nature genetics

ARTICLES
https://doi.org/10.1038/s41588-017-0030-7

Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming

Ralph Stadhouders[1,2,8]*, Enrique Vidal[1,2], François Serra[1,2,3], Bruno Di Stefano[1,2,4], François Le Dily[1,2,3], Javier Quilez[1,2], Antonio Gomez[1,2], Samuel Collombet[5], Clara Berenguer[1,2], Yasmina Cuartero[1,2,3], Jochen Hecht[2,6], Guillaume J. Filion[1,2], Miguel Beato[1,2], Marc A. Marti-Renom[1,2,3,7]* and Thomas Graf[1,2]*

Chromosomal architecture is known to influence gene expression, yet its role in controlling cell fate remains poorly understood. Reprogramming of somatic cells into pluripotent stem cells (PSCs) by the transcription factors (TFs) OCT4, SOX2, KLF4 and MYC offers an opportunity to address this question but is severely limited by the low proportion of responding cells. We have recently developed a highly efficient reprogramming protocol that synchronously converts somatic into pluripotent stem cells. Here, we used this system to integrate time-resolved changes in genome topology with gene expression, TF binding and chromatin-state dynamics. The results showed that TFs drive topological genome reorganization at multiple architectural levels, often before changes in gene expression. Removal of locus-specific topological barriers can explain why pluripotency genes are activated sequentially, instead of simultaneously, during reprogramming. Together, our results implicate genome topology as an instructive force for implementing transcriptional programs and cell fate in mammals.

Somatic cell reprogramming into PSCs is a widely studied model for dissecting how TFs regulate gene expression programs and shape cell identity[1,2]. Chromosomal architecture has recently been shown to be cell-type specific and to be critical for transcriptional regulation[3–5], but its importance in cell-fate decisions remains poorly understood.

Two major levels of topological organization have been identified in the genome[6–8]. The first level segregates the genome, at the megabase scale, into two subnuclear compartments: the A compartment, which corresponds to active chromatin typically associated with a more central nuclear position, and the B compartment, which represents inactive chromatin enriched at the nuclear periphery/lamina[9–14]. Compartmentalization is consistent among individual cells and is a potential driver of genome folding[15]. A second submegabase level consists of topologically associated domains (TADs)[16–18] and chromatin loops[11], which restrict or facilitate interactions between gene-regulatory elements[19,20]. Importantly, modifying chromatin architecture can lead to gene expression changes[19,21–24]. Moreover, de novo establishment of TAD structure during zygotic genome activation has been shown to be independent of ongoing transcription, thus demonstrating that chromatin architecture is not simply a consequence of transcription[25–27]. Genome topology could therefore be instructive for gene regulation[28,29], but whether this relationship reflects a general principle that occurs on a genome-wide scale in space and time is unknown.

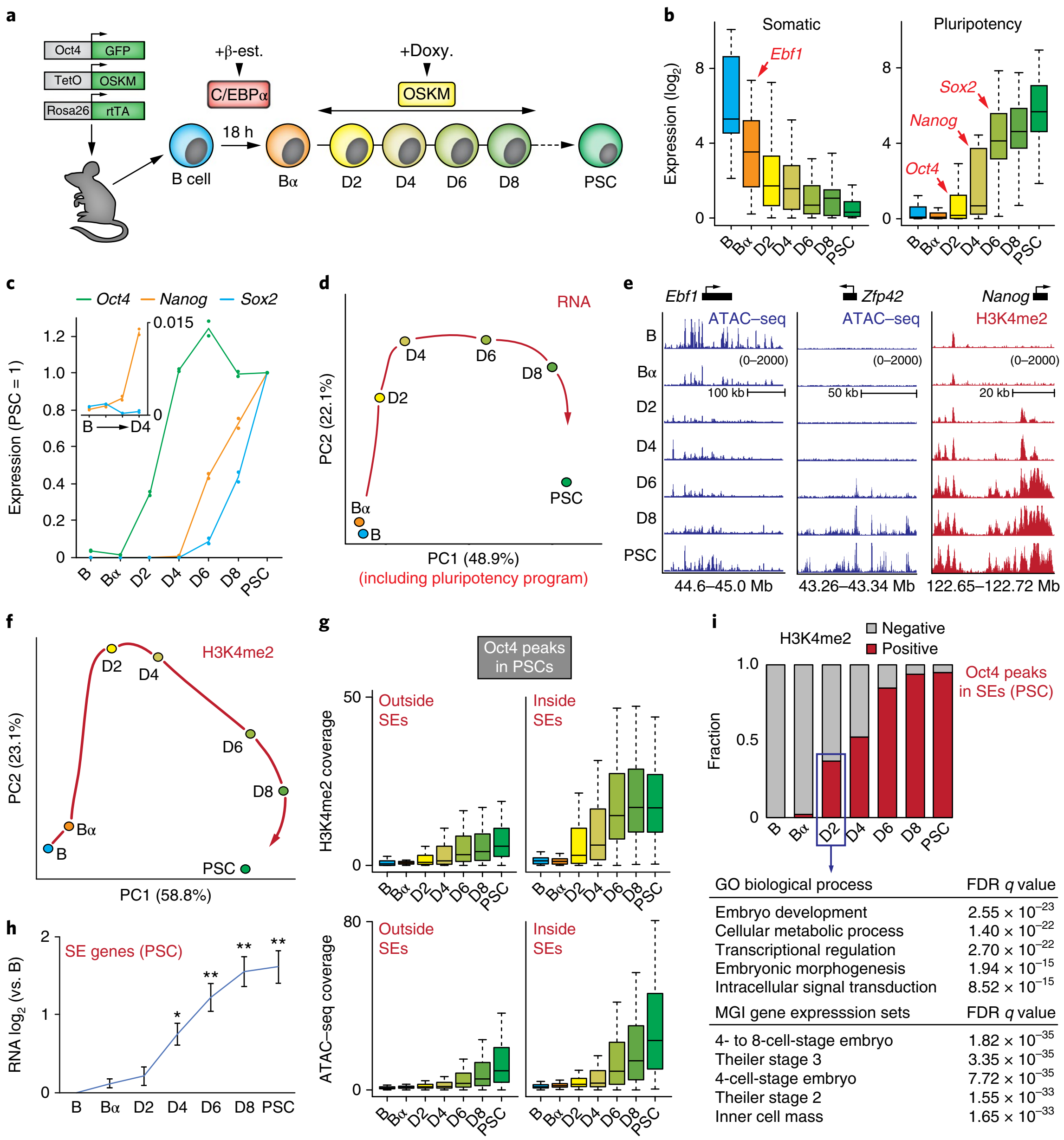
Mechanistic studies with mammalian cell reprogramming systems have been hampered by the typically small percentage of responding cells[1,30]. To overcome this shortcoming, we have recently developed a highly efficient and synchronous reprogramming system based on the transient expression of the TF C/EBPα before induction of the Yamanaka TFs OCT4, SOX2, KLF4 and MYC (OSKM)[31,32]. The OSKM TFs activate the endogenous core pluripotency TFs sequentially in the order of *Pou5f1* (also known as *Oct4*), *Nanog* and *Sox2*, thus suggesting that locus-specific barriers dictate gene activation kinetics[33–35]. Here, we studied how C/EBPα and OSKM affect genome topology, the epigenome and gene expression during reprogramming. We found that the TFs bind hotspots of topological reorganization at both the compartment and TAD levels. Dynamic reorganization of genome topology frequently preceded gene expression changes at all levels and provided an explanation for the sequential activation of core pluripotency genes during reprogramming. Together, our observations indicate that genome topology has an instructive role in implementing transcriptional programs relevant to cell-fate decisions in mammals.

Results

Transcription factors prime the epigenome for reprogramming. We exposed mouse bone marrow–derived pre-B cells to the myeloid TF C/EBPα to generate cells denoted Bα cells. Subsequent activation of OSKM has been found to induce reprogramming of nearly 100% of Bα cells into PSC-like cells within 4–8 d (refs [31,32]). To obtain a high-resolution map of changes in gene expression and chromatin structure, we examined six different reprogramming stages (B, Bα, D2,

---

[1] Gene Regulation, Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. [2] Universitat Pompeu Fabra (UPF), Barcelona, Spain. [3] Structural Genomics Group, CNAG-CRG, BIST, Barcelona, Spain. [4] Department of Stem Cell and Regenerative Biology, Harvard Stem Cell Institute, Harvard University and Harvard Medical School, Cambridge, MA, USA. [5] Institut de Biologie de l'Ecole Normale Supérieure (IBENS), CNRS UMR8197, INSERM U1024, Paris, France. [6] Genomics Unit, CRG, BIST, Barcelona, Spain. [7] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. Present address: [8] Department of Pulmonary Medicine, Erasmus MC, Rotterdam, the Netherlands. Ralph Stadhouders and Enrique Vidal contributed equally to this work. *e-mail: r.stadhouders@erasmusmc.nl; martirenom@cnag.crg.eu; thomas.graf@crg.eu



GO biological process	FDR q value
Embryo development	2.55×10^{-23}
Cellular metabolic process	1.40×10^{-22}
Transcriptional regulation	2.70×10^{-22}
Embryonic morphogenesis	1.94×10^{-15}
Intracellular signal transduction	8.52×10^{-15}
MGI gene expresssion sets	**FDR q value**
4- to 8-cell-stage embryo	1.82×10^{-35}
Theiler stage 3	3.35×10^{-35}
4-cell-stage embryo	7.72×10^{-35}
Theiler stage 2	1.55×10^{-33}
Inner cell mass	1.65×10^{-33}

Fig. 1 | Dynamics of the transcriptome and epigenome during reprogramming. a, Schematic overview of the reprogramming system. C/EBPα-ER in B cells translocates into the nucleus after β-estradiol (β-est.) treatment. After β-estradiol washout, *Oct4*, *Sox2*, *Klf4* and *Myc* (OSKM) were induced by doxycycline (Doxy.). **b**, Box plots of gene expression dynamics (normalized counts) of a set of core B cell ('somatic', $n=25$) and PSC ('pluripotency', $n=25$) identity genes. All box plots depict the first and third quartiles as the lower and upper bounds of the box, with a thicker band inside the box showing the median value and whiskers representing 1.5× the interquartile range. **c**, Average gene expression kinetics of *Oct4*, *Nanog* and *Sox2* during reprogramming ($n=2$, relative to the levels in PSCs). Inset showing that *Nanog* expression first appears at D4. **d**, PCA of gene expression dynamics ($n=16,332$ genes) during reprogramming. Red arrow, hypothetical trajectory. **e**, Representative examples of chromatin opening (measured by ATAC–seq) and H3K4me2 deposition (measured by ChIPmentation) at gene-regulatory elements controlling B cell (*Ebf1*) or pluripotency (*Zfp42* and *Nanog*) genes. **f**, PCA of H3K4me2 dynamics during reprogramming ($n=26,351$ 100-kb genomic bins). Red arrow, hypothetical trajectory. **g**, Box plots of dynamics of H3K4me2 deposition (top) and chromatin accessibility (bottom) at OCT4-binding sites outside ($n=31,869$) and inside ($n=821$) PSC SEs. **h**, Expression dynamics of genes associated with an SE in PSCs (mean values shown; $n=328$ genes). Error bars, 95% confidence intervals (*$P < 0.01$; **$P < 0.001$ versus B cells, unpaired two-tailed t test). **i**, Fraction of H3K4me2-positive OCT4-binding sites in PSC SEs ($n=821$) during reprogramming; table shows a Gene Ontology (GO) analysis for SE genes ($n=212$) associated with early OCT4 recruitment. Results shown are based on two independent reprogramming experiments; additional information in Methods.

D4, D6 and D8) and PSCs (Fig. 1a). We profiled the transcriptome through RNA-seq, profiled active chromatin deposition though dimethyl histone H3 K4 (H3K4me2) ChIP–seq using tagmentation (ChIPmentation)[36] and profiled chromatin accessibility through assay for transposase-accessible chromatin with high-throughput sequencing (ATAC–seq)[37] (Supplementary Fig. 1). Expression of half of all genes was significantly affected (false discovery rate (FDR) <0.01) between any two time points, starting with the rapid silencing of the core B cell program initiated by C/EBPα. Pluripotency genes were then activated sequentially, with the core pluripotency factors

Oct4, *Nanog* and *Sox2* being activated at D2, D4 and D6, respectively (Fig. 1b,c). RT–PCR measurements of primary *Nanog* and *Sox2* transcription confirmed their activation timing (Supplementary Fig. 1e).

