LETTER

doi:10.1038/nature12420

The pluripotent genome in three dimensions is shaped around pluripotency factors

Elzo de Wit[1]*, Britta A. M. Bouwman[1]*, Yun Zhu[1], Petra Klous[1], Erik Splinter[1], Marjon J. A. M. Verstegen[1], Peter H. L. Krijger[1], Nicola Festuccia[2], Elphège P. Nora[3], Maaike Welling[1], Edith Heard[3], Niels Geijsen[1,4], Raymond A. Poot[5], Ian Chambers[2] & Wouter de Laat[1]

It is becoming increasingly clear that the shape of the genome importantly influences transcription regulation. Pluripotent stem cells such as embryonic stem cells were recently shown to organize their chromosomes into topological domains that are largely invariant between cell types[1,2]. Here we combine chromatin conformation capture technologies with chromatin factor binding data to demonstrate that inactive chromatin is unusually disorganized in pluripotent stem-cell nuclei. We show that gene promoters engage in contacts between topological domains in a largely tissue-independent manner, whereas enhancers have a more tissue-restricted interaction profile. Notably, genomic clusters of pluripotency factor binding sites find each other very efficiently, in a manner that is strictly pluripotent-stem-cell-specific, dependent on the presence of Oct4 and Nanog protein and inducible after artificial recruitment of Nanog to a selected chromosomal site. We conclude that pluripotent stem cells have a unique higher-order genome structure shaped by pluripotency factors. We speculate that this interactome enhances the robustness of the pluripotent state.

In recent years, several technological advances have made it possible to delineate the three-dimensional shape of the genome[3]. Spatial organization of DNA has been recognized as an additional regulatory layer of chromatin, important for gene regulation and transcriptional competence[4,5]. In somatic cells active and inactive chromosomal regions are spatially segregated[6,7]. Recently, the genome was further shown to be subdivided into evolutionarily conserved topological domains[1,2].

4C (chromosome conformation capture combined with sequencing) is a genome-scale variant of the 3C technology[8], which examines the spatial organization of DNA and measures the contact frequencies of a chosen genomic site, or 'viewpoint', with the rest of the genome. To assess chromosome topology in mouse E14 embryonic stem cells (IB10), we generated high-resolution contact maps using 4C sequencing[9] for a series of individual sites representative of different chromosomal regions on various chromosomes (Supplementary Fig. 1 and Methods). All 4C experiments show the typical result of a chromosome conformation capture experiment, with the bulk of the signal close to the viewpoint, intrachromosomal captures outnumbering interchromosomal captures, and clustering of captures at distal sites[6,7] (Supplementary Fig. 2 and Supplementary Table 1). To identify significant intra- and interchromosomal contacts, we used a windowing approach in combination with a false discovery rate (FDR) analysis that determines significant clustering of independently captured sequences (FDR, $\alpha = 0.01$; ref. 10). Contacts in this case can mean either direct interactions between the chromatin of chromosomal regions or indirect contacts via intermediate protein complexes. 3D-DNA fluorescence *in situ* hybridization (FISH) experiments validated the 4C results (Supplementary Fig. 3).

Different from what is observed in somatic cells, in embryonic stem cells we find that transcriptionally inactive regions form low numbers of specific long-range contacts (Fig. 1a, b). This is not due to their inability to reach over large distances, but instead to a more random organization of their long-range captures (Supplementary Fig. 4), suggesting that inactive chromatin is spatially less organized in pluripotent nuclei. We confirmed these results in an independent, 129/Cast, embryonic stem (ES) cell line[11] (Supplementary Fig. 5a). We furthermore show that this is not an intrinsic feature of the selected regions as they do engage in many long-range contacts in astrocytes (Fig. 1c and Supplementary Fig. 6). For example, the chemoreceptor *Tas2r110* gene, part of a cluster of taste receptors that is specifically expressed in taste buds, engages in only three contacts in ES cells but shows 34 specific contacts in astrocytes (Fig. 1d).

We assessed whether the lack of long-range contacts is a global feature of ES cells, by analysing a recently published mouse ES cell Hi-C data set[1]. 'Virtual 4C' contact profiles extracted from the Hi-C data set (see Methods for details) correlate strongly to our 'true 4C' profiles (Supplementary Fig. 7), emphasizing the high level of agreement between the data sets. The Hi-C data confirm on a global scale that inactive and active chromatin differ in their propensity to form specific long-range contacts in ES cells. Similar to our 4C data, this difference is abolished in differentiated cells (cortex), where both active and inactive chromatin engage equally well in specific long-range contacts (Fig. 1e, f and Supplementary Fig. 8).

We next asked whether chromosomal organization is reversed during cellular reprogramming. 129/Cast neural precursor cells (NPCs)[9] were transduced with a lentivirus containing a multicistronic transcript encoding *Oct4* (also known as *Pou5f1*), *Klf4*, *Sox2* and *c-Myc*, to generate induced pluripotent stem (iPS) cells. Quantitative PCR (qPCR) expression analysis of several marker genes confirmed reprogramming (Supplementary Fig. 9). A reactivated gene (*Nanog*) gains contacts during iPS cell reprogramming, whereas a resilenced gene (*Ptprz1*) loses all but two contacts (Supplementary Fig. 10), demonstrating that cellular reprogramming is accompanied by the reemergence of a pluripotency-specific spatial organization of the genome.

A closer inspection of the intrachromosomal contacts made by the *Nanog* gene revealed another aspect of the 3D pluripotent genome; *Nanog* was found to interact with many genes that are known to have an important role in maintenance of ES cell pluripotency, including *Rybp*, *Ezh2*, *Tcf3* and *Smarcad1*. The ES-cell-specific nature of these contacts becomes obvious from a comparison between the ES cell and NPC contact profiles of the *Nanog* gene. Most of the ES-cell-specific interacting regions have a high density of binding sites for the pluripotency factors Oct4, Sox2 and Nanog (Fig. 2a, b). Importantly, we also find such preferential associations when we apply 4C to the *Sox2* enhancer (Supplementary Fig. 11). The ES cell Hi-C data[1] show that

[1]Hubrecht Institute-KNAW & University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands. [2]MRC Centre for Regenerative Medicine, Institute for Stem Cell Research, School of Biological Sciences, University of Edinburgh, 5 Little France Drive, Edinburgh EH16 4UU, UK. [3]Mammalian Developmental Epigenetics Group, Institut Curie CNRS UMR3215 INSERM U934, Paris, France. [4]Utrecht University, School of Veterinary Medicine, 3584 CL Utrecht, The Netherlands. [5]Department of Cell Biology, Erasmus Medical Center, Rotterdam 3015 GE, The Netherlands.
*These authors contributed equally to this work.

