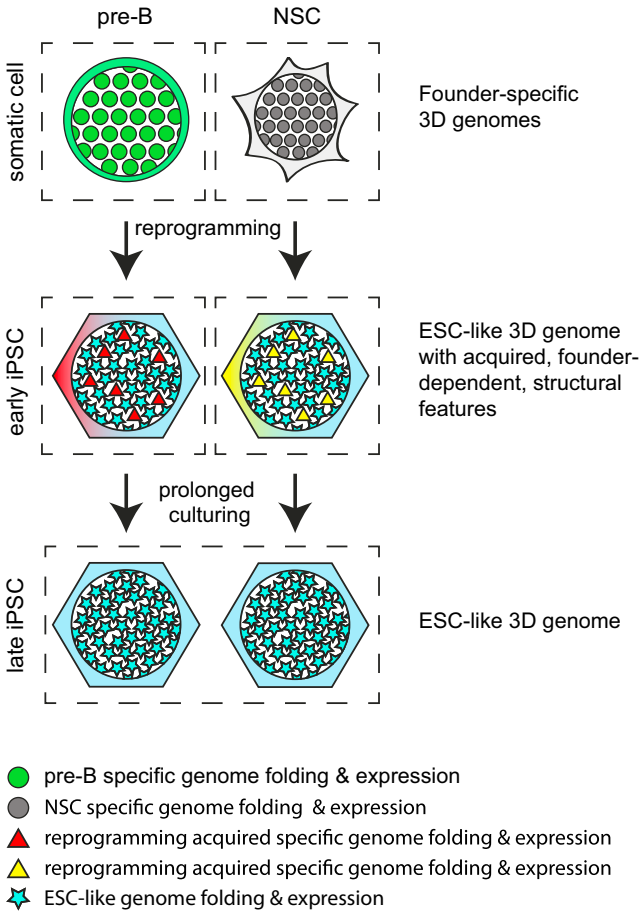


Figure 7. Graphical Summary of the Data

Somatic cell reprogramming is accompanied by massive changes in the somatic-cell-specific 3D genome and expression program, which, irrespective of the cell type of origin, induces convergence to an ESC-like 3D genome and expression program. Despite this, distinct topological features and differential gene expression separate p3 iPSCs according to their cell type of origin and those seem acquired during reprogramming in a cell type of origin dependent manner. These early iPSC distinct features are lost upon continued passaging.

scores were significantly enriched for genes with a founder-specific expression signature. Also, for some clusters, the overall transcriptional output differed specifically in the corresponding founder cells (Figure 6D).

Finally, we asked whether CTCF, a key chromatin architectural protein, could be involved in the establishment of specific structural features during reprogramming. For this, we performed CTCF ChIP-seq experiments, using a total of eight p3 and eight p20 iPSC lines, i.e., in duplicate for all founder tissues (Figure S6B). Already at p3, 50091 CTCF binding sites were shared between all eight iPSC lines. Reproducible, cell-of-origin-dependent differential CTCF binding sites were also observed, but these were relatively rare (327 for pre-B, 136 for NSC, 143 for MΦ, and 116 for MEF iPSCs) and showed low ChIP-seq scores (Figure S6C). Most of the differential cell-of-origin CTCF binding sites in p3 iPSCs were still conserved in the corresponding p20 iPSCs (Figure 6F). However, they were not detectable as CTCF bound sites in the corresponding founder cells, as judged from published ChIP-seq data. Thus, as also concluded for founder-associated expression patterns and domain structures, the founder-associated CTCF binding events observed in p3 iPSCs seem to have been acquired during reprogramming, rather than representing memory. Moreover, they were not enriched in any of the p3 passage structural domain clusters, implying that the cell-of-origin-specific 3D genome features are not related to rare and weak differential, cell-of-origin-associated, CTCF-binding events.

DISCUSSION

Our results demonstrate that somatic cell genomes have highly distinct 3D structures that after cell reprogramming into iPSCs converge into an ESC-like topology. We find that reprogramming induces (1) large-scale switching of TADs between the A and B compartments, (2) the disassembly of inter-TAD contacts mediated by founder-specific transcription factors and the construction of new contacts mediated by pluripotency factors, and (3) the disruption of founder-specific chromatin loops and de novo establishment of pluripotent chromatin loops. Interestingly, despite these massive structural reorganizations that reproducibly induce nearly identical, ESC-like, 3D genomes across all different iPS cell lines, p3 passage iPSCs derived from different somatic cell types can still be distinguished based on specific genome topology features (Figure 7). The cells can also be distinguished based on their expression signatures (Figure 2F), as reported before (Polo et al., 2010). In addition to the cell-of-origin-specific topological features and expression patterns in iPSCs, we observe founder cell-associated CTCF-binding events. However, differing from previous studies, our systematic analyses suggest that in general the cell-of-origin-specific expression patterns in iPSC lines do not reflect residual expression patterns inherited from the respective founders. Similarly, the cell-of-origin-specific topological features that we observe in iPSCs are not architectural features of the founder genomes that survived topological reorganization during reprogramming. The same is true for the cell-of-origin-dependent CTCF binding events that we observe in iPSCs; we find no evidence that they were already occupied by CTCF in the somatic founder cells. Instead, in all instances, these patterns seem to be acquired during reprogramming in a reproducible and cell-of-origin-specific manner. The finding that expression of most cell-type-specific genes can no longer be detected in iPSCs may reflect the fact that efficient erasure of somatic-cell-specific transcription programs is a prerequisite for the establishment of a pluripotent cell population (Buganim et al., 2013). Similarly, our data suggest that iPSCs can only tolerate architectural fluctuations and changes in CTCF binding events that do not challenge, via their impact on gene expression, the establishment, and maintenance of the pluripotent state. As such, we propose they represent innocuous side effects of reprogramming.

The observation that iPSCs derived from different somatic cell types have distinct transcriptional and genome topology features seems best explained by assuming that different somatic cell types follow distinct reprogramming trajectories. Consistent with this idea, single-cell expression analyses during reprogramming of MEFs into iPSCs identified the transient expression of a unique set of surface markers absent in both fibroblasts and iPS

cells, revealing transitions through distinct intermediate stages (Lujan et al., 2015; O'Malley et al., 2013). Furthermore, MEFs may follow different paths, depending on the reprogramming method used (Zunder et al., 2015). While little is known about the reprogramming paths in cell types other than fibroblasts, the reactivation sequence of pluripotency genes during reprogramming has been shown to differ between MEFs and pre-B cells (Buganim et al., 2012; Di Stefano et al., 2014; O'Malley et al., 2013), supporting the idea that different cell types may exhibit distinct reprogramming trajectories. Future research is necessary to uncover the molecular basis of the correlation between the distinct reprogramming trajectories, tissue-specific gene expression signatures, and genome topologies in the different cell-of-origin-derived iPSCs.

It has been reported that low passage iPSCs harbor residual DNA methylation signatures characteristic of their somatic tissue of origin, and these have been suggested to explain the cells' biases in differentiation potential (Kim et al., 2010). Whether differential methylation is also linked to the distinct topological features that we have observed in p3 iPSCs remains to be investigated. However, given the effective replacement of a founder-specific 3D genome by a pluripotent 3D genome during reprogramming, plus the fact that the distinctive structural features are subtle and not reflective of architectural hallmarks of the founder genome, we consider it unlikely that the cell-of-origin-specific differentiation bias seen in early iPSCs is a consequence of specific topological features.

EXPERIMENTAL PROCEDURES

Additional information and details regarding this work may be found in the Supplemental Experimental Procedures.

Reprogramming

Pre-B cells, MEFs, MΦ, and NSCs were isolated from a "reprogrammable mouse" line containing a doxycycline-inducible OSKM cassette, the reverse tetracycline transactivator (rtTA) (Carey et al., 2010), and an Oct4-GFP reporter transgene (Boiani et al., 2002), as described (Di Stefano et al., 2014). Reprogramming experiments with pre-B cells were performed as previously described (Di Stefano et al., 2014); with MEFs, macrophages and NSCs were conducted by plating 100,000 cells per well on gelatinized plates seeded with irradiated MEFs, using ESC medium supplemented with 2 μg/ml of doxycycline. For the isolation of iPSC lines, doxycycline was washed out after 15 days of reprogramming, and colonies with ESC-like morphology were picked at 20 days before further passaging. iPSC lines were expanded for an additional 9 days (three passages) to obtain P3 iPS cell lines or for 20 passages to obtain P20 iPS cell lines. Detailed information on cell culture and characterization of the iPSCs is provided in the Supplemental Experimental Procedures.

ChIP-Seq

ChIP experiments were performed as described previously (van Oevelen et al., 2008) using an antibody against H3K27ac (ab4729, Abcam) and CTCF (Millipore, 07-729). Detailed information is provided in the Supplemental Experimental Procedures.