Principal component analysis (PCA) revealed a trajectory along which B cells acquired a PSC gene expression program (Fig. 1d). Epigenome remodeling showed similar dynamics, with an early loss of chromatin accessibility at gene-regulatory elements controlling the B cell program induced by C/EBPα and the subsequent establishment of active and open chromatin at pluripotency genes by OSKM (Fig. 1e and Supplementary Fig. 1). OSKM induction led to a genome-wide expansion of active chromatin marked by H3K4me2, a mark known to be deposited at both primed and active gene-regulatory elements[38] (Supplementary Fig. 1f). The H3K4me2 landscape, compared with that of gene expression, more rapidly converged on a pluripotent state, thus suggesting that OSKM TFs prime regulatory elements for subsequent gene activation (Fig. 1f). Many regions bound by OCT4 in PSCs[39] had already acquired H3K4me2 at D2, and chromatin opening occurred progressively at OCT4-binding sites (Fig. 1g and Supplementary Fig. 1g–i). 37% of OCT4-binding sites in predicted PSC superenhancer (SE) elements[39] were already H3K4me2 positive by D2, whereas activation of most associated genes (assigned on the basis of in situ Hi-C data; additional information in Supplementary Methods) occurred 2 d later (Fig. 1i–h). These early targeted SEs were linked to genes involved in embryo development (for example, *Oct4*, *Nanog* and *Klf9*) and metabolism (for example, *Upp1* and *Uck2*), a gene signature strongly associated with four- to eight-cell-stage embryos (Fig. 1i).

Chromatin state, genome topology and transcription are dynamically linked. We used in situ Hi-C[11] to map 3D genome organization during cell reprogramming at high resolution and determined genome segmentation into A and B compartments (Supplementary Table 1). Quantitative changes in A–B-compartment association (on the basis of the PC1 values of a PCA on the Hi-C correlation matrix) during reprogramming were cumulative, widespread and highly reproducible (Pearson $R > 0.97$) (Fig. 2a,b and Supplementary Fig. 2a,b). Although the overall proportions assigned to A and B compartments (40% A to 60% B) remained unchanged throughout reprogramming, compartmentalization strength (as measured by the average contact enrichment within and between compartments) was dynamically altered (Supplementary Fig. 2c,d). OSKM induction initially (D2–D4) strengthened A–B compartment segregation, which was followed by substantial loss of compartmentalization due to reduced contact frequencies within the B compartment and increased intercompartment contacts.

Switching of loci between the A and B compartments was frequent, with 20% of the genome changing compartment at any time point during reprogramming. B-to-A and A-to-B switching each occurred in 10% of the genome, and 35% of these regions were involved in multiple switching events (Supplementary Fig. 2e). PCA highlighted a reprogramming trajectory of genome compartmentalization highly similar to that seen for the transcriptome (Fig. 2c and Supplementary Fig. 2f). Genes that stably switched compartment after reprogramming tended to change expression accordingly and were enriched in lineage-specific functions: A-to-B-switching genes were associated with immune-system processes, whereas B-to-A-switching genes were enriched in early developmental functions (Supplementary Fig. 2g,h). Compartment switching typically occurred in regions with low PC1 values at the edges of A- or B-compartment domains. At any time point, regions that switched also displayed the most substantial PC1 changes, thus suggesting that loci with a less pronounced compartment association are more likely to change their compartment status (Fig. 2d and Supplementary Fig. 2i–k).

Our dataset allowed us to monitor genome architecture and to study its interplay with chromatin state and gene expression changes over time. The core transcriptional network that defines B cell identity[40] resided primarily (88%) in the A compartment (for example, *Ebf1*, *Pax5* and *Foxo1*), of which 32% switched to B (Supplementary Table 2 and Supplementary Fig. 3a). Both switching and nonswitching genes were rapidly silenced, but switching genes were repressed to a larger extent. In contrast, 40% of core pluripotency genes[41] initially resided in the B compartment, of which 90% switched to A (Supplementary Table 2 and Supplementary Fig. 3b). Pluripotency genes already in the A compartment were activated early (D2–D4, for example, *Oct4*), whereas genes that underwent B-to-A switching were activated late (D6, for example, *Sox2*) (Fig. 2e). We next divided all genes that changed expression between endpoints ($\log_2$ fold change >0.5) into stable (nonswitching) and compartment-switching groups. Again, downregulated genes that changed compartment from A to B (21%) were silenced to a greater extent than were nonswitching genes in A (Supplementary Fig. 3c). Likewise, upregulated genes that switched from B to A (16%) were upregulated more substantially than genes already residing in A, albeit with slower kinetics. Interestingly, quantitative changes in compartment association occurred before transcriptional upregulation (Supplementary Fig. 3d). To further explore whether compartment switching can precede transcriptional changes, we examined four clusters of genes ($n = 5,467$ genes) stably upregulated at early, intermediate or late time points (Supplementary Fig. 3e). Nearly one-third of the genes ($n = 175/548$) that switched from B to A in these clusters did so before being upregulated (Fig. 2f and Supplementary Fig. 3f). Moreover, genes associated with predicted PSC SEs showed a substantial increase in A-compartment association at D2 before transcriptional upregulation at D4 (Supplementary Fig. 3g and Fig. 1h).

We performed *k*-means clustering on the PC1 values of the 20% of the genome ($n = 8,218$ genes) that switched compartment during reprogramming, identifying 20 clusters with a wide range of switching dynamics that included nonlinear and abortive trajectories (Fig. 2g). Eight of the 20 clusters displayed concomitant changes in compartmentalization and gene expression ($R > 0.9$; Fig. 2h). The remainder, although generally also showing strong correlations between gene expression and PC1 (average $R = 0.86$; range, 0.56–0.97), consisted of clusters with at least one time point at which this correlation was lost (Fig. 2h). Genes in these clusters were enriched in metabolic and secretory functions as well as developmental processes (Supplementary Fig. 4a,b). Strikingly, 9 of the 20 clusters showed changes in subnuclear-compartment status preceding changes in transcriptional output, involving over half of the genes that switched compartment (for example, cluster 2.I; Fig. 2h,j). Compartment modification lagged behind changes in gene expression in only a single cluster, and 2 of the 20 clusters displayed both preceding and lagging relationships. We furthermore observed a very strong overall correlation between chromatin state dynamics (gain or loss of H3K4me2) and genome compartmentalization (average $R = 0.95$; range, 0.93–0.98), with concomitant changes in H3K4me2 levels and gene expression occurring in 13 of the 20 clusters. However, in 7 of the 20 clusters, H3K4me2 dynamics preceded PC1 changes (Fig. 2i), thus implicating chromatin state as a driver of subnuclear compartmentalization. The extended *Nanog* locus provides a prime example of modifications to compartmentalization and chromatin state preceding transcriptional changes. It includes a region encompassing *Gdf3*, *Dppa3* and the −45 kb *Nanog* SE[39,42], which had already switched from B to A in Bα cells. OSKM induction strengthened A-compartment association of the entire locus, activated *Gdf3* expression and primed the *Nanog* and *Dppa3* regulatory elements (H3K4me2$^+$ATAC$^+$) at D2 for subsequent gene activation at D4–D6 (Fig. 2k).

These data showed that genome compartmentalization and chromatin state are dynamically reorganized during cell-fate conversion and are tightly coupled to global changes in gene expression. In addition, a substantial number of genes are subject to changes in compartmentalization before expression alterations.

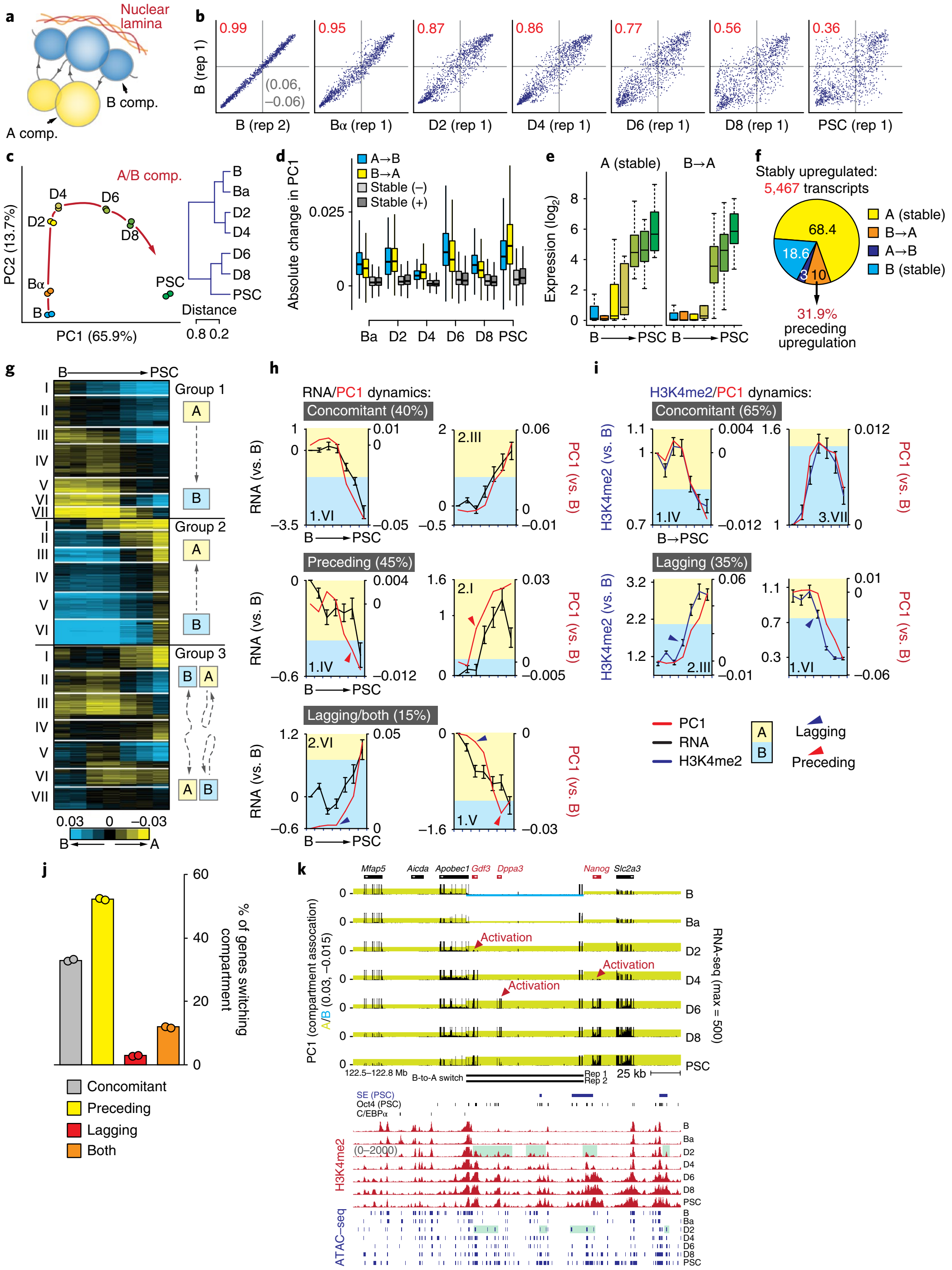


Fig. 2 | Kinetics of subnuclear compartmentalization, the transcriptome and epigenome. **a**, Schematic representation of chromosome compartments (comp.). **b**, Scatter plots of PC1 values ($n=26,370$ 100-kb bins) showing changes to initial B cell–genome compartmentalization for chromosome 13. Pearson correlation coefficient (R^2) is indicated in red. Rep, replicate. (**c**) PCA (left; red arrow, hypothetical trajectory) and unsupervised hierarchical clustering (right) of PC1 values ($n=26,370$ bins). **d**, Absolute PC1 changes per time point for regions ($n=35,348$) that switched compartment or did not switch ('stable') but increased (−) or decreased (+) in PC1 value. All box plots depict the first and third quartiles as the lower and upper bounds of the box, with a thicker band inside the box showing the median value and whiskers representing 1.5× the interquartile range. **e**, Box plots of normalized transcript counts for key pluripotency genes ($n=25$; as in Fig. 1b) that are stably associated with the A compartment or switch from B to A. **f**, Compartment switching at stably upregulated genes ($n=5,467$, across both independent biological replicates). **g**, k-means clustering ($k=20$) of PC1 values for 100-kb genomic bins that switch compartment at any time point. **h**, Examples of individual switching clusters with concomitant mean gene expression and PC1 changes (8/20), as well as clusters with PC1 changes preceding expression changes (9/20), clusters with expression changes preceding PC1 changes (1/20) or clusters with both phenomena (2/20). **i**, Examples of individual switching clusters that showed concomitant mean PC1/H3K4me2 changes (13/20) or H3K4me2 kinetics preceding PC1 modulation (7/20). **j**, Proportion of genes ($n=8,218$) located in the different categories of switching clusters. **k**, Genome-browser view of the *Gdf3*–*Dppa3*–*Nanog* locus. Top, integrated PC1 (shading denotes A/B compartment status) and RNA-seq values, with B-to-A-switch regions per replicate indicated below. Bottom part features SE location, Oct4 binding, C/EBPα binding, H3K4me2 dynamics and ATAC–seq peaks. Green shading indicates priming of *Dppa3*/*Nanog* enhancers at D2. Error bars throughout figure, s.e.m.