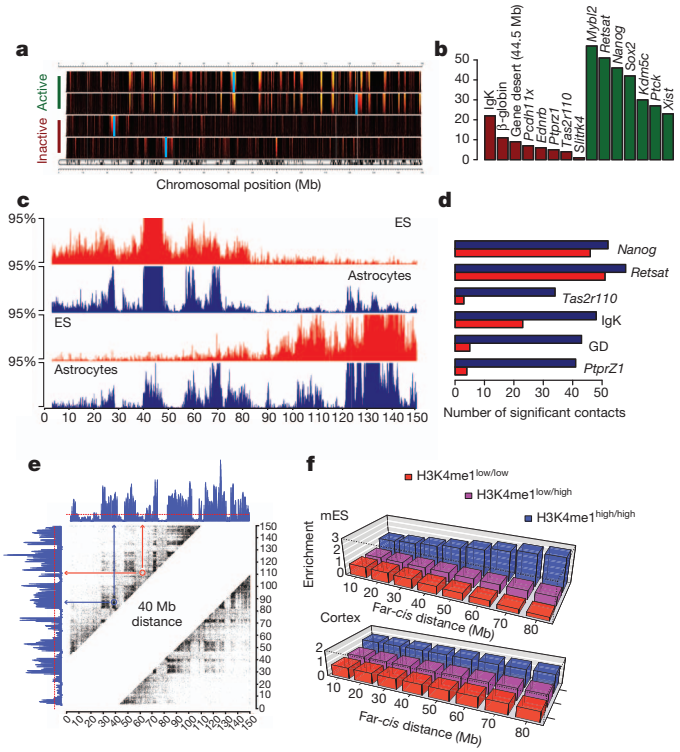

Figure 1 | Inactive regions lack specific long-range interactions in embryonic stem cells. **a**, Domainogram analysis (see Methods) shows 4C profiles of *Retsat*, *Nanog*, *Ptprz1* and gene desert (top to bottom) in ES cells. Plots represent contact profiles of active (dark green; $n = 7$) and inactive (dark red; $n = 8$) viewpoints. Below the domainograms, a map of the chromosomal position of the genes is plotted. **b**, Quantification of the number of significantly contacted regions for different viewpoints in ES cells. Green bars denote viewpoints in active regions, red bars denote viewpoints in inactive regions. **c**, Chromosomal maps show read count distribution for a gene desert (at 44.5 Mb) and for *Tas2r110*, for ES cells (red) and astrocytes (blue). The 4C signal is calculated using a sliding window average (running mean) of the read counts (window size is 51). The vertical axis is maximized at the ninety-fifth percentile. **d**, Quantification of the number of far-*cis* regions that are significantly contacted by a given viewpoint in ES cells (red) or astrocytes (blue). GD, gene desert. **e**, A pairwise contact matrix was generated to calculate disorganization scores from the Hi-C data (see Methods). Chromosome 6 was segmented into regions with high density of H3K4me1 and low density of H3K4me1, as a proxy for active and inactive chromatin. The pairwise contact matrix was subdivided into contacts between two regions of high H3K4me1 density ($H3K4me1^{high/high}$) or low H3K4me1 density ($H3K4me1^{low/low}$) or between a region with low H3K4me1 density and a region with high H3K4me1 density ($H3K4me1^{low/high}$). **f**, From the distribution of H3K4me1 high and low regions, we calculated an expected distribution of long-range contacts, under the null hypothesis that there is no difference between active and inactive regions with respect to their long-range contacts. An enrichment score is calculated by dividing the observed scores by these expected values.

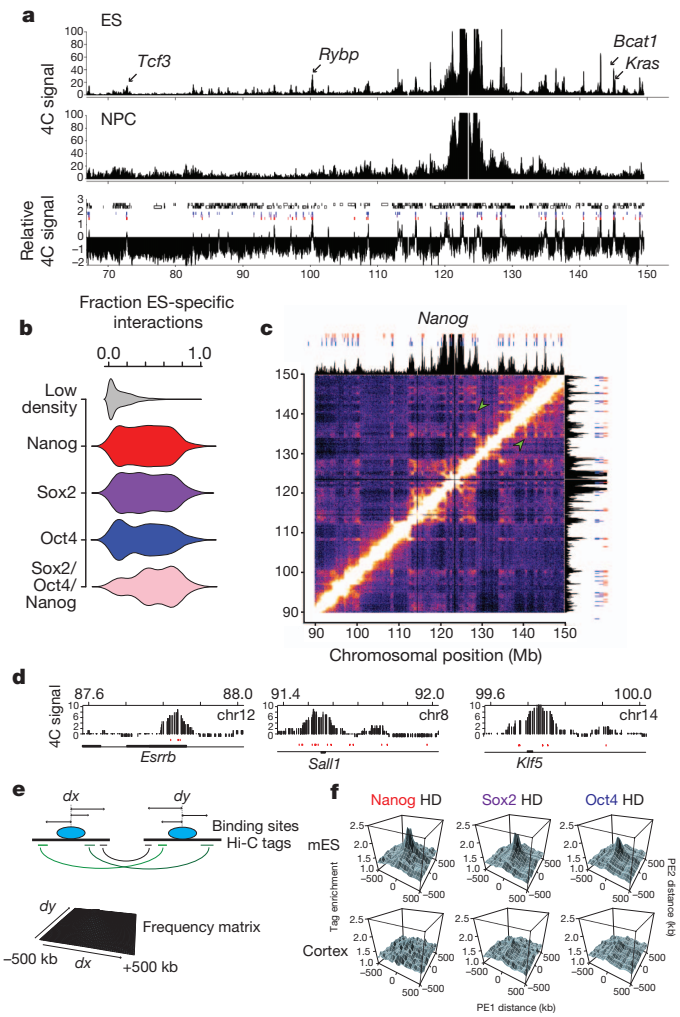

Figure 2 | Expressed *Nanog* gene shows preferential interaction with other pluripotency genes. **a**, Chromosomal map of 4C signal for the *Nanog* gene in ES cells and NPCs. Representative 4C data ($n = 6$ (ES cell) and $n = 2$ (NPC) biological replicates) is normalized to reads per million and plotted as a running mean with a window of 31. Bottom panel shows the ES cell to NPC ratio. Red, purple and blue rectangles denote the windows with a high density of Nanog-, Sox2- and Oct4-binding sites, respectively. High-density is defined as >5 sites per 100 kb. **b**, Violin plots show quantification of ES cell/NPC ratios for regions with a high density of binding sites for Nanog, Sox2, Oct4 and all three combined. **c**, Combined Hi-C–4C plot for the telomeric region of chromosome 6, shows a normalized Hi-C contact matrix (see Methods) with the 4C data for *Nanog* superimposed. Red, purple and blue rectangles show the high-density regions, as in **a**. Green arrowheads point to Nanog high-density (HD) Hi-C interactions other than with the *Nanog* enhancer. **d**, Examples of interchromosomal contacts made by *Nanog* with pluripotency genes *Esrrb*[27], *Sall1* (ref. 28) and *Klf5* (ref. 29). See methods for the definition of the 4C enrichment score. Nanog-, Sox2- and Oct4-binding sites are again highlighted with red, purple and blue rectangles. **e**, Schematic depiction of paired-end spatial chromatin analysis (PE-SCAn). Hi-C di-tags are sequentially aligned to ChIP-seq binding sites. From the total set of distances *dx* and *dy*, a normalized two-dimensional frequency matrix is calculated (see Methods). Hi-C pairs within 5 Mb were excluded to focus the analysis on contacts between, rather than within topological domains. **f**, PE-SCAn plots show the alignment of intrachromosomal Hi-C data to high-density clusters of pluripotency factors (≥5 sites in 50 kb, Nanog: $n = 423$, Sox2: $n = 607$, Oct4: $n = 1025$). Top row shows alignment of ES cell Hi-C data, bottom row shows alignment of cortex Hi-C data.

