ARTICLE

doi:10.1038/nature12593

# Single-cell Hi-C reveals cell-to-cell variability in chromosome structure

Takashi Nagano[1]*, Yaniv Lubling[2]*, Tim J. Stevens[3]*, Stefan Schoenfelder[1], Eitan Yaffe[2], Wendy Dean[4], Ernest D. Laue[3], Amos Tanay[2] & Peter Fraser[1]

**Large-scale chromosome structure and spatial nuclear arrangement have been linked to control of gene expression and DNA replication and repair. Genomic techniques based on chromosome conformation capture (3C) assess contacts for millions of loci simultaneously, but do so by averaging chromosome conformations from millions of nuclei. Here we introduce single-cell Hi-C, combined with genome-wide statistical analysis and structural modelling of single-copy X chromosomes, to show that individual chromosomes maintain domain organization at the megabase scale, but show variable cell-to-cell chromosome structures at larger scales. Despite this structural stochasticity, localization of active gene domains to boundaries of chromosome territories is a hallmark of chromosomal conformation. Single-cell Hi-C data bridge current gaps between genomics and microscopy studies of chromosomes, demonstrating how modular organization underlies dynamic chromosome structure, and how this structure is probabilistically linked with genome activity patterns.**

Chromosome conformation capture[1] (3C) and derivative methods (4C, 5C and Hi-C)[2–6] have enabled the detection of chromosome organization in the three-dimensional space of the nucleus. These methods assess millions of cells and are increasingly used to calculate conformations of a range of genomic regions, from individual loci to whole genomes[3,7–11]. However, fluorescence *in situ* hybridization (FISH) analyses show that genotypically and phenotypically identical cells have non-random, but highly variable genome and chromosome conformations[4,12,13], probably owing to the dynamic and stochastic nature of chromosomal structures[14–16]. Therefore, although 3C-based analyses can be used to estimate an average conformation, it cannot be assumed to represent one simple and recurrent chromosomal structure. To move from probabilistic chromosome conformations averaged from millions of cells towards determination of chromosome and genome structure in individual cells, we developed single-cell Hi-C, which has the power to detect thousands of simultaneous chromatin contacts in a single cell.

## Single-cell Hi-C

We modified the conventional or 'ensemble' Hi-C protocol[3] to create a method to determine the contacts in an individual nucleus (Fig. 1a and Supplementary Information). We used male mouse spleenic $CD4^+$ T cells, differentiated *in vitro* to T helper ($T_H1$) cells to produce a population of cells (>95% $CD4^+$), of which 69% have 2*n* genome content, reflecting mature cell withdrawal from the cell cycle. Chromatin crosslinking, restriction enzyme (BglII or DpnII) digestion, biotin fill-in and ligation were performed in nuclei (Fig. 1a and Extended Data Fig. 1a) as opposed to ensemble Hi-C, in which ligation is performed after nuclear lysis and dilution of chromatin complexes[3]. We then selected individual nuclei under the microscope, placed them in individual tubes, reversed crosslinks, and purified biotinylated Hi-C ligation junctions on streptavidin-coated beads. The captured ligation products were then digested with a second restriction enzyme (AluI) to fragment the DNA, and ligated to customized Illumina adapters with unique 3-bp (base pair) identification tags. Single-cell Hi-C libraries were then PCR amplified, size selected and characterized by multiplexed, paired-end sequencing.

De-multiplexed single-cell Hi-C libraries were next filtered thoroughly to systematically remove several sources of noise (Extended Data Fig. 1b–f and Supplementary Information). Hi-C in male diploid cells can theoretically give rise to at most two ligation products per autosomal restriction fragment end, and one product per fragment end from the single X chromosome. Using BglII, the total number of distinct mappable fragment-end pairs per single cell cannot therefore exceed 1,201,870 (Extended Data Fig. 1g and Supplementary Information). In practice, deep sequencing of the single-cell Hi-C libraries demonstrated that following stringent filtering our current scheme allows recovery of up to 2.5% of this theoretical potential, and has identified at least 1,000 distinct Hi-C pairings in half (37 out of 74) of the cells. Deep sequencing confirmed saturation of the libraries' complexity, and allowed elimination of spurious flow cell read pairings and additional biases (Extended Data Tables 1–3). On the basis of additional quality metrics we selected ten single-cell data sets, containing 11,159–30,671 distinct fragment-end pairs for subsequent in-depth analysis (Extended Data Fig. 1h–l). Visualization of the single-cell maps suggested that despite their inherent sparseness, they clearly reflect hallmarks of chromosomal organization, including frequent *cis*-contacts along the matrix diagonal and notably, highly clustered *trans*-chromosomal contacts between specific chromosomes (Fig. 1b).

## Single-cell and ensemble Hi-C similarity

We used the same population of $CD4^+$ $T_H1$ cells to generate an ensemble Hi-C library. Sequencing and analysis[17] of 190 million read pairs produced a contact map representing the mean contact enrichments within approximately 10 million nuclei. The probability of observing a contact between two chromosomal elements decays with linear distance following a power law regime for distances larger than 100 kilobases (kb)[3,18]. We found similar regimes for the ensemble, individual

[1]Nuclear Dynamics Programme, The Babraham Institute, Cambridge CB22 3AT, UK. [2]Department of Computer Science and Applied Mathematics and Department of Biological Regulation, Weizmann Institute, Rehovot 76100, Israel. [3]Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA, UK. [4]Epigenetics Programme, The Babraham Institute, Cambridge CB22 3AT, UK.
*These authors contributed equally to this work.

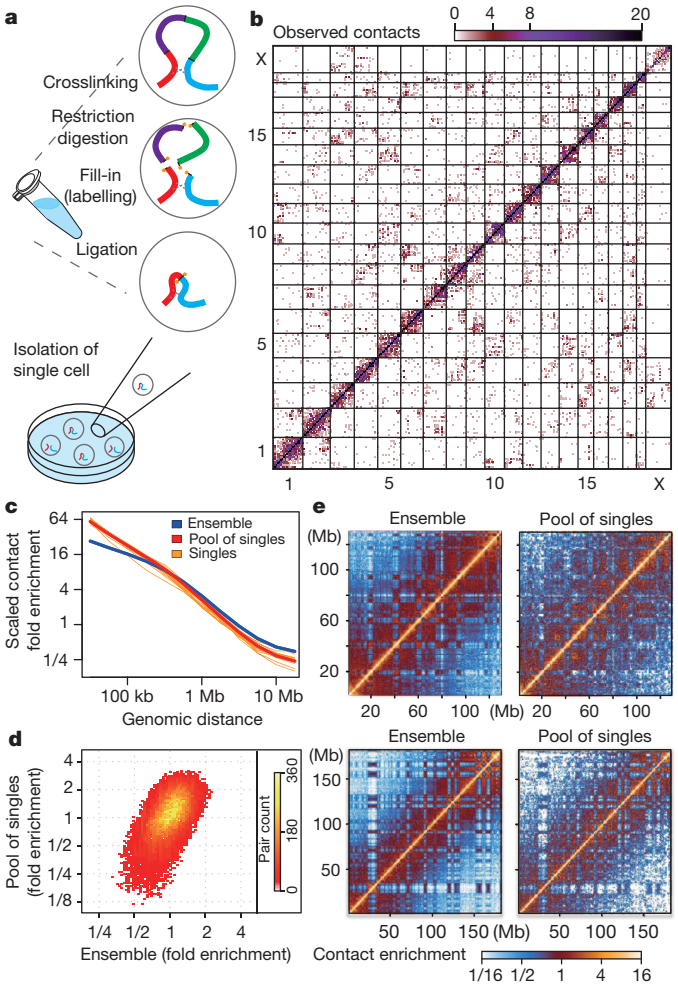



**Figure 1 | Single-cell and ensemble Hi-C.** **a**, Single-cell Hi-C method. **b**, Single-cell Hi-C heatmap (cell 5), coverage for 10-Mb bins. **c**, Contact enrichment versus genomic distance, from ensemble Hi-C, pool of 60 single cells and 10 individual cells, scaled to normalize sequencing depths. **d**, Normalizing by the trends in **c**, intra-chromosomal contact enrichments for 1-Mb square bins, comparing ensemble and pooled single-cell Hi-C (Spearman correlation = 0.56). **e**, Intra-chromosomal contact enrichment maps of ensemble and pooled single-cell Hi-C, for chromosome 10 (top) and chromosome 2 (bottom), using variable bin sizes.

cells and a pool of 60 single cells (Fig. 1c). Moreover, after normalizing the matrices given this canonical trend, comparison of intra-chromosomal interaction intensities for the pool and ensemble, by global correlation analysis of contact enrichment values at 1-megabase (Mb) resolution generates a highly significant correspondence (Fig. 1d). This is emphasized by the high similarity observed in comparisons of individual chromosomes from ensemble and pooled Hi-C maps (Fig. 1e). In summary, despite different experimental procedures and sparse nature of the single-cell matrices, the pooled matrix retains the most prominent properties of the ensemble map, confirming the validity of the approach and prompting us to explore further the similarities and differences among the individual cell chromosomal conformations.

## Intra- and interdomain contacts

A key architectural feature of ensemble Hi-C data sets is their topological domain structure[18–20]. As expected, 1,403 domains were identified in the $T_H1$ cell ensemble Hi-C map[18] (Supplementary Table 1 and Supplementary Information). We used the ensemble domains to ask whether the same domain structure can be observed at the single-cell level. Visual inspection of the domain structure overlaid on individual intra-chromosomal contact maps (Fig. 2a), and global statistical analysis of the ratios between intra- and interdomain contact intensities in individual cells (approximately twofold enrichment on scales of 100 Kb to 1 Mb; Fig. 2b and Extended Data Fig. 2a), both supported the idea that domains are observed consistently in the single-cell maps. To test whether domain structures are variable between individual cells, we estimated the distributions of intradomain contact enrichments across cells and compared it to the distributions derived from reshuffled maps. We reasoned that cell-to-cell variation in intradomain contact intensities would result in an increase of the variance of this distribution compared to the expected variance resulting from sampling contacts in uniformly (shuffled) intradomain contacts. However, the data (Fig. 2c) showed that the distributions for the intradomain enrichments in real cells are not more varied than expected (Kolmogorov–Smirnov $P < 0.52$). A similar observation was derived by comparison of the correlations between intradomain contact enrichments for pairs of real and pairs of reshuffled maps (Extended Data Fig. 2b). Although this analysis cannot quantify variability in the high-resolution internal structure of domains, the data suggest that domain intactness is generally conserved at the single-cell level.

Visual comparison of whole-chromosome contact maps (Fig. 2d) suggested that unlike intradomain interactions, interdomain contacts within single-cell chromosomes are structured non-uniformly. The maps showed large-scale structures as indicated, for example, by specific insulation points separating chromosomes into two or more mega domains in a cell-specific fashion. To rule out the possibility that this can be explained by sparse sampling of contacts in each single-cell map we again used reshuffled controls. In each map (real or randomized) we quantified the frequency of loci that strongly polarize the matrix into two weakly connected submatrices (using an insulation score; Supplementary Methods). We confirmed that single-cell maps indeed show many more such loci than reshuffled maps (Fig. 2e and Extended Data Fig. 2c). The reshuffled controls made by mixing contacts from different single-cell maps, are in fact similar to sparse versions of the ensemble map, which do not show specific structure at the intradomain level. Along similar lines, the correlation in contact intensities between domains on the same chromosome in pairs of single-cell maps is lower compared to reshuffled controls (Fig. 2f). Taken together, these data show that domains form a robust and recurrent conformational basis that is evident in each of the single cells. However, interdomain contacts are highly variable between individual cells, suggesting large-scale differences in higher-order chromosome folding that are obscured in ensemble maps, averaged over millions of such structures[21].

