doi:10.1038/nature21429

3D structures of individual mammalian genomes studied by single-cell Hi-C

Tim J. Stevens[1,2]*, David Lando[1]*, Srinjan Basu[1]*, Liam P. Atkinson[1], Yang Cao[1], Steven F. Lee[3], Martin Leeb[4]†, Kai J. Wohlfahrt[1], Wayne Boucher[1], Aoife O'Shaughnessy-Kirwan[1,4], Julie Cramard[4], Andre J. Faure[5], Meryem Ralser[4], Enrique Blanco[5], Lluis Morey[5]†, Miriam Sansó[5], Matthieu G. S. Palayret[3], Ben Lehner[5,6,7], Luciano Di Croce[5,6,7], Anton Wutz[4]†, Brian Hendrich[1,4], Dave Klenerman[3] & Ernest D. Laue[1]

The folding of genomic DNA from the beads-on-a-string-like structure of nucleosomes into higher-order assemblies is crucially linked to nuclear processes. Here we calculate 3D structures of entire mammalian genomes using data from a new chromosome conformation capture procedure that allows us to first image and then process single cells. The technique enables genome folding to be examined at a scale of less than 100 kb, and chromosome structures to be validated. The structures of individual topological-associated domains and loops vary substantially from cell to cell. By contrast, A and B compartments, lamina-associated domains and active enhancers and promoters are organized in a consistent way on a genome-wide basis in every cell, suggesting that they could drive chromosome and genome folding. By studying genes regulated by pluripotency factor and nucleosome remodelling deacetylase (NuRD), we illustrate how the determination of single-cell genome structure provides a new approach for investigating biological processes.

Our understanding of nuclear architecture has been built on electron and light microscopy studies that suggest the existence of territories pervaded by an inter-chromosomal space through which molecules diffuse to and from their sites of action[1]. In parallel, biochemical studies, in particular chromosome conformation capture experiments (such as 3C and Hi-C) in which DNA sequences in close spatial proximity in the nucleus are identified after restriction enzyme digestion and DNA ligation, have provided molecular information about chromosome folding[2]. At the megabase scale, Hi-C experiments have partitioned the genome into two (A or B) compartments[3]. In addition, they have provided evidence for 0.5–1.0-Mb topological-associated domains (TADs)[4–6], as well as smaller loops (hundreds of kilobases)[7]. 3C-type experiments have further shown that enhancers make direct physical interactions with promoters, and that these interactions are stabilized by a network of protein–protein interactions involving CTCF, cohesin and Mediator[8,9]. Although probabilistic methods can be used to calculate ensembles of low-resolution models that are consistent with population Hi-C data[10,11], understanding genome structure at higher resolution requires the development of single-cell approaches.

In mitotic cells, both TADs and A/B compartments disappear[12] and thus the structural complexity of interphase chromosomes is re-established during the G1 phase. To study interphase genome structure, we have combined imaging with an improved Hi-C protocol (Fig. 1a) to determine whole-genome structures of single G1-phase haploid mouse embryonic stem (ES) cells at the 100-kb scale. The structures allow us to study TAD and loop structure genome-wide, to analyse the principles underlying genome folding, and to understand which factors may be important for driving chromosome/genome structure. We also illustrate how combining single-cell genome structures with population-based RNA sequencing (RNA-seq) and chromatin immunoprecipitation followed by high-throughput sequencing (ChIP–seq)

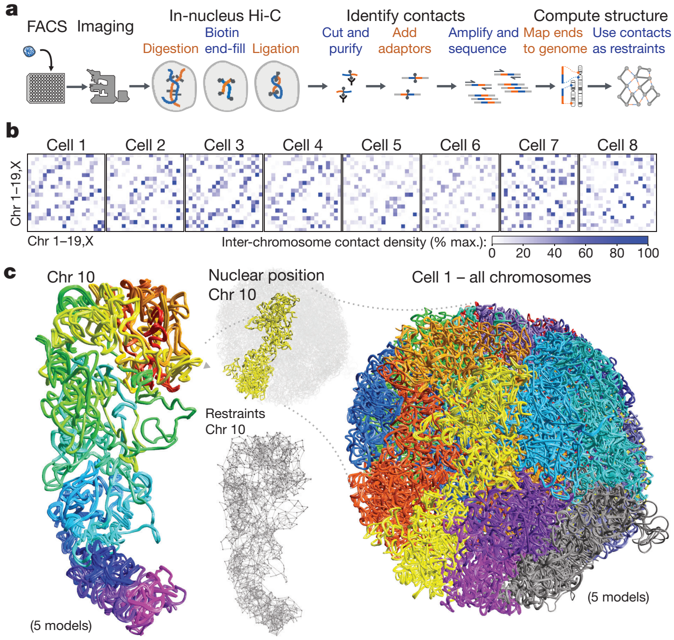


Figure 1 | Calculation of 3D genome structures from single-cell Hi-C data. a, Schematic of the protocol used to image and process single nuclei. **b**, Colour-density matrices representing the relative number of contacts observed between different pairs of chromosomes. **c**, Five superimposed structures from a single cell, from repeat calculations using 100-kb particles and the same experimental data, with the chromosomes coloured differently. An expanded view of chromosome 10 (Chr 10) is shown, coloured from red to purple (centromere to telomere), together with an illustration of the restraints determining its structure.

---

[1] Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK. [2] MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK. [3] Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. [4] Wellcome Trust – MRC Stem Cell Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK. [5] EMBL-CRG Systems Biology Unit, Centre for Genomic Regulation (CRG), 08003 Barcelona, Spain. [6] Universitat Pompeu Fabra, 08003 Barcelona, Spain. [7] Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain. †Present addresses: Max F. Perutz Laboratories, University of Vienna, Vienna Biocenter, Dr. Bohr-Gasse 9/3, 1030 Vienna, Austria (M.L.); Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Department of Human Genetics, Miami, Florida 33136, USA (L.M.); Inst. f. Molecular Health Sciences, ETH Zurich, HPL E 12, Otto-Stern-Weg 7, 8093 Zürich, Switzerland (A.W.).

*These authors contributed equally to this work.

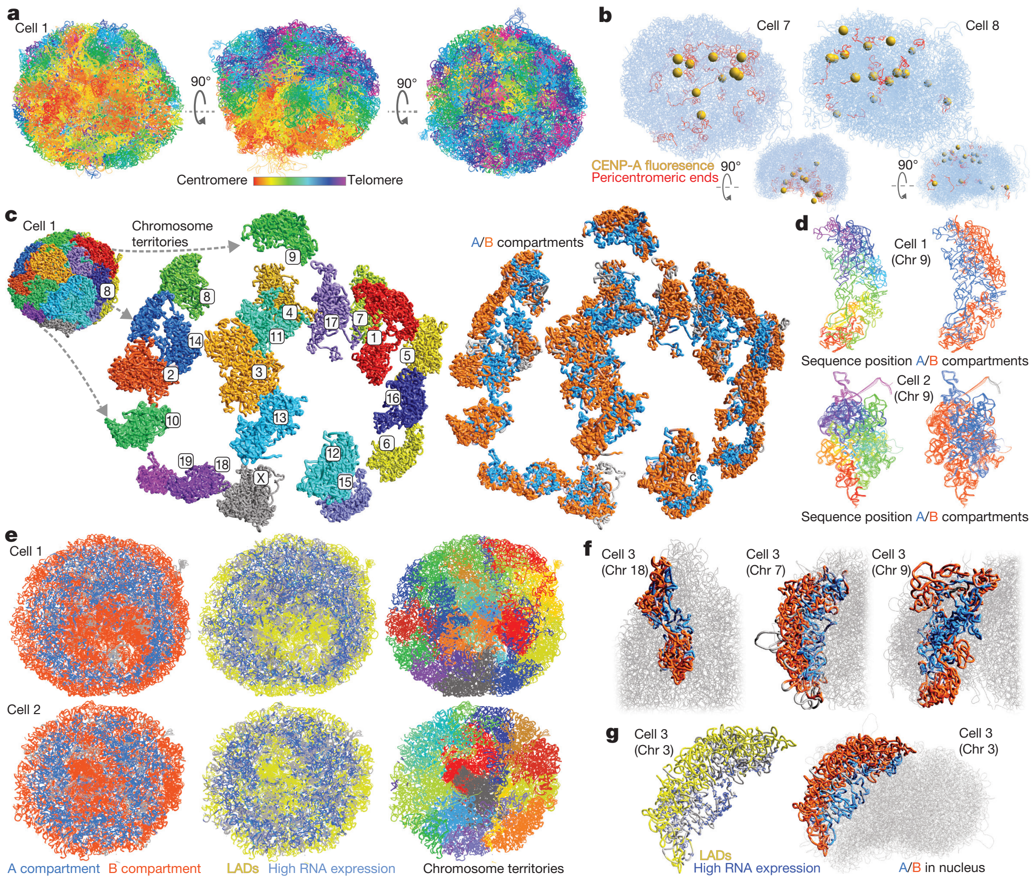


Figure 2 | Large-scale structure of the genome. a, Five superimposed structures from a single cell in three different orientations, with the chromosomes coloured from red to purple (centromere to telomere). **b**, Superposition of two single-cell structures with images of mEos3.2-tagged CENP-A recorded from the same single cells. The centromeres from the images are shown as yellow spheres and the centromeric ends of the chromosomes are coloured red. The same structures after rotation by 90° are shown below. **c**, 3D structure of a haploid mouse ES genome with expanded views of the separate chromosome territories (left), and the spatial distribution of the A (blue) and B (red) compartments (right). **d**, Structure of chromosome 9 from two different cells coloured from red to purple (centromere to telomere) (left), or according to whether the sequence is found in either the A (blue) or the B (red) compartments (right). **e**, Cross-sections through five superimposed 3D structures from two different cells, coloured according to: whether the sequence is in the A or B compartment (left); whether the sequence is part of a cLAD (yellow) or contains highly expressed genes (blue) (centre); and chromosome identity (right). **f**, Structures of selected chromosomes from a single cell illustrating the different ways chromosomes can contribute to the A and B compartments. **g**, Chromosome 3 from a single cell with the positions of highly expressed genes shown as blue circles (larger circles indicate higher expression) and lamina-associated regions shown in yellow (left), and in which the sequence is coloured according to whether it is in the A or B compartment (right).

data provides insights into the organization of pluripotency factor- and nucleosome remodelling deacetylase (NuRD)-regulated genes.