Nanog-contacting regions also form preferential contacts among each other (Fig. 2c). Among the interchromosomal contacts made by the *Nanog* gene are again a large number of pluripotency related genes: *Mybl2*, *Dppa5*, *Rex1* (also known as *Zfp42*), *Zfp281*, *Lefty1*, *Lin28a*, *Esrrb*, *Klf5*, *Sall1*, *Cbx5* and *Cbx7* (Fig. 2d and Supplementary Fig. 12; see Supplementary Table 2 for a full list of contacted regions). The contacts of *Nanog* with *Esrrb* and *Zfp281* were verified by 3D-DNA FISH (Supplementary Fig. 12). GREAT analysis[12] of these interchromosomal contacts reveals strong enrichment for genes involved in pluripotency and early embryogenesis, which is not observed for unrelated viewpoints or in other tissues (Supplementary Table 3). This suggests that pluripotency genes prefer to cluster with other pluripotency-specific genes.

We designed a computational strategy, paired-end spatial chromatin analysis (PE-SCAn) (Fig. 2e), which combines chromatin factor binding data and Hi-C data to analyse, on a global scale, whether given genomic sites (bound by a protein of interest) in different topological domains have a preference to interact among each other. PE-SCAn shows that individual Nanog-, Sox2- or Oct4-binding sites have little

preference to contact each other over such large chromosomal distances (Supplementary Fig. 13a). However, clusters of Nanog-, Oct4- or Sox- binding sites (5 or more per 50 kilobases (kb)) do show a strong preference to interact with each other in ES cells (Fig. 2f). When we circularly permute the positions of the Nanog, Sox2 or Oct4 clusters, this preference is not observed, confirming that these interactions are specific (data not shown). Moreover, these contacts are tissue-specific as they are absent in the cortex (Fig. 2f).

We also used PE-SCAn to investigate the contribution of other factors to the shape of the pluripotent genome. Although CTCF and cohesin have both been implicated in higher-order chromatin folding[13], CTCF has been suggested to predominantly form chromatin loops over shorter distances[14,15]. Indeed, we find that CTCF, but also cohesin-binding sites, contribute little to chromosome folding over larger distances (Fig. 3a and Supplementary Fig. 13b). Recent chromosome architecture experiments have revealed a central role for promoters in chromosome topology[16]. PE-SCAn for histone H3 trimethyl Lys 4 (H3K4me3) confirmed that active transcriptional start sites are engaged in specific long-range contacts (Fig. 3b). However, their contribution is largely tissue-invariant, because promoters marked by H3K4me3 in either ES cells or cortex also find each other equally well in the corresponding tissue (Fig. 3b). This is different for active enhancer sites (H3K27ac[17]), which contribute to genome topology in a more tissue-restricted manner (Fig. 3b). Pluripotency factors, but also cohesin, often bind to enhancer sequences. For Oct4, Sox2, Nanog and cohesin we find that 41%, 38%, 35% and 27%, respectively, of binding sites overlap with active enhancer sites. All intersected enhancer sites show an equal preference for homotypic contacts as the unselected enhancers (Fig. 3c). Importantly, the preferred contacts among *Nanog* enhancers were not dependent on cohesin (Fig. 3d). Finally, we assessed chromosomal contacts among, respectively, enhancer and cohesin clusters (5 or more per 50 kb). We found that they have no advantage over isolated sites to interact with each other, and that their contact preference is not as pronounced as seen for clusters of pluripotency-factor-binding sites (Supplementary Fig. 13b).

To investigate whether this pluripotency-specific genome configuration is dependent on pluripotency factors, we used ZHBTc4 (ref. 18) and RCNβH (ref. 19) ES cell lines, which allow the acute depletion of Oct4 and Nanog protein, respectively (Fig. 4a, b and Supplementary Fig. 14a, b). After Oct4 or Nanog protein removal, the overall chromosome topology is largely unaffected (Fig. 4c and Supplementary Fig. 14c). However, a close comparison between factor-depleted and wild-type contact profiles reveals a decrease in contact frequencies specifically at clusters where pluripotency factors normally bind (Fig. 4d, e). Quantification confirms that the regions with reduced contact frequency after removal of Oct4 or Nanog protein are those with a high density of cognate binding sites and not, for example, regions with a high density of CTCF-binding sites (Fig. 4f, see Methods for details). Of note, partial loss of Nanog by short interfering RNA (siRNA)-mediated knockdown (78%) has no effect (Supplementary Fig. 15), indicating that full knockout of Nanog protein is required to affect chromosome topology in ES cells.

To test further whether pluripotency factor binding has a direct role in this pluripotent-stem-cell-specific genome configuration, we made use of a C56Bl/6–129S1/SvImJ ES cell line with a 256× lacO repeat cassette integrated into the C56Bl/6 *Nfix* allele on chromosome 8 (Fig. 4g and Supplementary Fig. 16). We targeted green fluorescent protein (GFP)–LacR–Nanog fusion proteins to these lacO repeats and performed allele-specific 4C (ref. 20) to simultaneously analyse the contact spectra of the targeted C56Bl/6 and the non-targeted 129S1/SvImJ allele. Again, the overall chromosome topology for both alleles was highly similar, but several new specific contacts were found for the C56Bl/6 allele. Notably, these contacts coincide with high-density Nanog-binding sites around the pluripotency genes *Sall1* and *Klf2* and the *Irx* cluster of developmental regulators. Circular permutation