## Three-dimensional modelling of X chromosomes

To determine whether the single-cell Hi-C data are consistent with unique chromosome conformations we developed a modelling approach to reconstruct the conformations of the single-copy, male X chromosome. We used intra-chromosomal contacts as distance restraints and calculated structural models using a simulated annealing protocol to condense a particle-on-a-string representation of individual chromosomes from random initial conformations (Supplementary Information), to produce both fine-scale and low-resolution models, with backbone particles representing either 50 or 500 kb of the chromosome, respectively. For fine-scale calculations, each intra-chromosomal contact restrained its precise position on the chromosome, whereas low-resolution calculations combined contacts into larger bins. Tests of our simulation protocol demonstrated that restraint density was the most important parameter for modelling (Extended Data Fig. 3a, b). Hence, from the ten high-quality single-cell data sets, we selected six with the largest numbers of intra-chromosomal X contacts, plus one with a lower number of contacts (cell 9) for contrast.

Repeat calculations starting from random positions generated 200 X-chromosome models for each cell at both scales. The fine-scale models displayed very low numbers of restraint violations (Extended Data Fig. 3c). We introduced an estimated average unit DNA distance

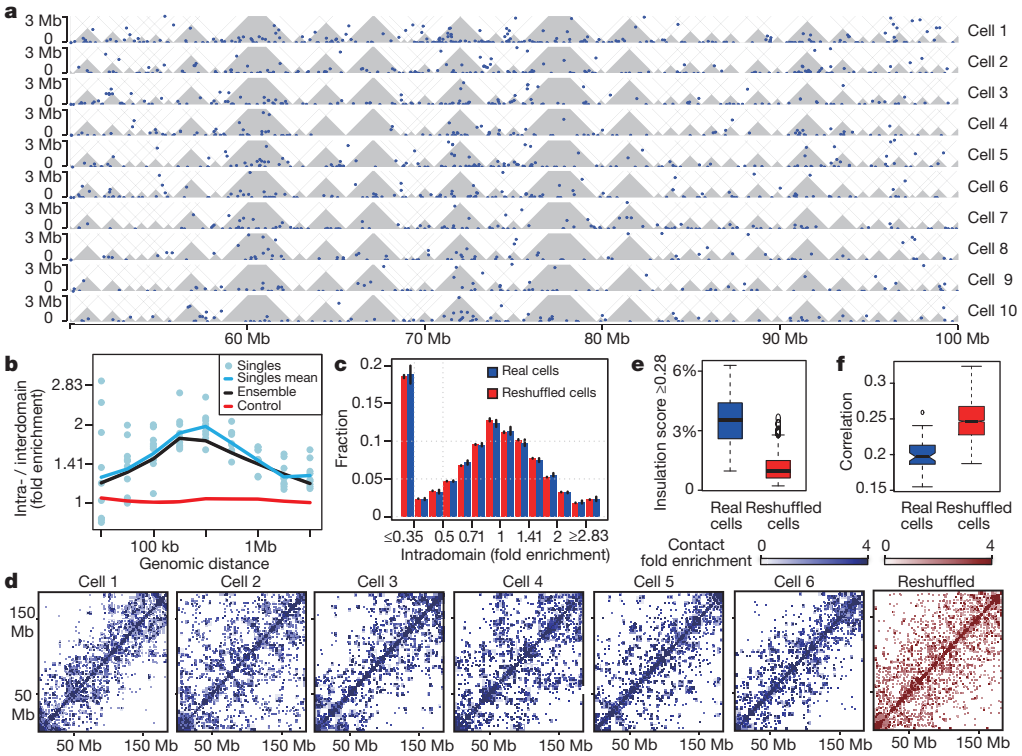



**Figure 2 | Conserved intradomain, but not interdomain structure in single cells.** **a**, Intra-chromosomal contact maps of 50-Mb region of chromosome 2 from ten single cells. Individual contacts up to 3-Mb distance are shown as blue dots. Domains inferred from the ensemble Hi-C maps in grey. **b**, Ratios between intradomain and interdomain contact enrichments over genomic distance. Control is combined trend of 10 single cells calculated by repeatedly shifting the domains randomly. **c**, Distribution of intradomain contact enrichments per domain from 9 cells (where BglII was used) and reshuffled data sets (black bars, s.e.m.). **d**, Maps of interdomain contact intensities for chromosome 2 from individual cells and reshuffled controls using variable bin sizes. **e**, Distribution of percentage of loci with high insulation scores in single versus reshuffled cells. **f**, For all pairs of single cells, the correlations between interdomain contact numbers of all pairs of domains within the same chromosome were computed. Shown are the distributions of these correlations in the real and reshuffled cells.

length[22] to approximate packaging of chromatin fibres (~0.15 μm per 50 kb) (Supplementary Information). This resulted in models with a mean X-chromosome territory diameter of 4.3 μm (range 3.3–5.9 μm), in good agreement with X-chromosome paint FISH in $T_H1$ cells (Fig. 3a; mean diameter 3.7 μm) and chromosome territory sizes in live cells[23]. We confirmed that the restrained points in a single cell are indeed close in the structures calculated from them (Extended Data Fig. 3c, d). Interestingly, the single-cell distance matrix demonstrates how the network of contacts in a model imparts further structural information beyond the directly observed contacts (Extended Data Fig. 3d).

Comparison of the low-resolution models demonstrated convergence towards a single conformation for each single-cell data set (Fig. 3b and Extended Data Fig. 3e). For fine-scale models, hierarchical clustering revealed four or five that were most representative of the data (Fig. 3c). In all cases models from a single cell were significantly more similar to each other than to models from different cells (Extended Data Fig. 4a, b).

Highlighting four regions of the X chromosome showed large-scale conformational differences between cells (Fig. 3d), supporting the finding of highly variable interdomain contacts. Models created by shuffling Hi-C contacts, or combining contacts from two cells resulted in structures smaller and more compact than observed chromosome territories (Extended Data Fig. 4c, d) with many restraints stretched towards or exceeding their upper bounds (Extended Data Fig. 3c). These results reaffirm that the variation in single-cell contacts is not the result of partial sampling of a single underlying structure.

We next asked whether despite their cell-to-cell variability, X-chromosome structures share common folding properties that could be tested in real cells. One such important property, which is often consistent within a cell population, and with multiple potential functional implications, is localization within the chromosomal territory relative to its surface. To predict loci with consistent positions within their chromosome territory we calculated the structural density along the X chromosome (Supplementary Information) and identified regions with consistently high or low structural density (Fig. 3e). We chose five such regions (P1–P5) with predicted positions near the surface (P1, P2, P5; low structural density) or inside (P3, P4; high structural density) the model X-chromosome territories using the 1,200 models from the six cells (Extended Data Fig. 4e). We then performed double label DNA FISH with X-chromosome paints and P1–P5 bacterial artificial chromosome (BAC) probes (Fig. 3f) to test directly these predictions. The distances between DNA FISH signals and edge of the chromosome territory in over one hundred $T_H1$ cells showed that probes P1, P2 and P5 were indeed found predominantly outside or towards the edge of the chromosome territory, whereas signals for probes P3 and P4 were found at internal positions (Fig. 3g). These data show that despite highly variable interdomain structure of the X-chromosomal territory, some of its key organizational properties are robustly observed across the cell population.

## Domains at the interface

Data from *trans*-chromosomal contacts were overlaid on the X-chromosome models, and this showed that *trans*-chromosomal contacting regions are strongly enriched towards the inferred surface of the models (Fig. 3h), providing further validation. These observations prompted us to explore further the structural characteristics of interfaces between chromosomal territories, and the relationships between such interfaces and the domain structure of the territory itself. We found that *trans*-chromosomal contact enrichments of domains vary across cells (Fig. 4a), showing a significant difference between the mean contact enrichment per domain in the real and reshuffled maps ($P < 1.2 \times 10^{-9}$, Kolmogorov–Smirnov test). The higher variance of the distribution for the real data suggests that some domains are more likely to contact elements on other chromosomes. Previous work has suggested that active genomic regions on the sub-domain scale often loop out of their chromosome territories[24], which may imply less defined local domain structures and disassociation from their chromosome territory. However, our analysis shows that *trans*-contacting domains retain domain organization, as demonstrated by the intradomain contact probabilities within them (Fig. 4b and Extended Data Fig. 5a, b). Conversely, *trans*-contacting domains show slightly reduced

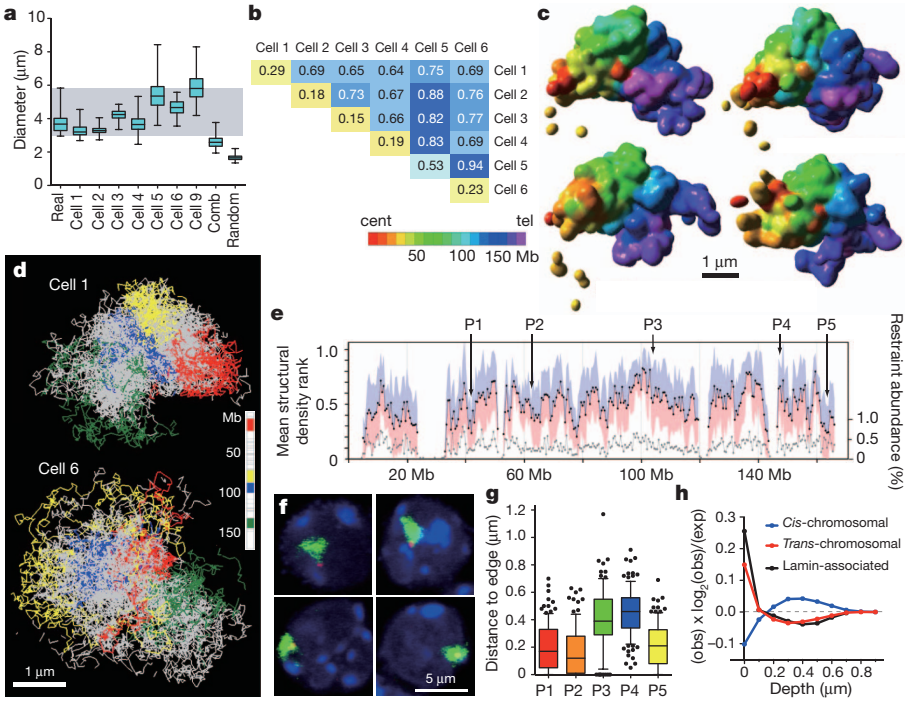



**Figure 3 | Structural modelling of X chromosomes.** **a**, Distribution of longest diameter of X-chromosome paint DNA FISH signals in 62 male $T_H1$ cells (real), 200 structural models calculated for each single cell (cell 1 to cell 9), 200 structures from combined data set (cell 1 and cell 2; comb) and 200 structures from 20 randomized cell 1 data sets (random; 10 calculations per data set). Whiskers denote minimum and maximum. **b**, Average coordinate root-mean-square deviation (r.m.s.d.) values in microns comparing 200 low-resolution structural models for each cell and between cells. **c**, Four surface-rendered models of the X chromosome from cell 1, which are most representative of the data based on hierarchical clustering of pair-wise r.m.s.d. values (Supplementary Information). Scale bar, 1 μm. **d**, Structural ensembles of the four most representative fine-scale models for cell 1 and cell 6, with four large regions coloured. Scale bar, 1 μm. **e**, Mean structural density rank for 500-kb regions (black) from 6 × 200 fine-scale models from cell 1 to cell 6. Standard deviation is shown in blue and pink. Abundance of intra-chromosomal restraints is shown in grey, right axis. DNA FISH probes (P1–P5) are indicated. **f**, DNA FISH on $T_H1$ cells. X-chromosome paint (green) and specific locus signals (red). **g**, Distribution of DNA FISH distance measurements between signal centres for probes P1–P5 and edge of the X-chromosome territory in $T_H1$ cells ($n = 114, 113, 105, 115$ and 108 for P1–P5, respectively). Whiskers denote 10 and 90 percentiles. **h**, Enrichment of *cis*-, *trans*-contacts and lamin-B1-associated domains at various depths on the chromosome models relative to null hypothesis of random positions.

contact intensity to other domains on the same chromosome (Fig. 4c and Extended Data Fig. 5c, d), consistent with localization on the interfaces of their territories rather than dissociation from them.