Hi-C

Pre-B cells, MEFs, NSCs, iPSCs, and E14 ESCs were cross-linked and further processed as DpnII 3C template, as previously described (Splinter et al., 2012). Libraries for paired-end sequencing were generated from sonicated, ~500- to 800-bp size-selected, 3C templates using the TruSeq DNA LT Sample Prep Kit (Illumina). Detailed information on Hi-C template generation and Hi-C data analysis is provided in the Supplemental Experimental Procedures. Hi-C eigenvalues, Domain segmentation in iPSC p3, and a list with genomic coordinates of all the called chromatin loops can be found in the Tables S2, S3, and S4, respectively.

ACCESSION NUMBERS

The accession number for the expression and sequencing data reported in this paper is GEO: GSE76481.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2016.01.007.

AUTHOR CONTRIBUTIONS

P.H.L.K., B.D.S., E.d.W., W.d.L., and T.G. conceived the study, participated in its design, and drafted the manuscript. P.H.L.K. performed the Hi-C experiments and analyzed the data. B.D.S. performed the cell culture and animal and molecular biology experiments and analyzed the data. E.d.W. computationally analyzed the data. F.L. performed cell culture experiments. C.v.O. performed the ChIP experiments.

ACKNOWLEDGMENTS

This work was supported by an NWO/CW TOP grant (714.012.002), an NWO VICI grant 724.012.003, a NanoNextNL grant, and a European Research Council Starting Grant (209700, "4C") to W.d.L.; a Ministerio de Educacion y Ciencia, SAF.2012-37167, Fundacio La Marató TV3 120410, AGAUR SGR 1136, and European Research Council Synergy Grant ("4D-Genome) to T.G.; and an ERC Stg (637587, "HAP-PHEN") to E.d.W.

Received: June 26, 2015
Revised: December 1, 2015
Accepted: January 13, 2016
Published: March 10, 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.

Bar-Nur, O., Russ, H.A., Efrat, S., and Benvenisty, N. (2011). Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells. Cell Stem Cell *9*, 17–23.

Boiani, M., Eckardt, S., Schöler, H.R., and McLaughlin, K.J. (2002). Oct4 distribution and level in mouse clones: consequences for pluripotency. Genes Dev. *16*, 1209–1219.

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.

Buganim, Y., Faddah, D.A., and Jaenisch, R. (2013). Mechanisms and models of somatic cell reprogramming. Nat. Rev. Genet. *14*, 427–439.

Carey, B.W., Markoulaki, S., Beard, C., Hanna, J., and Jaenisch, R. (2010). Single-gene transgenic mouse strains for reprogramming adult somatic cells. Nat. Methods *7*, 56–59.

Chandra, T., Ewels, P.A., Schoenfelder, S., Furlan-Magaril, M., Wingett, S.W., Kirschner, K., Thuret, J.Y., Andrews, S., Fraser, P., and Reik, W. (2015). Global reorganization of the nuclear landscape in senescent cells. Cell Rep. *10*, 471–483.

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 Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. Nature *502*, 499–506.

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.

Di Stefano, B., Sardina, J.L., van Oevelen, C., Collombet, S., Kallin, E.M., Vicent, G.P., Lu, J., Thieffry, D., Beato, M., and Graf, T. (2014). C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature *506*, 235–239.

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.

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, 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.

Lane, A.A., Chapuy, B., Lin, C.Y., Tivey, T., Li, H., Townsend, E.C., van Bodegom, D., Day, T.A., Wu, S.C., Liu, H., et al. (2014). Triplication of a 21q22 region contributes to B cell transformation through HMGN1 overexpression and loss of histone H3 Lys27 trimethylation. Nat. Genet. *46*, 618–623.

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.

Lodato, M.A., Ng, C.W., Wamstad, J.A., Cheng, A.W., Thai, K.K., Fraenkel, E., Jaenisch, R., and Boyer, L.A. (2013). SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. PLoS Genet. *9*, e1003288.

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.

Nishino, K., Toyoda, M., Yamazaki-Inoue, M., Fukawatase, Y., Chikazawa, E., Sakaguchi, H., Akutsu, H., and Umezawa, A. (2011). DNA methylation dynamics in human induced pluripotent stem cells over time. PLoS Genet. *7*, e1002085.

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.

O'Malley, J., Skylaki, S., Iwabuchi, K.A., Chantzoura, E., Ruetz, T., Johnsson, A., Tomlinson, S.R., Linnarsson, S., and Kaji, K. (2013). High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. Nature *499*, 88–91.

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.

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.

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

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

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.

van Oevelen, C., Wang, J., Asp, P., Yan, Q., Kaelin, W.G., Jr., Kluger, Y., and Dynlacht, B.D. (2008). A role for mammalian Sin3 in permanent gene silencing. Mol. Cell *32*, 359–370.

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.

Weissbein, U., Benvenisty, N., and Ben-David, U. (2014). Quality control: Genome maintenance in pluripotent stem cells. J. Cell Biol. *204*, 153–163.

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.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al.; Mouse ENCODE Consortium (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355–364.

Zhang, B., Kirov, S., and Snoddy, J. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res. *33*, W741–W748.

Zhao, X.Y., Li, W., Lv, Z., Liu, L., Tong, M., Hai, T., Hao, J., Guo, C.L., Ma, Q.W., Wang, L., et al. (2009). iPS cells produce viable mice through tetraploid complementation. Nature *461*, 86–90.

Zunder, E.R., Lujan, E., Goltsev, Y., Wernig, M., and Nolan, G.P. (2015). A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry. Cell Stem Cell *16*, 323–337.

Cell Stem Cell, Volume *18*

Supplemental Information

Cell-of-Origin-Specific 3D Genome Structure Acquired during Somatic Cell Reprogramming

Peter Hugo Lodewijk Krijger, Bruno Di Stefano, Elzo de Wit, Francesco Limone, Chris van Oevelen, Wouter de Laat, and Thomas Graf

Figure S1

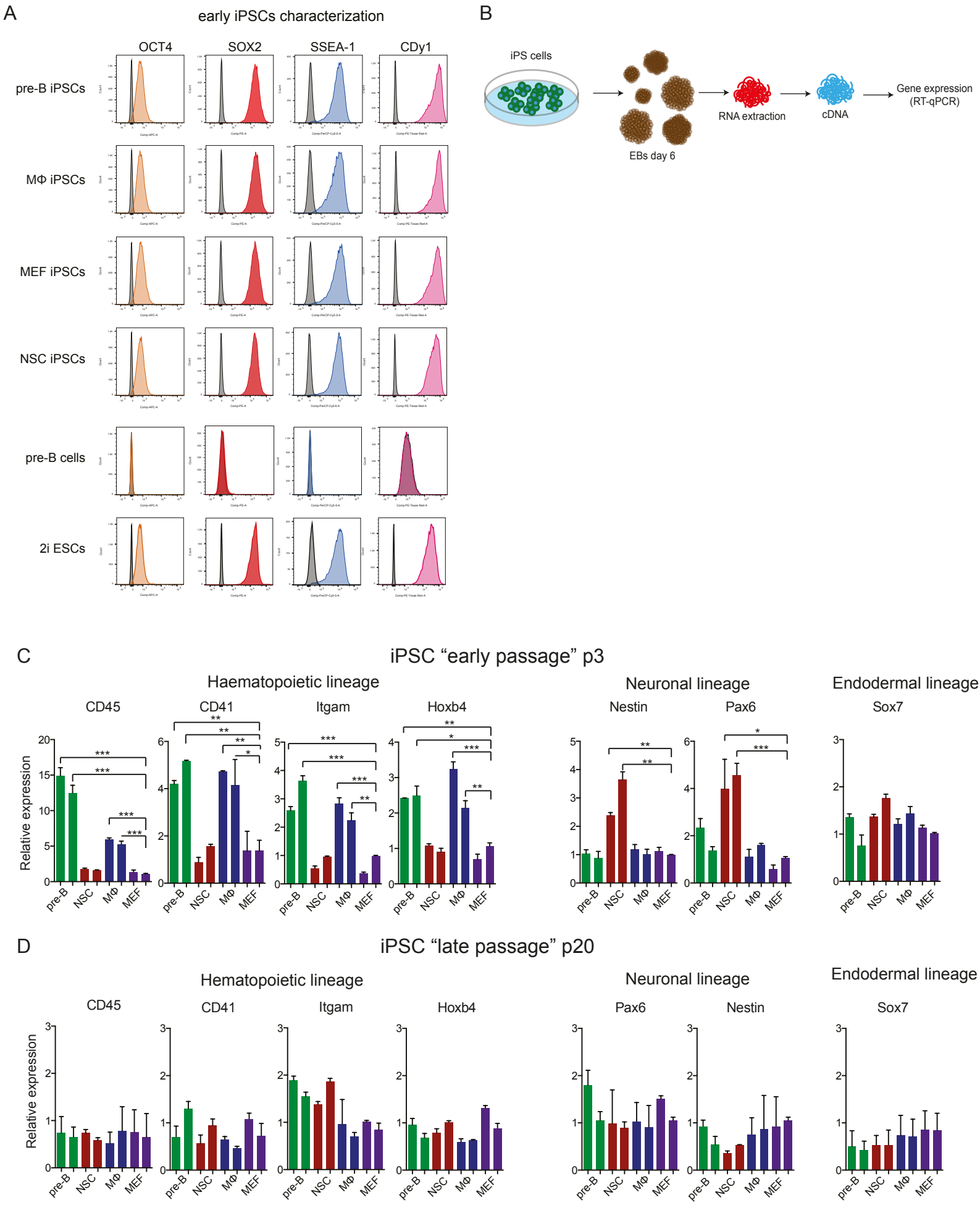
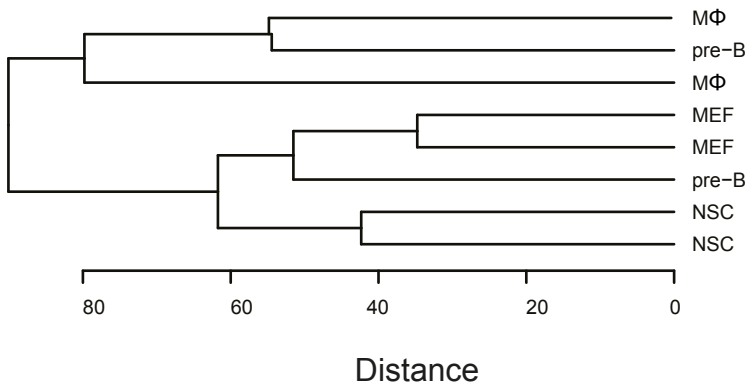
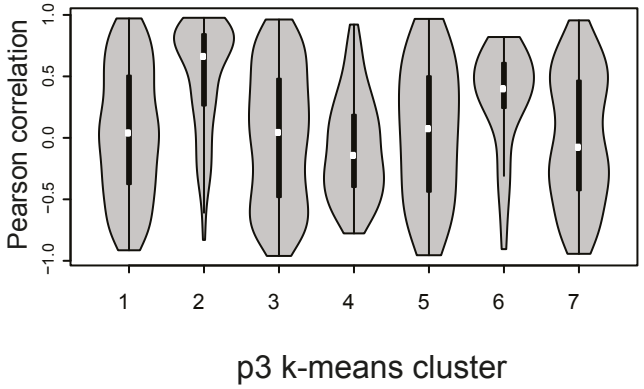