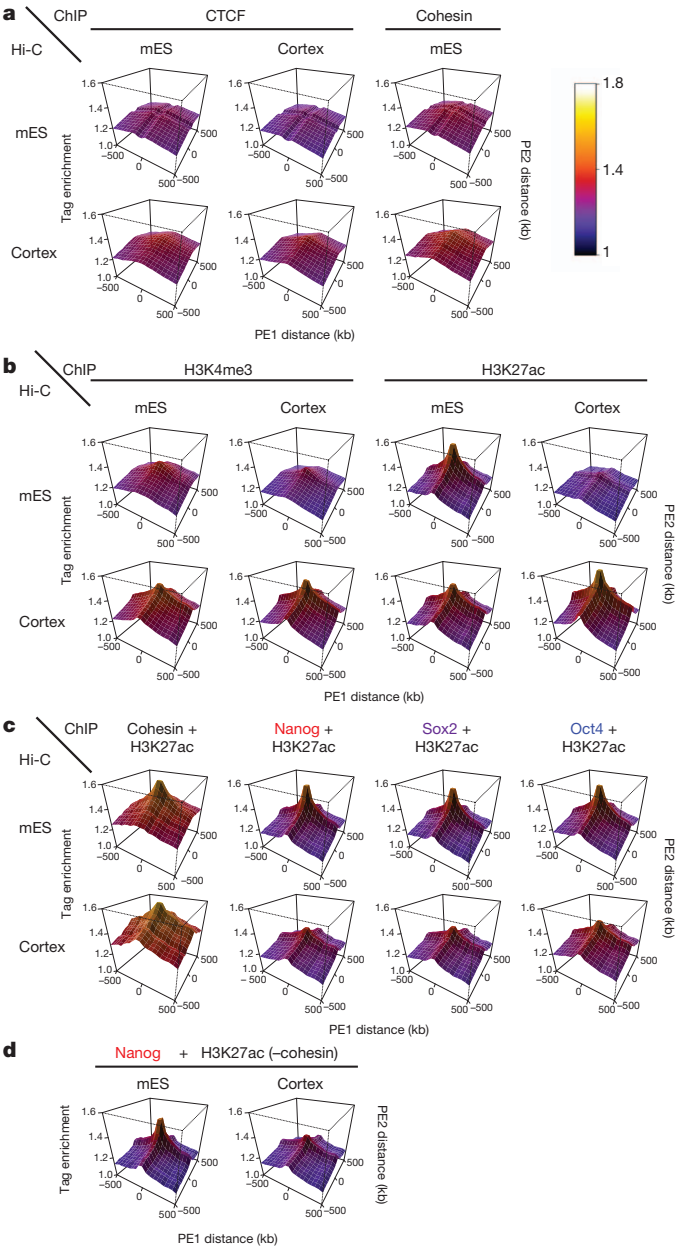

Figure 3 | Spatial interactome of chromatin factors is revealed by PE-SCAn. **a**–**d**, PE-SCAn plots for various chromatin factors. Plots are the same as in Fig. 2e, but with a different scale on the vertical axis. Note that the height of the data is colour-coded according to the colour bar shown in **a**. Top row represents mouse ES cell Hi-C data, bottom row represents cortex Hi-C data. **a**, PE-SCAn plots for known looping factors CTCF and cohesin (Smc1). **b**, PE-SCAn plots for promoter (H3K4me3) and active enhancer (H3K27ac) marks in mouse ES cells and cortex. **c**, PE-SCAn plots for active enhancer sites co-bound by cohesin (Smc1), Nanog, Sox2 or Oct4. **d**, PE-SCAn plots for genomic sites with active enhancer marks and Nanog binding, but which are devoid of cohesin. Left, mouse ES cell Hi-C data; right, cortex Hi-C data.

of the positions of the high-density Nanog clusters showed that increased contact frequency was significantly enriched at these sites ($P < 0.001$), demonstrating that Nanog has a direct role in bringing together distantly located clusters of Nanog-binding sites.

Our data show that pluripotency transcription factors shape the pluripotent genome via spatial intra- and interchromosomal gathering of high-density binding sites. It has been suggested previously that transcription factors position tissue-specific and co-regulated genes in somatic cells[21–23]. However, in contrast to previous studies, we validated this concept by comparing genome-wide contact maps

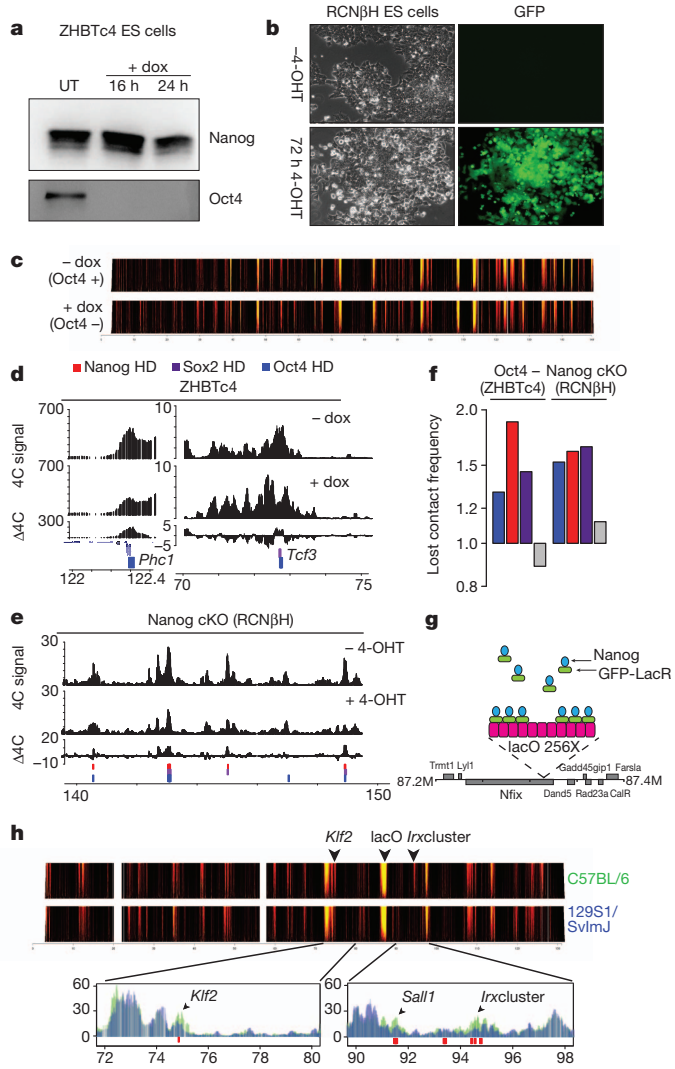

Figure 4 | Pluripotency factors influence the 3D organization of the genome. **a**, Immunoblot analysis before and after treatment of ZHBTc4 cells with 1 μg ml^{-1} doxycycline for the indicated times. Oct4 and Nanog proteins were detected using anti-Oct4 and anti-Nanog antibodies. UT, untreated. **b**, Morphology and GFP expression of RCNβH cells before and 72 h after treatment with 1 μM tamoxifen (4-OHT, 4-hydroxytamoxifen). **c**, 4C domainograms for Oct4-positive (−dox) and Oct4-negative (+dox) ZHBTc4 cells ($n = 2$ biological replicates) show that overall chromosome topology is maintained in Oct4-depleted cells. **d**, Zoomed-in regions show 4C signal (reads per million) for Oct4-positive (top) and Oct4-negative (middle) cells. Bottom, the difference (Δ) between the 4C signal of the Oct4-positive and -negative cells. Red, purple and blue rectangles show high-density Nanog, Sox2 and Oct4 regions, respectively. **e**, Same as **d** but for Nanog conditional knockout (cKO) ($n = 1$). Note that gene information is left out at this scale. **f**, Chromosome-wide analysis of differential 4C interactions for the *Nanog* enhancer viewpoint. Loss of 4C contact frequency is defined as a lower 4C signal in the knockout compared to the non-depleted reference cell line. Loss of contact frequency is determined for high-density Oct4, Nanog, Sox2 and CTCF (control) clusters for the Oct4-ablated and Nanog-conditional knockout cell lines and the enrichment over the background is calculated (see Methods for details). **g**, Schematic drawing depicting the integration site of the lacO repeat cassette in the C57Bl/6 allele of the *Nfix* locus and the targeting of GFP –LacR–Nanog fusion proteins. **h**, Domainograms showing allele-specific 4C (ref. 20) for the C56Bl/6 (containing the lacO cassette) and the 129 allele present in the hybrid ES cells ($n = 1$). Bottom panels show zoomed-in 4C profiles (C57Bl/6 green, 129 blue) for example differential regions. Red rectangles indicate high-density Nanog clusters (6 sites per 100 kb).