Analyses of ensemble Hi-C data have previously shown that active marks correlate with enrichment of *trans*-chromosomal contacts[3,17]. Using the single-cell maps combined with annotation of domains based on their enrichment for histone H3 lysine 4 tri-methylation (H3K4me3) hotspots[25] (Fig. 4d), we tested whether this correlation is the result of low frequency re-localization of active domains to other chromosome territories (looping out), or from frequent localization of active domains on territory interfaces. As shown in Fig. 4e, domains with high *trans*- to *cis*-chromosomal contact ratios (excluding intradomain) are highly correlated with H3K4me3 enrichment in all cells. However, the data show that domains (including active ones) retain their association with the territory in almost all cases. Very few domains with strong *trans*-contacts were found to lack association with their own territory (Fig. 4e; upper left points in graphs). Some of this lack of perfect territory re-localization can be explained by having two copies of each autosomal domain, but the overall reduction with territory association for *trans*-contacting domains is much smaller than the 50% expected by this explanation (reduction estimated at 15–20% and 10% for contacts across 1–5 Mb and 10 Mb, respectively, Fig. 4c). Comparison of active domain localization shows that different active domains are highly *trans*-contacting in each cell (Extended Data Fig. 5e). Together, these data show that preferential localization of active domains to territory interfaces is a hallmark of chromosome organization in all cells. Active domains maintain their intradomain organization, and only partially lose intra-chromosomal contacts with other domains. Our data are consistent with the concept that chromosomal territories are maintained robustly despite the *trans*-chromosomal contacts between active domains.

Interestingly, domains associated with lamin B1 (ref. 26), which are thought to be primarily inactive regions, are also found towards the surface of the models (Fig. 3h). However, these domains are highly anti-correlated with H3K4me3 domains (Spearman's correlation = −0.73) and typically depleted of *trans*-chromosomal contacts (Extended Data Fig. 5f–i). Superposition of H3K4me3, lamin-B1-enriched domains and *trans*-chromosomal contacts on the X-chromosome models illustrates spatial partitioning of the active, *trans*-contacting regions from those that are lamin-associated, although both types of domains tend towards the surface of the chromosome territory, supporting the above descriptions of differential positioning of domains (Extended Data Fig. 5j and Supplementary Videos 1 and 2).

Ensemble Hi-C maps generate a highly complex view of chromosomal contacts, including low-intensity contacts between all possible chromosomal pairings[3,8,17,19]. In contrast, years of single-cell analyses by microscopy have suggested that individual cells have much simpler and discrete chromosome structures involving a limited number of interfaces between spatially constrained chromosomal territories[27,28]. Our single-cell maps bridge the gap between the genomic and imaging techniques, showing cell-specific clusters of *trans*-chromosomal contacts associating some pairs of chromosomes, and a lack of contacts between other chromosome pairs (Fig. 5a, blue). Such organization is completely lacking in reshuffled maps (Fig. 5a, red) confirming it is not a consequence of sparse contact sampling (Extended Data Fig. 6a, b). *Trans*-chromosomal contact clusters bring pairs of domains together, as shown by comparing the enrichment in *trans*-contacts between pairs of elements connecting the same two domains and pairs connecting

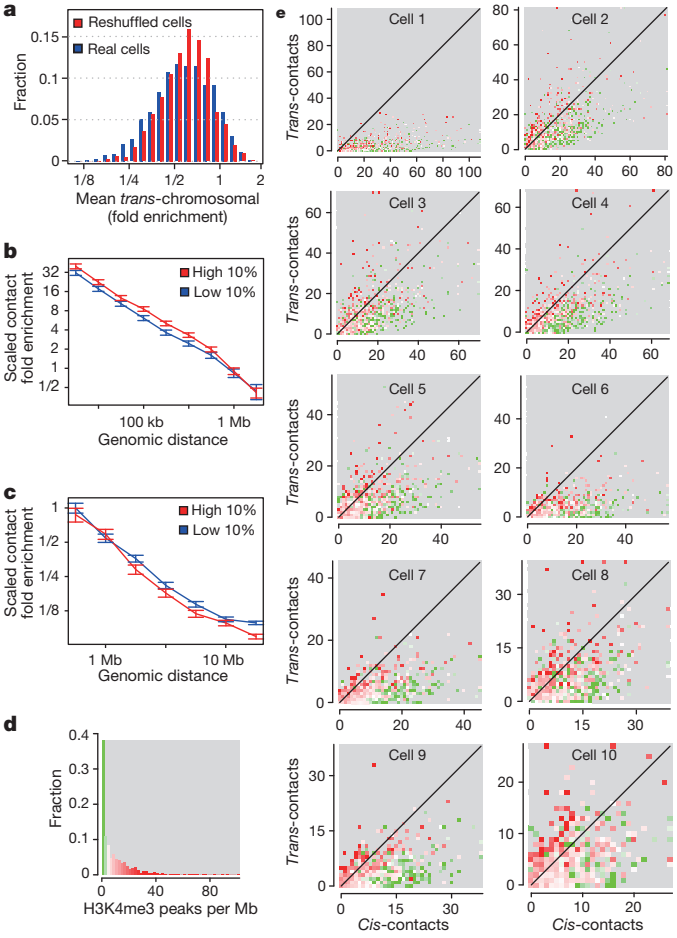



**Figure 4 | Active domains localize to territory interfaces.** **a**, Distribution of *trans*-chromosomal contact enrichments of each domain averaged across real and reshuffled cells. Reshuffling maintains the number of *cis*- and *trans*-contacts within each cell and chromosome. **b**, Intradomain contact enrichment over genomic distance for high versus low *trans*-chromosomal contacting domains selected independently in each cell, with 95% confidence intervals. **c**, Same sets as in **b** but plotting the enrichment of interdomain contacts. **d**, Distribution of H3K4me3 peak density in domains (number of peaks divided by size), colour-coded according to density. **e**, Domains plotted according to number of *trans*- and *cis*-chromosomal (excluding intradomain) contacts, colour coded for H3K4me3 density as in **d**.

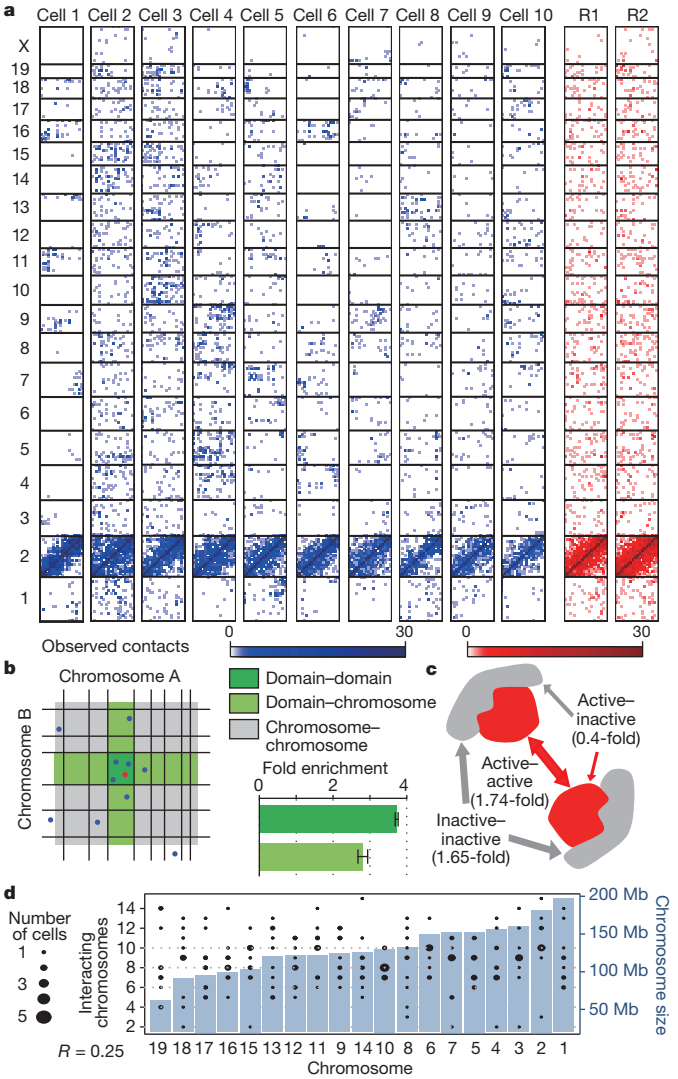



**Figure 5 | Chromosomal interfaces.** **a**, All *trans*-chromosomal contacts formed by chromosome 2 in real cells (blue) and reshuffled cells (red). **b**, Schematic diagram of a chromosomal interface between linearly adjacent domains, their borders marked in black on two chromosomes, A and B. We considered each of the two contacting fragments of every *trans*-chromosomal contact and classified every nearby *trans*-chromosomal contact as domain–domain, domain–chromosome and chromosome–chromosome, the latter being used as background for normalization (Supplementary Information). The contact under consideration is shown in red, and nearby contacts are shown in blue. Fold enrichments shown for each group type (error bars, standard deviation). **c**, *Trans*-chromosomal contacts are highly significantly enriched between active domains (H3K4me3 enriched) or between inactive domains, but not mixed interaction (chi-squared test; $P = 5.8 \times 10^{-18}$; even after taking account of the generally higher connectivity of active domains). **d**, Bar graph depicting mouse autosomes ordered by size with number of interacting chromosomes per single cell (black circles depict the distribution over individual cells). Mean number of interacting chromosomes changes modestly (30%) with chromosome size, suggesting a highly organized territory structure with surface that is not scaling with chromosome length.

one domain with two different domains (Fig. 5b). Such synergistic contacting preferentially brings together pairs of active domains, with interaction between active and inactive domains being underrepresented (Fig. 5c and Extended Data Fig. 6c). Although inactive domains are depleted as a group from *trans*-chromosomal interactions (Fig. 4e), inactive domains that engage in *trans*-contacts are more likely to interact with other inactive domains. Interestingly, analysis of interacting pairs of domains suggests that the number of chromosomes contacting each chromosome is relatively constant (less than 30% difference) despite the greater than threefold change in chromosome size, the total number of *trans*-chromosomal contacts in the map, or a number of other factors (Fig. 5d and Extended Data Fig. 7a–e). We note that even though the total number of chromosome–chromosome interfaces per single cell is bounded, the detailed interface between chromosome pairs can involve multiple domain–domain contacts reflecting higher-order organization (Extended Data Fig. 7f).

Overall, these results indicate that each chromosome contacts a discrete and fairly constant number of other chromosomes in a single cell, with little dependency on the chromosome size. At the single-cell level both the microscopic and genomic observations therefore indicate highly defined territory structures, which may harbour much of the chromosome within the territory, and expose a limited, relatively constant surface area engaged in chromosome-to-chromosome interfaces. As these interfaces are highly variable among different cells, their averaging by ensemble Hi-C contributes towards the relatively uniform *trans*-chromosomal contact matrices previously reported.