A
early iPSCs characterization
OCT4
SOX2
SSEA-1
CDy1
pre-B iPSCs
MΦ iPSCs
MEF iPSCs
NSC iPSCs
pre-B cells
2i ESCs
B
iPS cells
EBs day 6
RNA extraction
cDNA
Gene expression (RT-qPCR)
C
iPSC "early passage" p3
Haematopoietic lineage
Neuronal lineage
Endodermal lineage
CD45
CD41
Itgam
Hoxb4
Nestin
Pax6
Sox7
Relative expression
pre-B
NSC
MΦ
MEF
D
iPSC "late passage" p20
Hematopoietic lineage
Neuronal lineage
Endodermal lineage
CD45
CD41
Itgam
Hoxb4
Pax6
Nestin
Sox7
Relative expression
pre-B
NSC
MΦ
MEF

Figure S2

A



B



C

	GO class	GO ID	adjusted p-value	number of genes
cluster 1	regulation of response to stimulus	GO:0048583	0.0005	55
	regulation of cell proliferation	GO:0042127	0.0005	37
	cytokine production	GO:0001816	0.0005	20
	cellular response to interferon-beta	GO:0035458	0.0017	4
cluster 2	response to wounding	GO:0009611	1.16E-21	42
	immune system process	GO:0002376	1.74E-16	51
	fibrillar collagen	GO:0005583	1.25E-08	6
	collagen	GO:0005581	2.18E-12	14
	vasculature development	GO:0001944	1.81E-11	27
cluster 3	nucleus	GO:0005634	9.41E-06	80
cluster 4	NA			
cluster 5	zinc ion binding	GO:0008270	4.13E-07	76
	RNA biosynthetic process	GO:0032774	1.82E-05	116
cluster 6	NA			
cluster 7	vacuole	GO:0005773	0.0059	8