Intact genome structures from single-cell Hi-C data

We imaged haploid mouse ES cell nuclei, expressing fluorescently tagged CENP-A (the centromeric histone H3 variant) and histone H2B proteins, to select G1-phase cells (Extended Data Fig. 1a) and later validate the structures. Hi-C processing of eight individual mouse ES cells yielded 37,000–122,000 contacts (Extended Data Table 1), representing 1.2–4.1% recovery of the total possible ligation junctions. In single cells, unlike in population data, Hi-C contacts are observed between distinct and different sets of chromosomes (Fig. 1b and Extended Data Fig. 1b).

Using a particle-on-a-string representation and an extended simulated annealing protocol, we calculated highly consistent 3D genome structures (ensemble root mean square deviation (r.m.s.d.) < 1.75 particle radii) with discrete chromosome territories (Fig. 1c and Supplementary Videos 1, 2). The structures were calculated with an average of 1–3 Hi-C contact-derived restraints for each 100-kb particle (with a total of 26,000–75,000 restraints; Extended Data Table 2 and Extended Data Fig. 1c). Recalculation after randomly omitting 10–70% of the data reliably generated the same folded conformation (r.m.s.d. < 2.5 particle radii). Moreover, structure calculations after randomly merging half of the data from two different cells resulted in a vast increase in the number of violated experimental restraints (37.4% have a distance of more than 4 particle radii, compared with 5–6% for the separate data), and generated compacted, highly inconsistent structures (Extended Data Fig. 1d). Thus, single-cell Hi-C datasets cannot result from independent sampling of contacts from a single underlying conformation. In addition, cells with either a broken/recombined chromosome (Extended Data Fig. 1e) or a duplicated chromosome (Extended Data Fig. 1f) can be immediately recognized from the data.

Structure validation and contact coverage

A consistent Rabl configuration (with centromeres and telomeres clustered on opposite sides of the nucleus) was observed in all G1-phase ES cells, strongly validating the structures (Fig. 2a, Extended Data Fig. 2a and Supplementary Video 3). Figure 2b shows two examples of CENP-A image superposition with the corresponding genome structure from the same cell, providing independent evaluation of the reliability of the structures. Cell 7 shows typical clustering of the pericentromeric regions in a cavity on one side of the structure, which is clearly supported by the centromere positions in the CENP-A image. In cell 8, the centromeres are more diffusely distributed in both the image and the structure. The structures were also validated by comparison

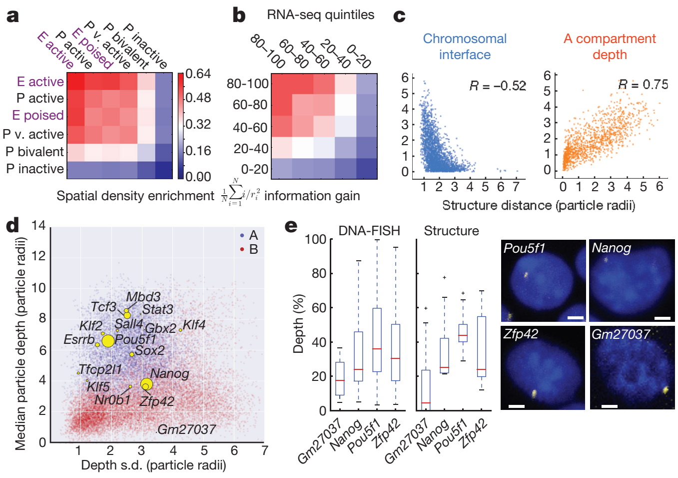


Figure 3 | Relationship between genome folding and gene expression. **a**, **b**, The enrichment in spatial density of: enhancers (E) and promoters (P) annotated using ChIP–seq data (**a**); and gene expression determined from nuclear RNA-seq data (**b**), with genes separated according to their relative level of expression. Data in **a** and **b** are presented in hierarchical cluster order, grouping the most similar datasets together. **c**, The enrichment in the spatial density of gene expression versus distance from the nearest inter-chromosomal interface (left) and the outer surface of the A compartment (right). **d**, Median versus standard deviation of the depth from the nuclear periphery for particles in the A (blue) or B (red) compartments. Particles containing pluripotency genes are indicated by yellow circles; the sizes illustrate relative levels of expression. **e**, Comparison of nuclear depth in either the 3D structures ($n = 8$) or DNA-FISH analysis of the *Nanog* ($n = 84$ cells) and *Zfp42* ($n = 142$ cells) genes, with *Pou5f1* ($n = 189$ cells) as a control. The pseudo-gene *Gm27037* ($n = 16$ cells) was an additional non-pluripotency factor control. Scale bars, 2 μm.

with previous imaging studies, and with both our own and previous DNA-FISH experiments, and by testing structural predictions using super-resolution microscopy (see below).

The single-cell Hi-C data shows fairly uniform coverage of long-range contacts across both the A and B compartments, suggesting similar restriction enzyme/ligase accessibility in each (Extended Data Fig. 2b). Notably, the contact probability is preserved for all nearby particles, showing that the entire structure is consistent with the Hi-C contact data (Extended Data Fig. 2c). We noticed an increase in contact density in some regions that coincided with sites of early DNA replication[13], but after studying violated experimental restraints we were unable to identify any region that cannot be described by a single structural conformation, that is, in which replication appeared to have begun (Extended Data Fig. 2b).

Comparison of haploid and diploid mouse ES cells using RNA-seq and ChIP–seq experiments, respectively, showed that the levels of gene expression are highly correlated with each other (Spearman's rho = 0.97, $P < 10^{-15}$) (Extended Data Fig. 2d) and that protein–genome interactions are highly similar (Extended Data Fig. 2e). This allowed us to use published ChIP–seq data when analysing the haploid structures.

3D genome architecture is conserved in all cells

Discrete chromosome territories can be seen in all the intact genome structures (Fig. 2c and Supplementary Video 1), although there is a considerable degree (5–10%) of chromosome intermingling (Extended Data Fig. 3a). While chromosome structure varies markedly from cell to cell, we find that regions belonging to the A or B compartments always cluster together and A segregates from B (Fig. 2d and Extended Data Fig. 3b). This is supported by recent imaging experiments showing that A- and B-compartment TADs are organized in a spatially polarized manner in single chromosomes[14], providing further validation of our structures. In all cells, the chromosomes then pack together to give an outer B compartment ring, an inner A compartment ring, and an internal B compartment region around the hollow nucleoli (Fig. 2e, Extended Data Fig. 3c and Supplementary Video 4). The nucleolus is often close to the nuclear membrane with the A compartment forming a bowl-like structure. To achieve this organization, chromosomes can fold in from the surface towards the nucleoli, or fold in and back out again, or go all the way through the nucleus (Fig. 2f and Supplementary Video 5). Chromatin states computed from the genome-wide association of post-translationally modified histones in mammalian cells[15] (a completely independent method) also show a similar organization (Extended Data Fig. 3d). Likewise, constitutive lamina-associated domain (cLAD) regions[16,17] are confined to either the nuclear membrane or the nucleolar periphery in every cell, consistent with reshuffling between these regions at each cell cycle[18,19]. Highly expressed genes, however, mostly lie in the inner A compartment ring (Fig. 2e, g, Extended Data Fig. 3c, e, f and Supplementary Videos 6, 7).

By mapping ChIP–seq data onto the single-cell genome structures, we observed 3D clustering of histones H3K4me1, H3K27ac and H3K4me3, consistent with the presence of enhancer/promoter clusters or transcription factories (Extended Data Fig. 4a, b). Annotating enhancers and promoters for activity (see Supplementary Methods) showed that active enhancers spatially associate most strongly with each other, followed by active enhancers with active promoters (Fig. 3a). We also found a pronounced clustering of highly expressed genes, in single cells, after mapping nuclear RNA-seq data onto the structures (Fig. 2g), and the greater the level of gene expression the larger the effect (Fig. 3b). Genome-wide analysis also showed that active/poised enhancers and active/bivalent promoters have a clear preference for being located at chromosomal interfaces (Extended Data Fig. 4c). Notably, there are very clear correlations between the expression level of a gene, and both localization to a chromosomal interface and depth within the A compartment (Fig. 3c and Extended Data Fig. 4d, e). We also related the preferred positions of pluripotency genes[20] to gene expression and found that two highly expressed genes, *Zfp42* (also known as *Rex1*) and *Nanog*, have variable positions in our structures (Fig. 3d). They are either found near the nuclear membrane or buried. DNA-FISH experiments, in which *Pou5f1* (also known as *Oct4*) is a typical highly expressed (and usually buried) gene control, verified these conclusions and provided further validation of the structures (Fig. 3e).

Notably, the A and B compartments, cLAD, ChIP–seq and RNA-seq data were all determined from populations of cells. Their consistent organization in every cell suggests that overall chromosome and genome conformation may be driven by a combination of interactions of LADs with the nuclear membrane/nucleolus and the clustering of active enhancers/promoters, which can be modulated by chromatin remodelling[21]. That genome structure is driven by transcription is supported by live-cell imaging of histone–GFP fusion proteins during *Caenorhabditis elegans* development, which shows that knockdown of RNA polymerase II leads to a collapse of chromatin to a ring inside the nuclear membrane[22].

Folding of chromosomes into TADs or CTCF/cohesin loops

As in previous studies[5,9,23], we observed an alignment between highly expressed genes and both A/B compartment and TAD boundaries (Fig. 4a and Extended Data Fig. 5a). Analysis of two pairs of TADs, either side of highly expressed genes (regions 1 and 2 in Fig. 4a), illustrates that in some cells a particular TAD is compacted, often such that its two boundaries are close enough to interact, whereas in others it is completely extended. This difference is not due to a lack of data because the structures obtained from repeated calculations using identical experimental restraints are very well defined (Fig. 4b and Extended Data Fig. 5b).