We have presented a new experimental strategy to create Hi-C contact maps from single cells. The approach allows for characterization of thousands of simultaneous contacts occurring in individual

cells, and provides unique insights into Hi-C technology and three-dimensional chromosomal architecture (Supplementary Videos 3 and 4). Single-cell contact maps reflect conservation of domain structure that was recently characterized[18–20], but show that interdomain and *trans*-chromosomal contact structure is highly variable between individual cells. Genome-wide statistical analysis and reconstruction of the single-copy X-chromosome models gave us the opportunity to quantify key features of chromosomal architecture. For example, active domains tend to locate on the boundaries of their chromosomal territories in the majority of nuclei, while maintaining associations with other domains on the same chromosome. Our results do not exclude chromosome territory intermingling[29], but argue against domains becoming completely immersed in other territories. Coupled with previous observations of small and large-scale chromatin mobility[30–32] a highly dynamic view of chromosomal organization emerges, where territories are continuously being remodelled, while maintaining some key local (domain) and global (depth from surface) organizational features.

## METHODS SUMMARY

$T_H1$ cells from male mice were fixed and subjected to modified Hi-C, in which nuclei were maintained through restriction-enzyme digestion, biotin fill-in labelling and ligation. Single nuclei were isolated and processed to prepare single-cell Hi-C libraries for paired-end sequencing.

Sequences were mapped to the mouse genome, and abnormal read pairs were discarded. Read pairs that occurred only once (without duplication) in the library sequencing were removed. We chose 10 single-cell data sets for further in-depth analyses based on several quality criteria (see Supplementary Information). To validate the single-cell Hi-C procedure, we pooled the single-cell Hi-C data sets and compared them to an ensemble Hi-C data set prepared from approximately 10 million cells essentially as described[3]. We created reshuffled data sets by randomly redistributing contacts of the analysed single cells to create the same number of cells with the same number of contacts in each cell as a control to analyse statistically the variation among single-cell data sets.

We reconstructed three-dimensional X-chromosome structure models using restrained molecular dynamics calculations employing a simulated annealing protocol. A combination of unambiguous distance restraints from the X intra-chromosomal contacts in the single-cell Hi-C data set and anti-distance restraints between regions that were found not to contact each other in the ensemble Hi-C data set was used. To assess the precision and accuracy of the structure generation process we used the protocol to generate synthetic Hilbert curve structures, and explored the impact of varying the number of restraints. For pair-wise comparison of the structures, we calculated the root-mean-square deviation (r.m.s.d.). To compare the X-chromosome models to X-chromosome structure *in vivo*, we selected five loci with consistently high or low structural density in the models, and compared distances between the loci and the X-chromosome territory surface in cells (DNA FISH).

Full description of the methods can be found in the Supplementary Information.

**Online Content** Any additional Methods, 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.

**Received 13 December 2012; accepted 27 August 2013.**

**Published online 25 September 2013.**

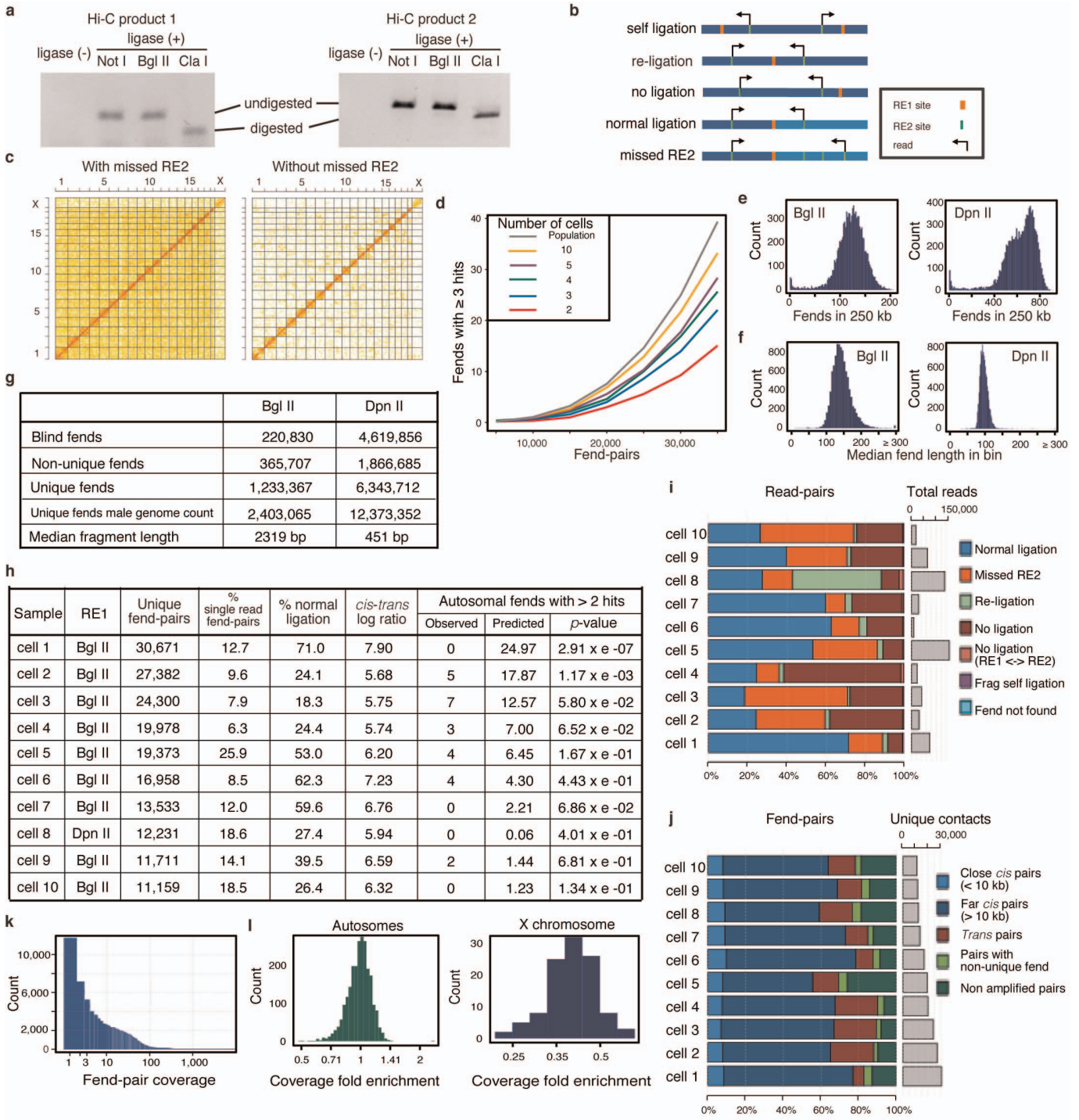
1. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295,** 1306–1311 (2002).
2. Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16,** 1299–1309 (2006).
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. Schoenfelder, S. *et al.* Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature Genet.* **42,** 53–61 (2010).
5. 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).
6. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genet.* **38,** 1341–1347 (2006).
7. Duan, Z. *et al.* A three-dimensional model of the yeast genome. *Nature* **465,** 363–367 (2010).
8. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nature Biotechnol.* **30,** 90–98 (2012).
9. Marti-Renom, M. A. & Mirny, L. A. Bridging the resolution gap in structural modeling of 3D genome organization. *PLOS Comput. Biol.* **7,** e1002125 (2011).
10. Tanizawa, H. *et al.* Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic Acids Res.* **38,** 8164–8177 (2010).
11. van de Werken, H. J. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nature Methods* **9,** 969–972 (2012).
12. Osborne, C. S. *et al.* Active genes dynamically colocalize to shared sites of ongoing transcription. *Nature Genet.* **36,** 1065–1071 (2004).
13. Rapkin, L. M., Anchel, D. R., Li, R. & Bazett-Jones, D. P. A view of the chromatin landscape. *Micron* **43,** 150–158 (2012).
14. Fraser, P. & Bickmore, W. Nuclear organization of the genome and the potential for gene regulation. *Nature* **447,** 413–417 (2007).
15. Lanctôt, C., Cheutin, T., Cremer, M., Cavalli, G. & Cremer, T. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nature Rev. Genet.* **8,** 104–115 (2007).
16. Osborne, C. S. *et al.* Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol.* **5,** e192 (2007).
17. Yaffe, E. & Tanay, A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nature Genet.* **43,** 1059–1065 (2011).
18. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148,** 458–472 (2012).
19. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
20. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).
21. Gibcus, J. H. & Dekker, J. The hierarchy of the 3D genome. *Mol. Cell* **49,** 773–782 (2013).
22. Jhunjhunwala, S. *et al.* The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell* **133,** 265–279 (2008).
23. Müller, I., Boyle, S., Singer, R. H., Bickmore, W. A. & Chubb, J. R. Stable morphology, but dynamic internal reorganisation, of interphase human chromosomes in living cells. *PLoS ONE* **5,** e11560 (2010).
24. Heard, E. & Bickmore, W. The ins and outs of gene regulation and chromosome territory organisation. *Curr. Opin. Cell Biol.* **19,** 311–316 (2007).
25. Deaton, A. M. *et al.* Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Res.* **21,** 1074–1086 (2011).
26. Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **38,** 603–613 (2010).
27. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Rev. Genet.* **2,** 292–301 (2001).
28. Misteli, T. Beyond the sequence: cellular organization of genome function. *Cell* **128,** 787–800 (2007).
29. Branco, M. R. & Pombo, A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.* **4,** e138 (2006).
30. Chuang, C. H. *et al.* Long-range directional movement of an interphase chromosome site. *Curr. Biol.* **16,** 825–831 (2006).
31. Chubb, J. R., Boyle, S., Perry, P. & Bickmore, W. A. Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr. Biol.* **12,** 439–445 (2002).
32. Dundr, M. *et al.* Actin-dependent intranuclear repositioning of an active gene locus *in vivo*. *J. Cell Biol.* **179,** 1095–1103 (2007).

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

**Acknowledgements** The authors thank I. Clay, S. Wingett, K. Tabbada, D. Bolland, S. Walker, S. Andrews, M. Spivakov, N. Cope, L. Harewood and W. Boucher for assistance. This work was supported by the Medical Research Council, the Biotechnology and Biological Sciences Research Council (to P.F.), the MODHEP project, the Israel Science Foundation (to A.T.) and the Wellcome Trust (to E.D.L.).

**Author Contributions** T.N. and P.F. devised the single-cell Hi-C method. T.N. performed single-cell Hi-C and DNA FISH experiments. S.S. carried out ensemble Hi-C experiments. W.D. microscopically isolated single cells. Y.L., E.Y. and A.T. processed and statistically analysed the sequence data. T.J.S. and E.D.L. developed the approach to structural modelling and analysed X-chromosome structures. T.J.S. wrote the software for three-dimensional modelling, analysis and visualisation of chromosome structures. T.N., Y.L., T.J.S., E.D.L., A.T. and P.F. contributed to writing the manuscript, with inputs from all other authors.