Genome partitioning into TADs is largely stable. We next used chromosome-wide insulation potential to identify TAD borders and define TADs[43], detecting ~2,800–3,400 borders per time point. Border calls were highly reproducible between biological replicates (Jaccard index >0.8) and enriched in binding sites for the transcriptional repressor CTCF and transcription start sites[17,44] (Supplementary Fig. 5). Borders not called in both biological replicates were excluded from all subsequent analyses. Partitioning of the genome into TADs was largely stable during reprogramming, because most TAD borders (>75%) were detected at all stages. Nevertheless, 18% of the 3,100 TAD borders were stably acquired ($n=431$) or lost ($n=124$) during reprogramming, thus resulting in a net increase in the number of borders and a decrease in the average TAD size from 891 kb to 741 kb (Supplementary Fig. 5). Unexpectedly, no correlation was found between the stable gain or loss of TAD borders (referred to hereafter as qualitative TAD-border changes) and CTCF binding. In fact, newly acquired TAD borders were relatively depleted in CTCF binding, and CTCF enrichment levels did not significantly change during reprogramming at borders gained or lost (Fig. 3a). However, we did observe specific regions where qualitative TAD changes clearly correlated with CTCF binding, for example, at the *Sox2* locus, where acquisition of a new border and chromatin-loop formation (described below) was paralleled by a substantial gain in CTCF-binding sites (Supplementary Fig. 5g).

The gain or loss of TAD borders did not correlate with overall increased or decreased local gene expression, respectively, thus suggesting that changes in the level of transcription per se are not a main driver of TAD border dynamics (Supplementary Fig. 5h). Gene expression changes during reprogramming at dynamic border regions were highly context dependent, and there was no apparent correlation between border gain or loss and the direction of transcriptional change (Supplementary Fig. 5i). Moreover, these border regions rarely switched compartment (3–9% versus 17% for all borders). Interestingly, at borders that showed transcriptional changes (>0.5 $\log_2$ change between endpoints) gene expression was often not significantly altered until after TAD borders were newly acquired or lost (Fig. 3b and Supplementary Fig. 5j).

Quantitative changes in TAD border strength occur early in reprogramming. Local chromatin insulation by TAD borders can also be approached quantitatively by calculating an insulation strength score ('I score'; R^2 >0.87 between biological replicates) for each border[43,45] (Fig. 3c). Compared with qualitative border changes (i.e., a gain or loss of border detection), quantitative changes in TAD insulation were more abundant: half of all borders showed a > 20% difference in I score between the first three and last three time points of reprogramming (Fig. 3d; green, red and blue clusters). Stably acquired or lost borders often had lower average I scores than did invariant TAD borders (Supplementary Fig. 6a). CTCF occupancy correlated with I scores, and metaborder plots confirmed that the I-score dynamics reflected actual contact maps (Supplementary Fig. 6b,c). PCA of I-score kinetics showed a reprogramming trajectory grossly resembling the transcriptome, PC1 and H3K4me2 trajectories determined earlier (Supplementary Fig. 6d).

Border regions contained a large number of genes with cell-type-specific functions (for example, 'immune system' and 'developmental biology'), in addition to the expected housekeeping genes[17] (Supplementary Fig. 6e,f). Pluripotency genes, including *Nanog* and *Sox2*, were often found at or near border regions. Both of these loci showed rapid I-score changes that preceded their transcriptional activation (Fig. 3e–g). In B and Bα cells, *Nanog* was separated from *Dppa3* by a strong border in a region containing the −45 kb *Nanog* SE and *Gdf3* (Figs. 3e and 2k), which probably interfered with the reported spatial clustering of these genes and enhancers in PSCs[46]. I scores were considerably lower at D2 after OSKM induction (Fig. 3g), a process facilitating the interactions between genes and their enhancers required for subsequent transcriptional activation (D4–D6). Furthermore, both Hi-C-derived virtual chromosome conformation capture on chip (4C) data obtained at 5-kb resolution and conventional 4C–seq analyses showed higher (cross-)border contact frequencies of the *Nanog* promoter as early as D2 (Fig. 3h and Supplementary Fig. 7a). Within the *Sox2* TAD, a new internal border and several chromatin loops appeared between the Bα and D4 stages, before *Sox2* activation at D6. High-resolution virtual 4C analysis showed that early border emergence progressively skewed interactions of *Sox2* toward its key downstream SE[47,48], thus resulting in the formation of an insulated *Sox2*-SE subdomain at D6 that is likely to be critical for *Sox2* activation (Fig. 3i,j and Supplementary Fig. 7b).

To further understand the relationship between I-score changes and local gene expression, we analyzed transcriptional changes at the 184 most dynamic borders regions that increased in insulation strength (>75% change in I score). Gene expression was altered at many of these borders ($n=88$, >0.5 $\log_2$ change between endpoints) during reprogramming, and again there was no clear bias for activation or repression. At 49% of these borders ($n=43/88$) I scores increased before transcriptional changes (Fig. 3k), whereas for the remaining borders, a mix of concomitant ($n=15$), lagging ($n=15$) and more complex ($n=15$) kinetics was observed. Likewise, I-score changes also preceded modulation of chromatin state and subnuclear compartmentalization (Supplementary Fig. 7). Thus, altered insulation strength at TAD borders is an early reprogramming event often preceding transcriptome rewiring.

Topological plasticity increases late in reprogramming. TADs differ in their propensity to form contacts with other TADs[49,50]. To quantify this 'connectivity' within a given TAD, we computed a domain score (D score) defined by the ratio of intra-TAD interactions over all cis interactions[49] (Fig. 4a). Whereas the I score measures a border's ability to prevent interactions between two neighboring TADs, the D score quantifies a TAD's tendency to self-interact. D scores positively correlated with gene expression and A-compartment association (Supplementary Fig. 8), as previously noted[49,50]. The correlations among D scores, gene expression and compartment association seen at early time points progressively weakened after D4 (Supplementary Fig. 8a,b). Whereas TADs explained a greater proportion of expression variability than linear neighborhoods when we estimated the overall effects of TAD structure on gene expression (Supplementary Information), this proportion progressively decreased during reprogramming (Supplementary Fig. 8c). Together with the observed decrease in overall A- compartment segregation (Supplementary Fig. 2d), and in line with the previously reported decreased organization of inactive chromatin in PSCs[51], these data suggested that at the topological level cells gradually acquire a plastic state characteristic of the pluripotent genome[52].

Altered TAD connectivity frequently precedes transcriptional changes. PCA of D-score kinetics identified a reprogramming trajectory for TAD connectivity similar to those for compartmentalization, transcription, active chromatin and I score (Fig. 4b and Supplementary Fig. 8d). *k*-means clustering showed that 79% of TADs exhibited D-score changes (i.e., >20% change between endpoints) (Fig. 4c). The D-score kinetics correlated closely with compartmentalization (PC1) changes (R >0.84; Supplementary Fig. 8e). TADs with the most dynamic connectivity pattern frequently switched compartment and contained genes enriched in immune-cell and stem-cell-related functions (Supplementary Fig. 8f,g). These TADs were highly biased in their compartment association: 88% of TADs that showed a rapid increase in D scores initially localized to the B compartment, whereas 83% of the TADs with substantially lower D scores initially resided in the A compartment (Supplementary Fig. 8f).