We systematically studied compaction in chromosome 12 TADs (Extended Data Fig. 5a) by computing the radius of gyration $\left(R_g^2\right)$ after excluding possible sites of early DNA replication where TAD structure might be disrupted. As with previous studies of the Tsix TAD[24],

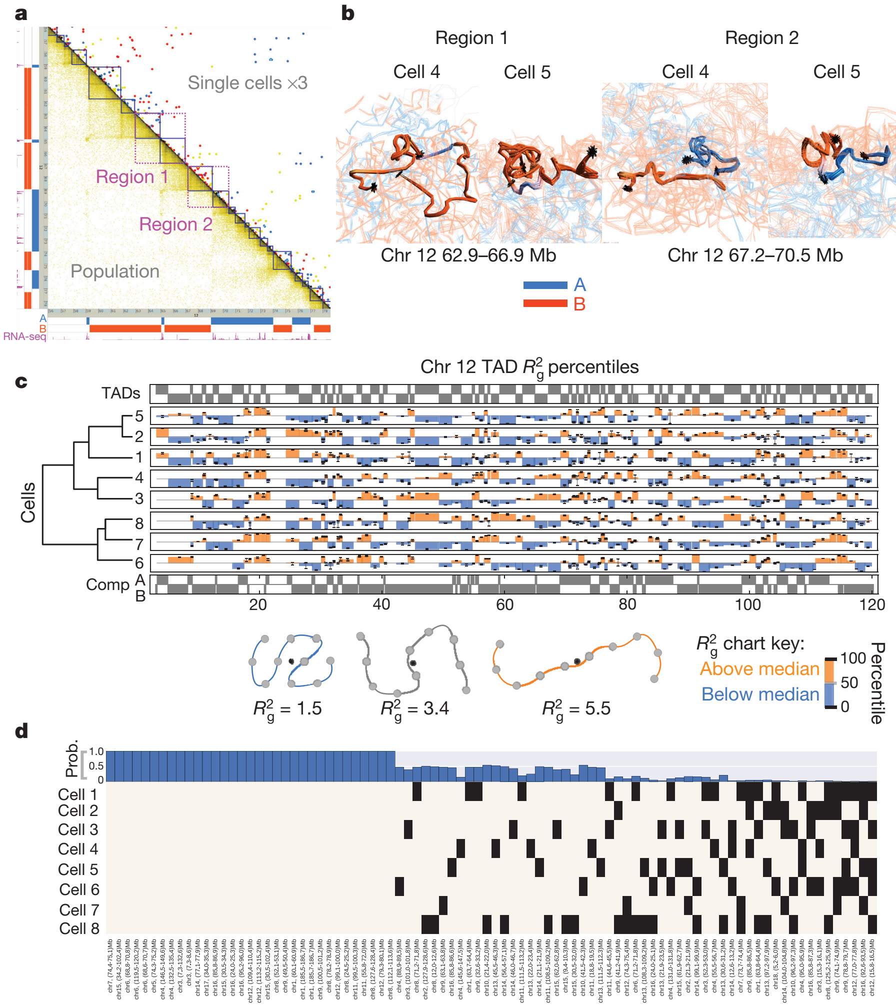


Figure 4 | Structure of TADs and CTCF/cohesin loops. a, Part of the Hi-C contact map from chromosome 12 showing: contacts observed in three different single cells (coloured red, yellow and blue; above the diagonal); and the corresponding population Hi-C data (below the diagonal). TADs identified previously[5] are shown in dark blue, and the two regions analysed in **b** are shown in magenta. **b**, Ensembles of five superimposed structures showing: two B-compartment TADs (region 1 in **a**; left); and TADs either side of an A/B-compartment boundary (region 2 in **a**; right). The TADs are coloured according to whether they are in the A (blue) or B (red) compartments, with white indicating a transitional segment (between A and B). Boundaries are marked by asterisks. **c**, The mean radius of gyration $\left(R_g^2\right)$ of chromosome 12 TADs ± s.e.m. Data are scaled according to TAD size, and presented as quantile values for the chromosome. Values below and above the fiftieth percentile value are coloured blue and red, respectively. The R_g^2 values for multiple cells are presented in hierarchical cluster order, grouping the most similar cell traces together. A schematic illustrating the calculation of the R_g^2 as a measure of the compaction of a particle chain is shown below. **d**, Analysis illustrating whether CTCF/cohesin loops with sequence separation of more than 600 kb identified previously[7] could be formed in the different single cells. Black squares indicate that a loop could be formed; white squares indicate that the two relevant particles are too far apart in the structure. The bar chart across the top shows the probability, for each loop, of random particles (pairs with the same sequence separation) forming the same number of contacts, or better.

individual TAD compaction varies widely from highly extended to compacted states (Fig. 4c), consistent with ligation occurring between almost every site in population Hi-C data. The structures of both compact and extended TADs are well defined and there is little correlation between the R_g^2 and Hi-C contact density (Extended Data Fig. 5c), further showing that extended TAD structures do not result from a lack of experimental contacts. Analysis of TAD structure in all the other chromosomes gave analogous results (Extended Data Fig. 6). It is noteworthy that compaction in the structures often appears to involve the formation of loops within a TAD (see Fig. 4b, Extended Data Fig. 5b and Supplementary Videos 8–11), and it will be interesting to investigate whether these structures are related to supercoiling[25,26] or loop extrusion[27–29].

We found that CTCF/cohesin loops identified in high-resolution Hi-C data from mouse B lymphoblasts[7] mostly involve interactions in which at least one end of the loop is in (or very near to) the A compartment (Extended Data Fig. 5d). Considering the 88 largest loops from 2,823 in total (with sequence separation of greater than 600 kb), we found that 33% do not form in any of the cells, whereas the boundaries of the remainder contact each other in 12–62% of the cells (Fig. 4d). Extending this analysis to all 2,823 loops in 8 cells showed that the boundaries interact in 62.1% of cells (Extended Data Fig. 7). Our genome-wide results suggest that TADs and CTCF/cohesin loops do not form in all cells, in agreement with previous DNA-FISH experiments[7] in which four representative loops were shown to form in only a proportion of cells.

Our structures provide snapshots of genome folding at a particular time in different cells, and thus do not provide information about dynamics. They are, however, markedly consistent with what is expected from recently proposed loop-extrusion models, in which TADs and CTCF/cohesin loops might be expected to have highly dynamic and variable structures as cohesin rings are driven to stable

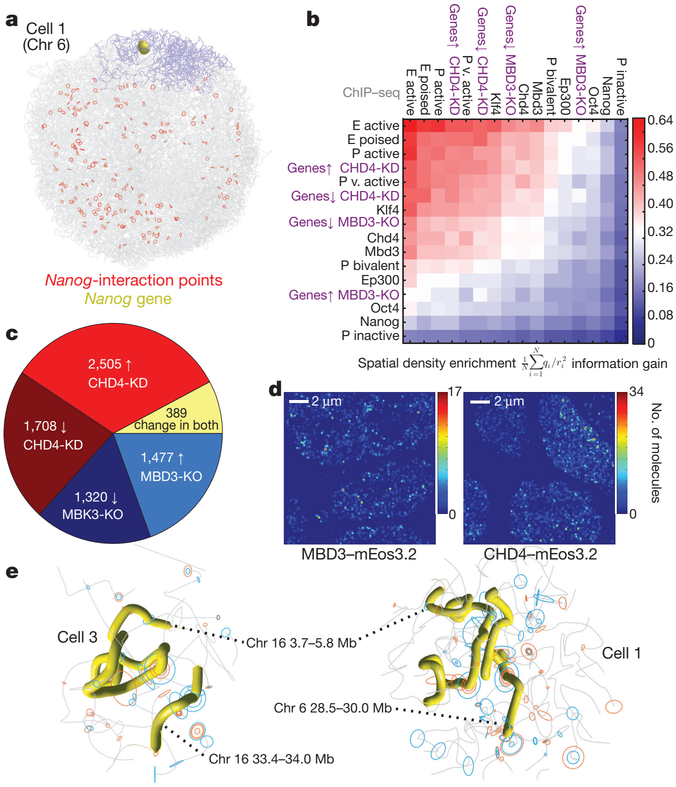


Figure 5 | Understanding the nature of gene networks in mouse ES cells. a, Structure of an individual cell illustrating the interactions identified between the *Nanog* gene (highlighted in yellow) in chromosome 6 (coloured blue) and other regions of the genome (red circles) in a population 4C experiment[35]. **b**, The spatial density enrichment of NuRD components (CHD4 and MBD3), pluripotency factors and NuRD-regulated genes, as well as annotated enhancers and promoters defined using ChIP–seq data. **c**, Pie chart showing the numbers of NuRD-regulated genes in different classes. CHD4-KD, CHD4-knockdown; MBD3-KO, MBD3-knockout. **d**, Heat map showing clustering of CHD4 and MBD3 molecules in 2D super-resolution photo-activated light microscopy in fixed mouse ES cells. **e**, Structures of a region of chromosome 16 in two different cells, showing clustering of regions containing genes that are highly regulated by NuRD (highlighted in yellow). The positions of genes in either the CHD4-knockdown or MBD3-knockout cells that are downregulated (red circles) or upregulated (blue circles) are indicated (larger circles for more highly regulated).

binding sites[7,27–29]. It is not known what drives the movement of cohesin rings in mammalian cells, but previous studies in yeast suggest that it might be RNA polymerase molecules and transcription[30]. This would be consistent with our observation that CTCF/cohesin loops[7] are mostly found in the A compartment (in which transcription levels are higher), studies in *Drosophila* suggesting that TADs result from the compaction of chromatin due to transcription[31,32], and recent studies of the inactive mouse X chromosome that show a global loss of TAD structure except at expressed genes[33,34].

Understanding the gene networks in single ES cells

In addition to CTCF, cohesin and Mediator, previous studies have implicated key pluripotency factors as well as the Polycomb complexes (PRC1 and PRC2) in organizing 3D genome structure in mouse ES cells. Analysis of one of the published 4C *Nanog* gene-interaction networks[35] showed that only one (or two) of the previously identified 4C contacts can be identified in each single-cell structure, showing that the propensity for particular genes to interact is low (Fig. 5a and Extended Data Fig. 8a, b). Analysis of *Pou5f1* gene-interacting regions[36] gave very similar results (Extended Data Fig. 8c).

We mapped ChIP–seq data for different pluripotency factors onto the single-cell genome structures and showed that, in single cells, Klf4 spatially clusters strongly with itself, H3K4me1, H3K27ac and H3K4me3, that is, with active enhancers and promoters (Extended Data Fig. 4a, b). This analysis also suggested 3D clustering of histone H3K27me3 (a marker for Polycomb complexes), but lower levels of 3D clustering of Nanog, both with itself and with H3K27me3. These results are consistent with previous ES-cell imaging experiments[36,37], and strongly validate our single-cell structures. They support the proposal that Klf4 organizes long-range chromosomal interactions[36], and suggest that the observed large-scale 3D segregation of Nanog and H3K27me3 (ref. 37) mostly results from Nanog and PRC complexes binding to separated sequences in chromosomes. However, although they suggest that Klf4-bound genes cluster, they also show that there is little propensity for particular Klf4-bound genes to interact with each other.