D

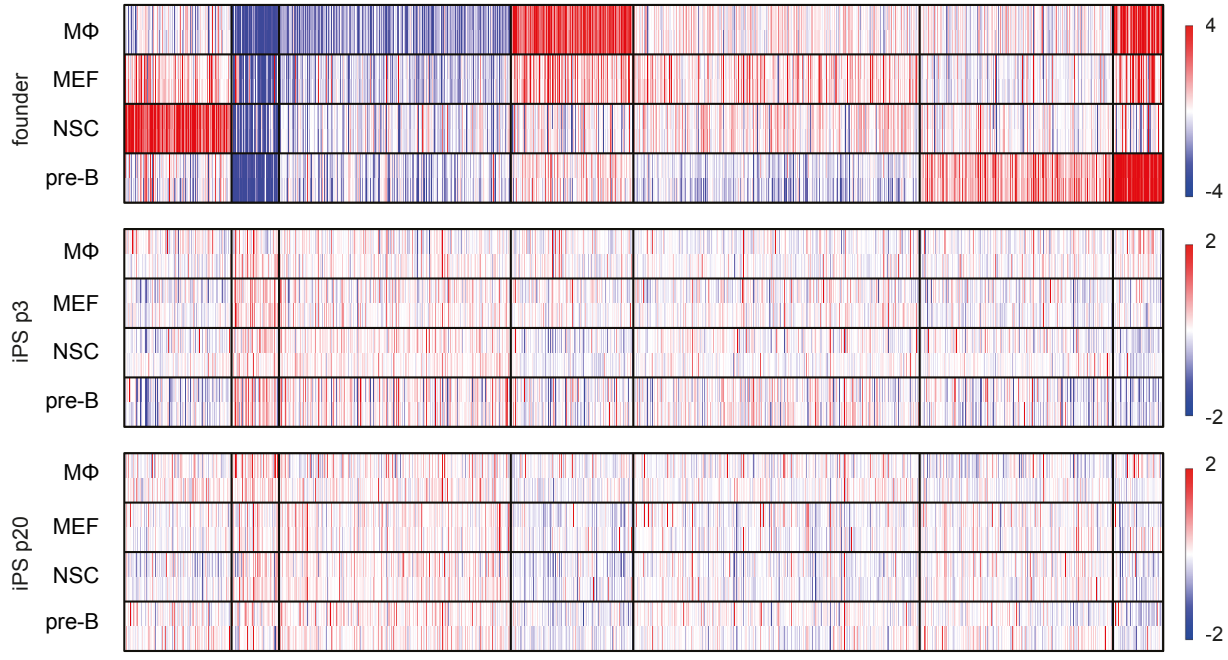


Figure S3

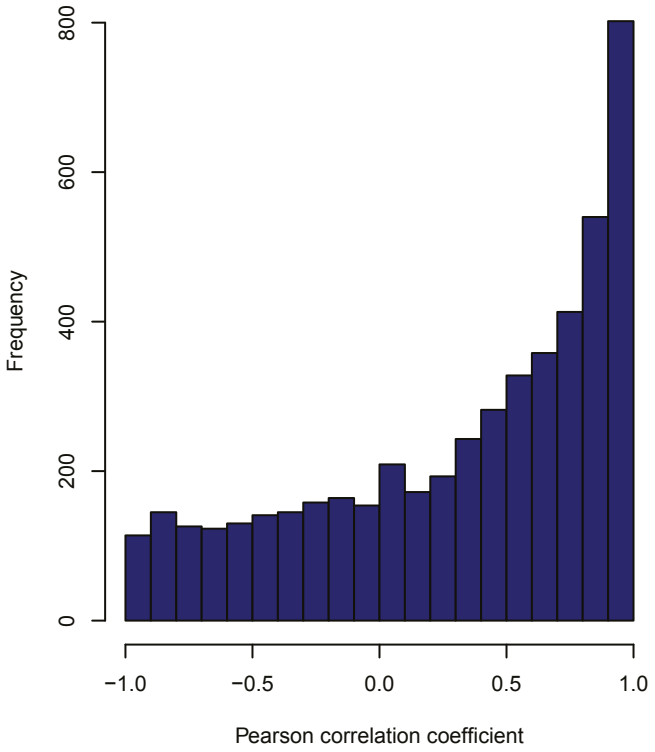
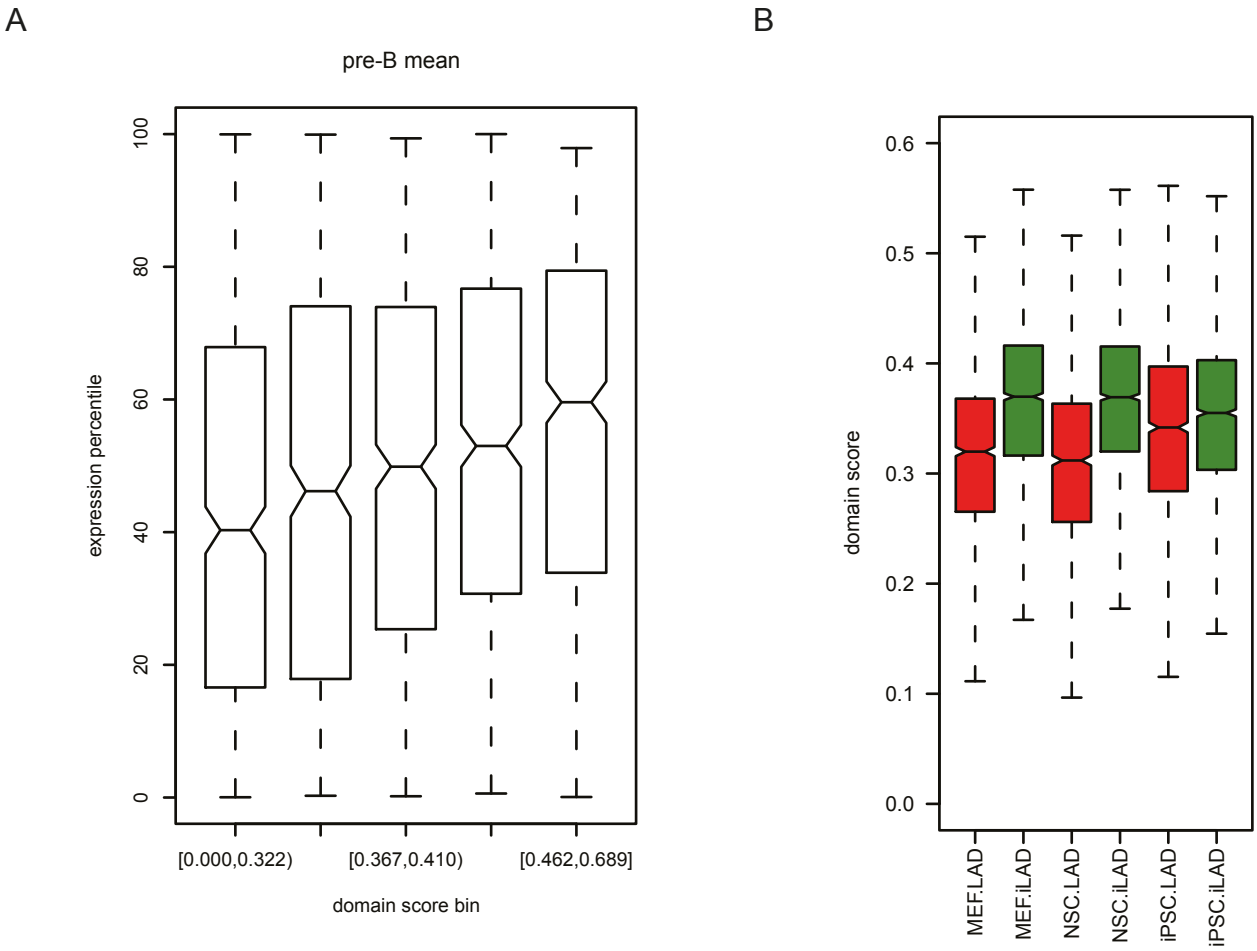

Frequency
0
200
400
600
800
−1.0
−0.5
0.0
0.5
1.0
Pearson correlation coefficient

Figure S4


A
pre-B mean
expression percentile
0
20
40
60
80
100
[0.000,0.322)
[0.367,0.410)
[0.462,0.689]
domain score bin
B
domain score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
MEF.LAD
MEF.iLAD
NSC.LAD
NSC.iLAD
iPSC.LAD
iPSC.iLAD

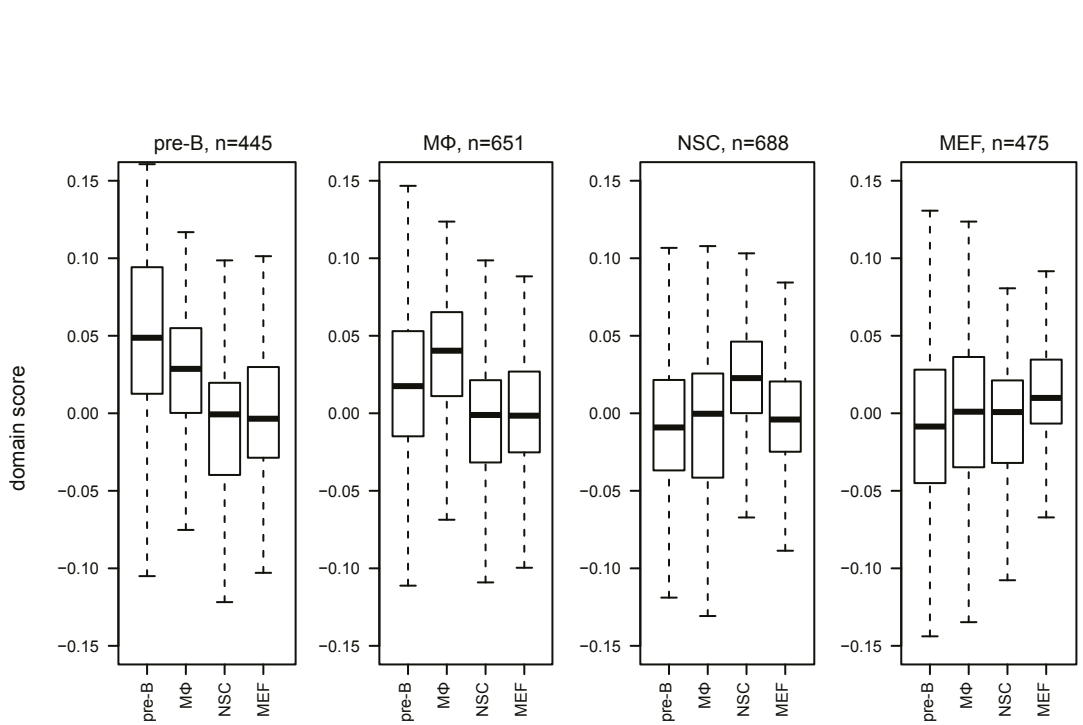

C
pre-B, n=445
MΦ, n=651
NSC, n=688
MEF, n=475
domain score
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
pre-B
MΦ
NSC
MEF