generated in wild-type and transcription factor knockout cells and by studying an artificially induced cluster of binding sites. Our observation that targeting or removing a given factor to or from the genome only changes specific contacts while the overall folding of chromosomes remains intact is in accordance with a recently proposed model for chromosome topology. This 'dog-on-a-lead' model predicts that chromosomes are dominant over their individual segments (genes, domains, enhancers) in dictating the overall shape of the genome, but that segments can search the nuclear subvolumes they occupy for preferred contact partners[24]. There is accumulating evidence that stochastically determined nuclear environments can influence the transcriptional output of resident genes, leading to cell-to-cell variability[25,26]. We propose that the observed spatial clustering of pluripotency factor binding sites in pluripotent stem cells enhances the transcription efficiency of nearby genes and thereby contributes to the robustness of the pluripotent state.

METHODS SUMMARY

4C sequencing and mapping. 4C sequencing was performed as previously described[9]. We used HindIII as the first restriction enzyme to generate the 3C template, which was further trimmed with DpnII. Sequencing was performed on Illumina GAII and HiSeq 2000 over multiple runs. Raw sequencing data and mapped wig files can be found under Gene Expression Omnibus (GEO) accession GSE37275.

PE-SCAn. To assess which factors are associated with genome organization, we aligned ChIP data to the Hi-C data. For this, intrachromosomal captures that are >5 Mb from each other are aligned to transcription-factor -binding sites. Only captures where both di-tags mapped within 500 kb of a ChIP peak were considered in the analysis. As a result we get for every pair of ChIP peaks on the same chromosome a set of two distances (dx, dy), to all the Hi-C di-tags that are found within 500 kb of these peaks. From the distribution of dx and dy a frequency matrix is calculated with a bin size of 50 kb, which is normalized by dividing by a randomized data set that is calculated by aligning the Hi-C data to a circularly permuted ChIP-seq data set, that is, the ChIP peaks are linearly shifted 10 Mb along the chromosome. In this manner the structure of the Hi-C data is preserved; the structure of the ChIP data is also preserved, only shifted.

Depletion of pluripotency factors. RCNβH cells were treated with tamoxifen and replated the next day. Seventy-two hours after initial tamoxifen treatment, cells were collected for 4C template preparation and analyses. ZHBTc4 cells were collected after 48 h of treatment with 1 μg ml^{-1} doxycycline.

Full Methods and any associated references are available in the online version of the paper.

Received 4 April 2012; accepted 26 June 2013.
Published online 24 July 2013.

1. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
2. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).
3. de Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* **26,** 11–24 (2012).
4. van Steensel, B. & Dekker, J. Genomics tools for unraveling chromosome architecture. *Nature Biotechnol.* **28,** 1089–1095 (2010).
5. Splinter, E. & de Laat, W. The complex transcription regulatory landscape of our genome: control in three dimensions. *EMBO J.* **30,** 4345–4355 (2011).
6. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genet.* **38,** 1348–1354 (2006).
7. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).
8. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295,** 1306–1311 (2002).
9. Splinter, E. *et al.* The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev.* **25,** 1371–1383 (2011).
10. Splinter, E., de Wit, E., van de Werken, H. J., Klous, P. & de Laat, W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* **58,** 221–230 (2012).
11. Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nature Genet.* **30,** 167–174 (2002).
12. McLean, C. Y. *et al.* GREAT improves functional interpretation of *cis*-regulatory regions. *Nature Biotechnol.* **28,** 495–501 (2010).
13. Apostolou, E. *et al.* Genome-wide chromatin interactions of the *Nanog* locus in pluripotency, differentiation, and reprogramming. *Cell Stem Cell* **12,** 699–712 (2013).
14. Handoko, L. *et al.* CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature Genet.* **43,** 630–638 (2011).
15. Lin, Y. C. *et al.* Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nature Immunol.* **13,** 1196–1204 (2012).

16. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148,** 84–98 (2012).
17. Shen, Y. *et al.* A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488,** 116–120 (2012).
18. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genet.* **24,** 372–376 (2000).
19. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450,** 1230–1234 (2007).
20. Holwerda, S. J. *et al.* Allelic exclusion of the immunoglobulin heavy chain locus is independent of its nuclear localization in mature B cells. *Nucleic Acids Res.* http://dx.doi.org/10.1093/nar/gkt491 (7 June 2013).
21. Schoenfelder, S. *et al.* Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature Genet.* **42,** 53–61 (2010).
22. Xu, M. & Cook, P. R. Similar active genes cluster in specialized transcription factories. *J. Cell Biol.* **181,** 615–623 (2008).
23. Dhar, S. S. & Wong-Riley, M. T. Chromosome conformation capture of transcriptional interactions between cytochrome c oxidase genes and genes of glutamatergic synaptic transmission in neurons. *J. Neurochem.* **115,** 676–683 (2010).
24. Krijger, P. H. & de Laat, W. Identical cells with different 3D genomes; cause and consequences? *Curr. Opin. Genet. Dev.* **23,** 191–196 (2013).
25. Noordermeer, D. *et al.* Variegated gene expression caused by cell-specific long-range DNA interactions. *Nature Cell Biol.* **13,** 944–951 (2011).
26. Kind, J. *et al.* Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153,** 178–192 (2013).
27. Feng, B. *et al.* Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nature Cell Biol.* **11,** 197–203 (2009).
28. Karantzali, E. *et al.* Sall1 regulates embryonic stem cell differentiation in association with nanog. *J. Biol. Chem.* **286,** 1037–1045 (2011).
29. Parisi, S. *et al.* Klf5 is involved in self-renewal of mouse embryonic stem cells. *J. Cell Sci.* **121,** 2629–2634 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank C. Vermeulen and S. Holwerda for counting FISH slides; G. Geeven for the analysis of sequencing data; J. Brandsma for technical assistance; P. Verschure for the LacR–GFP backbone construct; and H. Niwa for providing ZHBTc4 ES cells. We also thank the Netherlands Institute for Regenerative Medicine (NIRM) network for supporting the R.A.P. laboratory and the Medical Research Council UK for supporting the I.C. laboratory. This work was financially supported by grants from the Dutch Scientific Organization (NWO) to E.d.W. (700.10.402, 'Veni') and W.d.L. (91204082 and 935170621), InteGeR FP7 Marie Curie ITN (PITN-GA-2007-214902) and a European Research Council Starting Grant (209700, '4C') to W.d.L.