Next, we used the structures to study genes regulated by the NuRD complex, which has a key role in controlling the earliest stages of differentiation of mouse ES cells[38]. ChIP–seq experiments showed that although the chromatin-remodelling component CHD4 and the deacetylase component MBD3 (ref. 39) are widely distributed (data not shown), marked 3D clustering of NuRD-regulated genes occurs (Fig. 5b, c). Super-resolution microscopy and single-particle tracking using photo-activated light microscopy in fixed and live cells, respectively, showed clustering of both the chromatin-remodelling and deacetylase sub-modules (as illustrated by the mEos3.2-tagged CHD4 and MBD3 proteins, respectively), consistent with the 3D clustering of NuRD-regulated genes (Fig. 5d and Extended Data Fig. 8d). Notably, although our structures show that the regions containing highly NuRD-regulated genes cluster, the actual regions that interact vary from cell to cell (Fig. 5e). In addition, we found that most genes are upregulated or downregulated in either the CHD4-depletion experiment or the MBD3-knockout cells, but not in both (Fig. 5c), suggesting that the chromatin-remodelling and deacetylase sub-modules may function separately. However, despite regulating different sets of genes, it is notable that genes that are downregulated in the MBD3-knockout cells cluster more strongly than those that are upregulated, and genes that are downregulated in the MBD3-knockout cluster more strongly with genes that are upregulated in the CHD4-knockdown (and vice versa) (Fig. 5b). Although further work is necessary to understand what drives the formation of NuRD clusters, the 3D clustering of CHD4 and MBD3 with active enhancers and promoters is noteworthy (Fig. 5b).

Conclusion

The structures allow one of the first genome-wide analysis of 3D interactions of individual regulatory elements and genes in single cells. In combination with 3D imaging the data show that although Klf4- and NuRD-regulated genes interact and cluster to form foci, the genes they bring together are very variable. Our combination of imaging with determination of genome structure will allow further studies of these and many other biological processes. In addition, the finding that chromosomes have a Rabl configuration in mammalian G1-phase cells may underlie slight preferences in long-range chromosomal interactions, such as those leading to translocation events involved in disease[40].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Data Availability The ChIP–seq, RNA-seq and Hi-C data, structures and images reported in this study have been made available at the Gene Expression Omnibus (GEO) repository under accession code GSE80280.

Received 18 March 2016; accepted 26 January 2017.

Published online 13 March; corrected online 5 April 2017 (see full-text HTML version for details).

1. Cremer, T. *et al.* The 4D nucleome: Evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments. *FEBS Lett.* **589** (20 Pt A), 2931–2943 (2015).
2. Bickmore, W. A. & van Steensel, B. Genome architecture: domain organization of interphase chromosomes. *Cell* **152,** 1270–1284 (2013).
3. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).
4. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).
5. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
6. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148,** 458–472 (2012).
7. 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).
8. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl Acad. Sci. USA* **111,** 996–1001 (2014).
9. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153,** 1281–1295 (2013).
10. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* **30,** 90–98 (2011).
11. Tjong, H. *et al.* Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. *Proc. Natl Acad. Sci. USA* **113,** E1663–E1672 (2016).
12. Naumova, N. *et al.* Organization of the mitotic chromosome. *Science* **342,** 948–953 (2013).
13. Foti, R. *et al.* Nuclear architecture organized by Rif1 underpins the replication-timing program. *Mol. Cell* **61,** 260–273 (2016).
14. Wang, S. *et al.* Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353,** 598–602 (2016).
15. Julienne, H., Zoufir, A., Audit, B. & Arneodo, A. Human genome replication proceeds through four chromatin states. *PLOS Comput. Biol.* **9,** e1003233 (2013).
16. Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **38,** 603–613 (2010).
17. Meuleman, W. *et al.* Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res.* **23,** 270–280 (2013).
18. van Koningsbruggen, S. *et al.* High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. *Mol. Biol. Cell* **21,** 3735–3748 (2010).
19. Kind, J. *et al.* Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153,** 178–192 (2013).
20. Dunn, S. J., Martello, G., Yordanov, B., Emmott, S. & Smith, A. G. Defining an essential transcription factor program for naïve pluripotency. *Science* **344,** 1156–1160 (2014).
21. Therizols, P. *et al.* Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science* **346,** 1238–1242 (2014).
22. Krüger, A. V. *et al.* Comprehensive single cell-resolution analysis of the role of chromatin regulators in early *C. elegans* embryogenesis. *Dev. Biol.* **398,** 153–162 (2015).
23. Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* **32,** 3119–3129 (2013).
24. Giorgetti, L. *et al.* Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **157,** 950–963 (2014).
25. Naughton, C. *et al.* Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat. Struct. Mol. Biol.* **20,** 387–395 (2013).
26. Kouzine, F. *et al.* Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat. Struct. Mol. Biol.* **20,** 396–403 (2013).
27. Guo, Y. *et al.* CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162,** 900–910 (2015).
28. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci. USA* **112,** E6456–E6465 (2015).
29. Fudenberg, G. *et al.* Formation of chromosomal domains by loop extrusion. *Cell Reports* **15,** 2038–2049 (2016).
30. Lengronne, A. *et al.* Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* **430,** 573–578 (2004).
31. Eagen, K. P., Hartl, T. A. & Kornberg, R. D. Stable chromosome condensation revealed by chromosome conformation capture. *Cell* **163,** 934–946 (2015).
32. Zhimulev, I. F. *et al.* Genetic organization of interphase chromosome bands and interbands in *Drosophila melanogaster*. *PLoS One* **9,** e101631 (2014).
33. Minajigi, A. *et al.* Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349,** aab2276 (2015).
34. Giorgetti, L. *et al.* Structural organization of the inactive X chromosome in the mouse. *Nature* **535,** 575–579 (2016).
35. de Wit, E. *et al.* The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature* **501,** 227–231 (2013).
36. 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).
37. Denholtz, M. *et al.* Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell* **13,** 602–616 (2013).
38. Reynolds, N. *et al.* NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell Stem Cell* **10,** 583–594 (2012).
39. Zhang, W. *et al.* The nucleosome remodeling and deacetylase complex NuRD is built from preformed catalytically active sub-modules. *J. Mol. Biol.* **428,** 2931–2942 (2016).
40. Zhang, Y. *et al.* Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148,** 908–921 (2012).

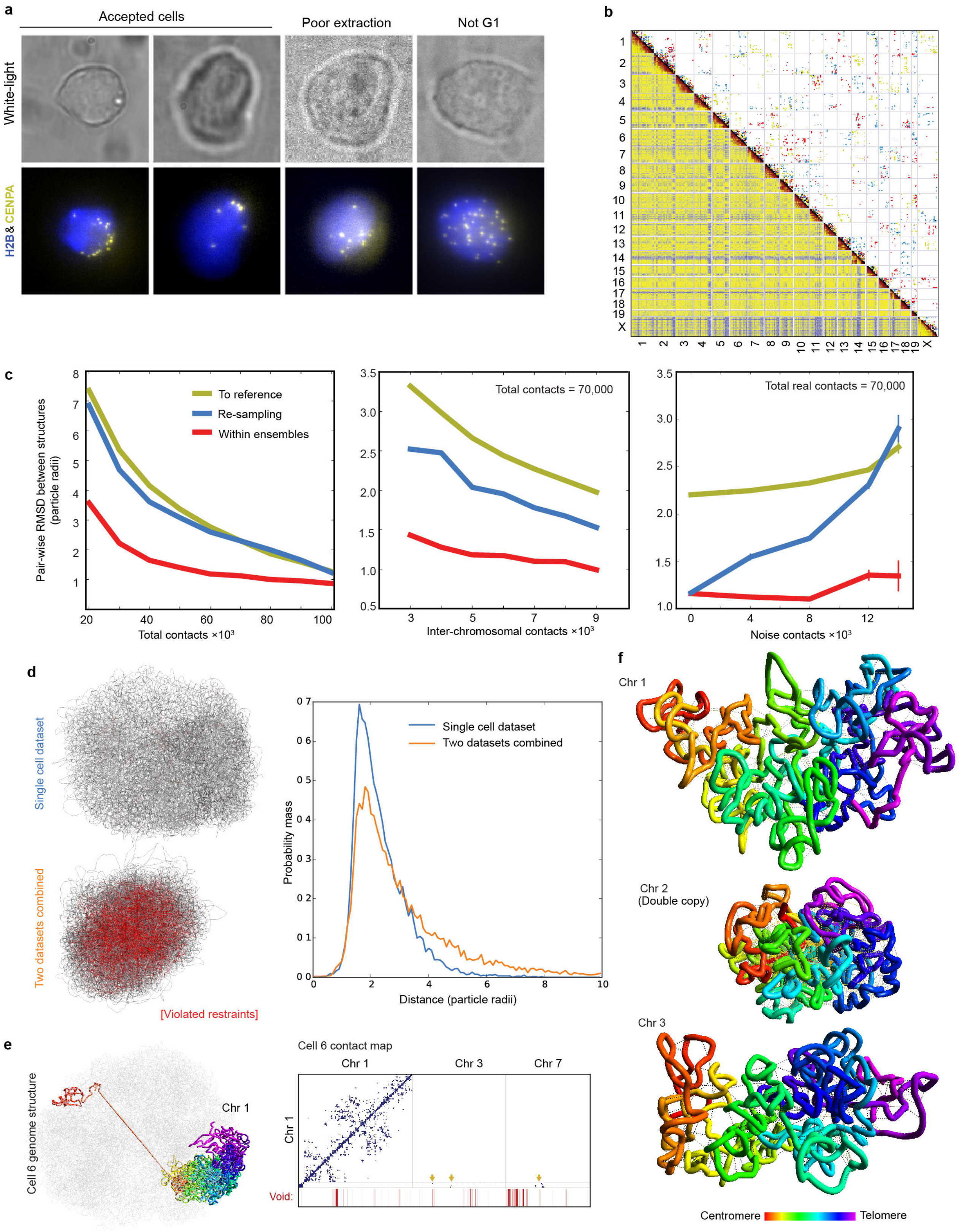
Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Riddell for cell sorting, P. Humphreys for confocal microscopy, A. Peter Gunnarson for the density mapping software, the CRUK Cambridge Institute for DNA sequencing, T. Nagano and P. Fraser for processing the preliminary haploid mouse ES cells, and W. Dean, S. Schoenfelder and S. Wingett for advice. We thank the Wellcome Trust (082010/Z/07/Z), the EC FP7 4DCellFate project (277899) and the MRC (MR/M010082/1) for financial support.