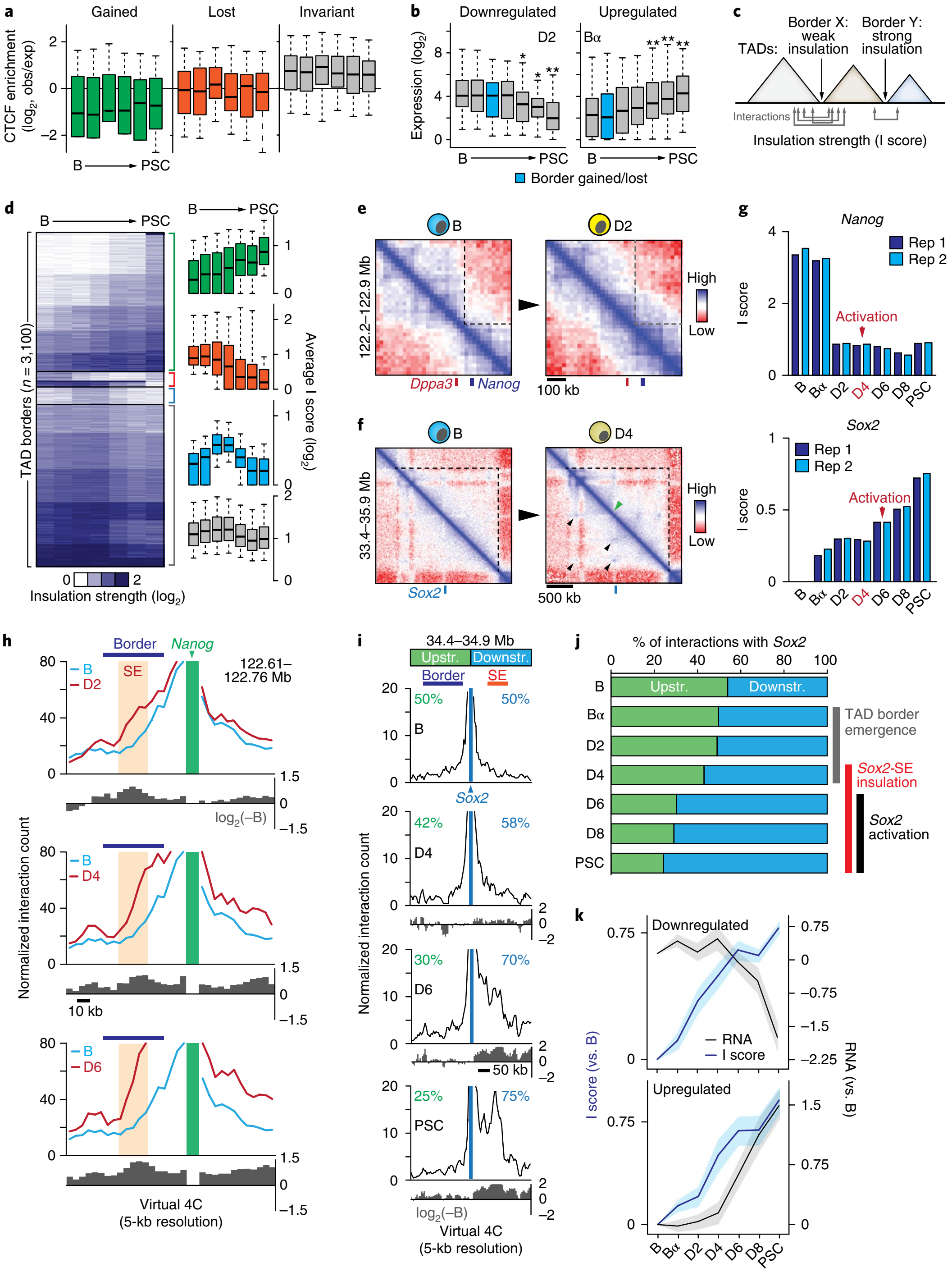


Fig. 3 | Kinetics of domain insulation during reprogramming. a, CTCF enrichment dynamics (from ChIP–seq experiments during reprogramming) at TAD borders that were gained ($n = 431$), lost ($n = 124$) or invariant ($n = 2{,}185$) during reprogramming. All box plots depict the first and third quartiles as the lower and upper bounds of the box, with a thicker band inside the box showing the median value and whiskers representing 1.5× the interquartile range. Obs, observed; exp, expected. **b**, Gene expression dynamics at transcriptionally modulated border regions (divided into upregulated or downregulated groups per time point) gained or lost at D2 or Bα stages (*$P < 0.05$; **$P < 0.005$ versus B cells; unpaired two-tailed t test). Sample sizes are indicated in Supplementary Fig. 5. **c**, Cartoon illustrating the concept of the insulation strength score (I score). **d**, k-means clustering ($k = 20$) of I scores. Bar graphs show I-score kinetics for groups that increased ($n = 1{,}291$), decreased ($n = 141$), transiently increased ($n = 159$) or did not change ($n = 1{,}509$). **e**,**f**, Representative (from two independent biological-replicate experiments) in situ Hi-C contact maps (20-kb resolution) of the *Dppa3*–*Nanog* border (**e**) or the internal *Sox2* border (**f**). Black arrows, loops; green arrow, border formation. **g**, I-score kinetics of the *Nanog* and *Sox2* borders. **h**,**i**, Representative (from two independent biological-replicate experiments) virtual 4C analysis with *Nanog* (**h**) or *Sox2* (**i**) used as viewpoints. TAD borders and SEs are indicated. $\log_2$ ratios (over B) are shown below each line graph; percentages shown in *i* depict proportions of all interactions with *Sox2*. **j**, Proportion of interactions with *Sox2* from the immediate upstream or downstream region (indicated in *i*). Timing of key events involved in *Sox2* activation is indicated. **k**, Gene expression and I-score kinetics at dynamic border regions where I-score changes preceded transcriptional modulation (49%, $n = 43/88$ borders). Line graphs depict I-scores and gene expression dynamics for those borders where gene expression was downregulated or upregulated. Error bars/shading, 95% confidence interval.

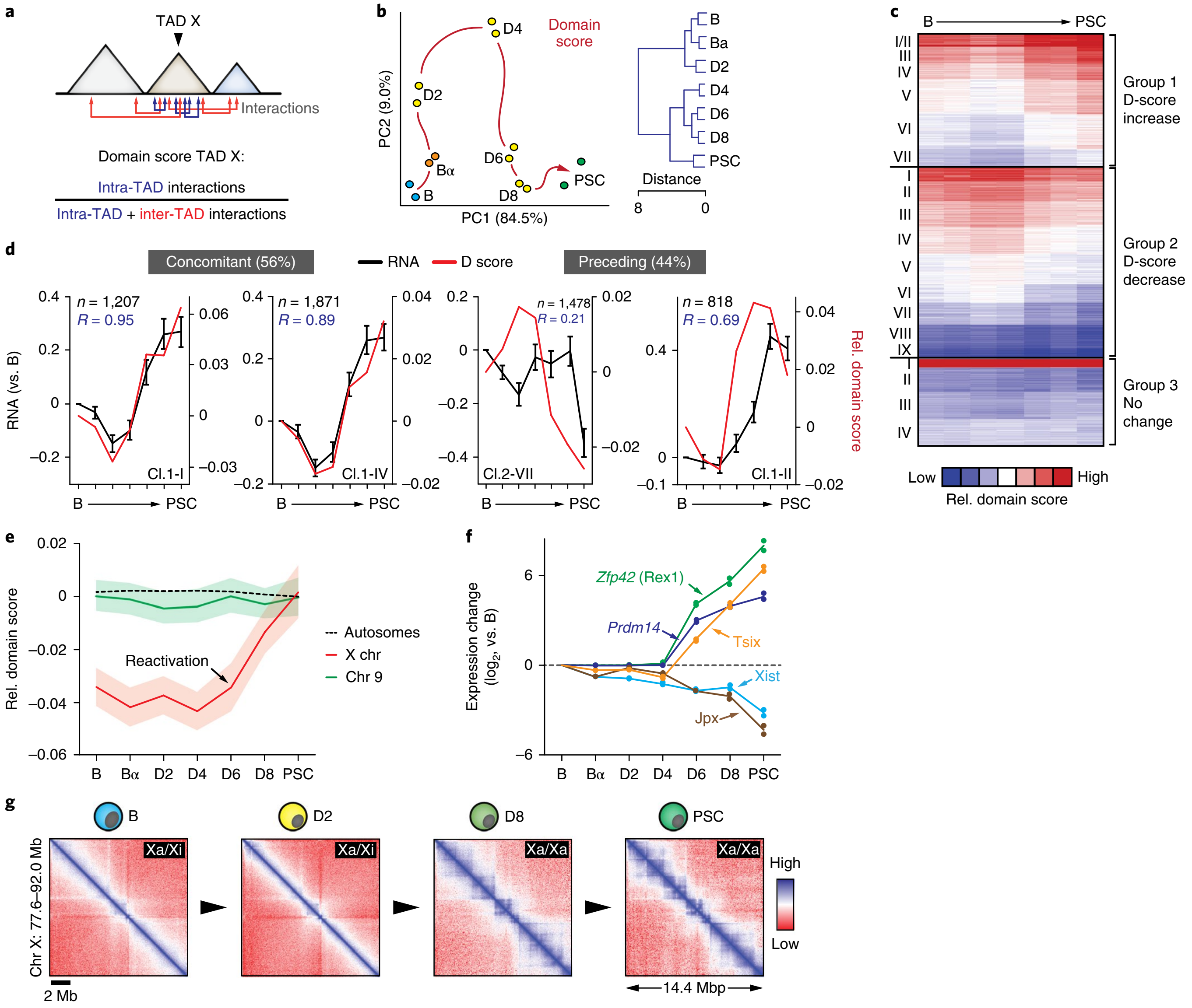


Fig. 4 | Dynamics of TAD connectivity during reprogramming. a, Cartoon depicting D-score calculation. Arrows indicate intra- or inter-TAD interactions. **b**, PCA (left) and unsupervised hierarchical clustering (right) of D-score kinetics ($n=2{,}153$ TADs). Red arrow, hypothetical trajectory. **c**, *k*-means clustering ($k=20$) of genome-wide relative D scores (centered on mean). **d**, Examples of individual dynamic D-score clusters for which gene expression and D-score kinetics (mean values presented; number of genes per cluster (*n*) indicated) were concomitant or where D-score changes preceded transcriptional changes. Error bars, s.e.m. *R* values denote Pearson correlation coefficients. **e**, Average relative D-score changes for chromosome (chr) 9 ($n=115$ TADs), all autosomes combined ($n=1{,}959$ TADs) and the X chromosome ($n=106$ TADs). Shading denotes 95% confidence interval. **f**, Mean gene expression changes (versus B cells, $n=2$ independent biological replicate reprogramming experiments) of key regulators of X-chromosome re/inactivation during reprogramming. **g**, Representative (from two independent biological-replicate experiments) in situ Hi-C contact maps (50-kb resolution) of a 14.5-Mb region on the X chromosome during reprogramming. B-D2 cells carry one inactive X (Xi) and one active X (Xa) chromosome; D8-PSCs carry two Xa.

To assess the correlation between TAD connectivity and gene expression, we compared D scores with intra-TAD gene expression kinetics for the 16 dynamic D-score clusters (Fig. 4c). In 9 of 16 clusters, D-score changes coincided with alterations in gene expression (Fig. 4d), particularly for TADs that showed both higher D scores and intra-TAD expression ($R=0.78$). However, 7 of 16 clusters showed D-score changes preceding transcriptional changes, and no clusters showed the opposite pattern (Fig. 4d). Thus, changes in TAD connectivity frequently precede intra-TAD transcriptional modulation.

X-chromosome reactivation evokes TAD reorganization. X-chromosome reactivation in PSCs is a classic model for studying the relationship between chromosome structure and gene expression[53]. The B cells used were derived from female mice carrying one inactive X chromosome, thus allowing us to study this process within our dataset. Whereas the average TAD connectivity for each time point remained similar on autosomes, X-chromosome TADs displayed substantial gains in D scores after D4 (Fig. 4e). We reasoned that the observed chromosome-wide D-score increase might have been caused by a reactivation of the largely TAD-devoid inactive X chromosome[11,54–56]. Indeed, after D4, TAD structures were fully reestablished, and key regulators of X reactivation were activated (*Zfp42*, *Prdm14* and *Tsix*), whereas X-chromosome repressors (*Xist* and *Jpx*) were downregulated (Fig. 4f,g).