Author Contributions E.d.W. conceived the study, analysed the data and wrote the manuscript. B.A.M.B. designed and performed reprogramming and knockout experiments, and helped to write the manuscript. Y.Z. and P.H.L.K. designed and performed LacR–Nanog experiments. E.S. and P.H.L.K. performed cell culture and 4C experiments. M.J.A.M.V., E.P.N. and E.H. designed, performed and analysed FISH experiments. M.W. and N.G. assisted with reprogramming experiments. R.A.P. shared Oct4 conditional knockout cells and assisted with depletion experiments. N.F. and I.C. shared conditional knockout cells and assisted with Nanog depletion experiments. W.d.L. conceived the study and wrote the manuscript

Author Information 4C sequencing data and mapped wig files have been submitted to the Gene Expression Omnibus (GEO) under accession number GSE37275. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.d.L. (w.delaat@hubrecht.eu).

METHODS

Cell culture. E14 ES cells (129/Ola background) and C56Bl/6-129 ES cells were grown in BRL-conditioned DMEM (high glucose, Gibco) supplemented with 15% FBS, 1× non-essential amino acids (NEAA; Gibco), 1× penicillin–streptomycin (Gibco), 1:1,000 β-mercaptoethanol (Invitrogen), 1× L-glutamine (Gibco) and 1,000 U ml^{-1} leukaemia inhibitory factor (LIF; Gibco). Independently derived 129/Cast ES cells (129SVJ/Castaneus background) were grown on irradiated mouse embryonic fibroblasts in DMEM supplemented with 15% FBS, 1× NEAA, 1× penicillin–streptomycin, 1:1,000 β-mercaptoethanol and 1,000 U ml^{-1} LIF. RCNβH cells were cultured in GMEM, β-mercaptoethanol, 10% FCS and LIF as described previously[19,30]. ZHBTc4 (ref. 18) cells were cultured in GMEM, β-mercaptoethanol, 15% FCS, sodium bicarbonate and LIF. Culture medium was supplemented with 1 μg ml^{-1} doxycycline or 1 μM tetracycline when applicable. NPCs (E14 and 129/Cast) were grown in DMEM/F12 supplemented with 1:100 N2 (Gibco), 20 ng ml^{-1} bFGF (Peprotech), 20 ng ml^{-1} murine EGF (Peprotech). For the 129/Cast NP cells 1× B-27 (Gibco) was added[9]. We generated astrocytes by growing E14NP cells to confluency and washing twice with DMEM before adding astrocyte medium (DMEM/F12 supplemented with 1:100 N2 and 2% FBS)[31]. The culture medium was changed twice and cells were grown for 5 days to make sure differentiation was complete, which was confirmed by immunofluorescence.

Generation of iPS cells. For generation of iPS cells, 10,000 129/Cast NPCs were seeded on gelatin-coated dishes in N2B27 medium (StemCell Resources). Cells were infected overnight with lentivirus expressing a multicistronic reprogramming cassette, encoding the iPS factors Oct4, Klf4, Sox2 and c-Myc[32]. After 5 days, cells were collected and plated on irradiated mouse embryonic fibroblasts. On day 6, N2B27 medium was replaced with mouse ES cell medium (DMEM with 15% FBS, 1× NEAA, 1× penicillin–streptomycin, 1:1,000 β-mercaptoethanol and 1,000 U ml^{-1} LIF). iPS cell colonies were picked for clonal expansion on days 20–22 after infection. At passage 11 after colony picking, proper iPS cell reprogramming was examined by qPCR analysis on a panel of marker genes on total RNA (pluripotency markers: *Nanog*, *Zic3*, *Dppa4*, *Sall4*, *Cer1*, *Sox17* and *Fgf5*, neuronal lineage markers: *Olig2*, *Nestin*, *Blbp* and *Glast*). Cells were collected for 4C at passage 11.

siRNA knockdown of *Nanog*. For our knockdown experiments we used a pool of siRNA oligonucleotides targeting *Nanog* (M-057004-01) and a control pool containing non-targeting siRNAs (D-001206-13, siGENOME SMARTpool, Dharmacon). 129/Cast ES cells were seeded without feeders in 100-mm culture dishes at ~20% confluency on the day before transfection. Cells were transfected according to the manufacturer's protocol using 25 nM final siRNA concentration combined with 50 μl DharmaFECT 1. Transfection mixtures were added directly into the culture medium and plates were incubated at 37 °C with 5% CO_2. Forty-eight hours after transfection, cells were collected for protein level analysis and 4C template preparation.

Conditional ablation of Nanog and Oct4. RCNβH cells were treated with tamoxifen and replated the next day. Seventy-two hours after initial tamoxifen treatment, cells were collected for 4C template preparation and analyses. ZHBTc4 cells were collected after 48 h of treatment with 1 μg ml^{-1} doxycycline.

Protein analysis. Protein levels before and after conditional deletion were analysed in cells collected at the time points as described above. Immunoblot analysis was carried out on nuclear extracts that were made as described in[33]. Extracts were subjected to SDS–PAGE[34], and proteins were transferred to a methanol-activated PVDF membrane. Blots were blocked in blocking buffer (5% non-fat dry milk in TBST (50 mM Tris, pH 7.4, 150 mM NaCl, 0.1%Tween)) for 1 h at room temperature or overnight at 4° C, while tumbling. Primary antibody was diluted in blocking buffer and incubated for 1–3 h at room temperature or overnight at 4° C, while tumbling. Blots were washed four times with TBST and incubated with secondary antibody for 1 h in blocking buffer. Membranes were then incubated with SuperSignal West Pro (Thermo Scientific) and digitally analysed using an LAS 4000 ECL ImageQuant imager and ImageJ software. Used antibodies: anti-Nanog (A300-397A, Bethyl Laboratories) at 1:5,000, anti-Oct4 (C30A3, Cell Signaling Technology) at 1:1,000, anti-histone H3 (Abcam 1791) at 1:2,000.

Flow cytometry. Tamoxifen-treated and -untreated RCNβH cells were trypsinized and pellets were resuspended as single cells in regular ES cell medium at about 10^6 per ml. For each condition, 50,000 live cells were analysed for GFP fluorescence, using a Becton Dickinson FACSCalibur flow cytometer and FloJo software.