Author Contributions D.L., S.B. and Y.C. developed the protocol and carried out imaging/Hi-C processing. T.J.S. developed the software with assistance from L.P.A., W.B. and K.J.W. A.O'S.-K., J.C., M.R. and B.H. carried out the CHD4/MBD3 depletion experiments, associated RNA-seq and ChIP–seq, and created the mEos3.2-Halo tagged ES cell lines. M.L. and A.W. provided the initial samples of haploid mouse ES cells. S.F.L., M.G.S.P. and D.K. designed and built the microscope. L.M., M.S. and L.D.C. carried out ChIP–seq and RNA-seq experiments, while A.J.F., E.B. and B.L. carried out bioinformatics analysis. T.J.S. and E.D.L. designed experiments, analysed the results and wrote the manuscript with contributions from all the other authors.

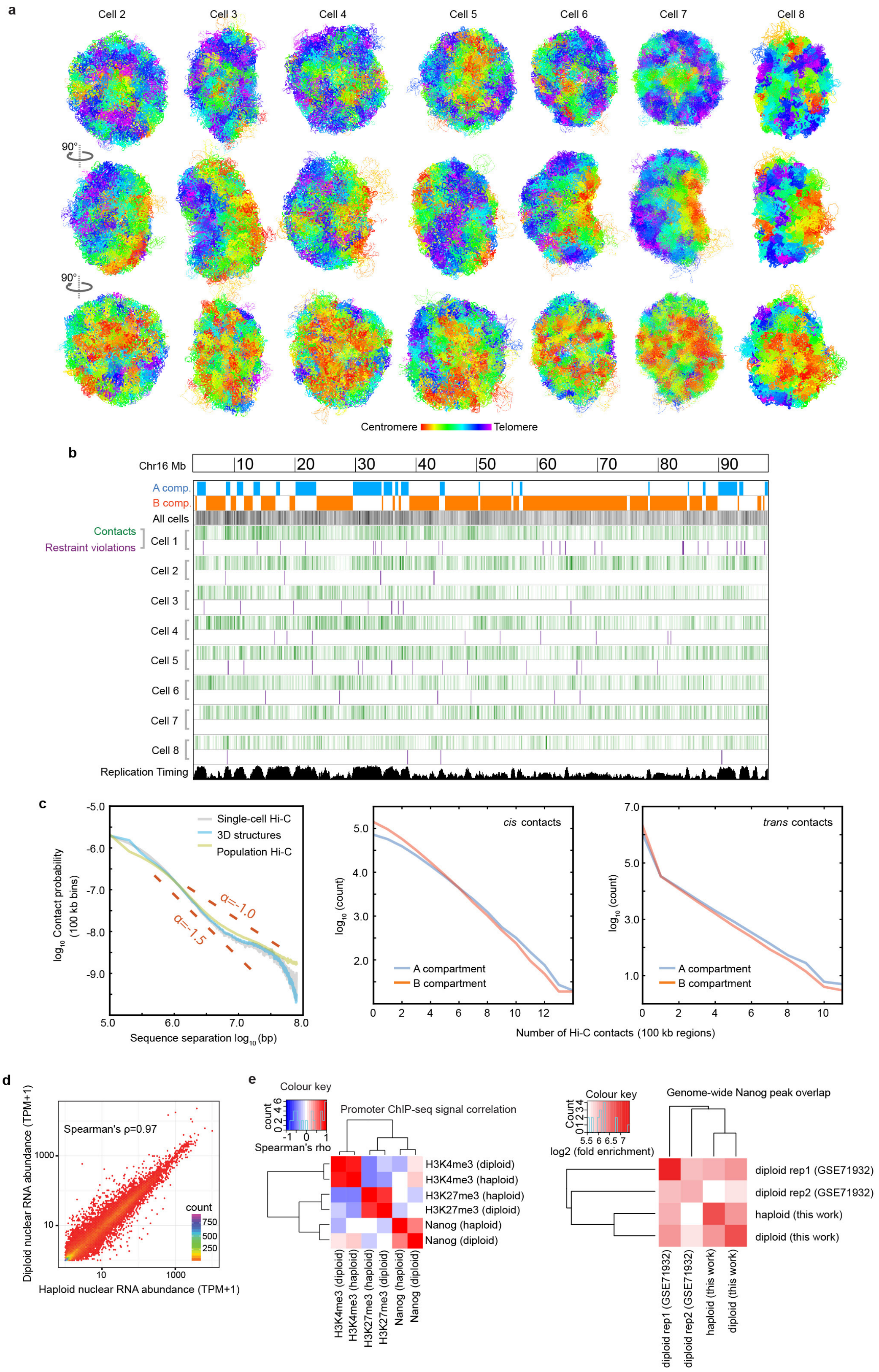
Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.D.L. (e.d.laue@bioc.cam.ac.uk).

Reviewer Information *Nature* thanks W. Huber and the other anonymous reviewer(s) for their contribution to the peer review of this work.



Extended Data Figure 1 | See next page for caption.

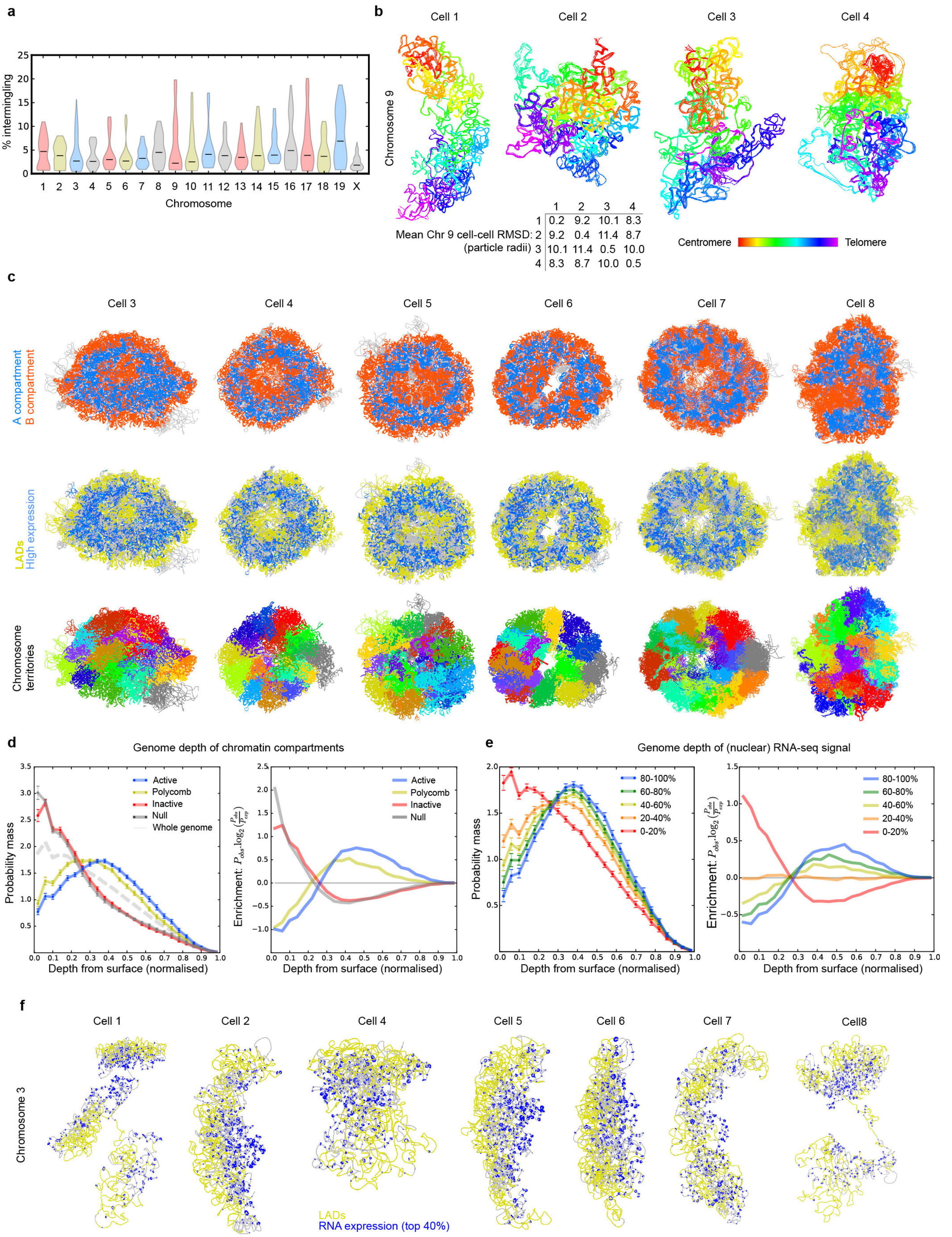
Extended Data Figure 1 | Quality control for Hi-C processing and 3D structure calculation. a, Comparison of 3D images of CENP-A in haploid mouse ES cell nuclei, expressing mEos3.2-tagged CENP-A and tandem infrared fluorescent protein (iRFP)-tagged histone H2B, with their corresponding white-light images. **b**, Comparison of three single-cell Hi-C contact maps (above the diagonal; contacts coloured red, yellow and blue) with the population Hi-C map (below the diagonal). **c**, An analysis of the accuracy and precision of the 100-kb structure calculation procedure for cell 1. The graphs show how the global (dis)similarity of structures is affected by: the total number of contacts (left); the number of inter-chromosomal contacts (middle); and the number of random noise contacts (right). Mean r.m.s.d. values for all pairs of conformations ± s.e.m. are shown for: the precision within ensembles arising from ten re-calculations using the same contacts (red); the variation across ensembles that arises from different random resampling (blue); and, as a measure of accuracy, the similarity to the best ensemble of structures (yellow). **d**, An example of a structure calculation carried out using either a single dataset, or after randomly merging 50% of the data from two different cells. Strongly violated experimental restraints (>4 particle radii apart) are shown in red. The plot (right) shows the probability of any two particles connected by an experimental restraint being violated to different degrees. **e**, Left, the structure of chromosome 1 from cell 6, where part of the chromosome lies at the opposite side of the genome structure, with no intermediate chromosome folding, illustrating the presence of a chromosomal break or recombination event. Right, the contact map shows that there are no contacts from the disconnected region to any other part of chromosome 1, but clear contacts to chromosomes 3 and 7. **f**, An example of an attempted calculation of the haploid genome structure for a cell containing a duplicated chromosome 2 shows many violations of the experimental restraints for that chromosome and a much more compacted structure (here compared with chromosomes 1 and 3). The structures are coloured according to position in the chromosome sequence from red to purple (centromere to telomere).



Extended Data Figure 2 | See next page for caption.