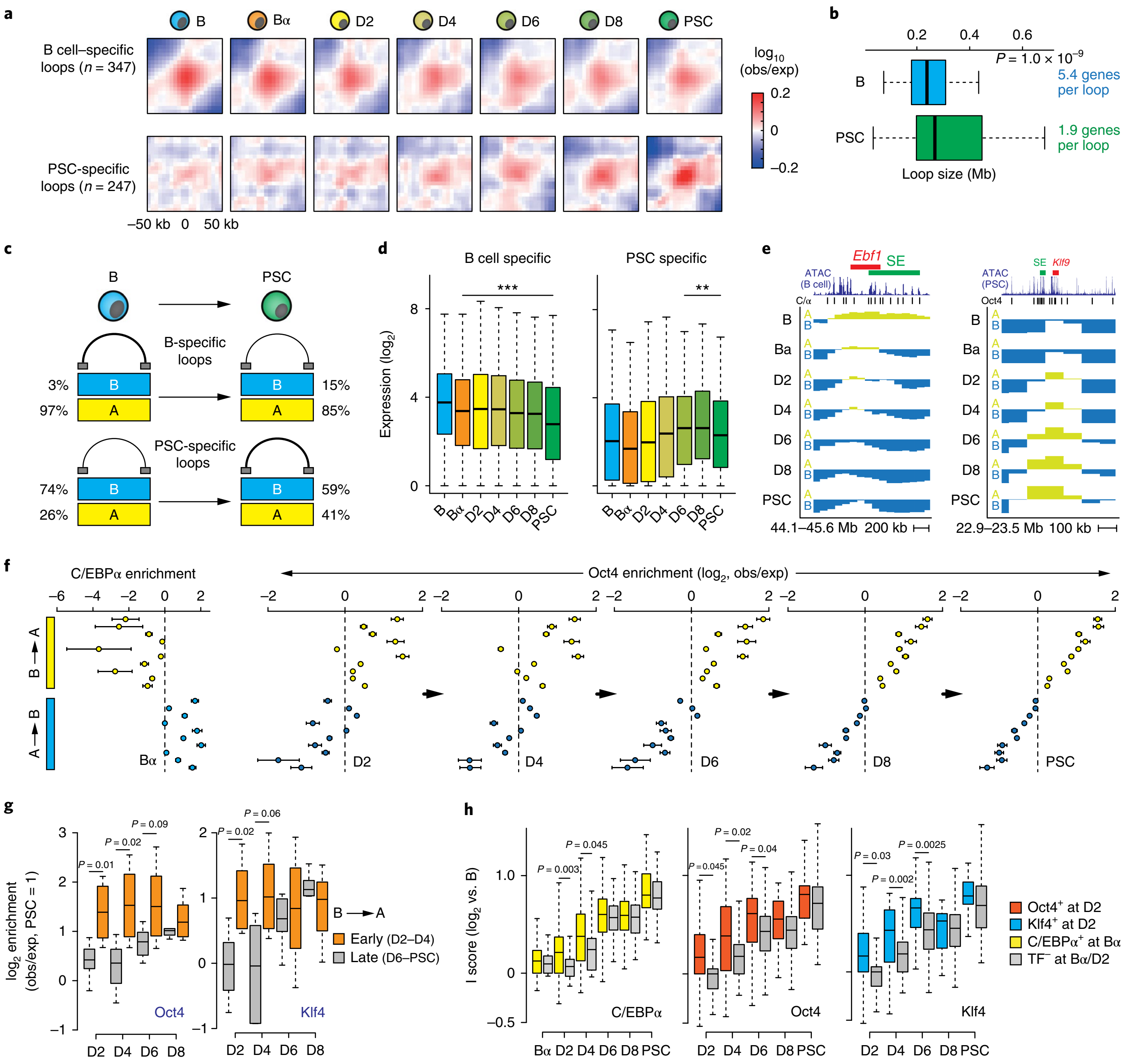


Fig. 5 | Chromatin looping and transcription-factor-binding dynamics. a, Metaloop analysis at 5-kb resolution of B cell– or PSC-specific loops[49]. Area shown is centered on the respective TF-binding sites (±50 kb). **b**, Box plot showing median loop size ($P = 1.0 \times 10^{-9}$, two-sided Wilcoxon rank-sum test) and average number of genes per loop for B cells ($n = 347$ loops) or PSCs ($n = 247$ loops). All box plots depict the first and third quartiles as the lower and upper bounds of the box, with a thicker band inside the box showing the median value and whiskers representing 1.5× the interquartile range. (**c**) Cartoon depicting percentage of B cell– or PSC-specific loops within A or B compartments in reprogramming end stages. **d**, Box plot showing gene expression dynamics of genes within B cell-specific (left, $n = 1{,}874$ genes)) or PSC-specific (right, $n = 469$ genes) loops ($^{**}P < 0.005$; $^{***}P < 0.001$ versus B cells; two-sided Wilcoxon rank-sum test). **e**, Examples of C/EBPα-mediated A-to-B switching (*Ebf1* locus) and OSKM-mediated B-to-A switching (*Klf9* locus). SE location is indicated. **f**, C/EBPα and OCT4 binding enrichment (inferred from ChIP-seq and ATAC–seq, respectively; additional data in Supplementary Information) relative to the genome-wide average at the 20 switching clusters shown in Fig. 2g. Mean values with 95% confidence interval are shown. Obs, observed; exp, expected. **g**, Box plots showing OCT4 and KLF4 binding enrichment in clusters ($n = 10$ clusters) that switch from the B to A compartment early (D2–D4) or late (D6–PSCs). Statistical significance was assessed with an unpaired two-tailed *t* tests. **h**, Insulation strength (I-score) dynamics at hyperdynamic borders ($n = 184$) bound ($n = 123$ borders for C/EBPα; $n = 37$ borders for OCT4; $n = 22$ borders for KLF4) or not bound ($n = 61$ borders for C/EBPα; $n = 147$ borders for OCT4; $n = 162$ borders for KLF4) by the indicated TFs. Statistical significance was assessed with unpaired two-tailed *t* tests.

Cell-type-specific changes in chromatin loops. Chromatin loops appear as foci in high-resolution Hi-C maps and represent particularly strong interactions between two distant regions[11]. We visualized chromatin-loop dynamics during reprogramming by performing metaloop analyses at 5-kb resolution of a previously identified set of loops in primary B cells and PSCs[49]. Similarly to the TADs that they often demarcate[11], chromatin loops generally behave as remarkably stable topological structures during reprogramming

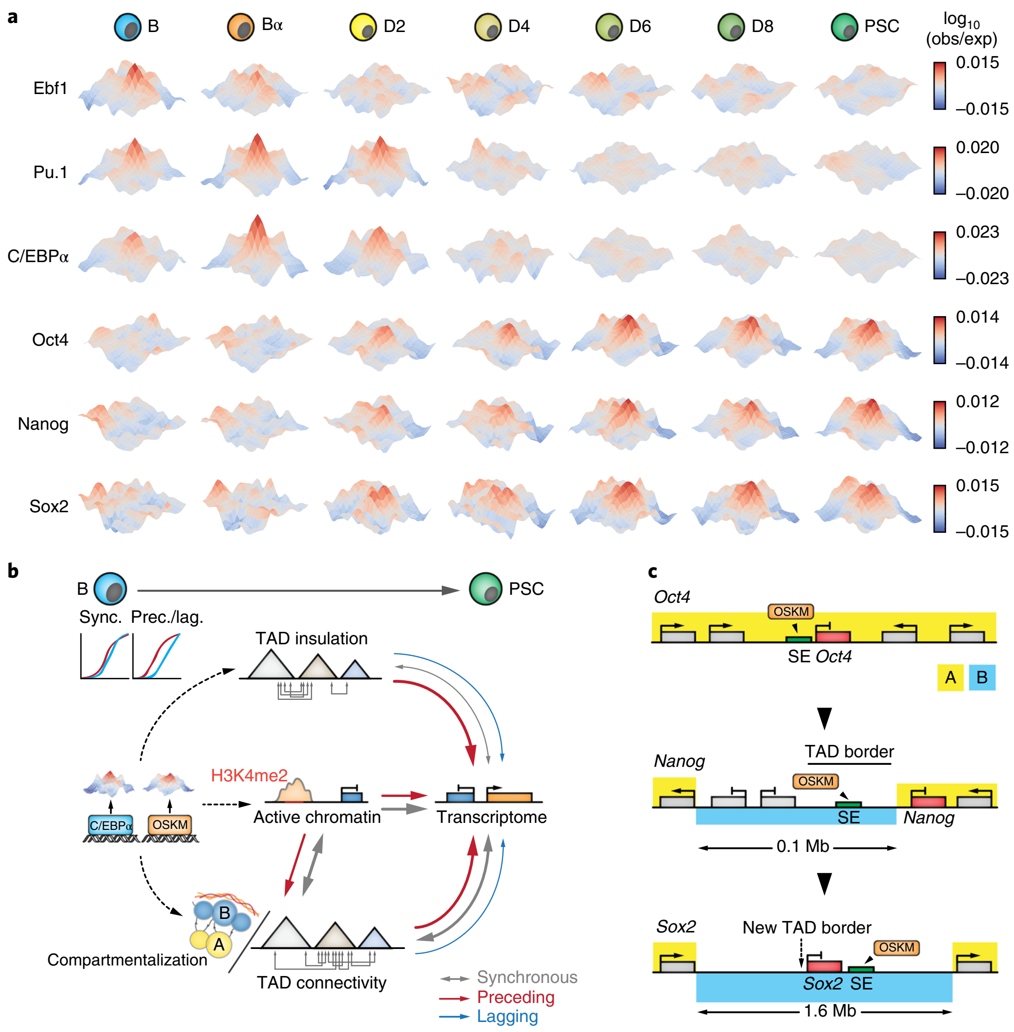


Fig. 6 | Dynamics of 3D cross-talk between transcription-factor target sites and model schemes. a, 3D interaction metaplots (5-kb resolution) depicting interaction frequencies of sites bound by the indicated TFs during reprogramming. Hubs visualize inter-TAD cross-talk between TF-binding sites 2–10 Mb apart. Area shown is centered on the respective TF-binding sites (including 50 kb upstream and downstream). (**b**) Summary scheme depicting the interplay among TF binding, chromatin state, various aspects of genome topology and gene regulation during cell reprogramming. Arrows denote synchronous (sync.) or preceding or lagging (prec./lag.) relationships. Arrow thickness indicates prevalence. **c**, Activation scenarios for the pluripotency factors *Oct4*, *Nanog* and *Sox2*. *Oct4* activation does not appear to require major topological modifications, because the gene and its SE already reside in the A compartment in B cells and TAD border strength is unaltered. In contrast, *Nanog* activation is preceded by B-to-A compartment switching of its nearby SE as well as a decrease in TAD border strength that facilitates *Nanog*–SE interaction. *Sox2* activation is preceded by the formation of a new TAD border through chromatin-loop formation that progressively insulates the gene and its SE into a smaller subdomain. The complete 1.6-Mb *Sox2* region switches from the B to the A compartment, in a manner concomitant with activation of the gene at D6.

(Supplementary Fig. 9a). However, cell-type-specific loops, representing a minor fraction of all loops (13% for B cells, 5% for PSCs[49]), showed a dynamic behavior: B cell–specific loops lost interaction strength, whereas PSC-specific loops were established de novo from D4 onward (Fig. 5a). Intriguingly, the nature of these somatic and pluripotent cell-type-specific loops appeared to be different: PSC-specific loops were larger than B cell–specific loops and localized mostly to the B compartment (whereas virtually all B cell–specific loops localized to A); contained fewer genes that showed lower average gene expression levels; and were less enriched in cell-type-specific genes (Fig. 5b,c and Supplementary Fig. 9b). However, in both cases, the presence of a loop positively correlated with gene expression changes, thus indicating that both the formation and loss of cell-type-specific loops are dynamically linked to gene regulation (Fig. 5d).

Transcription factors drive topological genome reorganization. We investigated the effects of C/EBPα and OSKM on genome topology. Approximately 5% of the genome switched compartment during the C/EBPα-induced B-to-Bα transition, and 5% switched compartment during the OSKM-induced Bα-to-D2 transition. Of these early switching regions, only 29% (B to Bα) and 36% (Bα to D2) represented stable switches (Supplementary Fig. 10a). C/EBPα had a largely repressive effect (66% A-to-B switches, for example, *Ebf1*), whereas OSKM TFs operated predominantly as activators (70% B-to-A switches, for example, *Klf9*) (Fig. 5e and Supplementary Fig. 10a). Both C/EBPα and OSKM evoked A-to-B switching and transcriptional silencing of B cell–related loci. At D2, OSKM induced B-to-A switching and activation of known target genes of pluripotency factors involved in developmental processes (Supplementary Fig. 10b). However, genes undergoing stable B-to-A switching in

Bα cells were upregulated only after OSKM activation; many of these genes have been implicated in early embryonic development (for example, *Dppa3*; Supplementary Fig. 10c). Globally, C/EBPα binding was strongly enriched in the previously identified A-to-B-switching clusters and depleted in B-to-A-switching clusters (Fig. 5f). In contrast, OCT4 and KLF4 binding (as inferred from ATAC–seq data) was concentrated in B-to-A-switching regions (Fig. 5f and Supplementary Fig. 10d). This biased genomic distribution was already apparent at D2 and was stably maintained or reinforced. Hereby, early-switching clusters (D2–D4) were rapidly targeted by OCT4 and KLF4, whereas late-switching clusters (D6-PSCs) became more gradually enriched (Fig. 5g).

We next examined TF action at TAD borders. OCT4-target sites within ~30% of all border regions were already accessible at D2 (Supplementary Fig. 10e). OCT4 and KLF4 recruitment to the most dynamic borders at D2, as compared with borders bound at later time points, correlated with accelerated I-score gains (Fig. 5h and Supplementary Fig. 10f). I scores of C/EBPα-bound borders increased more rapidly only after OSKM activation at D2 (Fig. 5h), and OCT4 enrichment was significantly higher at borders previously bound by C/EBPα (Supplementary Fig. 10g), thus suggesting that C/EBPα primes border regions for subsequent OSKM-induced topological changes. In agreement with this possibility, OCT4, KLF4 and C/EBPα were frequently recruited to the same dynamic borders early in reprogramming (Supplementary Fig. 10h).

Because TF-bound sites cluster over large distances[14,51,57,58], we sought to address the dynamics of such 3D cross-talk during reprogramming by measuring inter-TAD spatial connectivity between TF-bound genomic sites at 5-kb resolution (within a 2- to 10-Mb window, a procedure analogous to paired-end spatial chromatin analysis (PE-SCAn)[51]). We observed strong interactions between EBF1- or PU.1-binding sites in B cells, in agreement with their function as key B cell regulators (Fig. 6a). These interaction networks largely disappeared for EBF1 in Bα and for PU.1 in D4 cells. Spatial clustering of C/EBPα targets was already present in B cells (Fig. 6a), thus indicating that C/EBPα exploits existing 3D interaction hubs, such as those formed by PU.1. Alongside interaction hubs mediated by hematopoietic TFs, OCT4-binding sites clustered from D2 onward, thus establishing 3D cross-talk among PSC-specific regulatory elements and showing that interaction hubs mediated by lineage-specific and pluripotency TFs can coexist (Fig. 6a). Moreover, NANOG-targeted regions formed interaction hubs as early as D2, before the gene became expressed at D4 (Fig. 6a), thereby suggesting that late pluripotency factors ‘hitchhike’ onto an OSKM-mediated interaction hub and consequently lock in PSC fate.

In summary, binding of C/EBPα and OSKM correlates with accelerated topological remodeling of compartmentalization and TAD insulation. In addition, computing inter-TAD 3D cross-talk among TF targets enabled us to visualize the stage-specific formation and disassembly of TF interaction hubs during reprogramming.