Generation of lacO targeted cell line. Homology arms were excised (KpnI digest) from bacterial artificial chromosome (BAC) RP24-136A15, and ligated into a low-copy bluescript plasmid. A total of 256 copies of a lacO array were inserted into a unique AatII site of the homology arms. F_1 ES cells derived from C56Bl/6 and 129 mouse strains were transfected with the linearized targeting construct by electroporation. After 14 days of selection with neomycine, positive colonies were picked and screened by Southern blotting. The GFP–LacR–Nanog construct was generated in the backbone of pHAGE2-IRES-puro with an EF1α promoter[35]. LacO cells were stably transduced with the GFP–LacR–Nanog construct, and positive cells were selected with puromycine for 10 days after which cells were collected and tested for purity of by flow cytometry (90% GFP-positive). Allelic paired-end 4C technology was performed as described[20], using HindIII-DpnII digestion and the following 4C primers: 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCGGAACTAAATGGAGGATC-3′ and 5′-CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTTACCAGGACCCCTGGGACCC-3′.

3D-DNA FISH. 3D-DNA FISH for interchromosomal interaction was performed essentially as described in ref. 2. For slide preparation, ES cells were spotted on polylysine microscopy slides after which slides were washed in PBS. Cells were fixed in 3% paraformaldehyde/PBS and washed twice with PBS, after which cells were permeabilized on ice using ice-cold 0.5% Triton X-100 for 6 min. Slides were then washed for 3 min with 70% ethanol and stored in 70% ethanol at −20 °C. For preparation of probes, 10 μl of both labelled BACs was combined with 5 μl mouse Cot1 DNA and mixtures were speedvacced until pellets were dry. Pellets were resuspended in 12.5 μl 50+ hybmix, incubated for 5 min at 95 °C, cooled on ice, and incubated for 30 min at 37 °C.

For FISH hybridization, slides were dehydrated for 3 min in 70% ethanol, 3 min in 90% ethanol, 3 min in 100% ethanol, after which slides were air-dried. One-hundred microliters of 70+ hybmix was then added to the dried slides, and slides were covered with a coverslip and incubated for 3 min at 85 °C. Slides were washed on ice, using ice-cold 2× saline-sodium citrate buffer (SSC) for 5 min, then using ice cold 70% ethanol for 5 min, after which slides were dehydrated again as described above. After air drying, 10 μl probe was added and covered with a coverslip and hybridizing slides were incubated overnight at 37 °C in a humid box containing 50% formamide/2× SSC. After hybridization, slides were washed in 2× SSC for 5 min, which also removes the coverslip. Subsequently, slides were washed three times for 10 min in 50% formamide/2× SSC at 37 °C. Slides were then dehydrated as described above, and air-dried slides were mounted using 40 μl DAPI/Vectashield. Slides were covered with new coverslips and sealed with transparent nail polish. We performed manual distance measurements in ImageJ using the Image5D plugin.

General 4C template information. For high-quality 4C experiments library complexity is crucial; by applying 4C to 1 million genome equivalents (3 μg DNA), we analyse a large number of ligation products per viewpoint. The generated DNA contact profiles are therefore a true population average[3]. The observed ligation products are the result of spatial proximity. Note that these ligation products can be a reflection of direct DNA contacts (such as promoter–enhancer interactions) or indirect contacts mediated by large macromolecular complexes or nuclear particles.

Experimental and primer design is done as previously described[10]. For the allele-specific 4C we have used a paired-end 4C strategy[20]. To this end, we designed forward and reverse primers compatible with the Illumina flow cell. The forward primer analyses the ligation product and the reverse primer is selected such that it sequences an SNP that distinguishes the C57Bl/6 allele from the 129S1/SvImJ allele. After sequencing, this SNP is used to demultiplex the two alleles, to create two separate 4C profiles.

Definitions. To make this methods section clearer to non-experts we present the following definitions. Fragment: a genomic region (or sequence) that is generated after the first restriction. In this case, the first restriction enzyme, that generates the 3C template, is always HindIII. Fragment end: to generate the 4C template, the 3C template is further digested with a frequent cutter, in our case DpnII. The resulting HindIII-DpnII restriction fragment is referred to as the fragment end, because this restriction fragment represents the end of the 3C fragment. Capture frequency: captures are defined as the ligations in the 3C protocol resulting from 3D genome conformation. The 4C primers directly interrogate the ligation junction. Therefore the resulting capture frequency can be estimated from counting the number of reads coming from a given fragment end.

4C sequencing and mapping. 4C sequencing was performed as previously described[9]. We used HindIII as the first restriction enzyme to generate the 3C template, which was further trimmed with DpnII. Sequencing was performed on Illumina GAII and HiSeq 2000 over multiple runs. Primer sequence (internal barcode) is removed from the sequence and the trimmed reads were aligned to a reduced genome consisting of sequences that flank HindIII restriction sites. The mouse mm9 genome was used as the reference genome for mapping 4C sequence captures. Non-unique sequences (repeats) that flank a restriction site were removed from the analysis. From the mapping a frequency distribution along the genome is calculated, which is the input for all downstream analyses. Raw sequencing data and mapped wig files can be found under GEO accession GSE37275.

Statistical analysis of 4C data. Statistical analyses of 4C data (that is, domainograms and target identification) was performed as described previously[9,10]. For formal definitions we refer the reader to these articles. Here we will briefly describe underlying principles of the data analysis. An inherent challenge of 4C data (and

genome-wide chromosome capture data in general) is the highly non-uniform data distribution. Close to the viewpoint the signal is very high, whereas the signal rapidly decreases as a function of the distance from the viewpoint. Therefore, we statistically define significant interactions as regions that have an increased number of captures compared to the local background. To this end we must estimate the local background capture frequency. To minimize potential PCR artefacts we transform the 4C-seq read count at HindIII-DpnII fragment ends to binary data (that is, captured or not captured). From this it is clear that 0 s play an essential role in determining significant interactions. Local background is then determined as the frequency of captured fragment ends (1 s) in a large window, typically 3,000 fragment ends. Following the binomial distribution, we can estimate μ and σ (for details see ref. 10, which is used to determine a z-score for a window of fragment ends of fixed size.

To visualize the 4C data using domainograms, z-scores are calculated using windows with a range of size (from 3 to 200). The z-scores are subsequently transformed to P values with a one-tailed normal test. The $-\log_{10}$-tranformed P values are colour-coded and visualized along the linear chromosome. As such regions can be visualized with a high likelihood of interaction with the viewpoint.

To distill discrete regions of significant interaction we choose a fixed window size of 100 fragment ends and calculate the z-scores for this window size across the chromosome. To select significant regions we determine the z-score threshold based on a FDR level of 0.01. The FDR is determined based on the z-score distribution in 100 randomly permuted chromosomes. The windows exceeding the z-score threshold are selected as significantly contacted regions.