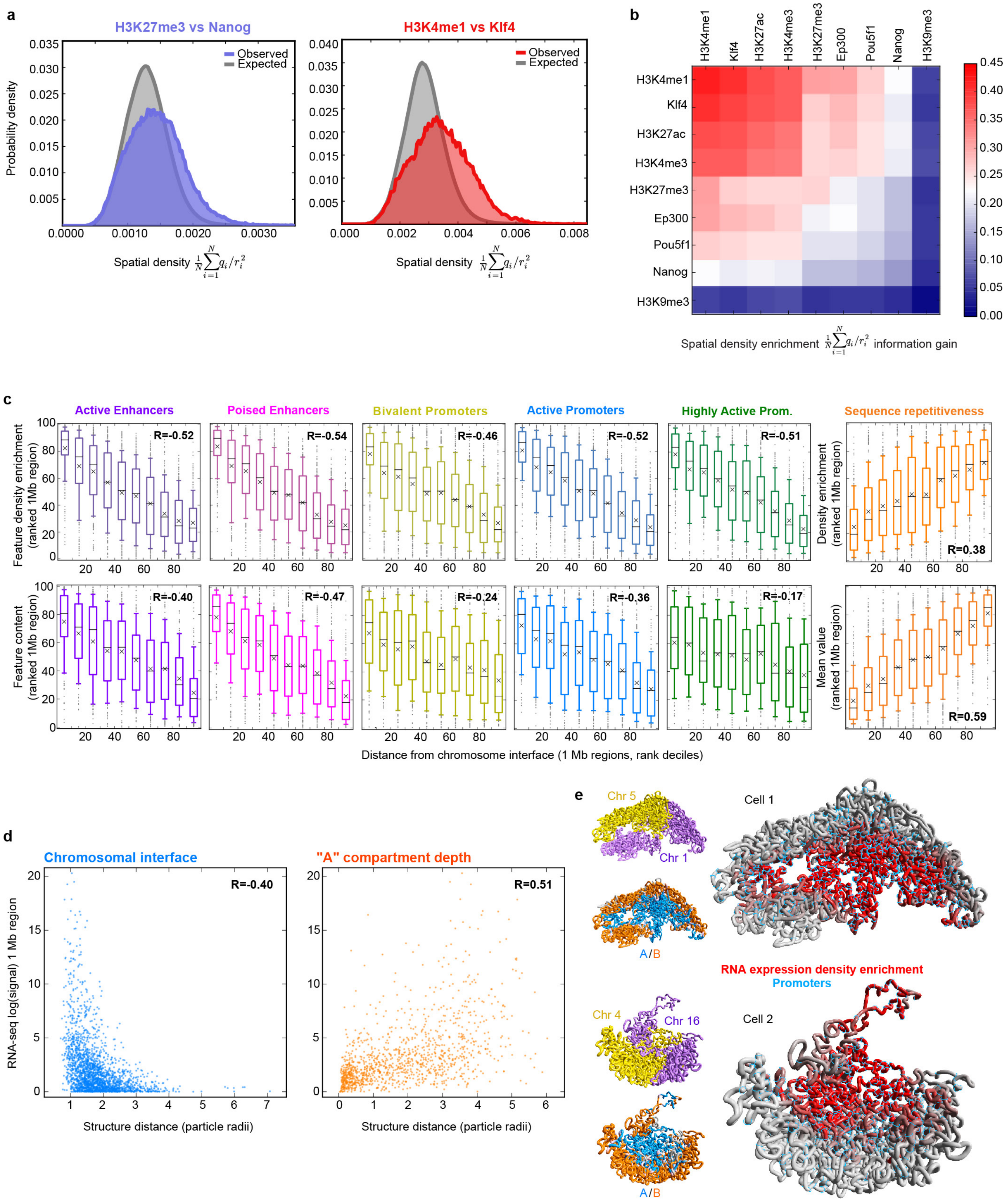
Extended Data Figure 2 | Validation and analysis of single-cell contacts. a, Structure of the entire haploid mouse ES cell genome from cells 2 to 8. The structural ensemble is represented by five superimposed conformations from repeat calculations, and is shown in three different orientations (after rotation through 90° relative to each other), with the chromosomes coloured according to their position in the chromosome sequence from red to purple (centromere to telomere). **b**, Correspondence between the distribution of Hi-C contacts (both *cis* and *trans*), violations of the distance restraints in the 3D structures, and DNA replication timing[13] for a representative chromosome (chromosome 12). **c**, Left, log-scale plots of contact probability (P_{cont}) against sequence separation (S). The slopes for a power law relationship ($P_{\mathrm{cont}} \propto S^{\alpha}$) in which α is either -1.0 or -1.5 are also indicated. Data are shown for the combined single-cell Hi-C contact data, for all of the non-sequential particles that are close to each other in the structures (<2 particle radii apart), and for the population Hi-C data. Right, the distribution in the number of intra-chromosomal (*cis*) or inter-chromosomal (*trans*) contacts between 100-kb regions in the single-cell Hi-C data are shown for both the A and B compartments. **d**, Correlation of gene expression levels (left), and hierarchically clustered heat maps showing the pairwise enrichment of ChIP–seq peak overlaps between haploid and diploid ES cells (centre), and Nanog ChIP–seq peak overlaps between haploid and diploid ES cells used in this study, as well as that previously published from diploid ES cells (right)[41].

41. Murakami, K. *et al.* NANOG alone induces germ cells in primed epiblast *in vitro* by activation of enhancers. *Nature* **529,** 403–407 (2016).



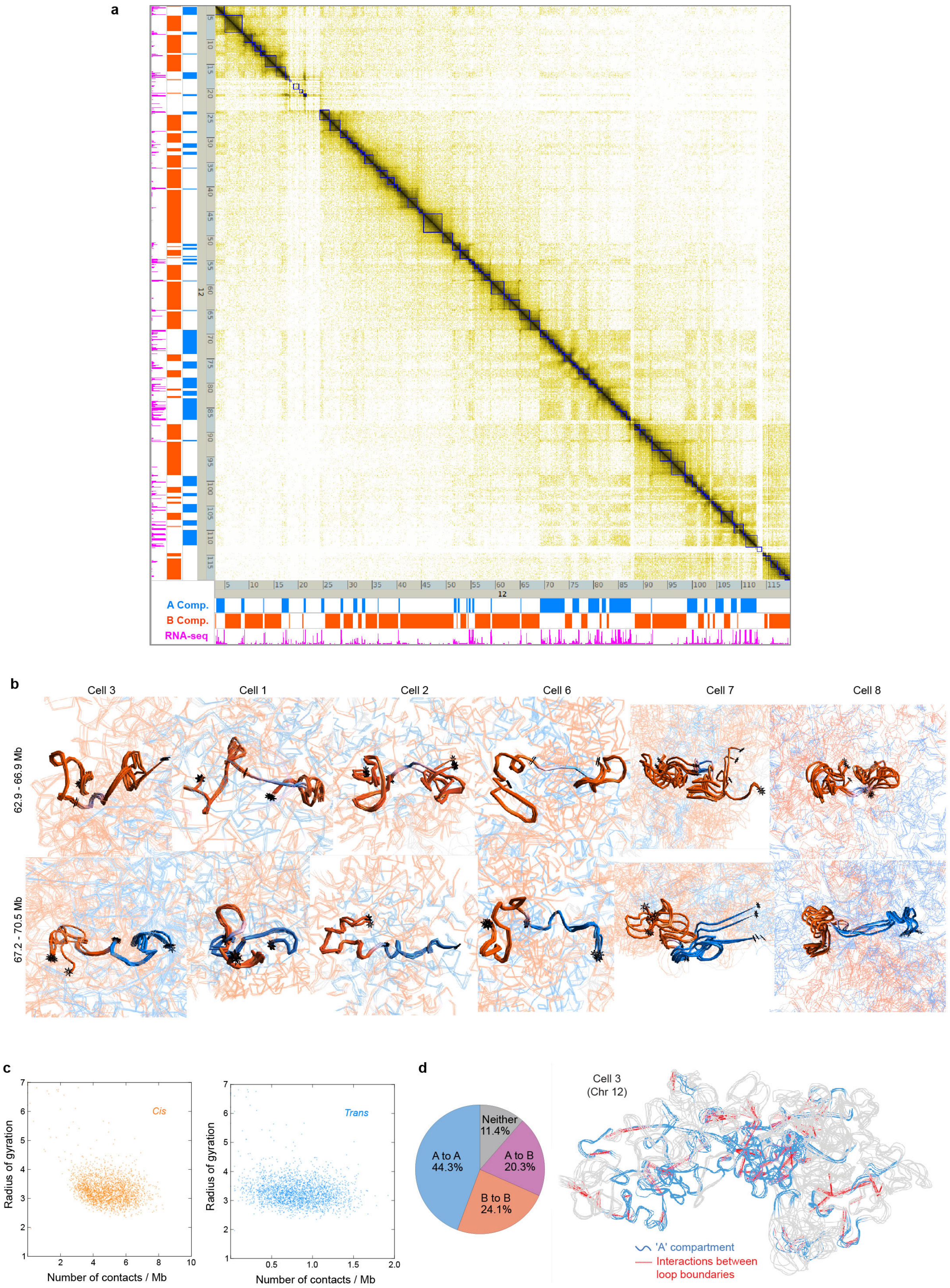
Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Chromosome interactions. a, Violin plot showing the proportion of each chromosome that intermingles with other chromosomes. **b**, Pairwise comparison of the chromosome structure in different cells by r.m.s.d. analysis. Four models of chromosome 9 from a selection of different cells are shown, coloured according to the chromosome sequence (from red to purple, centromere to telomere), together with a table showing the r.m.s.d. values between the chromosomal 3D coordinates for each cell (bottom). **c**, Further cross-sections through the structures of haploid genomes from cells 3–8 through the structures of haploid genomes (see Fig. 2e), coloured according to: whether the sequence is in the A or B compartment (top); whether the sequence is part of a cLAD or contains highly expressed genes (coloured yellow and blue, respectively) (centre); and the identity of the chromosomes (bottom). In each case, an ensemble of five superimposed conformations arising from repeat calculations starting from different randomly generated sets of coordinates is shown. **d**, An analysis of the genome depth of various chromatin class categories, determined by *k*-means clustering of 100-kb segments according to the presence of histone H3 ChIP–seq data[15]. The active class is associated with H3K4me3, Polycomb with H3K27me3, the inactive class with H3K9me3, and null denotes the remainder. Left, the probability distribution for each of the categories at different normalized nucleus depths. Right, the divergence of the probability distribution for each category from the whole-genome average. Data are shown for the genome structures of all cells. **e**, An analysis of the genome depth for regions with differing levels of gene expression, as measured by nuclear RNA-seq. Here, RNA-seq signal peaks were ranked and split into five classes. As in **d**, the probability distribution for each class with regard to genome depth is shown (left), together with the divergence of each distribution from the genome as a whole (right). **f**, Further comparisons of the structure of chromosome 3 from different cells, coloured according to whether the sequence is part of the cLAD domains (yellow), with the positions of highly expressed genes indicated by the blue rings (larger circles indicate higher expression).