Figure S5

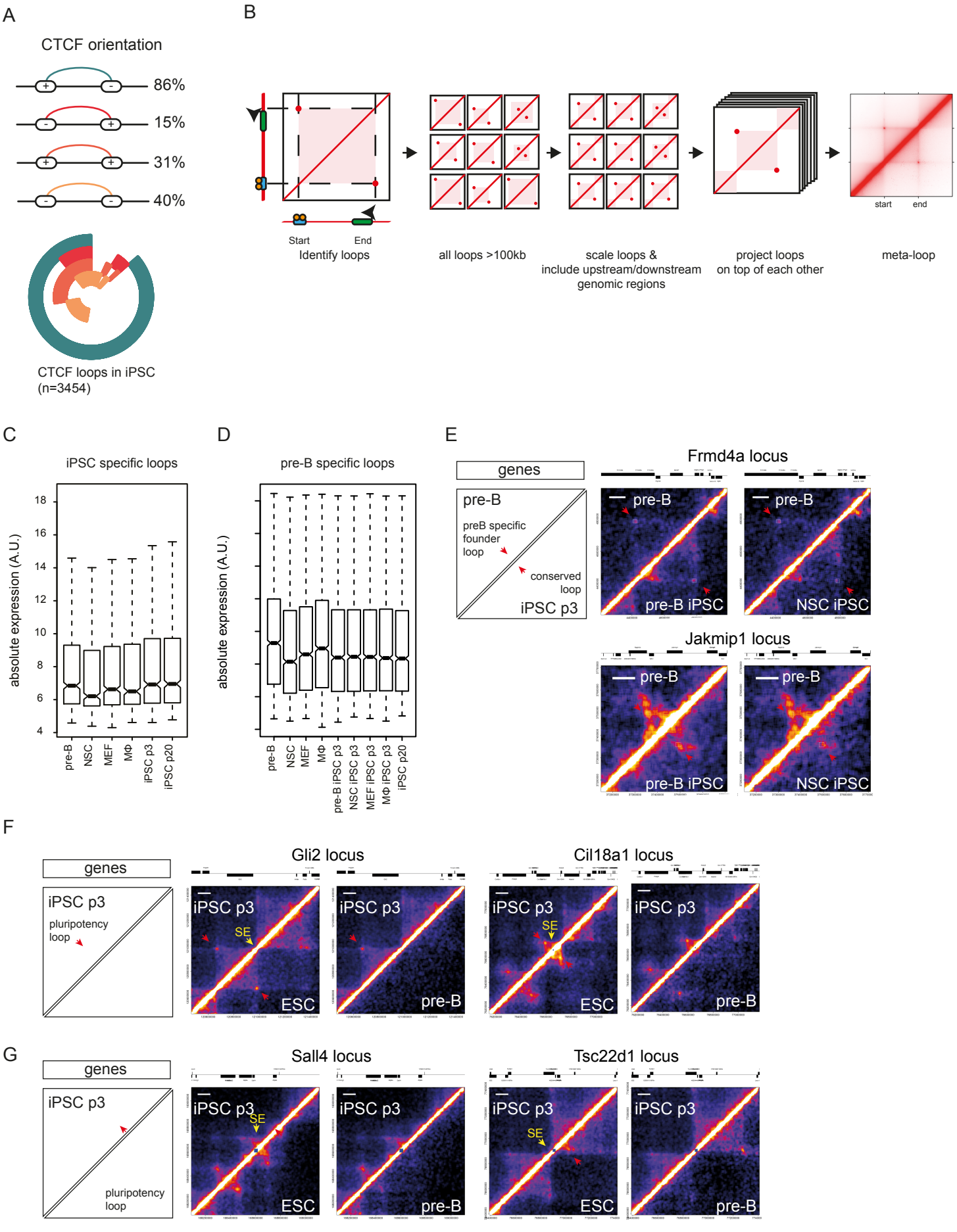
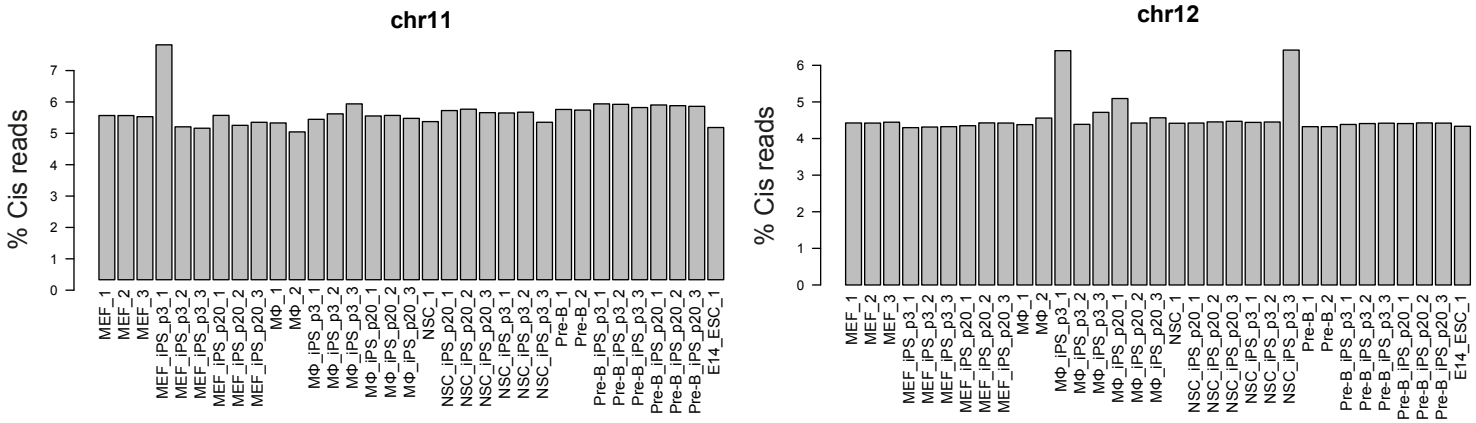

A
CTCF orientation
86%
15%
31%
40%
CTCF loops in iPSC
(n=3454)
B
Start
End
Identify loops
all loops >100kb
scale loops &
include upstream/downstream
genomic regions
project loops
on top of each other
start
end
meta-loop
C
iPSC specific loops
absolute expression (A.U.)
pre-B
NSC
MEF
MΦ
iPSC p3
iPSC p20
D
pre-B specific loops
absolute expression (A.U.)
pre-B
NSC
MEF
MΦ
pre-B iPSC p3
NSC iPSC p3
MEF iPSC p3
MΦ iPSC p3
iPSC p20
E
genes
pre-B
preB specific
founder
loop
conserved
loop
iPSC p3
Frmd4a locus
pre-B
pre-B iPSC
pre-B
NSC iPSC
Jakmip1 locus
pre-B
pre-B iPSC
pre-B
NSC iPSC
F
genes
iPSC p3
pluripotency
loop
Gli2 locus
iPSC p3
SE
ESC
iPSC p3
pre-B
Cil18a1 locus
iPSC p3
SE
ESC
iPSC p3
pre-B
G
genes
iPSC p3
pluripotency
loop
Sall4 locus
iPSC p3
SE
ESC
iPSC p3
pre-B
Tsc22d1 locus
iPSC p3
SE
ESC
iPSC p3
pre-B

Figure S6

A



B

Shared CTCF peaks
iPSC p3

Shared CTCF peaks iPSC p3	MEF #1	MEF #2	Mo #1	Mo #2	NSC #1	NSC #2	pre-B #1	pre-B #2
MEF #1	69152	63001	53970	59026	66341	67832	65647	64723
MEF #2	63001	71707	54260	61396	70196	71250	69664	67580
Mo #1	53970	54260	61074	52518	56437	58149	56375	56507
Mo #2	59026	61396	52518	64721	63784	64410	63336	61943
NSC #1	66341	70196	56437	63784	90396	87422	80209	75848
NSC #2	67832	71250	58149	64410	87422	109839	85347	80581
pre-B #1	65647	69664	56375	63336	80209	85347	87678	76727
pre-B #2	64723	67580	56507	61943	75848	80581	76727	84428

C

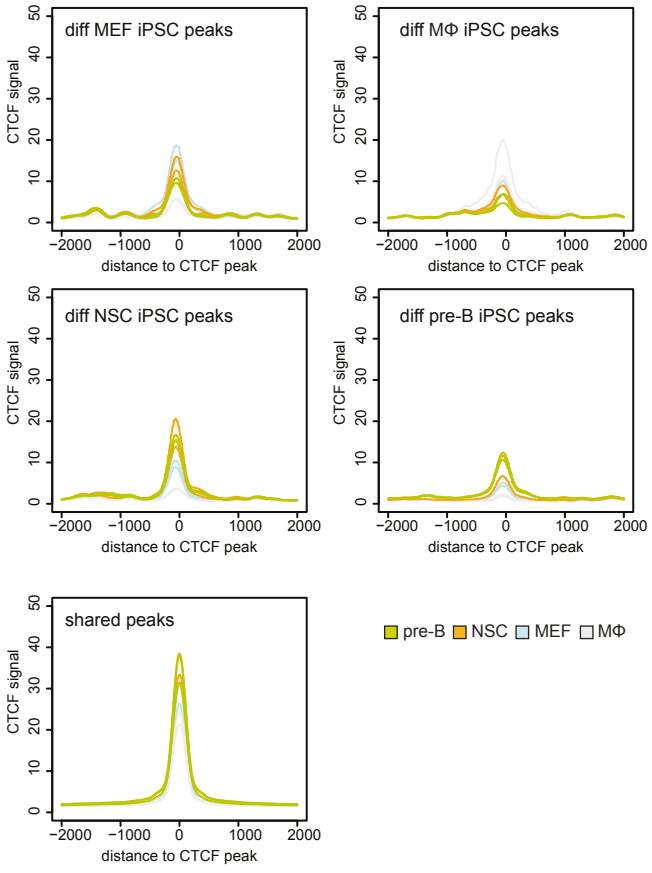


Figure S1. Differentiation bias in early passage iPSCs. Related to Figure 1.

(A) FACS analyses of early passage iPSCs, pre-B and naïve ESC for the expression of the pluripotency transcription factors OCT4 and SOX2, the surface marker SSEA-1 and for the stem cell specific dye CDy1. Essentially 100% of the different P3 iPS cell populations stained positive for these markers, similar to the ESCs.
(B) Schematic representation of the expression analysis of selected lineage specific genes in day 6 embryoid bodies (EBs) derived from iPSCs.
(C) Expression analysis of EBs derived from early passage iPSCs derived from pre-B (blue), NSC (green), MΦ (red) and MEF (purple). Normalized against PGK. Error bars indicate s.d (n=3). Student's t-test *p<0.05, **p<0.01, ***p<0.001
(D) Expression analysis of EBs derived from late passage iPSCs derived from pre-B (blue), NSC (green), MΦ (red) and MEF (purple). Normalized against PGK. Error bars indicate s.d (n=3). Student's t-test *p<0.05, **p<0.01, ***p<0.001

Figure S2. Cell-of-origin influences gene expression in early passage iPSCs in an indirect manner. Related to Figure 2.