Discussion

Our analysis of somatic cell reprogramming (summarized in Supplementary Fig. 10i) showed that the overall dynamics of genome topology, chromatin state and gene expression is closely linked. Nevertheless, this linkage often occurs in a nonsynchronous manner: changes in subnuclear compartmentalization, TAD connectivity and TAD-border insulation strength frequently precede transcriptional changes, and the reverse situation occurs only at low frequencies. We propose that transcription factors induce successive changes in chromatin state and genome architecture and consequently enable gene-regulatory rewiring during cell reprogramming (Fig. 6b). Genome topology as an instructive force that facilitates transcriptional changes may represent a general principle for cell-fate decisions.

Our findings also provide an explanation for the sequential activation of the genes encoding the pluripotency factors OCT4, NANOG and SOX2 in spite of the cells’ continuous exposure to the Yamanaka factors (Fig. 6c). The embedding of *Oct4* and its enhancers within an A-compartment domain, surrounded by genes highly expressed in B cells, may explain its almost immediate activation by OSKM without detectable topological alterations. In contrast, the late activation of *Nanog* and *Sox2* is preceded and accompanied by substantial changes in compartmentalization and TAD structure, thus indicating that the removal of topological barriers creates new opportunities for gene regulation. That active chromatin dynamics often anticipates changes in subnuclear compartmentalization suggests that it plays a major role in mediating switching between the active A and the inactive B compartments (Fig. 6b), a possibility supported by imaging and local chromatin conformation analyses[59,60]. Given the strong correlation between compartmentalization and DNA-replication timing[61], it will be of interest to attempt coupling changes in replication timing with the dynamics of genome topology and gene regulation. A preliminary analysis suggested that replication timing in the starting cell state is not a strong predictor of ordered A–B-compartment switching (Supplementary Fig. 11). Perturbation experiments aimed at demonstrating causality between specific topological changes and their effects on reprogramming represent the next frontier in dissecting the relationships among genome form, genome function and cell fate.

Previous studies have defined TADs as stable topological structures with little cell-type specificity[17,50]. At a qualitative level (i.e., present or not present), we indeed found that only a minor portion of TAD borders were altered during reprogramming. However, there were notable exceptions (for example, de novo border establishment near *Sox2*), thus cautioning against using TAD definitions from unrelated cell types for interpreting gene-regulatory processes. However, quantitative aspects of TADs, namely their connectivity and insulation potential, are subject to substantial changes during reprogramming and therefore are more cell-type specific in nature.

How do TFs drive 3D genomic changes? C/EBPα and OCT4 are selectively enriched in different regions destined to switch compartment. Here, TFs can act by inducing the subnuclear repositioning of specific loci[62], for example by initiating modification of local chromatin states. In addition, the TFs rapidly induce insulation-strength changes at the most dynamic TAD borders, independently of major changes in compartmentalization or chromatin state. Separate modes of action for TFs at these two topological levels seem plausible, because compartmentalization and TAD organization have been suggested to depend on distinct mechanisms[63,64]. Mechanistically, intrinsic abilities (for example, via TF dimerization[24]) or interactions with canonical architectural proteins[65–67] could allow TFs to modify genome topology. Here, inter-TAD hubs of TF target regions may contribute to topological reorganization by TFs, possibly through exploiting architecture previously established by other factors. As early targets (Fig. 1), SE regions may provide key platforms for TFs to achieve topological genome remodeling[68]. The ability of lineage-instructive regulators to alter genome topology raises the possibility that, in addition to their classical role as transcriptional regulators, they may possess unappreciated architectural functions at distinct topological layers.

URLs. FastQC quality-control tool, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; Picard tool; http://broadinstitute.github.io/picard/; MACS2 tool, https://github.com/taoliu/MACS/.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-017-0030-7.

Received: 29 April 2017; Accepted: 1 December 2017;

References

1. Buganim, Y., Faddah, D. A. & Jaenisch, R. Mechanisms and models of somatic cell reprogramming. *Nat. Rev. Genet.* **14**, 427–439 (2013).
2. Apostolou, E. & Hochedlinger, K. Chromatin dynamics during cellular reprogramming. *Nature* **502**, 462–471 (2013).
3. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499–506 (2013).
4. Gorkin, D. U., Leung, D. & Ren, B. The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell* **14**, 762–775 (2014).
5. Dekker, J. & Mirny, L. The 3D genome as moderator of chromosomal communication. *Cell* **164**, 1110–1121 (2016).
6. Denker, A. & de Laat, W. The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev.* **30**, 1357–1382 (2016).
7. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin domains: the unit of chromosome organization. *Mol. Cell* **62**, 668–680 (2016).
8. Cavalli, G. & Misteli, T. Functional implications of genome topology. *Nat. Struct. Mol. Biol.* **20**, 290–299 (2013).
9. Boettiger, A. N. et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).
10. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
11. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
12. Wang, S. et al. Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016).
13. Vieux-Rochas, M., Fabre, P. J., Leleu, M., Duboule, D. & Noordermeer, D. Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. *Proc. Natl. Acad. Sci. USA* **112**, 4672–4677 (2015).
14. Lin, Y. C. et al. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nat. Immunol.* **13**, 1196–1204 (2012).
15. Stevens, T. J. et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).
16. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
17. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
18. Sexton, T. et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
19. Lupiáñez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
20. Symmons, O. et al. Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400 (2014).
21. Andrey, G. et al. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* **340**, 1234167 (2013).
22. Montavon, T. et al. A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147**, 1132–1145 (2011).
23. Symmons, O. et al. The Shh topological domain facilitates the action of remote enhancers by reducing the effects of genomic distances. *Dev. Cell* **39**, 529–543 (2016).
24. Deng, W. et al. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**, 1233–1244 (2012).
25. Hug, C. B., Grimaldi, A. G., Kruse, K. & Vaquerizas, J. M. Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell* **169**, 216–228.e19 (2017).
26. Ke, Y. et al. 3D chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis. *Cell* **170**, 367–381.e20 (2017).
27. Du, Z. et al. Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* **547**, 232–235 (2017).
28. Apostolou, E. et al. Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. *Cell Stem Cell* **12**, 699–712 (2013).
29. Ghavi-Helm, Y. et al. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100 (2014).
30. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994–1004 (2012).
31. Di Stefano, B. et al. C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature* **506**, 235–239 (2014).
32. Di Stefano, B. et al. C/EBPα creates elite cells for iPSC reprogramming by upregulating Klf4 and increasing the levels of Lsd1 and Brd4. *Nat. Cell Biol.* **18**, 371–381 (2016).
33. Buganim, Y. et al. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209–1222 (2012).
34. Bar-Nur, O. et al. Small molecules facilitate rapid and synchronous iPSC generation. *Nat. Methods* **11**, 1170–1176 (2014).
35. Ruetz, T. & Kaji, K. Routes to induced pluripotent stem cells. *Curr. Opin. Genet. Dev.* **28**, 38–42 (2014).
36. Schmidl, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat. Methods* **12**, 963–965 (2015).
37. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
38. Heinz, S., Romanoski, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* **16**, 144–154 (2015).
39. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
40. Nutt, S. L. & Kee, B. L. The transcriptional regulation of B cell lineage commitment. *Immunity* **26**, 715–725 (2007).
41. Martello, G. & Smith, A. The nature of embryonic stem cells. *Annu. Rev. Cell Dev. Biol.* **30**, 647–675 (2014).
42. Blinka, S., Reimer, M. H. Jr., Pulakanti, K. & Rao, S. Super-enhancers at the Nanog locus differentially regulate neighboring pluripotency-associated genes. *Cell Rep.* **17**, 19–28 (2016).
43. Crane, E. et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).
44. Ong, C. T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* **15**, 234–246 (2014).
45. Van Bortle, K. et al. Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol.* **15**, R82 (2014).
46. Levasseur, D. N., Wang, J., Dorschner, M. O., Stamatoyannopoulos, J. A. & Orkin, S. H. Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells. *Genes Dev.* **22**, 575–580 (2008).
47. Li, Y. et al. CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells. *PLoS One* **9**, e114485 (2014).
48. Zhou, H. Y. et al. A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes Dev.* **28**, 2699–2711 (2014).
49. Krijger, P. H. et al. Cell-of-origin-specific 3D genome structure acquired during somatic cell reprogramming. *Cell Stem Cell* **18**, 597–610 (2016).
50. Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
51. de Wit, E. et al. The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature* **501**, 227–231 (2013).
52. Meshorer, E. et al. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* **10**, 105–116 (2006).
53. Pasque, V. & Plath, K. X chromosome reactivation in reprogramming and in development. *Curr. Opin. Cell Biol.* **37**, 75–83 (2015).
54. Giorgetti, L. et al. Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575–579 (2016).
55. Deng, X. et al. Bipartite structure of the inactive mouse X chromosome. *Genome Biol.* **16**, 152 (2015).
56. Minajigi, A. et al. Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, aab2276 (2015).
57. Schoenfelder, S. et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.* **42**, 53–61 (2010).
58. Liu, Z. et al. 3D imaging of Sox2 enhancer clusters in embryonic stem cells. *eLife* **3**, e04236 (2014).
59. Therizols, P. et al. Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science* **346**, 1238–1242 (2014).
60. Wijchers, P. J. et al. Cause and consequence of tethering a SubTAD to Different Nuclear Compartments. *Mol. Cell* **61**, 461–473 (2016).
61. Pope, B. D. et al. Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**, 402–405 (2014).
62. Zullo, J. M. et al. DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. *Cell* **149**, 1474–1487 (2012).
63. Schwarzer, W. et al. Two independent modes of chromosome organization are revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
64. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944 (2017). e22.
65. van den Berg, D. L. et al. An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* **6**, 369–381 (2010).
66. Donohoe, M. E., Silva, S. S., Pinter, S. F., Xu, N. & Lee, J. T. The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature* **460**, 128–132 (2009).
67. Wei, Z. et al. Klf4 organizes long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency. *Cell Stem Cell* **13**, 36–47 (2013).

68. Beagrie, R. A. et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).

Acknowledgements

We thank D. Higgs, J. Hughes, J. Davies and Z. Duan for advice on Hi-C technology; C. Schmidl for ChIPmentation advice; C. van Oevelen for help with CTCF ChIP–seq; C. Segura for mouse-colony management; T. Tian for bone marrow collection; the CRG Genomics Core Facility and the CRG-CNAG Sequencing Unit for sequencing; and members of the laboratory of T.G. for discussions. This work was supported by the European Research Council under the 7th Framework Programme FP7/2007-2013 (ERC Synergy Grant 4D-Genome, grant agreement 609989 to T.G., G.J.F., M.A.M.-R. and M.B.) and the Ministerio de Educacion y Ciencia, SAF.2012-37167. R.S. was supported by an EMBO Long-term Fellowship (ALTF 1201-2014) and a Marie Curie Individual Fellowship (H2020-MSCA-IF-2014). We also acknowledge support from 'Centro de Excelencia Severo Ochoa 2013-2017' (SEV-2012-0208) and AGAUR to the CRG.