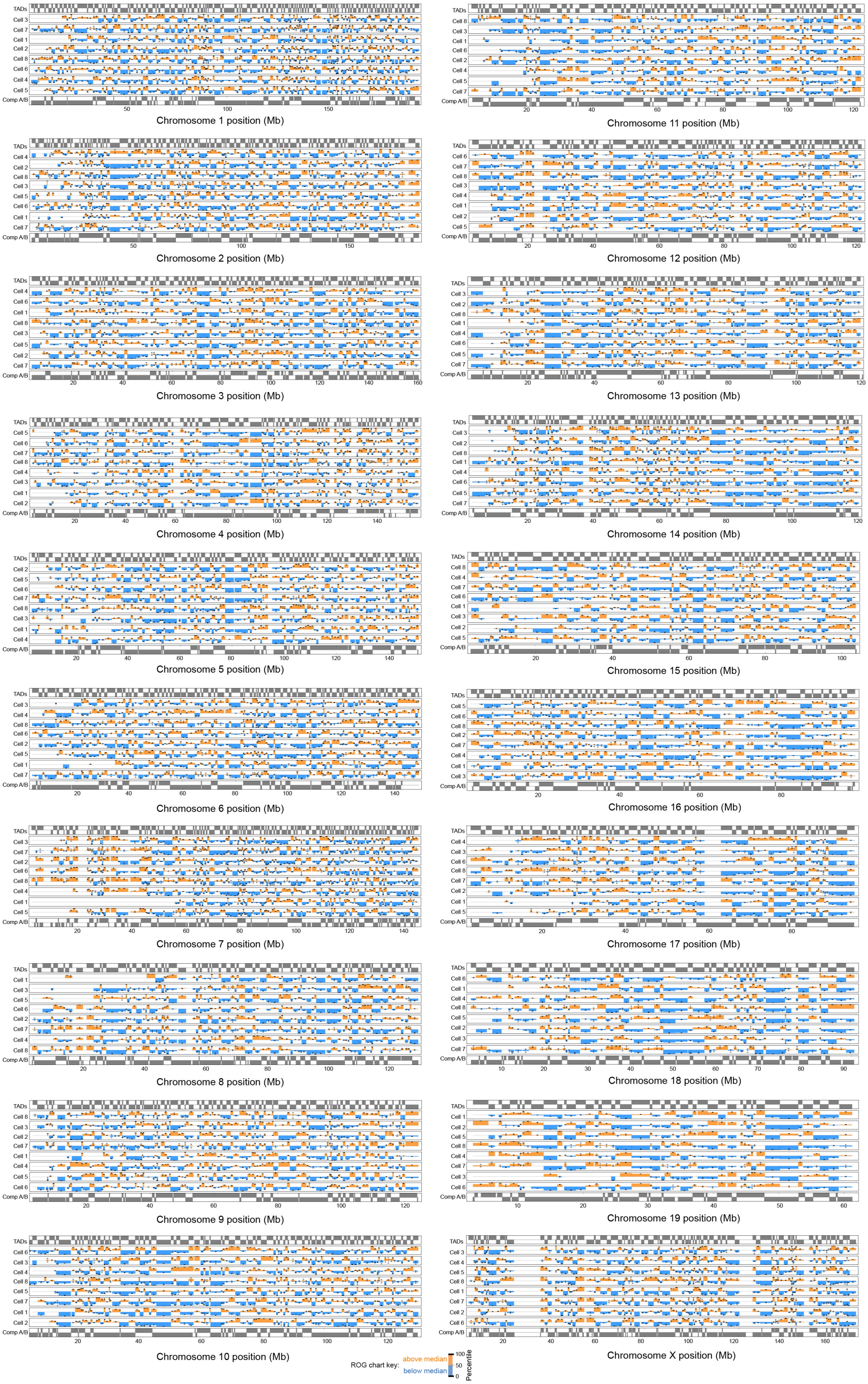
Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | Relationship between genome folding and gene expression. a, Calculation of 3D spatial clustering compared to a random hypothesis in which the same data were circularly permuted around the sequence, and repeating the calculations, using the same structure. Two examples, with strong (Klf4/H3K4me1) and weaker (Nanog/H3K27me3) spatial co-localization compared to random, are shown. **b**, The enrichment in spatial density (after removal of any clustering expected from their being located in the same chromosome sequence) of histone H3 with various post-translational modifications and selected pluripotency factors as determined by ChIP–seq data. The enrichment is calculated over all cells as the Kullback–Liebler divergence of the normalized spatial density distribution from a random, circularly permuted, expectation (see Supplementary Methods for more details), and the data are presented in hierarchical cluster order, grouping the most similar datasets together. **c**, Box and whisker plots showing enhancer, promoter and repetitive sequence content (bottom), and the enrichment in spatial density of different types of enhancer, promoter and repetitive sequence (top), after the data have been divided into ten groups based on increasing distance from the nearest inter-chromosomal interface. Whiskers represent the tenth and ninetieth percentiles, boxes represent the range from the twenty-fifth to the seventy-fifth percentile, and outliers are shown as dots. Mean and median values are shown by black crosses and bars, respectively. The *R* values are the Pearson's correlation coefficient on the underlying, unranked data. **d**, Plots of the level of gene expression as measured by the nuclear RNA-seq signal within 1-Mb regions against distance from the nearest inter-chromosomal interface (left) and the outer surface of the A compartment (right). **e**, Examples of inter-chromosomal interfaces from two different cells in which the chromosomes are coloured increasingly bright red for higher enrichment in the density of gene expression, compared to what would be expected for a given sequence separation. The remainder of the two chromosomes is coloured grey, and the positions of promoters are indicated by blue circles. The same views are shown with the two different chromosomes coloured yellow and blue (top), or with their regions in the A and B compartments coloured blue and red (bottom).

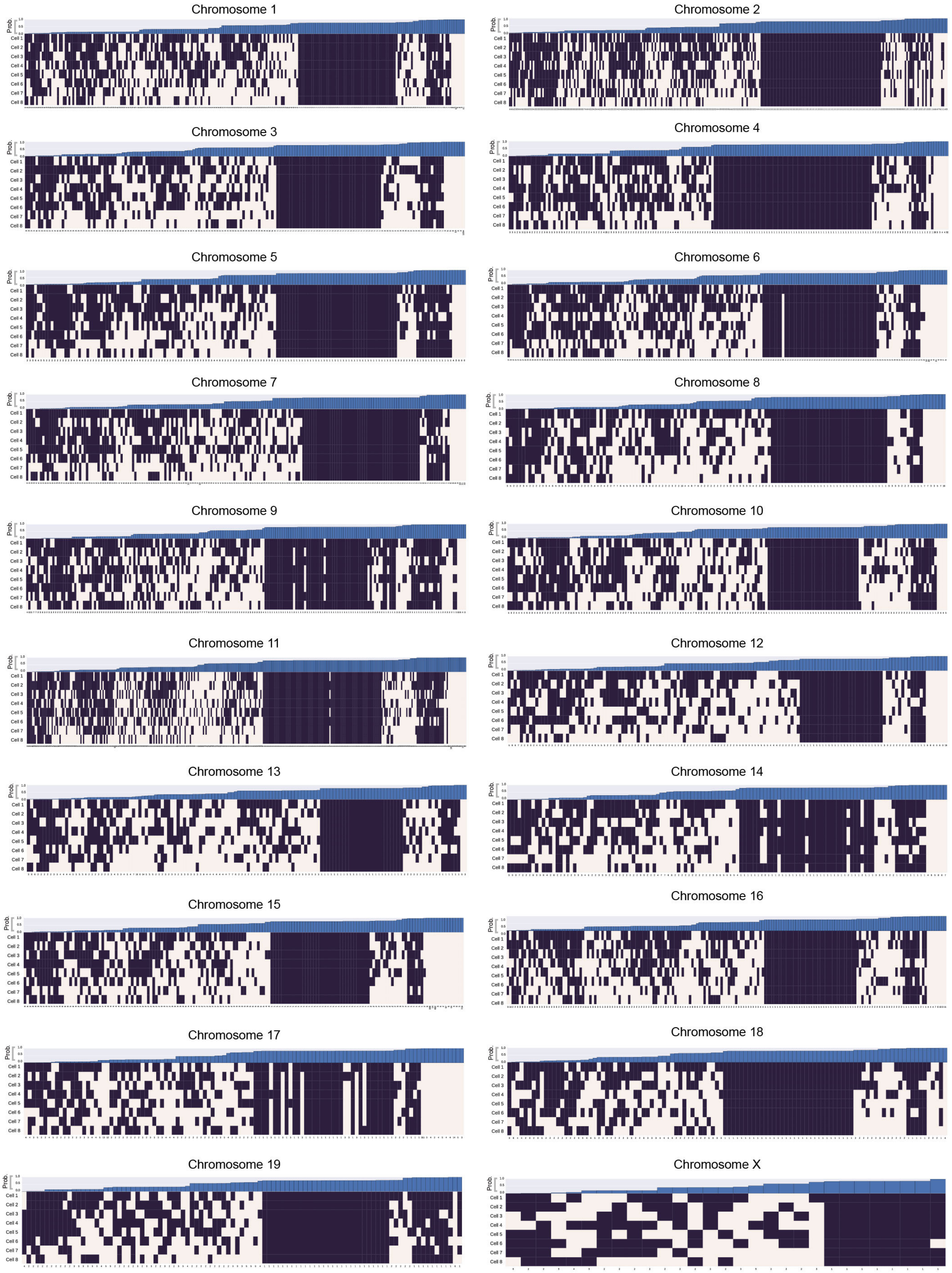


Extended Data Figure 5 | See next page for caption.