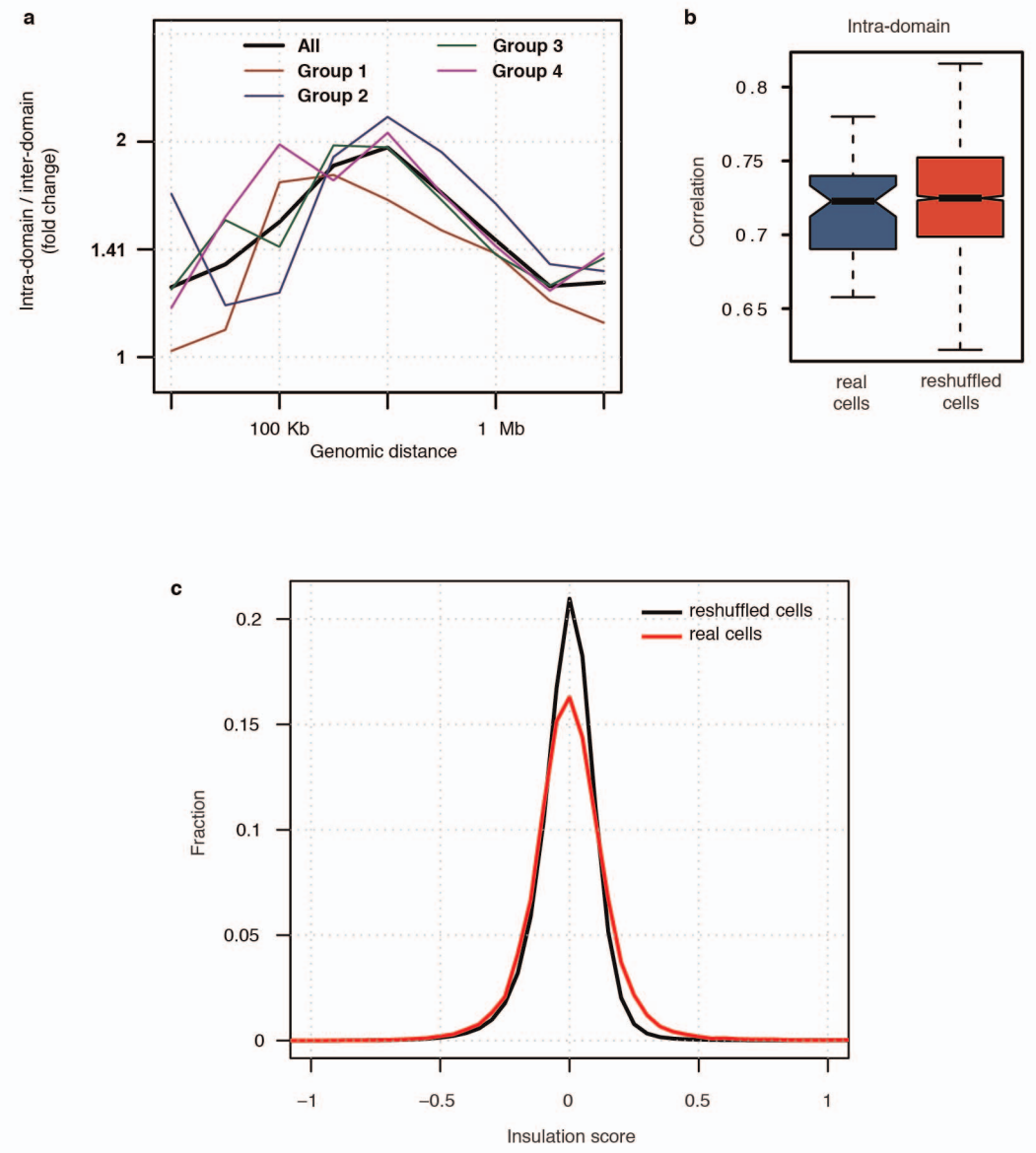
**Author Information** Data deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE48262. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.F. (peter.fraser@babraham.ac.uk) for the single-cell Hi-C method, A.T. (amos.tanay@weizmann.ac.il) for the statistical analysis, or E.D.L. (e.d.laue@bioc.cam.ac.uk) for the structural modelling.



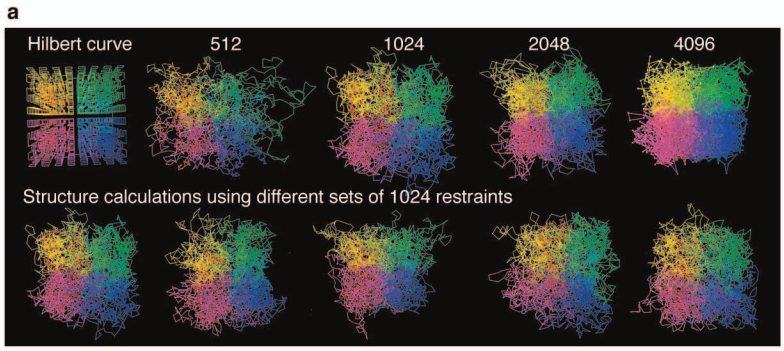

| | Bgl II | Dpn II |
|---|---|---|
| Blind fends | 220,830 | 4,619,856 |
| Non-unique fends | 365,707 | 1,866,685 |
| Unique fends | 1,233,367 | 6,343,712 |
| Unique fends male genome count | 2,403,065 | 12,373,352 |
| Median fragment length | 2319 bp | 451 bp |

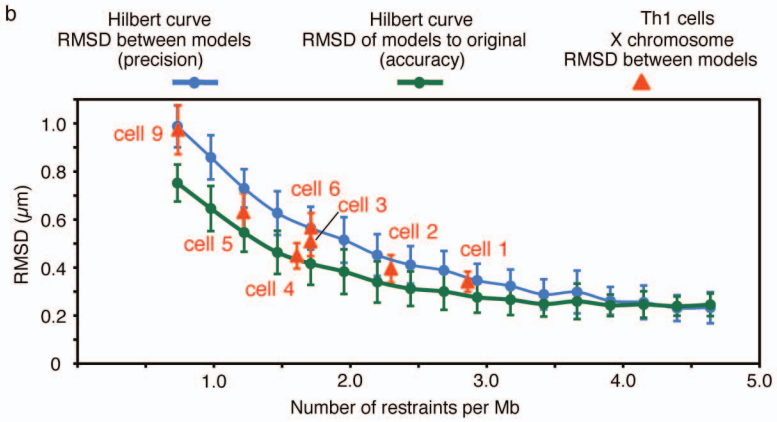
| Sample | RE1 | Unique fend-pairs | % single read fend-pairs | % normal ligation | *cis-trans* log ratio | Autosomal fends with > 2 hits | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Observed | Predicted | *p*-value |
| cell 1 | Bgl II | 30,671 | 12.7 | 71.0 | 7.90 | 0 | 24.97 | 2.91 x e -07 |
| cell 2 | Bgl II | 27,382 | 9.6 | 24.1 | 5.68 | 5 | 17.87 | 1.17 x e -03 |
| cell 3 | Bgl II | 24,300 | 7.9 | 18.3 | 5.75 | 7 | 12.57 | 5.80 x e -02 |
| cell 4 | Bgl II | 19,978 | 6.3 | 24.4 | 5.74 | 3 | 7.00 | 6.52 x e -02 |
| cell 5 | Bgl II | 19,373 | 25.9 | 53.0 | 6.20 | 4 | 6.45 | 1.67 x e -01 |
| cell 6 | Bgl II | 16,958 | 8.5 | 62.3 | 7.23 | 4 | 4.30 | 4.43 x e -01 |
| cell 7 | Bgl II | 13,533 | 12.0 | 59.6 | 6.76 | 0 | 2.21 | 6.86 x e -02 |
| cell 8 | Dpn II | 12,231 | 18.6 | 27.4 | 5.94 | 0 | 0.06 | 4.01 x e -01 |
| cell 9 | Bgl II | 11,711 | 14.1 | 39.5 | 6.59 | 2 | 1.44 | 6.81 x e -01 |
| cell 10 | Bgl II | 11,159 | 18.5 | 26.4 | 6.32 | 0 | 1.23 | 1.34 x e -01 |

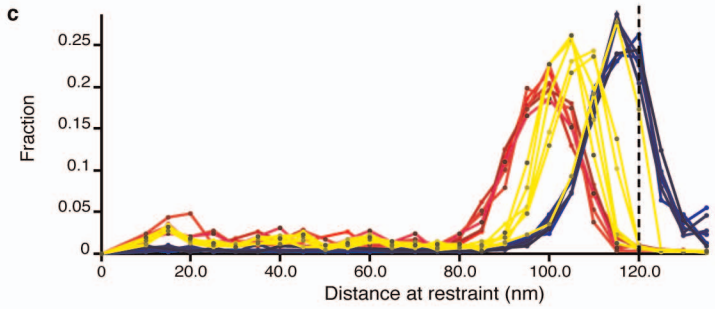
**Extended Data Figure 1 | Single-cell Hi-C quality controls.** **a**, Efficiency of biotin labelling at Hi-C ligation junctions for two Hi-C ligation products, showing 90–95% efficiency (Supplementary Information). **b**, Read-pair classification. **c**, Discarding the missed RE2 read-pairs removes a uniform 'blanket' of non-specific contacts from the map. **d**, Estimating numbers of multiple covered fends. Shown is the dependency between the number of fend pairs in a sample and the estimated number of autosomal fends covered by more than two fend pairs under different models. The binomial model (grey line) distributes fend pairs to fends randomly without any constraint, as if sampling fend pairs from an infinite number of chromosomes. **e**, Single-cell Hi-C fragments coverage. Number of fends in each 250-kb genomic bin for BglII or DpnII as RE1. Tail of bins with few fends is for bins of low mappability and near the chromosomes edges. **f**, Median fend length (distance from RE1 to the first upstream RE2) in each 250-kb genomic bin for BglII or Dpn II as RE1. Values larger than 300 bp are of poorly mappable bins. **g**, Information on the two restriction enzymes we used for RE1, BglII (6 cutter, which we used predominantly) and DpnII (4 cutter, only used for cell 8). Blind fends do not have a RE2 site in their fragment. Fends in which their first RE2 site starts a non-unique 36-bp sequence are marked as non-unique fends. We discarded both blind and non-unique fends and used only the unique fends. The number of actual fends in a male mouse genome, which have two copies of each autosome and a single X chromosome are shown as well as the median fragment length (chromosome Y and mitochondrial genome were ignored throughout the analysis). **h**, Information on the ten single-cell data sets that successfully passed the quality control filters. *P* value of the number of autosomal fends with more than two covering fend-pairs was calculated from the binomial model (panel **d** and Supplementary Information). **i**, Percentages of read-pair types. **j**, Percentage of fend-pair types. **k**, Distribution of fend-pair coverage (number of read pairs that support each fend pair) in the ten single-cell data sets. **l**, Distribution of mean contacts per fend calculated for each mappable 1 Mb, normalized by the mean value in each cell, and averaged across autosomal or X chromosomes from the ten single cells.