Author contributions

R.S. and T.G. conceived the study and wrote the manuscript with input from all coauthors. R.S. performed molecular biology, RNA-seq, ChIP–seq, ChIPmentation, ATAC–seq, 4C–seq and in situ Hi-C experiments. R.S., E.V., F.S., J.Q., A.G., S.C. and M.A.M.-R. performed bioinformatic analyses. R.S., E.V., F.S. and M.A.M.-R. integrated and visualized data. B.D.S. performed reprogramming experiments with help from R.S. and C.B.; R.S., F.L.D. and Y.C. optimized and implemented in situ Hi-C technology. J.H. performed high-throughput sequencing. F.L.D., G.J.F., M.B. and M.A.M.-R. provided valuable advice, and T.G. supervised the research.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-017-0030-7.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to R.S. or M.A.M.-R. or T.G.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Mice. We crossed 'reprogrammable mice' containing a doxycycline-inducible OSKM cassette and the tetracycline transactivator[69] with an Oct4-GFP reporter strain[70], as previously described[31,32]. B cells were isolated from 8- to 16-week-old female mice ($n=6$ mice per biological replicate). Mice were housed in standard cages under 12-h 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 culture and somatic cell reprogramming. Mouse embryonic stem cells (E14TG2a) and short-term-induced PSCs were cultured on gelatinized plates or mitomycin C–inactivated mouse embryonic fibroblasts (MEFs) in N2B27 medium (50% DMEM-F12, 50% Neurobasal medium, N2 (100×), B27 (50×)) supplemented with small-molecule inhibitors PD (1 µM PD0325901) and CHIR (3 µM CHIR99021), as well as LIF (10 ng ml^{-1}). Reprogramming of primary B cells isolated from the bone marrow of reprogrammable/Oct4-GFP mice was performed as previously described[32]. Two independent biological-replicate reprogramming experiments were used for data generation. Briefly, pre-B cells were infected with C/EBPαER-hCD4 retrovirus, plated at 500 cells cm^{-2} in gelatinized 12-well plates on mitomycin C–inactivated MEF feeders in RPMI medium. C/EBPα was activated by the addition of 100 nM β-estradiol (E2) for 18 h. After E2 washout, the cultures were switched to N2B27 medium supplemented with IL-4 (10 ng ml^{-1}), IL-7 (10 ng ml^{-1}) and IL-15 (2 ng ml^{-1}). OSKM was activated by the addition of 2 µg ml^{-1} of doxycycline. Harvesting was done at the indicated time points by trypsinization followed by a 20 min preplating step to remove feeder cells. All cell lines were routinely tested for mycoplasma contamination.

RNA isolation, quantitative RT–PCR and RNA-seq. RNA was extracted with a miRNeasy mini kit (Qiagen) and quantified with a NanoDrop spectrophotometer. cDNA was produced with a High Capacity RNA-to-cDNA kit (Applied Biosystems) and was used for qRT–PCR analysis in triplicate reactions with SYBR Green QPCR Master Mix (Applied Biosystems). Primers are available upon request. Libraries were prepared with an Illumina TruSeq Stranded mRNA Library Preparation Kit, and paired-end sequencing (2 × 125 bp) was performed on an Illumina HiSeq2500 instrument.

Assay for transposase-accessible chromatin with high throughput sequencing (ATAC–seq). ATAC–seq was performed as previously described[32]. 100,000 cells were washed once with 100 µl PBS and resuspended in 50 µl lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$ and 0.2% IGEPAL CA-630). Cells were centrifuged for 10 min at 500*g* (4 °C), supernatant was removed, and nuclei were resuspended in 50 µl transposition reaction mix (25 µl TD buffer, 2.5 µl Tn5 transposase and 22.5 µl nuclease-free water) and incubated at 37 °C for 45 min. DNA was isolated with a MinElute DNA Purification Kit (Qiagen). Library amplification was performed by PCR with two sequential reactions (eight then five cycles). Library quality was checked on a Bioanalyzer and was followed by paired-end sequencing (2 × 75 bp) on an Illumina HiSeq2500 instrument.

Chromatin immunoprecipitation followed by high-throughput sequencing (ChIP–seq). ChIP–seq using tagmentation (ChIPmentation) was performed as previously described[36] with 100,000 cross-linked cells and1 µl of anti-H3K4me2 (Abcam, ab32356) per IP. Tagmentation of immobilized H3K4me2-enriched chromatin was performed for 2 min at 37 °C in 25 µl transposition reaction mix (12.5 µl TD buffer, 1.0 µl Tn5 transposase and 11.5 µl nuclease-free water). Library amplification was performed as described for ATAC–seq. Library quality was checked on a Bioanalyzer, and sequencing (1 × 75 bp) was performed on an Illumina NextSeq500 instrument. Conventional ChIP–seq was performed as previously described[71] with 300,000 cross-linked cells and 5 µl of anti-CTCF (Millipore, 07-729). Libraries were prepared with an Illumina TruSeq ChIP Library Preparation Kit and sequenced (1 × 50 bp) on an Illumina HiSeq2500 instrument.

Chromosome conformation capture followed by high-throughput sequencing (4C–seq). 4C–seq was performed as described previously[72,73]. Briefly, 0.5 million to 1 million cross-linked nuclei were digested with Csp6I, then ligated under dilute conditions. After de-cross-linking and DNA purification, samples were digested overnight with DpnII and were again ligated under dilute conditions. Column-purified DNA was directly used as input for inverse PCR with primers (available upon request) with Illumina adaptor sequences as overhangs, and several reactions were pooled, purified and sequenced (1 × 75 bp) on an Illumina HiSeq2500 instrument.

Gene Ontology (GO) analysis. GO analyses were performed with the Molecular Signatures Database (MSigDB)[74] for gene lists or GREAT[75] for peak lists. Only statistically significant (FDR <0.01) terms and pathways were used.

In situ Hi-C library preparation. In situ Hi-C was performed as previously described[11] with the following modifications: (i) two million cells were used as starting material; (ii) chromatin was initially digested with 100 U MboI (New England BioLabs) for 2 h, and then another 100 U (2 h incubation) and a final 100 U were added before overnight incubation; (iii) before fill-in with bio-dATP, nuclei were pelleted and resuspended in fresh 1× NEB2 buffer; (iv) ligation was performed overnight at 24 °C with 10,000 cohesive end units per reaction; (v) de-cross-linked and purified DNA was sonicated to an average size of 300–400 bp with a Bioruptor Pico (Diagenode; seven cycles of 20 s on and 60 s off); (vi) DNA fragment-size selection was performed only after final library amplification; (vii) library preparation was performed with an NEBNext DNA Library Prep Kit (New England BioLabs) with 3 µl NEBNext adaptor in the ligation step; (viii) libraries were amplified for 8–12 cycles with Herculase II Fusion DNA Polymerase (Agilent) and were purified/size-selected with Agencourt AMPure XP beads (>200 bp). Hi-C library quality was assessed through ClaI digestion and low-coverage sequencing on an Illumina NextSeq500 instrument, after which every technical replicate ($n=2$) of each biological replicate ($n=2$) was sequenced at high coverage on an Illumina HiSeq2500 instrument. Data from technical replicates were pooled for downstream analysis. We sequenced >18 billion reads in total to obtain 0.78–1.21 billion valid interactions per time point per biological replicate (dataset statistics in Supplementary Table 1).

Gene expression analysis of RNA-seq data. Reads were mapped with STAR[76] (-outFilterMultimapNmax 1 -outFilterMismatchNmax 999 -outFilterMismatchNoverLmax 0.06 -sjdbOverhang 100 --outFilterType BySJout -alignSJoverhangMin 8 -alignSJDBoverhangMin 1 -alignIntronMin 20 -alignIntronMax 1000000 -alignMatesGapMax 1000000) and the Ensembl mouse genome annotation (GRCm38.78). Gene expression was quantified with STAR (--quantMode GeneCounts). Sample scaling and statistical analysis were performed with the R package DESeq2 (ref. [77]) (R 3.1.0 and Bioconductor 3.0), and vsd counts were used for further analysis unless stated otherwise. Standard RPKM values were used as an absolute measure of gene expression. Genes whose expression changed significantly at any time point were identified with the nbinomLRT test (FDR <0.01) and a greater than twofold change between at least two time points (average of two biological replicates, vsd values). Clustering was performed with the R package Mfuzz (2.26.0).

Chromatin accessibility analysis of ATAC–seq data. Reads were mapped to the UCSC mouse genome build (mm10) in Bowtie2 (ref. [78]) with standard settings. Reads mapping to multiple locations in the genome were removed in SAMtools[79]; PCR duplicates were filtered in Picard. Bam files were parsed to HOMER[80] for downstream analyses and browser visualization. Peaks in ATAC–seq signals were identified with *findPeaks* (-region -localSize 50000 -size 250 -minDist 500 -fragLength 0, FDR <0.001).

ChIPmentation/ChIP–seq data analysis. Reads were mapped and filtered as described for ATAC–seq. H3K4me2-enriched regions were identified with HOMER *findpeaks* (findPeaks -region -size 1000 -minDist 2500, by using a mock IgG experiment as background signal). H3K4me2 coverage per 100-kb genomic bin was computed in BEDTools[81] and normalized for differences in sequencing depth (normalized coverage = coverage/(number of unique mapped reads in dataset/1×10^6)). CTCF peaks were identified in MACS2 (ref. [82]) with *callpeak*–nolambda–nomodel -g mm--extsize 100 -q 0.01.

4C–seq data analysis. The sequence of the 4C–seq reading primer was trimmed from the 5′ of reads with the demultiplex.py script from the R package fourCseq[83] (allowing four mismatches). Reads in which this sequence could not be found were discarded. Reads were mapped in STAR and processed with fourCseq to filter out reads not located at the end of a valid fragment and to count reads per fragment. Signal tracks were made after smoothing of RPKM counts per fragment with a running mean over three fragments.

In situ Hi-C data processing and normalization. We processed Hi-C data by using an in-house pipeline based on TADbit[84]. First, the quality of the reads was checked with FastQC to discard problematic samples and detect systematic artifacts. Trimmomatic[85] with the recommended parameters for paired-end reads was used to remove adaptor sequences and poor-quality reads (ILLUMINACLIP:TruSeq3-PE.fa:2:30:12:1:true; LEADING:3; TRAILING:3; MAXINFO:targetLength:0.999; and MINLEN:36).

For mapping, a fragment-based strategy implemented in TADbit was used, which was similar to previously published protocols[86]. Briefly, each side of the sequenced read was mapped in full length to the reference genome (mm10, Dec 2011 GRCm38). After this step, if a read was not uniquely mapped, we assumed that the read was chimeric, owing to ligation of several DNA fragments. We next searched for ligation sites, discarding those reads in which no ligation site was found. The remaining reads were split as often as ligation sites were found. Individual split read fragments were then mapped independently. These steps were repeated for each read in the input FASTQ files. Multiple fragments from a single uniquely mapped read resulted in a number of contacts identical to the number of possible pairs between the fragments. For example, if a single read was mapped through three fragments, a total of three contacts (all-versus-all) was represented in the final contact matrix. We used the TADbit filtering module to remove noninformative contacts and to create contact matrices. The different categories of filtered reads applied were:

1. Self-circle: reads coming from a single restriction enzyme (RE) fragment and pointing to the outside.
2. Dangling end: reads coming from a single RE fragment and pointing to the inside.
3. Error: reads coming from a single RE fragment and pointing in the same direction.
4. Extra dangling end: reads coming from different RE fragments but that were sufficiently close and point to the inside; the distance threshold used was left to 500 bp (default), which was between percentiles 95 and 99 of average fragment lengths.
5. Duplicated: the combination of the start positions and directions of the reads was repeated, thus suggesting a PCR artifact; this filter removed only extra copies of the original pair.
6. Random breaks: the start position of one of the reads was too far from RE cutting site, possibly because of noncanonical enzymatic activity or random physical breaks; the threshold was set to 750 bp (default), >percentile 99.9.

From the resulting contact matrices, low-quality bins (those presenting low contact numbers) were removed, as implemented in TADbit's 'filter_columns' routine. A single round of ICE normalization[87], also known as 'vanilla' normalization[11], was performed. That is, each cell in the Hi-C matrix was divided by the product of the interactions in its columns and the interactions in its row. Finally, all matrices were corrected to achieve an average content of one interaction per cell.

Identification of subnuclear compartments and topologically associated domains (TADs). To segment the genome into A/B compartments, normalized Hi-C matrices at 100-kb resolution were corrected for decay as previously described, by grouping diagonals when the signal-to-noise ratio was below 0.05 (ref. [11]). Corrected matrices were then split into chromosomal matrices and transformed into correlation matrices by using the Pearson product-moment correlation. The first component of a PCA (PC1) on each of these matrices was used as a quantitative measure of compartmentalization, and H3K4me2 ChIPmentation data were used to assign negative and positive PC1 categories to the correct compartments. If necessary, the sign of the PC1 (which was randomly assigned) was inverted so that positive PC1 values corresponded to A-compartment regions, and negative values corresponded to the B compartment.

Normalized contact matrices at 50-kb resolution were used to define TADs, through a previously described method with default parameters[43,54]. First, for each bin, an insulation index was obtained on the basis of the number of contacts between bins on each side of a given bin. Differences in the insulation index between both sides of the bin were computed, and borders were called, searching for minima within the insulation index. The insulation score of each border was determined as previously described[43], by using the difference in the delta vector between the local maximum to the left and the local minimum to the right of the boundary bin. This procedure resulted in a set of borders for each time point and replicate. To obtain a set of consensus borders along the time course, we proceeded in two steps: (i) merging borders of replicates and overlapping merged borders (that is, for each pair of replicates, we expanded the borders one bin on each side and kept only those borders present in both replicates as merged borders) and (ii) further expanding two extra bins (100 kb) on each side and determining the overlap to obtain a consensus set of borders common to any pair of time points.