(A) Unsupervised hierarchical clustering of the transcription profiles of late passage iPSCs.
(B) Violin plots showing the Pearson correlation scores comparison of expression profiles in founder cells and early passage iPSCs for all differentially expressed genes in the early passage iPSCs. Correlation scores are stratified based on the clusters from the k-means clustering in p3 iPSCs,
(C) Gene Ontology analysis of genes located in k-means clusters indentified in early passage iPSCs. Enrichment anlysis was done in WebGestalt (Zhang et al., 2005).
(D) k-means clustering of differentially expressed genes between pre-B, NSC, MΦ and MEF. Expression change of each differentially expressed gene is indicated for the founder cells, early passage iPSC and late passage iPSC (n=2).

Figure S3. Pearson correlation coefficients of the eigenvector with differential expression. Related to Figure 3.

A histogram shows the distribution of the correlation coefficient between the differential expression in the founders and the eigenvector score from the Hi-C in the genomic region that contained them. The analysis shows an intersection of regions that switch from A to B compartment or vice versa and genes that are significantly differentially expressed.

Figure S4. Higher domain scores are associated with increased expression and internal nuclear position. Related to Figure 4.

(A) Average expression of genes in pre-B plotted as the expression percentile of all TADs binned for the domain score in pre-B.
(B) Domain scores of MEF, NSC and iPS for LAD and iLAD identified in MEF, NSC and ESC (Peric-Hupkes et al., 2010).
(C) Domain scores of TADs containing tissue specific genes (having four-fold higher expression in both replicates of a given founder compared to the replicates of all other somatic cells) in pre-B, MΦ, NSC and MEF. From left to right tissue specific genes in pre-B, MΦ, NSC and MEF.

Figure S5. Characterization and validation of iPSC loops. Related to Figure 5.

(A) CTCF motif orientation of all iPSC loops identified between two ESC CTCF binding sites.
(B) Schematic representation of the meta-loop analysis.

(C) Absolute expression of genes (A) located in pre-B specific loops in pre-B, NSC, MEF, MΦ, iPSC p3 derived from pre-B, NSC, MEF, MΦ and iPSC p20.
(D) Absolute expression of genes located in iPS specific loops in pre-B, NSC, MEF, MΦ, iPSC p3 and iPSC p20.
(E) Hi-C interaction heatmap showing chromatin loops in pre-B cells (upper triangle) and iPSC derived from pre-B and NSC (lower triangle) at the *Frmd4a* (top) and *Jakmip1* (bottom) locus. pre-B specific founder loops conserved in iPSC are indicated with a red arrowhead. Scale bar indicates 100kb.
(F) Hi-C interaction heatmap showing chromatin loops in p3 iPSC (upper triangle) and ESC (lower triangle) at the *Gli2* (left) and *Cil18a1* (bottom) locus. Pluripotency loops are indicated with a red arrowhead. Super enhancers identified in ESC are indicated with a yellow arrowhead. Scale bar indicates 100kb.
(G) Hi-C interaction heatmap showing chromatin loops in p3 iPSC (upper triangle) and ESC (lower triangle) at the *Sall4* (left) and *Tsc22d1* (bottom) locus. Pluripotency loops are indicated with a red arrowhead. Super enhancers identified in ESC are indicated with a yellow arrowhead. Scale bar indicates 100kb.

Figure S6. 3D genome of iPSCs memorizes its cell of origin. Related to Figure 6.

(A) %cis reads for chromosome 11 (top) and chromosome 12 (bottom) for each Hi-C library
(B) Number of CTCF peaks shared between early passage iPSCs.
(C) CTCF ChIPseq scores for the cell-of-origin dependent differential and shared CTCF binding sites in early passage iPSCs.

Table S1. Number of Valid Hi-C Reads. Related to Figure 3.

Experiment	Valid Hi-C Reads	Valid Cis Reads	% Valid Cis Reads
MEF_1	14420095	10884256	75.5
MEF_2	13533085	9928328	73.4
MEF_3	11263329	8954602	79.5
Total MEF	39216509	29767186	75.9
MEF_iPS_p3_1	27702764	21645276	78.1
MEF_iPS_p3_2	16075757	11056408	68.8
MEF_iPS_p3_3	14370735	10619560	73.9
Total MEF iPS p3	58149256	43321244	74.5
MEF_iPS_p20_1	19167057	16217187	84.6
MEF_iPS_p20_2	9760872	8213675	84.1
MEF_iPS_p20_3	14731426	12561757	85.3
Total MEF iPS p20	43659355	36992619	84.7
MΦ_1	23243455	15315162	65.9
MΦ_2	32445564	21033790	64.8
Total MΦ	55689019	36348952	65.3
MΦ_iPS_p3_1	14524297	11065086	76.2
MΦ_iPS_p3_2	17716188	12314282	69.5
MΦ_iPS_p3_3	14942672	10996057	73.6
Total MΦ iPS p3	47183157	34375425	72.9
MΦ_iPS_p20_1	14942764	12207532	81.7
MΦ_iPS_p20_2	17926833	13980168	78.0
MΦ_iPS_p20_3	15112059	12472334	82.5
Total MΦ iPS p20	47981656	38660034	80.6
NSC_1	52575404	37706443	71.7
NSC_iPS_p20_1	15709417	13087382	83.3
NSC_iPS_p20_2	16485267	13955056	84.7
NSC_iPS_p20_3	18140176	15071380	83.1
Total NSC iPS p3	50334860	42113818	83.7
NSC_iPS_p3_1	18054935	15309925	84.8
NSC_iPS_p3_2	15899555	12904538	81.2
NSC_iPS_p3_3	17064723	12475987	73.1
Total NSC iPS p20	51019213	40690450	79.8
Pre-B_1	32448132	24918448	76.8
Pre-B_2	37414976	28567354	76.4
Total pre-B	69863108	53485802	76.6
Pre-B_iPS_p3_1	22309511	17535333	78.6
Pre-B_iPS_p3_2	33129155	25614782	77.3
Pre-B_iPS_p3_3	25033777	18835842	75.2
Total pre-B_iPS_p3	80472443	61985957	77.0
Pre-B_iPS_p20_1	37891651	32772065	86.5
Pre-B_iPS_p20_2	15601950	12973052	83.2
Pre-B_iPS_p20_3	18931024	14959635	79.0
Total pre-B_iPS_p20	72424625	60704752	83.8
E14_ESC_1	29065859	22671989	78.0

Table S2. Eigenvector values. Related to Figure 3.

Table S3. Domain segmentation in iPSC p3. Related to Figure 4.

Table S4. List with genomic coordinates of all the called chromatin loops. Related to Figure 5.

Table S5. List of primers used for RT-qPCR. Related to Figure 1.

Gene	Forward	Reverse
Mac1	TCGTATGTGAGGTCTAAGACA	CAGCAGTGATGAGAGCCAAG
Sox1	GCAGCGTTTCCGTGACTTT	GGCAGAACCACAGGAAAGA
CD45	TTGTCACAGGGCAAACACCT	TTGGGGGTGTGGATTCAGTG
CD41	AAGGACAAACATGGAAGCGTG	CTCTTGACTTGCGTTTAGGGC
Hoxb4	CCTGGATGCGCAAAGTTCA	CGTCAGGTAGCGATTGTAGTGA
Nestin	CTCTTGGCTTTCCTGACCCC	AGGCTGTCACAGGAGTCTCA
Pax6	TTCCCGAATTCTGCAGACCC	GTCGCCACTCTTGGCTTACT
Sox7	TCAGGGGACAAGAGTTCGGA	CCTTCCATGACTTTCCCAGCA
Pgk	ATGTCGCTTTCCAACAAGCTG	GCTCCATTGTCCAAGCAGAAT

Supplemental Experimental Procedures

Mice

We used a 'reprogrammable mouse' line containing a doxycycline-inducible OSKM cassette, the reverse tetracycline transactivator (rtTA), (Carey et al., 2010) and an Oct4-GFP reporter transgene (Boiani et al., 2002) as described (Di Stefano et al., 2014). Mice were housed in standard cages under 12 hour light-dark cycles and fed ad libitum with a standard chow diet. All experiments were approved by the Ethics Committee of the Barcelona Biomedical Research Park (PRBB) and performed according to Spanish and European legislation.

Cell cultures

ESCs and iPSCs were cultured on Mitomycin-C treated MEFs in KO-DMEM (Invitrogen) supplemented with 1% nonessential amino acids (Invitrogen), 0.1 mM β-mercaptoethanol (Invitrogen), 1,000 U/ml LIF (Millipore) and 15% fetal bovine serum (FBS, Invitrogen) (ESC medium). Naïve ESCs were cultured in serum-free N2B27 medium on gelatin-coated dishes. N2B27 medium (500ml) was generated by inclusion of the following: 240ml DMEM/F12 (Invitrogen), 240ml Neurobasal (Invitrogen), 5ml N2 supplement (Invitrogen), 10ml B27 supplement (Invitrogen), 1,000 U/ml LIF (Millipore), 1% nonessential amino acids (Invitrogen), 0.1mM β-mercaptoethanol (Invitrogen), and the small molecules PD0325901 (Stemgent, 1μM) and CHIR (Stemgent, 3μM). $CD19^+$ pre-B cells and $Mac1^+$ macrophages were isolated from bone marrow with monoclonal antibodies to CD19 and Mac-1 (BD Pharmingen) respectively, using MACS sorting (Miltenyi Biotech). Pre-B cells were grown in RPMI medium with 10% FBS and 10ng/ml IL-7 (Peprotech); macrophages in DMEM with 10% FBS and 10ng/ml each of CSF1 and IL-3 (Peprotech). MEFs were established from day 13.5 mouse embryos and cultured in DMEM containing 10% FBS. NSC were isolated and cultured as previously described (Di Stefano et al., 2009). All media were supplemented with L-glutamine and penicillin/streptomycin (GIBCO).