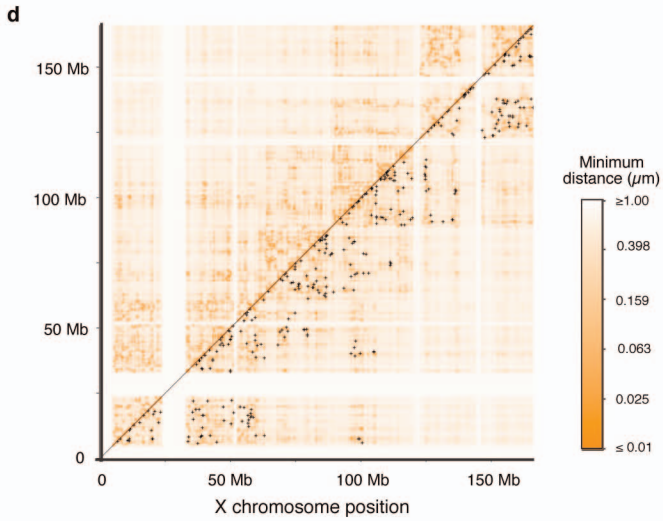


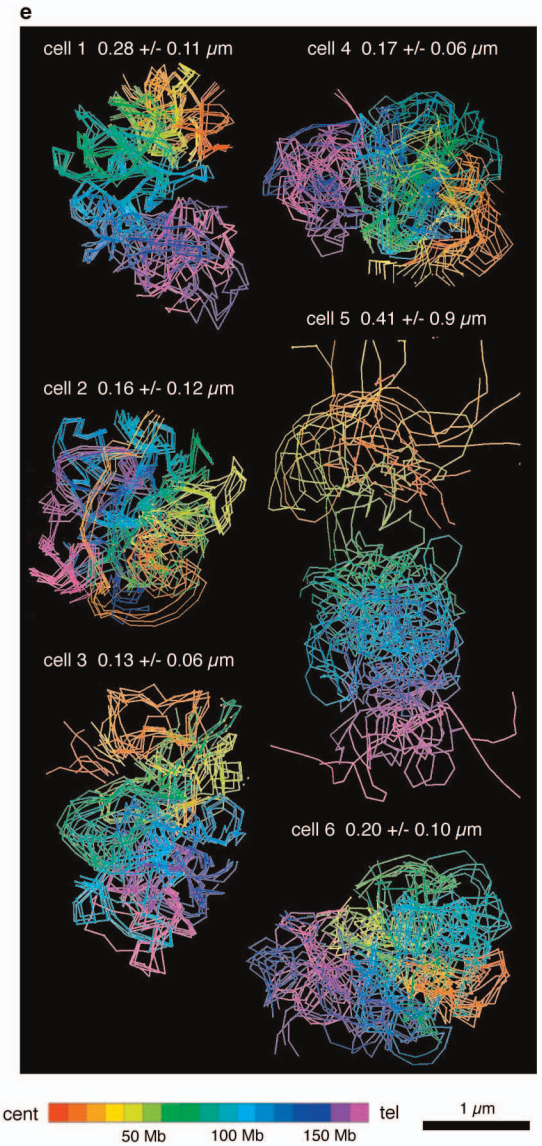

**Extended Data Figure 2 | Chromosomal domains.** **a**, Ratios between intradomain and interdomain contact enrichments over genomic distance. The mean single-cell trend is shown in black. Chromosomes are grouped into four groups: group 1 (chromosomes 1, 8, 15, 16 and X), group 2 (chromosomes 2, 6, 10, 13 and 18), group 3 (chromosomes 3, 5, 11, 14 and 17) and group 4 (chromosomes 4, 7, 9, 12 and 19). The intra- over interdomain enrichment is persistent in all chromosome groups and does not seem to stem from peculiar chromosomes. **b**, Distribution of correlations between intradomain contact numbers of all domains from pairs of real and reshuffled controls. **c**, Distribution of the insulation score at each fend in nine single-cell Hi-C data sets (where RE1 is BglII; real cells) is shown in red. Fifty sets of reshuffled cells were produced (see Supplementary Methods) and their insulation score distribution is shown in black. Real cells have a heavier tail of highly insulating loci, which is indicative of non-uniform and cell-specific interdomain contact structure.