Domain scores were obtained by averaging cells over parts of the Hi-C matrix. In nature, this metric is sensitive to outlier cells with many counts and is fairly insensitive to missing data. For this analysis (and for the metaloop analysis below), we thus used a more stringent strategy to remove low-coverage bins by fitting a logistic function to the distribution of the sum of interactions in each bin:

$$f(x) = \frac{N}{1 + e^{-a(x-b)}} + c$$

where f is the logistic function optimized by the variables a, b and c. N is the number of bins in the matrix, and x is the number of interactions in a given bin. This fit was implemented by weighting bins with higher values of interactions, as we considered bins with lower count artifacts. We set the weight function as being dependent on the bin index, in the context of bins sorted by their sum of interactions:

$$W_i = \frac{\log(i)}{\log(N)}$$

with i representing the index of the bin and W representing the weight applied to the fitting. After the logistic function was fitted, we used it to define a threshold. We removed bins with fewer counts than x when $f(x)$ was equal to zero. The resulting filtered matrices were ICE normalized (1 round, described above). Finally, domain scores were calculated with matrices binned at 50 kb by dividing the sum of intra-TAD contacts by the sum of all contacts involving the TAD.

Expression variability explained by TADs. To estimate expression variability, we fitted a hierarchical regression model per gene expression values for each time point, including three levels of organization: the gene itself, the local neighborhood (the 50-kb TSS bin) and the TAD. We used the variance associated with each level and the total variance of the model to assess the proportion of variability explained by each factor. To test whether topology played a role beyond the linear proximity of genes, we repeated the estimation, replacing actual TADs by a fixed segmentation of the genome in domains with the same size as the average TAD (i.e., 'fake TADs' constructed by placing a border at fixed distances corresponding to the average size of TADs). Model estimation was performed with the lme4 R package.

Inter- and intracompartment strength measurements. We followed a previously reported strategy to measure overall interaction strengths within and between A and B compartments[63]. Briefly, we based our analysis on the 100-kb bins showing the most extreme PC1 values, discretizing them by percentiles and taking the bottom 20% as the B compartment and the top 20% as the A compartment. We classified each bin in the genome according to PC1 percentiles and gathered contacts between each category, computing the $\log_2$ enrichment over the expected counts by distance decay. Finally, we summarized each type of interaction (A–A, B–B and A–B/B–A) by taking the median values of the $\log_2$ contact enrichment.

Meta-analysis of borders, loops and interactions among TF-binding sites. To assess whether particular parts of the Hi-C interaction matrices had common structural features, we performed meta-analyses by merging individual submatrices into an average metamatrix in a manner similar to a previously published method[51]. Three types of meta-analysis were performed. First, we studied TAD border dynamics at 50-kb resolution by extracting interaction counts 1.25 Mb up- and downstream of the TAD border. Extracted matrices were averaged for each group of clustered TAD borders, including those whose insulation scores increased, decreased or did not change during reprogramming. Second, using 5-kb resolution contact maps, we investigated the dynamics of a previously identified set of chromatin loops in primary B cells and PSCs[49] by extracting interaction counts 50 kb up- and downstream of loop-anchor regions. Metaloop matrices were then calculated by averaging individually subtracted loop matrices into a single one per group. Third, we studied whether two regions bound by a given TF were likely to find each other more frequently than expected within a genomic distance ranging from 2 to 10 Mb. All submatrices at 5-kb resolution between pairs of TF-binding sites and 50 kb up- and downstream of a TF peak were extracted and averaged into a single metamatrix. For OCT4, NANOG and SOX2 meta-analyses, we used those TF-binding sites that overlapped with an ATAC–seq peak (described above) at the D2 stage. All meta-analyses were performed with the observed/expected Hi-C matrices, which were filtered, ICE normalized and corrected for decay. For visualization proposes, the resulting meta-analysis matrices were smoothed with a Gaussian filter of sigma = 1.

Virtual 4C analysis (v4C) and promoter–superenhancer linking. For the generation of v4C profiles, we first chose a bait region (for example, *Sox2*) and (optionally) a window size around the bait (with the final viewpoint window centered on the bait). We then extracted the observed Hi-C matrix at 5-kb resolution for that specific region. Rows overlapping the bait were divided into subsets, after which we summed all bait rows to obtain the number of observed contacts per bin (column). Aiming to reduce noise, we performed a moving-average smoothing (5 bins) to obtain v4C profiles. Count numbers per bin were normalized for differences in sequencing depth between time points. For visualization purposes, we removed all data overlapping the bait extended with one bin per side.

We took advantage of this approach to link promoters to SEs. For each SE, we set a window of 2 Mb around the SE bait and extracted the corresponding Hi-C matrix at 5-kb resolution, removing low-count and/or low-mappability bins. Using the full interchromosomal matrix, we computed an expected Hi-C matrix, averaging all pairs of loci at the same distance per chromosome. After merging the two replicates, we generated virtual 4C profiles for each SE with the observed and corresponding expected number of counts. These profiles allowed us to rank nearby promoters according to their contact enrichment (observed/expected), designating the two highest-ranking genes as putative SE targets. Using this method, we detected a larger number of genes associated with the superenhancer subset (372 versus the 210 assigned by GREAT), which included half of the genes identified with GREAT. The GO analyses and gene expression analyses on the GREAT gene set or this extended target gene set were similar, although the Hi-C-based gene set showed stronger enrichment in GO terms associated with embryonic development. Analyses on the Hi-C-based gene set were used in Fig. 1.

Integration of B cell replication-timing data. We partitioned the genome into 100-Kb bins, labeling the compartment (A, B or 0) for each time point and biological replicate. Then we identified the bins with more than one compartment type (i.e., switching bins). For each bin, the residence time in A or B was the number of consecutive time points in A or B before a switch. The results presented are the grand sum per compartment, residence time and biological replicate over all switching bins.

Statistics and reproducibility. In situ Hi-C data throughout the paper were generated by analysis of two independent B-to-iPS replicate reprogramming experiments. Representative data are shown only if results were similar for both independent biological-replicate experiments. All box plots depict the first and third quartiles as the lower and upper bounds of the box, with a thicker band inside the box showing the median value and whiskers representing 1.5× the interquartile range. Wilcoxon rank-sum tests were performed with the wilcox.test() function in R in a two-sided manner. *t* tests were performed with the t.test() function in R in an unpaired and two-sided fashion with (n – 2) degrees of freedom.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability. All data generated have been deposited in the Gene Expression Omnibus (GEO) database under accession code GSE96611. Accession codes of published datasets used in this study are as follows: CTCF ChIP–seq in pre-B cells, Sequence Read Archive SRR397837 (ref. [88]); CTCF ChIP–seq in induced PSCs, GEO GSE76478 (ref. [49]); OCT4 and NANOG ChIP–seq in PSCs, GEO GSE44286 (ref. [39]); KLF4 ChIP–seq in PSCs, GEO GSE11431 (ref. [89]); C/EBPα and PU.1 ChIP–seq in Bα cells, GEO GSE71215 (ref. [32]); EBF1 V5-ChIP–seq in pro-B cells, GEO GSE53595 (ref. [90]). CH12 Repli-chip data were obtained from ENCODE Biosample ENCBS789HDO.

References

69. Carey, B. W., Markoulaki, S., Beard, C., Hanna, J. & Jaenisch, R. Single-gene transgenic mouse strains for reprogramming adult somatic cells. *Nat. Methods* **7**, 56–59 (2010).
70. Boiani, M., Eckardt, S., Schöler, H. R. & McLaughlin, K. J. Oct4 distribution and level in mouse clones: consequences for pluripotency. *Genes Dev.* **16**, 1209–1219 (2002).
71. van Oevelen, C. et al. C/EBPα activates pre-existing and de novo macrophage enhancers during induced Pre-B cell transdifferentiation and myelopoiesis. Stem. *Cell Rep.* **5**, 232–247 (2015).
72. Stadhouders, R. et al. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat. Protoc.* **8**, 509–524 (2013).
73. Brouwer, R. W., van den Hout, M. C., van IJcken, W. F., Soler, E. & Stadhouders, R. Unbiased interrogation of 3D genome topology using chromosome conformation capture coupled to high-throughput sequencing (4C-Seq). *Methods Mol. Biol.* **1507**, 199–220 (2017).
74. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
75. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
76. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
77. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
78. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
79. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
80. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
81. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
82. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
83. Klein, F. A. et al. FourCSeq: analysis of 4C sequencing data. *Bioinformatics* **31**, 3085–3091 (2015).
84. Serra, F. et al. Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS Comput. Biol.* **13**, e1005665 (2017).
85. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
86. Ay, F. et al. Identifying multi-locus chromatin contacts in human cells using tethered multiple 3C. *BMC Genomics* **16**, 121 (2015).
87. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
88. Ribeiro de Almeida, C. et al. The DNA-binding protein CTCF limits proximal Vκ recombination and restricts κ enhancer interactions to the immunoglobulin κ light chain locus. *Immunity* **35**, 501–513 (2011).
89. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
90. Schwickert, T. A. et al. Stage-specific control of early B cell development by the transcription factor Ikaros. *Nat. Immunol.* **15**, 283–293 (2014).

nature research

Corresponding author(s): Ralph Stadhouders, Marc Marti-Renom, Thomas Graf

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.	No statistical methods were used to determine sample sizes.

2. Data exclusions

Describe any data exclusions.	TAD borders not called in both independent biological replicates were excluded in all subsequent analyses. All read and bin filtering strategies used for Hi-C data analysis are described in detail in the Online Methods.

3. Replication

Describe whether the experimental findings were reliably reproduced.	No attempts to reproduce results failed.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.	Randomization is not relevant to this study because no comparisons between experimental groups were made.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.	Blinding was not relevant to this study because all metrics were derived from absolute quantitative methods without human subjectivity.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	STAR (Dobin et al., 2013) DESeq2 (Love et al., 2014) Bowtie2 (Langmead and Salzberg, 2012) SAMtools (Li et al., 2009) Picard (http://broadinstitute.github.io/picard) HOMER (Heinz et al., 2010) TADbit (Serra et al., 2017) fourCseq (Klein et al., 2015) GREAT (McLean et al., 2010) BEDTools (Quinlan and Hall, 2010) FastQC (http://www.bioinformatics.babraham.ac.uk /projects/fastqc/) Trimmomatic (Bolger et al., 2014) Mfuzz R package (2.26.0) fourCseq R package (Klein et al. 2015) MACS2 (Zhang et al. 2008) MSigDB (Liberzon et al. 2011) ICE (Imakaev et al. 2012) wilcox.test() R function t.test() R function

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	anti-H3K4Me2: Abcam, ab32356 (validated in Egelhofer et al. Nat Struct Mol Biol 2011, see http://compbio.med.harvard.edu/antibodies/) anti-CTCF: Millipore, 07-729 (validated by ENCODE, see https://www.encodeproject.org/antibody-characterizations/890eca82-62f8-406c-868a-c94a5a3d748e/@@download/attachment/human_CTCF_07-729_validation_Snyder.pdf)

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	The EG14TG2a embryonic stem cell line was obtained from ATCC (CRL-1821).
b. Describe the method of cell line authentication used.	The cell line was purchased as a validated cell line from ATCC. Sequencing data was compared to published data with the cell line to verify identity.
c. Report whether the cell lines were tested for mycoplasma contamination.	Cell line was tested for mycoplasma contamination upon receipt and periodically (once a month) thereafter.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No commonly misidentified cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	We used a previously described (Di Stefano et al. Nature Cell Biology 2016) cross between reprogrammable mice (Carey et al. Nature Methods 2010) and Oct4-GFP reporter mice (Boiani et al. Genes & Development 2002). Mice were kept on a C57BL/6 background and housed in standard cages under 12h light–dark cycles and fed ad libitum (standard chow). Developing B cells were obtained from bone marrow of female mice (8-16 weeks old) for reprogramming experiments in vitro.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	Study did not involve human research participants.