Analysis of 4C *trans*-interactions. In our data set we find highly specific inter-chromosomal interactions. Like the intrachromosomal profiles we calculate an enrichment score over the background capture frequency. However, because the background capture frequency is distributed more or less uniformly across the chromosome, we can use a single background frequency per chromosome. The 4C enrichment score is calculated in the following way:

$$E_{\mathrm{trans},i} = \frac{w \cdot p_{w,i} - (p_{\mathrm{chrom}} \cdot w)}{\sqrt{w \cdot p_{\mathrm{chrom}} \cdot (1 - p_{\mathrm{chrom}})}} \quad (1)$$

in which w is the window size, and i is window index along the chromosome. p_{chrom} is defined as follows:

$$p_{\mathrm{chrom}} = \frac{N_{\mathrm{captured}}}{N_{\mathrm{chrom}}} \quad (2)$$

in which N_{captured} is the number of fragment ends captured on the chromosome and N_{chrom} is the total number fragment ends on the chromosome. $p_{w,i}$ is defined as follows:

$$p_{w,i} = \frac{n_{w,i,\mathrm{captured}}}{w} \quad (3)$$

in which $n_{w,i,\mathrm{captured}}$ is the number of fragments captured in genomic window i of size w.

Windows with an $E_{\mathrm{trans},i}$ larger than 6 were chosen for subsequent analysis in GREAT.

4C/Hi-C alignment to ChIP profiles. To test enrichment of 4C signal along ChIP peaks we aligned the *trans* fragments to nearest ChIP peak. We used several ES cell ChIP-seq profiles from various sources. For Oct4, Sox2 and Nanog we used the data described in ref. 36 (GSE11724). H3K27ac was taken from ref. 37 (GSE24165). Smc1 data[38] (GSE22557), H3K4me3 (ref. 39) (GSE12241), RNA PolII and CTCF (mES cell and cortex), H3K27ac and H3K4me3 (cortex)[17] (GSE29218).

4C data was binarized because *in trans* the capture frequency is so low that read count more likely represent differences in PCR efficiency rather than genuine unique captures. For the binarized data the distance to the nearest ChIP peak was calculated. To calculate enrichment scores, the distances to the nearest ChIP peak were sorted (that is, aligned) and a sliding average was calculated. The window size of the sliding average was set to 1% of the total data set.

Hi-C normalization and analysis. Hi-C data[1] was downloaded from GEO (accession GSE35156). We removed all read pairs that are mapped within 500 bp of each other on the chromosome, because these read pairs are probably genomic background sequence, rather than bona fide Hi-C captures. For the virtual 4C and disorganization analysis we average the data to bins of 100 kb, which results in a matrix of pairwise capture frequencies between all the genomic bins. A proper analysis of the Hi-C data requires that we correct the Hi-C matrix for genomic biases. For this normalization we assume that the capture probability in a given genomic bin is dependent on the number of restriction sites in this bin. The strong positive correlation between the restriction site density and the number of captures for that given bin is evidence that this assumption is correct (data not shown). We therefore normalize the bins by dividing by the capture probability. First we calculate the restriction density in 100-kb bins along the chromosome, which gives us a capture probability for in a given bin ($p_{\mathrm{capture},i}$). The capture probability between two bins on the chromosome ($P_{\mathrm{capture},i,j}$) can now be calculated by taken the product of the capture probabilities of the two single bins ($p_{\mathrm{capture},i}$, $p_{\mathrm{capture},j}$). Before normalization the correlation between the diagonals of the the Hi-C matrices for the NcoI and HindIII experiments from mouse ES cells is 0.32. However, after normalization this correlation has jumped to 0.86.

For the virtual 4C based on the Hi-C data, we combine the three normalized matrices (2× HindIII, 1× NcoI). Because the data are too sparse to perform a virtual 4C analysis for a single fragment, we analyse a single row from the Hi-C interaction matrix. For comparison, we also calculate the average 4C signal in 100-kb genomic bins.

For the analysis of genomic disorganization we use the two HindIII experiments for mouse ES cell and cortex (GSE35156). In this analysis we want to compare the propensity of active and inactive regions to contact regions over large genomic distances. To this end we segment the chromosomes in active and inactive bins of 100 kb, based on the density of H3K4me1 sites[40]. On the basis of this segmentation we can create a matrix with similar in size to the Hi-C matrix. In this matrix three classes of interaction bins can be created: (1) H3K4me1 high in both: interaction bin between two active genomic regions; (2) H3K4me1 low in both: interaction bin between two inactive genomic regions; and (3) H3K4me1 low/H3K4me1 high: interaction bin between and active and inactive.

Because we perform a 50/50 segmentation, the classes H3K4me1$^{\mathrm{high/high}}$ and H3K4me1$^{\mathrm{low/low}}$ will both be 25% of the interaction bins, H3K4me1$^{\mathrm{high/low}}$ class will be 50% of the interaction bins. In addition, the Hi-C matrix is segmented into high-contact and low-contact bins by setting an arbitrary threshold (75% quantile value of the entire matrix). Next, we overlay the segmented Hi-C matrix and the contact bins to determine the number of long-range contacts made for each of the classes. We use various minimal distance cut-offs running from 10–70 Mb with step sizes of 10 Mb. This process is schematically explained in Fig. 1e.

Alignment of Hi-C data to ChIP peaks (PE-SCAn). To assess which factors are associated with genome organization, we aligned ChIP data to the Hi-C data. To this end we selected the intrachromosomal captures, however, because of the strongly non-uniform distribution we removed the captures that lie within 5 Mb of each other. This has the effect that we only analyse interactions between, rather than within, topological domains. The Hi-C pairs were aligned to the ChIP data in two iterations. First, one end of the paired reads was aligned to the ChIP data. Only reads that mapped within 500 kb up- or downstream of the ChIP peaks were selected for further analysis. Of this reduced set the corresponding read was also aligned to the ChIP peaks within 500 kb. As a result we get for every intrachromosomal pair of ChIP peak a set of two distances (dx, dy), to all the Hi-C ditags that are found within 500 kb of these peaks. From the distribution of dx and dy a frequency matrix is calculated, which is the result of our two-dimensional alignment, with a bin size of 50 kb. To calculate whether the binding sites of a given factor show preferential spatial contacts, we calculate an enrichment score over a randomized data set. The randomized data set is calculated by aligning the Hi-C data to a circularly permuted ChIPseq data set, that is, the ChIP peaks are linearly shifted 10 Mb along the chromosome. It is important to note that in this manner the structure of the Hi-C data are preserved; the structure of the ChIP data are also preserved, only shifted. The resulting frequency matrix serves as an internal normalization for the observed Hi-C data alignment scores.

30. Smith, A. G. Culture and differentiation of embryonic stem cells. *J. Tissue Cult. Methods* **13,** 89–94 (1991).
31. Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **38,** 603–613 (2010).
32. Warlich, E. *et al.* Lentiviral vector design and imaging approaches to visualize the early stages of cellular reprogramming. *Mol. Ther.* **19,** 782–789 (2011).
33. Andrews, N. C. & Faller, D. V. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res.* **19,** 2499 (1991).
34. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680–685 (1970).
35. Wilson, A. A. *et al.* Sustained expression of α_1-antitrypsin after transplantation of manipulated hematopoietic stem cells. *Am. J. Respir. Cell Mol. Biol.* **39,** 133–141 (2008).
36. Marson, A. *et al.* Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134,** 521–533 (2008).
37. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107,** 21931–21936 (2010).
38. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467,** 430–435 (2010).
39. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).
40. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454,** 766–770 (2008).