Reprogramming

Reprogramming experiments with pre-B cells were performed as previously described (Di Stefano et al., 2014); with MEFs, macrophages and NSCs were conducted by plating 100.000 cells/well on gelatinized plates seeded with irradiated MEFs, using ESC medium supplemented with 2 μg/ml of doxycycline. For the isolation of iPSC lines, doxycycline was washed out after 15 days of reprogramming and colonies with ESC-like morphology were picked at 20 days before further passaging. iPSC lines were expanded for an additional 9 days (3 passages) to obtain P3 iPS cell lines or for 20 passages to obtain P20 iPS cell lines.

Immunofluorescence assays

The cells were fixed with 4% paraformaldehyde, blocked with 5% of goat serum solution in PBS and incubated with primary antibodies overnight at 4°C. On the next day, the cells were exposed to secondary antibodies (all Alexa Fluor from Invitrogen) at RT for one hour. The primary antibodies used were Nanog (Calbiochem, SC1000) and SSEA-1 (Santa Cruz, SC-21702). Nuclear staining was performed with DAPI (Invitrogen).

Flow cytometry

For FACS analysis of the pluripotency markers Oct4, Sox2, and SSEA-1, P3 iPSCs and naïve ESCs were fixed and stained using the Multicolor Flow Cytometry kit (R&D Systems, Minneapolis, MN), according to the manufacturer instructions. Live staining of the P3 iPSCs and naïve ESCs with the Stem Cell CDy1 Dye (Active Motif) was performed by incubating the cells with the CDy1 Dye for 1 hour at 37C, followed by washout of the dye and FACS analysis. For Oct4GFP expression, iPSCs were resuspended in PBS and analyzed by FACS. Cells were analyzed with an LSR Fortessa FACS machine (BD Biosciences) using Diva v6.1.2 (BD Biosciences) and FlowJo software v10 (TreeStar).

Chimeric mice

For the chimera formation assay, 10 to 15 iPS cells (Agouti color coat) were injected into 3.5dpc blastocysts of C57BL/6J mice (black coat color) and transferred into pseudo-pregnant CD1 females. Chimerism of the transplanted offspring was assessed by the presence of agouti coat color derived from the iPS cells.

Differentiation of iPSCs

Embryoid bodies were derived by plating iPSCs at a concentration of $1.3x10^6$ cells/ml in bacterial grade dishes in ES medium without LIF. After 6 days in culture they were harvested for total RNA extraction.

Expression analysis

To remove the feeder cells, iPS cells were cultured on gelatinized plates for 2 passages before RNA extraction. RNA was isolated with the miRNeasy mini kit (Qiagen), eluted with RNase-free water and quantified by Nanodrop. cDNA was produced with the High Capacity RNA-to-cDNA kit (Applied Biosystem). RNA samples with a RIN greater than 9 were analyzed by expression arrays. 500ng of total RNA per sample were labeled using Agilent's QuickAmp labeling kit and hybridized to Agilent 8X60K expression arrays. RT-qPCR reactions were set up in triplicate with the SYBR Green QPCR Master Mix (Applied Biosystem), using primers listed in Supplementary Table 2. Reactions were run on an AB7900HT PCR machine with 40 cycles of 30s at 95 °C, 30s at 58 °C and 30s at 72 °C. Raw array data was processed using limma (Ritchie et al., 2015). We perform "normexp" background correction with an offset of 16. We normalize between arrays using quantile normalization. To identify differently expressed genes we create a contrast matrix in which all pairwise comparisons are made. For each probe an empirical Bayes moderated t-statistic test is performed. An FDR correction is applied to the nominal p-values. All statistical analyses were performed in R. (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.)

ChIPseq analysis

ChIP experiments were performed as described previously (van Oevelen et al., 2008) using an antibody against H3K27ac (ab4729, Abcam) and CTCF (Millipore, 07-729). DNA libraries were prepared using Illumina's reagents and instructions. All libraries were sequenced on the Hiseq2000 sequencer. Mapping of the raw fastq files was performed with bwa aln to the mm9 genome. Peak calling was performed with macs v. 1.4.2. (Zhang et al., 2008). Coverage scores for the creation of heatmaps were calculated with the compEpiTools package in R. To determine tissue-specific peaks in the founder tissues we calculated the coverage 600bp surrounding (300bp +/-) the summit of the peaks for a given tissue. We required that the coverage score was above 1 for the tissue of interest and below 1 for the three other tissues.

Differential peak calling

For the differential CTCF peaks between iPS lines we made use of THOR (Allhoff et al., 2014), which allows differential peak calling using replicate experiments using standard settings. After differential peak calling we performed a stringent selection of regions that were differentially bound. We selected regions that had a p-value below 10e-30 and had a fold change of at least 3.

Hi-C template generation & mapping

pre-B cells, MEF, NSC, iPSC and E14 ESC were cross-linked and further processed as DpnII 3C template as previously described (Splinter et al., 2012). Hi-C libraries from macrophages were incubated at 65°C during the lysis and subsequent SDS step to inactivate nucleases present in lysates from these cells. Ligation of crosslinked DNA fragments was performed without incorporating biotin, as described previously (Sexton et al., 2012). Libraries for paired-end sequencing were generated from sonicated, ~500-800bp size selected, 3C templates using the TruSeq DNA LT Sample Prep Kit (Illumina). Sequencing of the Hi-C libraries was performed on an Illumina HiSeq 2000. ESC Hi-C data used for the visual inspection of super enhancer loci were obtained from (Geeven et al., 2015).

Paired-end FASTQ files are mapped independently to the mouse genome (mm9) using bwa mem. Reads mapping to repetitive regions in the genome are filtered from the dataset. The resulting files are further filtered and deduplicated using HiCUP (http://www.bioinformatics.babraham.ac.uk/projects/hicup/). Reads that are within 1kb of each other are filtered from the dataset as they likely represent uncut genomic fragments. Reads mapped to the X chromosome were excluded in all analyses.

Eigenvector analysis

To segment the genome into A and B compartments we use the method described by Lieberman-Aiden et al. (Lieberman-Aiden et al., 2009). We divided the chromosomes into bins of 300kb and created contact frequency matrices for every chromosome. Next we performed vanilla coverage normalization (Rao et al., 2014) on the contact matrix. To this end we calculate a vector V_c representing the total intrachromosomal coverage for every 300kb window. A normalization matrix M_{norm} is calculated by taking the outer product of V_c . A normalized contact matrix is calculated by dividing the raw contact matrix by the square root of M_{norm}. The normalized contact matrix is transformed to an observed/expected matrix, which is used to calculate a spearman rank correlation matrix. The first eigenvector of the correlation matrix is used as the A/B segmentation.

Domain segmentation

To segment the genome into topologically associating domains (TADs) we used the R package HiCseg (Levy-Leduc et al., 2014). The software package requires a contact matrix as input. To this end we the combined the data for all the early iPS cell lines into an early dataset. The tag positions were transferred into fragment space. We binned the fragments into windows of 50 fragments, corresponding to a median genomic size of 20kb and calculated a fragment window matrix and performed vanilla coverage normalization (see above). The resulting matrix was used as input for the HiCseg algorithm.

Domain score

Given the TADs identified in the domain segmentation, we calculate an intradomain contact score (simplified to domain score). The domains score is the ratio of the number of contacts that occur between regions within the same TAD (intraTAD contacts) over the total number of intrachromosomal contacts for a TAD (intraTAD + interTAD). To identify TADs that have a differential domain score we used the bioconductor package limma (Ritchie et al., 2015). We perform quantile normalization (using the normalizeQuantiles function in the limma package) on the domain scores. To identify TADs with a differential domain score between the iPS lines we create a contrast matrix in which all pairwise comparisons are made. For each TAD an empirical Bayes moderated t-statistic test is performed. Nominal p-values are corrected using the FDR method. TADs with an FDR value < 0.05 were selected for subsequent analysis.

Loop calling

For the loop calling we initially follow a similar procedure as for the domain segmentation, i.e. mapping Hi-C tags to fragments and creating a coverage normalized contact frequency matrix in fragment space. For the contact matrix we use overlapping windows with a size of 50 fragments and a step size of 5 fragments. The resulting matrix is used as the input for the looping calling procedure.