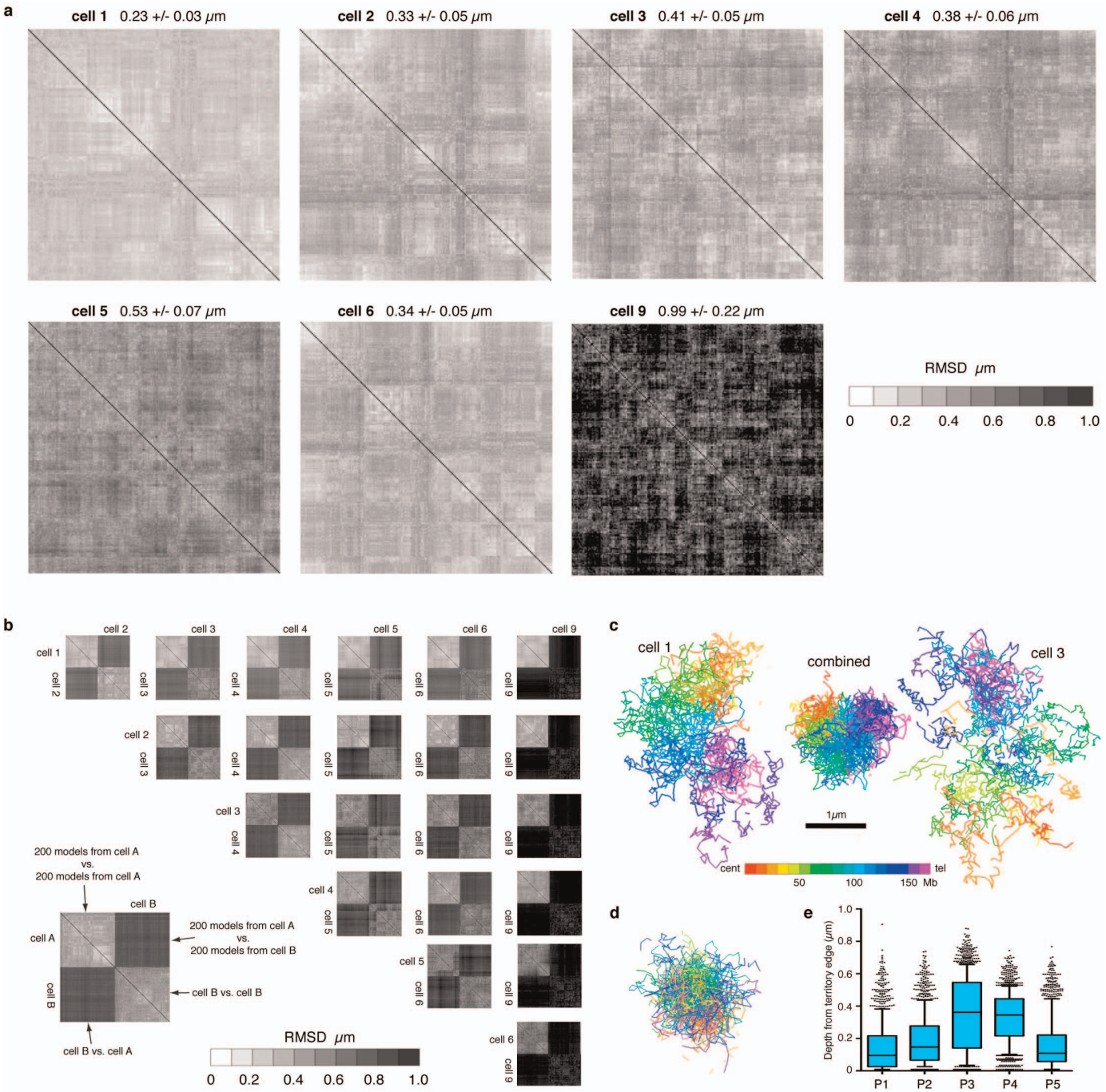



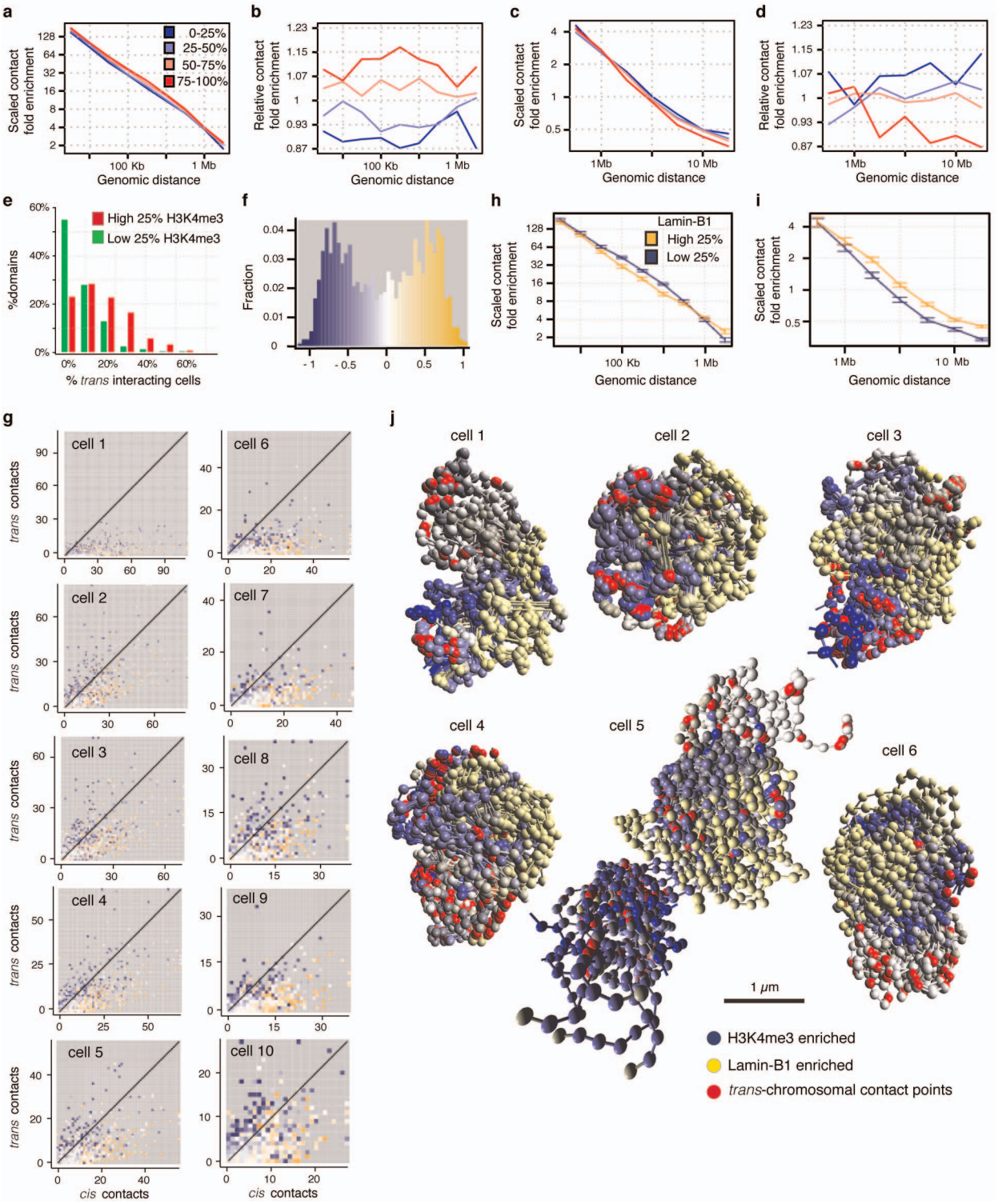



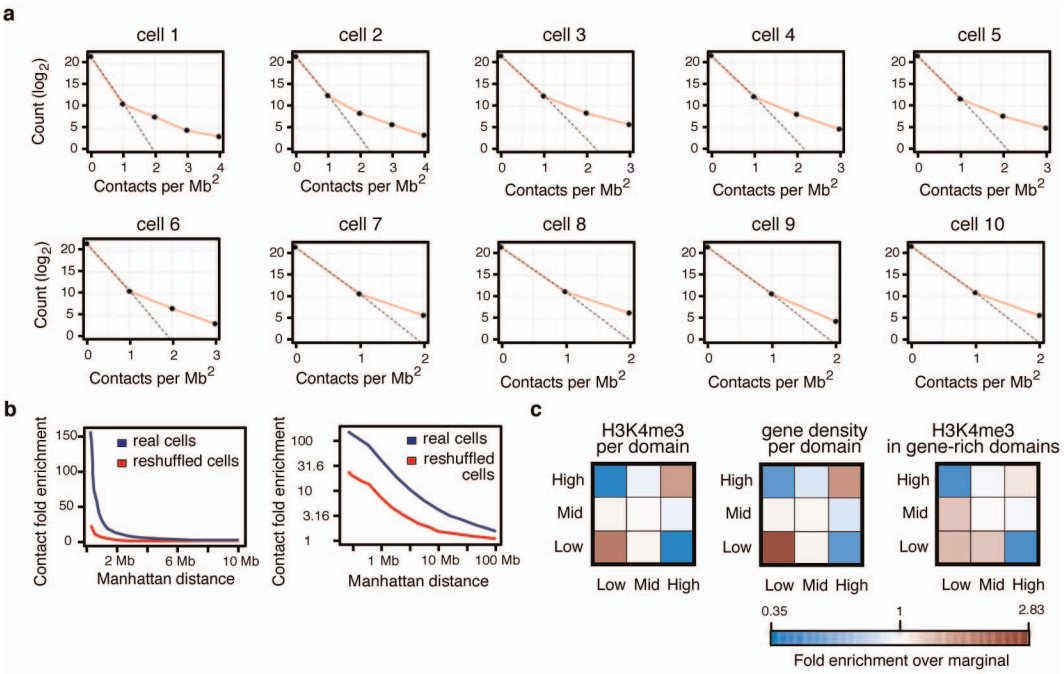
**Extended Data Figure 3 | Modelling protocol quality controls.** **a**, Results of structure calculations using restraints from a space-filling Hilbert curve test structure with 4,096 particles and four typical results of structure modelling using different numbers of restraints are shown (upper panels). Structure calculations of the Hilbert curve from random positions using different sets of 1,024 restraints (lower panel). **b**, Comparison of r.m.s.d. values from Hilbert curve and single-cell X-chromosome models. Structure calculations for Hilbert curves were repeated 100 times with variable numbers of restraints as shown. The root mean square deviation (r.m.s.d.) values between 100 models (precision) using the indicated number of restrains (mean ± s.d.) are plotted in blue. The r.m.s.d. values between the original Hilbert curve and each of the 100 models (accuracy) for the same numbers of restraints are plotted in green (mean ± s.d.). r.m.s.d. values from 100 repeated calculations of fine-scale (50-kb backbone) X-chromosome structure from the seven single-cell data sets are also plotted (red; mean ± s.d.). **c**, Restraint violation analysis. The distances between directly restrained positions in fine-scale (50-kb backbone) X-chromosome models are shown. Models for the six single-cell data sets (cell 1 to cell 6; red) show no values exceeding the upper bound (dashed line). Calculations with six shuffled interaction maps (created from cell 1 data set; blue) show significant violations. Structure calculations performed on merged pairs of data sets (yellow; all possible combinations of cell 1 to cell 4) have a few violations and are significantly closer to the upper limit. **d**, Comparison of structure-derived distance matrix from 200 fine-scale X-chromosome models from cell 1 (orange) and its single-cell Hi-C contacts (black crosses). The orange colour indicates the minimum distance between backbone particles. **e**, Comparison of X-chromosome structural models for six cells computed using low-resolution (500-kb binned) single-cell Hi-C interaction data. The bundles shown represent minimised structural alignments of five models from repeat calculations for each cell. Colours indicate chromosomal positions as shown. Scale bar, 1 μm.



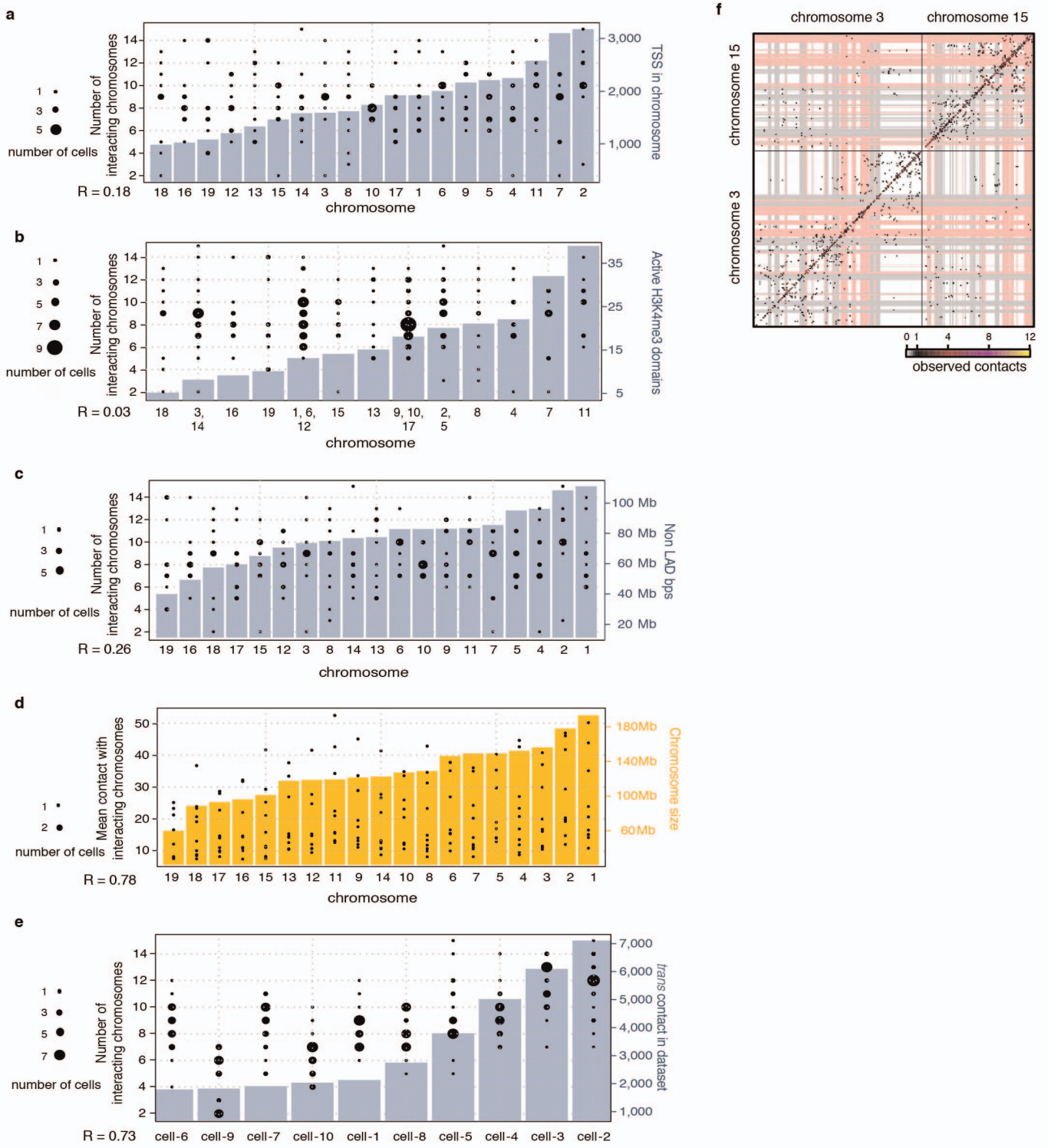

**Extended Data Figure 4 | Comparison and investigation of models.** **a**, Pair-wise comparison of fine-scale X chromosome structural models by r.m.s.d. analysis. Each pixel represents an r.m.s.d. value for a pair-wise comparison of two models. Lighter pixels indicate structures of higher similarity (low r.m.s.d.). Diagonal elements have been excluded. The order of 200 models in each panel was determined by hierarchical clustering of the r.m.s.d. values. Numbers shown are the mean r.m.s.d. values and the standard deviations for all the comparisons for each cell calculated by comparing the Hi-C contact particles. **b**, Cell-to-cell comparison of 200 fine-scale X chromosome structural models by r.m.s.d. analysis. Each pixel represents an r.m.s.d. value for a pair-wise comparison of two models. **c**, Fine-scale X-chromosome structures calculated from cell 1 and cell 3 data sets, and a structure from the combined data set. Colours indicate chromosomal positions as shown. Scale bar, 1 μm. **d**, Typical structure calculated using a randomized data set, where the interacting points for cell 1 have been shuffled with a pairing probability proportional to one over the square root of the sequence separation. Colours and scale as shown in **c**. **e**, Distribution of measurements of depth from the surface for five loci P1–P5 (Fig. 3e) in 1,200 X-chromosome models (200 fine-scale models for each of the six cells). Whiskers on box plots define 10th and 90th percentiles and the outliers are shown as individual dots.



**Extended Data Figure 5 | Epigenomic landscape of chromosomes.** **a**, Intradomain contact enrichment for each quartile of *trans*-chromosomal contacting domains. **b**, Same as **a** but subtracting the mean quartile enrichment in each genomic distance emphasizing the differences shown in **a**. **c**, Using the same sets as in **a** but plotting the enrichment of interdomain contacts within the same chromosome. **d**, Same as **c** but subtracting the mean quartile enrichment in each genomic distance. **e**, Percentage of cells in which high and low H3K4me3 enriched domains are *trans*-interacting. For each cell the domains with top 10th percentile *trans* intensity were defined as *trans*-interacting in that cell. We then counted for each domain the fraction of cells in which that domain was *trans*-interacting. Shown are the distributions of these fractions for H3K4me3 enriched and non-enriched domains (the top and bottom 25th percentiles, respectively). **f**, Distribution of the average lamin B1 (ref. 26) enrichment in chromosomal domains, colour coded according to the enrichment value. **g**, Domains plotted according to their number of *trans*- and *cis*-chromosomal (but excluding intradomain) contacts, colour-coded as in **f**. The domain lamin B1 enrichment and H3K4me3 peak density are highly anti-correlated (Spearman's correlation $= -0.73$). **h**, Intradomain contact enrichment for high versus low quartile of domains stratified by their mean lamin B1 enrichment. Error bars indicate 95% confidence intervals. **i**, Using the same sets as in **h** but plotting the enrichment of interdomain contacts within the same chromosome. Error bars as in **h**. **j**, Lamin B1 domains show a minor decrease in intradomain contact intensities that might suggest less compacted domains, and significantly increased *cis*-interdomain contact, maybe owing to lack of *trans*-chromosomal contacts. Topology of lamin B1, H3K4me3 and *trans*-contacts on five-model bundles of low-resolution X-chromosome models. Regions of low mappability have been excluded. Scale bar, 1 μm.




**Extended Data Figure 6 | Interchromosomal contacts.** **a**, Comparison of observed coverage of the *trans*-chromosomal 1-Mb square bins of each cell (red lines), versus predicted coverage assuming a binomial model (random uniform distribution of contacts to bins; black dashed line). Observed coverage is consistently higher than the uniform model, indicating the highly non-random distribution of *trans*-chromosomal contacts to genomic bins. **b**, *Trans*-chromosomal contact enrichment around observed *trans*-contacts as a function of the contacts total distance on both chromosomes (Manhattan distance in the contact map). Observed and expected (by random uniform contact distribution) numbers of contacts are counted around each *trans*-contact, and their ratio is shown for the 9 real cells (blue; where RE1 is BglII) and reshuffled cells (red), at two different scales. **c**, Left panel, *trans*-contacts were classified according to H3K4me3 density of the domains they associate: High and low for top and bottom 25th percentiles, respectively, mid for 25th–75th percentiles. Shown is the log ratio of the contingency table counts with the expected counts generated by multiplying the corresponding marginal probabilities for each group (chi-squared test; $P = 5.8 \times 10^{-18}$). To make sure these phenomena are not caused by the *trans* enrichment of active domains and depletion of non-active ones, only the top 15th percentile trans enriched domains from each cell were used. Middle panel, similar to left panel but contacts are classified by their associated domain gene density (chi-squared test; $P = 2.3 \times 10^{-12}$). Right panel, similar to left panel but using domains in the top 40th percentile of gene density, classifying by their H3K4me3 density, to test H3K4me3 enrichment beyond gene density (chi-squared test; $P = 3.6 \times 10^{-06}$). In all cases active or gene-rich domains preferentially interact with each other, although active domains (high H3K4me3 density) show greater interaction than expected from their gene density.