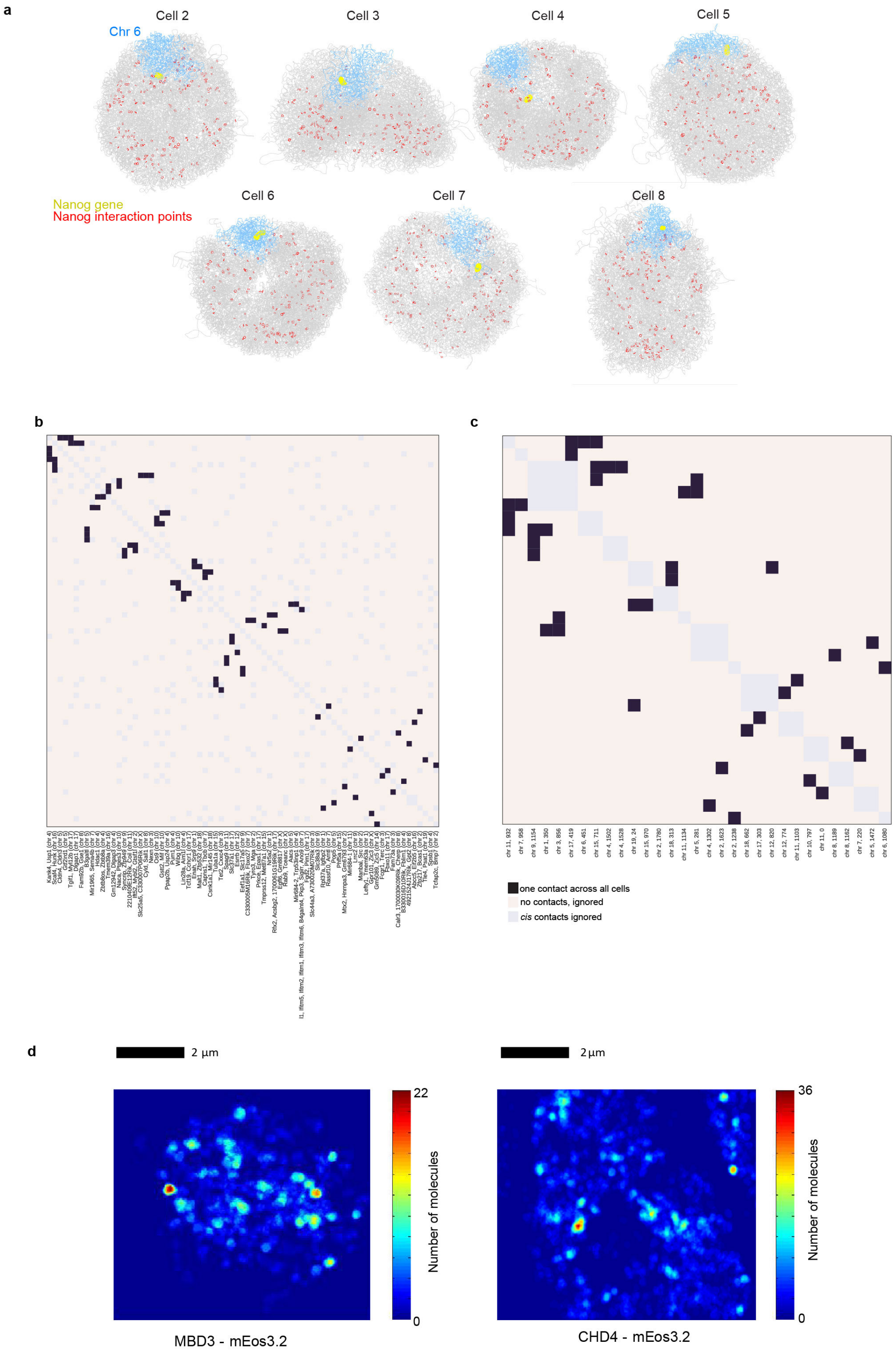
Extended Data Figure 5 | Chromosome folding into compartments, TADs and loops. a, A contact map showing the population Hi-C data for chromosome 12 with TADs identified using the directionality index[5] as blue squares. On the left and at the bottom, data tracks are shown identifying the A and B compartments (in blue and red, respectively), and highly expressed genes (in magenta). **b**, Further comparisons (see Fig. 4b) showing the structures (and their variability) of two B compartment TADs either side of a highly expressed gene(s) in a short region of A compartment (top), or at a boundary between the A and B compartments (bottom). Ensembles of five superimposed conformations, from repeat calculations using the same experimental data, are shown with pairs of TADs highlighted and coloured according to whether they are in the A or B compartments (blue and red, respectively), with white indicating a transitional segment (between A and B). TAD boundaries are marked by asterisks. **c**, Scatter plots of the mean radius of gyration for 1-Mb regions of genome structure compared to the average number of single-cell Hi-C contacts, within the same region, considering a 1-Mb sliding analysis window. Data are shown for all genome structures and split according to *cis* (left) and *trans* (right) contacts. **d**, Structure of chromosome 12, with the A compartment coloured blue and positions of CTCF/cohesin loops identified previously[7] indicated by dotted red lines. The pie chart shows the numbers of loops between sequences in the A and B compartments.



Extended Data Figure 6 | Chromosome folding into TADs. Bar charts of the mean R_g^2 values of TADs identified using the directionality index[5] for all the different chromosomes. The data are mean values over all structure conformations, scaled according to TAD size, and presented as quantile values for the chromosome. The fiftieth percentile value corresponds to the central grey line. Values below and above this are coloured blue and red respectively. TADs that contain both regions of early replication timing (above the ninetieth percentile) and moderate restraint violation (see Extended Data Fig. 2b) are excluded from the calculation. The errors in the R_g^2 are the percentiles ± the s.e.m. Values for multiple cells are presented in hierarchical cluster order, grouping the most similar cells together.



Extended Data Figure 7 | Chromosome folding into loops. A genome-wide analysis illustrating whether CTCF/cohesin loops[7] could be formed in the different single cells, in each chromosome. A black square indicates that the two boundaries in the loop could interact, a white square indicates that the two relevant particles are too far apart in the structure. The loop boundary separation, in particles, is shown along the *x* axis. The bar chart across the top shows the probability, for each loop, of random particles (pairs with the same sequence separation) forming the same number of contacts, or better. The probability of choosing a set of loop boundary points, which interact more frequently than we observed is 0.00072 (see Supplementary Methods).



Extended Data Figure 8 | Understanding the nature of gene networks in mouse ES cells. a, Structures of cells 2–8 illustrating the interactions identified between the *Nanog* gene and other regions of the genome by population 4C (ref. 34). Chromosome 6 is coloured blue, with the position of the *Nanog* gene highlighted in yellow, and the remainder of the chromosomes are coloured grey. Interacting positions in the genome are indicated by red circles. **b**, Heat map showing the number of times a particular interaction is detected between two of the 4C *Nanog* gene-interacting points[35]. **c**, Heat map showing the number of times a particular interaction is detected between two of the 4C *Pou5f1* gene-interacting points[36]. In **b** and **c**, the interaction points are presented in hierarchical order, grouping the regions that show the most interactions together. **d**, 2D single-molecule tracking using photo-activated light microscopy in live mouse ES cells shows clustering of CHD4 and MBD3. In both cases, a heat map of a single cell is shown in which the pixels have been colour-coded according to the density of molecules detected in that region.

Extended Data Table 1 | Summary of data for the eight cells analysed, and statistics for the sequence analysis

Cell	Input read pairs	Unique mapped pairs	Primary contacts	Final contacts*	Normal ligation %	Single read %	Trans %	Promiscuous ends	Mean redundancy
1	1,969,076	1,235,949	110,042	122,475	89.36	21.5	11.65	4342	7.88
2	1,621,648	944,140	65,636	84,129	92.29	21.7	6.30	2320	7.70
3[§]	1,937,061	1,326,834	32,534	37,604	68.67	14.0	22.66	1826	24.15
4	1,517,614	704,831	75,740	92,748	89.99	23.7	6.64	771	6.39
5	1,592,161	883,678	61,855	74,417	89.86	21.2	9.67	1833	10.12
6	1,493,430	643,721	60,334	75,352	92.21	21.7	5.21	699	7.70
7[†]	4,796,232	1,191,086	46,438	55,792	80.60	16.5	5.51	1655	17.26
8[†]	1,776,396	810,661	35,157	42,586	90.40	20.13	7.42	1423	16.00

**Final contact counts include those with genome sequence mapping ambiguities that were resolved by initial 100-kb resolution genome structure calculations. [†]These cells have fluorescent image data. [§]Contacts in this cell were sequenced using the Nextera protocol.*

Extended Data Table 2 | Summary of statistics from the genome structure calculation process for the eight cells analysed#

Cell	400-kb Particles				200-kb Particles				100-kb Particles (without ambiguous restraints)				100-kb Particles (with ambiguous restraints)			
	RMSD*	N_{Rest}†	%V§ >3	%V§ >4	RMSD*	N_{Rest}†	%V§ >3	%V§ >4	RMSD*	N_{Rest}†	%V§ >3	%V§ >4	RMSD*	N_{Rest}†	%V§ >3	%V§ >4
1	0.13	26,453	2.94	0.76	2.32	43,150	1.29	0.08	0.56	71,729	2.81	0.57	0.41	74,945	2.29	0.22
2	0.29	22,838	0.16	0.00	0.50	35,244	0.29	0.00	0.88	47,598	1.04	0.15	0.63	56,159	0.74	0.13
3	0.24	17,944	0.81	0.28	0.51	21,049	1.66	0.08	0.89	22,822	4.82	1.36	0.70	25,952	2.87	0.42
4	0.33	25,126	2.35	0.68	0.58	37,594	1.26	0.14	1.03	50,516	2.49	0.94	0.74	60,420	1.14	0.17
5	0.32	23,062	1.18	0.30	0.65	34,103	0.86	0.10	1.08	44,298	2.12	0.73	0.86	51,824	1.24	0.21
6	0.38	25,084	0.60	0.10	0.69	35,526	0.70	0.06	1.09	44,687	0.99	0.22	0.91	53,710	0.74	0.17
7	0.81	18,301	0.07	0.04	1.26	26,585	0.03	0.00	1.91	33,588	0.11	0.01	1.75	39,257	0.13	0.02
8	0.49	15,877	0.00	0.00	0.93	22,956	0.00	0.00	1.52	28,083	0.02	0.02	1.26	32,482	0.03	0.02

*$^{\#}$ The separate statistics for calculations involving 400-kb to 100-kb particles relate to the different stages of the hierarchical structure calculation protocol (see Supplementary Methods for details), with the last section representing the final, highest quality structures. *The precision of structure ensembles in terms of mean pairwise all-particle RMSDs, in units of particle radii. †Restraint numbers include all experimental Hi-C contacts between non-adjacent particles, and thus increase as the particle size decreases. (NB – A single restraint may correspond to multiple Hi-C contacts, i.e. between the same two particle regions.) §The percentage of violated restraints was calculated for thresholds of 3.0 and 4.0 particle radii, respectively.*