A chromatin loop in a 4C or Hi-C experiment is appreciated as an increase in signal over the background. For these data the background is not distributed randomly, rather it follows a monotonically decreasing pattern, which is a function of the distance to the site of interest. To identify chromatin loops we explicitly model this background distribution by performing monotonic regression on the columns of the contact matrix. For this we use the R package isotone (Mair, 2009). We use the gpava function with the solver weighted.fractile to calculate the background distribution. Values that reach above a certain threshold value over the background and with a minimal amount of reads are selected as loops. Loops represent a combination of two windows in the genome, effectively forming a rectangle. Due to the nature of the input matrix and the fact that loops can span multiple windows, we often find overlapping rectangles for a single loop. We collapse overlapping rectangles and the select the rectangle with the highest coverage as the chromatin loop.

Loops were called in every founder tissue separately and in a combined dataset containing all early iPS datasets. To call tissue specific loops in a given founder tissue we first identified the loops in the selected founder dataset. Next we calculated the coverage for all the founders over the loops called in the selected founder. Coverage scores were normalized to total amount of valid ditags per experiment. Tissue specific loops were loops that have at least a 2-fold higher coverage in the selected tissue compared to the other tissue. The procedure for the iPS specific loops is similar, except that the coverage of the iPS loops has to be 2-fold higher compared to all founder tissues.

Contact frequency plots

To visualize the interaction between loci we create 2D heatmaps that show the contact frequency along a color scale. For the heatmap we generate a windowed interaction matrix with the pairwise coverage between genomic windows of 20kb within a given region. We use a sliding window approach with a step size of 2kb. The contact matrix is normalized using vanilla coverage (VC) normalization (see above). The VC normalization matrix is divided by the median, so that the scale of the resulting normalized contact matrix is preserved. Because different Hi-C experiments had different numbers of reads we needed to scale them in order to directly compare them in our visualizations of chromatin loops. Our scaling method is similar to the calculation of the RPKM score in RNAseq experiments (Mortazavi et al., 2008). We scale our dataset to 1 million intrachromosomal contacts per 100 Mb of sequence.

Meta-loop analysis

In order to study the distribution of Hi-C tags around a set of loops we performed created a meta-loop. In essence, this is a 2D variation of a meta-gene analysis, commonly performed for ChIPseq data. In the meta-loop analysis we align Hi-C tags on a set of loops. Because these are of varying size we scale the loops to have the same size. In addition to the loop itself we also analyse 50% of the genome upstream of the loop and 50% of the genome downstream of the loop. The genomic regions that interact in the loop are used as anchor points for the alignment. A frequency matrix is subsequently created in which the tag frequency in the loop regions and the flanking regions is stored. The frequency matrix is represented as a heatmap.

Alignment of Hi-C data to ChIP peaks (PE-SCAn)

Intrachromosomal Hi-C captures located more than 5MB of each other were aligned to ChIP data as previously described (de Wit et al., 2013). Briefly, one of the paired Hi-C reads was aligned to the ChIP data. Only reads that mapped within 500 kb up- or downstream of the ChIP peaks were selected for further analysis. Of this reduced set the corresponding read was also aligned to the ChIP peaks within 500 kb, resulting in a set of two distances (dx, dy) to all the Hi-C di-tags that are found within 500 kb of these peaks, for every intrachromosomal pair of ChIP peak. From the distribution of dx and dy a frequency matrix was calculated with a bin size of 50 kb. To calculate whether the binding sites of a given factor show preferential spatial contacts anenrichment score was calculated over a randomized data set. The randomized data set is calculated by aligning the Hi-C data to a circularly permuted ChIPseq data set, that is, the ChIP peaks are linearly shifted 10 Mb along the chromosome.

Data sources

H3K27ac data BMDM, H3K27ac data MEF from ENCODE (Yue et al., 2014), gene annotation from GENCODE (Harrow et al., 2012), H3K27ac data pre-B cells from (Lane et al., 2014), H3K27ac data NSC from (Creyghton et al., 2010), Sox2 data NPC and ESC from (Lodato et al., 2013), Pu.1 B cells from (Heinz et al., 2010), Oct4 data ESC from (Marson et al., 2008). LAD data from (Peric-Hupkes et al., 2010). CTCF data NPC (Phillips-Cremins et al., 2013), CTCF data MEF (Tedeschi et al., 2013), CTCF data macrophages (Daniel et al., 2014), CTCF data pre-B cells (Ribeiro de Almeida et al., 2011).

Supplemental References

Allhoff, M., Sere, K., Chauvistre, H., Lin, Q., Zenke, M., and Costa, I.G. (2014). Detecting differential peaks in ChIP-seq signals with ODIN. Bioinformatics *30*, 3467-3475.

Boiani, M., Eckardt, S., Scholer, H.R., and McLaughlin, K.J. (2002). Oct4 distribution and level in mouse clones: consequences for pluripotency. Genes & development *16*, 1209-1219.

Carey, B.W., Markoulaki, S., Beard, C., Hanna, J., and Jaenisch, R. (2010). Single-gene transgenic mouse strains for reprogramming adult somatic cells. Nature methods *7*, 56-59.

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. Proceedings of the National Academy of Sciences of the United States of America *107*, 21931-21936.

Daniel, B., Nagy, G., Hah, N., Horvath, A., Czimmerer, Z., Poliska, S., Gyuris, T., Keirsse, J., Gysemans, C., Van Ginderachter, J.A., *et al.* (2014). The active enhancer network operated by liganded RXR supports angiogenic activity in macrophages. Genes & development *28*, 1562-1577.
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.
Di Stefano, B., Prigione, A., and Broccoli, V. (2009). Efficient genetic reprogramming of unmodified somatic neural progenitors uncovers the essential requirement of Oct4 and Klf4. Stem cells and development *18*, 707-716.
Di Stefano, B., Sardina, J.L., van Oevelen, C., Collombet, S., Kallin, E.M., Vicent, G.P., Lu, J., Thieffry, D., Beato, M., and Graf, T. (2014). C/EBPalpha poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature *506*, 235-239.
Geeven, G., Zhu, Y., Kim, B.J., Bartholdy, B.A., Yang, S.M., Macfarlan, T.S., Gifford, W.D., Pfaff, S.L., Verstegen, M.J., Pinto, H., *et al.* (2015). Local compartment changes and regulatory landscape alterations in histone H1-depleted cells. Genome Biol *16*, 289.
Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., *et al.* (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome research *22*, 1760-1774.
Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell *38*, 576-589.
Lane, A.A., Chapuy, B., Lin, C.Y., Tivey, T., Li, H., Townsend, E.C., van Bodegom, D., Day, T.A., Wu, S.C., Liu, H., *et al.* (2014). Triplication of a 21q22 region contributes to B cell transformation through HMGN1 overexpression and loss of histone H3 Lys27 trimethylation. Nature genetics *46*, 618-623.
Levy-Leduc, C., Delattre, M., Mary-Huard, T., and Robin, S. (2014). Two-dimensional segmentation for analyzing Hi-C data. Bioinformatics *30*, i386-392.
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.
Lodato, M.A., Ng, C.W., Wamstad, J.A., Cheng, A.W., Thai, K.K., Fraenkel, E., Jaenisch, R., and Boyer, L.A. (2013). SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. PLoS genetics *9*, e1003288.
Mair, P., Hornik, K. & de Leeuw, J. (2009). Isotone optimization in R: pool-adjacent-violators algorithm (PAVA) and active set methods. Journal of statistical software *32*, 1-24.
Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., *et al.* (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521-533.
Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature methods *5*, 621-628.
Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Graf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., *et al.* (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Molecular 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.
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., *et al.* (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-1680.
Ribeiro de Almeida, C., Stadhouders, R., de Bruijn, M.J., Bergen, I.M., Thongjuea, S., Lenhard, B., van Ijcken, W., Grosveld, F., Galjart, N., Soler, E., *et al.* (2011). The DNA-binding protein CTCF limits proximal Vkappa recombination and restricts kappa enhancer interactions to the immunoglobulin kappa light chain locus. Immunity *35*, 501-513.
Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research *43*, e47.
Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458-472.

Splinter, E., de Wit, E., van de Werken, H.J., Klous, P., and de Laat, W. (2012). Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods *58*, 221-230.
Tedeschi, A., Wutz, G., Huet, S., Jaritz, M., Wuensche, A., Schirghuber, E., Davidson, I.F., Tang, W., Cisneros, D.A., Bhaskara, V., *et al.* (2013). Wapl is an essential regulator of chromatin structure and chromosome segregation. Nature *501*, 564-568.
van Oevelen, C., Wang, J., Asp, P., Yan, Q., Kaelin, W.G., Jr., Kluger, Y., and Dynlacht, B.D. (2008). A role for mammalian Sin3 in permanent gene silencing. Molecular cell *32*, 359-370.
Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., *et al.* (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355-364.
Zhang, B., Kirov, S., and Snoddy, J. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic acids research *33*, W741-748.
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology *9*, R137.