**Extended Data Figure 7 | Chromosomal interfaces.** **a**, The number of interacting chromosomes per chromosome is depicted in circles sized according to the number of single cells the value was observed in, and the order of chromosomes is shown by the number of transcription start sites (TSSs) in each chromosome (blue bars). Only autosomes are displayed. Spearman correlation between the number of TSSs and the mean value of the number of interacting chromosomes per chromosome is 0.18. Two chromosomes were defined as interacting when they had at least one domain–domain interaction (see main text) supported by two or more contacts. The number of interacting chromosomes per chromosome rises together with the number of TSSs. However, the change is small, and the number of interacting autosomal chromosomes per chromosome (the plotted value divided by two) remains between 4 and 6. **b**, Same as **a** except that chromosomes ordered by the number of active H3K4me3 domains (the top 25th percentile H3K4me3 peak density domains). **c**, Same as **a** except that chromosomes are ordered by the number of non-lamin-associated domains (non-LAD) base pairs in the chromosome. The fraction of a chromosome covered by LADs ranges from 31% to 53% and is correlated with chromosome size (0.52 Spearman). Thus, chromosome lengths span a range of 3.2-fold change, while their non-LAD fraction spans a smaller range of 2.8-fold change. **d**, Examination of the number of contacts between two chromosomes and the chromosomes sizes. The mean number of contacts of each chromosome with others it interacts with is shown for the ten single cells, with chromosomes ordered by their size. Chromosome size is correlated with the number of contacts it has, but the dynamic range of this number is small. **e**, The number of interacting chromosomes per chromosome is depicted in circles sized as the number of single cells the value was observed in, and the ten single-cell data sets are ordered by the number of *trans*-contacts in each data set, shown by blue bars. Only autosomes are displayed. Spearman correlations between the number of *trans*-contacts in each data set and the mean value of the number of interacting chromosomes per chromosome is 0.73. The number of interacting chromosomes per chromosome rises together with the coverage. However, the change is small, and the number of interacting autosomal chromosomes per chromosome (the plotted value divided by two) remains between 4 and 6. **f**, Example of multi-way chromosomal interfaces. Contact map of chromosomes 3 and 15 in cell 5. Shown is the number of contacts in 1-Mb size bins. Top and bottom 30th percentiles of H3K4me3 peak density domains are marked in light pink and light grey, respectively. Note the grid-like *trans*-contacts arrangement, and the correspondence between the two large *trans*-contact clusters and the organization of *cis*-contacts in both chromosomes to large 'mega domains'.

**Extended Data Table 1 | Testing sequencing saturation.**

**a**

| Sample | Read-pairs | | Fend-pairs | | | %valid non-amplified |
|---|---|---|---|---|---|---|
| | Original | Re-sequenced | Original | Combined | Addition | |
| cell 5 | 1,485,494 | 13,882,742 | 13,118 | 19,373 | 47.68 % | 70.20 % |
| cell 1 | 1,253,954 | 6,210,472 | 26,318 | 30,671 | 16.54 % | 51.85 % |
| cell 9 | 2,413,574 | 4,079,969 | 10,040 | 11,711 | 16.64 % | 34.05 % |
| cell 8 | 1,132,934 | 12,437,493 | 10,043 | 12,231 | 21.79 % | 71.71 % |
| cell 3 | 1,420,807 | 4,096,785 | 21,414 | 24,300 | 13.48 % | 47.04 % |

**b**

| cell 5 | | Re-sequenced coverage | | | |
|---|---|---|---|---|---|
| | | 0 | 1 | ≥ 2 | |
| Original coverage | 0 | 0 | 7088 | 5635 | |
| | 1 | 170 | 69 | 563 | 5.76 % |
| | ≥ 2 | 28 | 26 | 13078 | 94.24 % |
| | | | 27.15 % | 72.85 % | |

| cell 8 | | Re-sequenced coverage | | | |
|---|---|---|---|---|---|
| | | 0 | 1 | ≥ 2 | |
| Original coverage | 0 | 0 | 2754 | 1331 | |
| | 1 | 220 | 87 | 778 | 9.74 % |
| | ≥ 2 | 33 | 28 | 9997 | 90.26 % |
| | | | 19.16 % | 80.84 % | |

| cell 1 | | Re-sequenced coverage | | | |
|---|---|---|---|---|---|
| | | 0 | 1 | ≥ 2 | |
| Original coverage | 0 | 0 | 3908 | 3240 | |
| | 1 | 821 | 337 | 1247 | 8.36 % |
| | ≥ 2 | 1092 | 295 | 24962 | 91.64 % |
| | | | 13.36 % | 86.64 % | |

| cell 3 | | Re-sequenced coverage | | | |
|---|---|---|---|---|---|
| | | 0 | 1 | ≥ 2 | |
| Original coverage | 0 | 0 | 1270 | 1718 | |
| | 1 | 1010 | 413 | 1264 | 11.13 % |
| | ≥ 2 | 1141 | 534 | 19777 | 88.87 % |
| | | | 8.88 % | 91.12 % | |

| cell 9 | | Re-sequenced coverage | | | |
|---|---|---|---|---|---|
| | | 0 | 1 | ≥ 2 | |
| Original coverage | 0 | 0 | 1378 | 1204 | |
| | 1 | 724 | 260 | 508 | 12.93 % |
| | ≥ 2 | 663 | 261 | 9126 | 87.07 % |
| | | | 14.91 % | 85.09 % | |

Several single-cell libraries were extensively re-sequenced. **a**, Shown are the numbers of read pairs in the original and re-sequenced runs, the number of fend pairs in the original run, the number of fend pairs when combining the sequences of the two runs, and the addition to fend pairs that the re-sequencing contributed. The percentages of singly covered fend pairs in the original sample that were supported by more read-pairs in the re-sequenced one are shown (%valid non-amplified). These fend pairs were discarded as potential spurious pairs in the original run, but proved by the re-sequencing to be valid pairs. This gives a sense of the fraction of valid pairs we discard when removing the read pairs suspected to be sequencing pairing errors. **b**, Shown are fend-pair coverage contingency tables of the original and the re-sequenced runs for the five single cells.

**Extended Data Table 2 | Testing intercellular spurious ligations.**

| Library | Total read -pairs | %mm9-mm9 | mean% unexpected | %hg18-hg18 | %hg18-mm9 | mean% hg18-mm9 |
|---|---|---|---|---|---|---|
| Group A_1 | 2232565 | 99.452 | → 0.103 | 0.543 | 0.005 | 0.007 |
| Group A_2 | 2181145 | 99.955 | | 0.039 | 0.007 | |
| Group A_3 | 3170076 | 99.987 | | 0.006 | 0.007 | |
| Group A_4 | 4528440 | 99.988 | | 0.004 | 0.008 | |
| Group A_5 | 2214693 | 99.967 | | 0.025 | 0.009 | |
| Group A_6 | 2887996 | 99.992 | | 0.002 | 0.006 | |
| | | | | | | |
| Group B_7 | 3058729 | 64.053 | | 35.936 | 0.011 | 0.009 |
| Group B_8 | 4341817 | 99.976 | | 0.017 | 0.007 | |
| Group B_9 | 3262211 | 61.122 | | 38.869 | 0.009 | |
| Group B_10 | 3063179 | 99.986 | | 0.005 | 0.009 | |
| Group B_12 | 3662647 | 99.326 | | 0.668 | 0.006 | |
| | | | | | | |
| Group C_1 | 574035 | 0.132 | ← 0.057 | 99.859 | 0.010 | 0.008 |
| Group C_2 | 2186006 | 0.146 | | 99.847 | 0.007 | |
| Group C_3 | 653496 | 0.010 | | 99.983 | 0.007 | |
| Group C_4 | 965577 | 0.025 | | 99.971 | 0.004 | |
| Group C_5 | 1578045 | 0.031 | | 99.957 | 0.012 | |
| Group C_6 | 891922 | 0.001 | | 99.992 | 0.007 | |
| Group C_7 | 1117686 | 99.975 | → 0.027 | 0.015 | 0.010 | 0.010 |
| Group C_8 | 970213 | 99.974 | | 0.019 | 0.007 | |
| Group C_9 | 1338728 | 99.966 | | 0.028 | 0.006 | |
| Group C_10 | 1043818 | 99.982 | | 0.007 | 0.011 | |
| Group C_11 | 2351657 | 99.906 | | 0.083 | 0.010 | |
| Group C_12 | 3202531 | 99.975 | | 0.012 | 0.013 | |
| | | | | | | |
| Group D_1 | 1736070 | 99.962 | → 0.074 | 0.033 | 0.006 | 0.007 |
| Group D_2 | 1754172 | 99.796 | | 0.192 | 0.012 | |
| Group D_4 | 1093540 | 99.813 | | 0.180 | 0.008 | |
| Group D_5 | 1554973 | 99.962 | | 0.034 | 0.005 | |
| Group D_6 | 1958313 | 99.885 | | 0.112 | 0.003 | |
| Group D_7 | 1730004 | 99.991 | | 0.000 | 0.009 | |
| Group D_8 | 2097397 | 99.954 | | 0.042 | 0.004 | |
| Group D_9 | 2105599 | 99.963 | | 0.030 | 0.007 | |
| Group D_12 | 2059016 | 99.946 | | 0.047 | 0.007 | |

Mouse and human nuclei or single-cell Hi-C samples were mixed in different stages of the experiment (group A, before fixation; group B, before library construction (so all the mouse and human samples in each library have the same identification tag); group C, before library amplification (so mouse and human samples in each library have different identification tags)). We created single-cell (for group A) or human and mouse two-cell (for groups B and C) Hi-C libraries and analysed them. The table shows the percentages of the three possible read pairs: mouse–mouse (mm9–mm9), human–human (hg18–hg18) and human–mouse (hg18–mm9). The expected pair type in each library is marked in blue. Mean percentage of unexpected read pairs per lane are also shown. For group A, we selected mouse cells based on morphology. In Group A, all six libraries contain almost exclusively mouse–mouse read pairs with insignificant human–human or mouse–human pairs. Each group B library has both human–human and mouse–mouse read pairs as expected, and the number of spurious human–mouse read pairs is extremely low. In each group C library, which was created by amplifying the distinctly tagged human (C1–C6) and mouse (C7–C12) single-cell samples in the same tube (for example, C1 and C7, C2 and C8, etc.), the fractions of foreign pairs (human reads with a mouse tag and vice versa) and of spurious pairs (human–mouse) were consistently extremely low.

To estimate the fraction of foreign and spurious pairs that could have originated simply from mapping a truly pure mouse library to a concatenated human-mouse genome, libraries from pure mouse cells (group D) were mapped to such a genome. The mean percentages of both foreign and spurious fend pairs in this lane are the same as those found in the different human–mouse mixed lanes, suggesting there is no intercellular contamination.

**Extended Data Table 3 | Sequencing pairing errors.**

| | lane_A | lane_B | lane_C | lane_D |
|---|---|---|---|---|
| % phiX loaded / lane capacity | 10 % | 10 % | 25 % | 1 % |
| % phiX reads | 12 % | 15 % | 40 % | 2 % |
| Total number of read pairs | 24.4 M | 21.5 M | 12.8 M | 25.3 M |
| phiX – phiX pairs | 3 M | 3.3 M | 5 M | 0.5 M |
| phiX – mm 9 pairs | 674 | 553 | 323 | 40 |
| mm 9 – phiX pairs | 3266 | 1647 | 3159 | 200 |
| Estimated spurious mm 9 - mm 9 fend-pairs with the same identification tag | ~3000 | ~1500 | ~300 | ~1000 |

PhiX174 DNA library was added to four lanes of single-cell Hi-C multiplexed libraries. In theory, no mixed mouse–phiX174 read pair is expected, but in fact a small number were detected. Shown are the fraction of phiX174 DNA loaded to each lane capacity, the percentage of phiX174 read ends in the lane, and the observed number of read pairs by type. The pairing probability was crudely estimated from these figures, and from it the number of expected spurious mouse–mouse read-pairs was calculated. Most of these spurious pairs are discarded due to mismatching unique identification tags at the beginning of each read end. Shown is the estimated number of spurious mouse pairs that coincidently have matching identification tag and are therefore not detected and removed.