Published in final edited form as:

Nature. 2017 July 05; 547(7661): 61–67. doi:10.1038/nature23001.

Cell-cycle dynamics of chromosomal organisation at single-cell resolution

Takashi Nagano^{1,*}, Yaniv Lubling^{2,*}, Csilla Várnai^{1,#,§}, Carmel Dudley^{2,§}, Wing Leung¹, Yael Baran², Netta Mendelson-Cohen², Steven Wingett¹, Peter Fraser^{1,3,#}, and Amos Tanay^{2,#}

¹Nuclear Dynamics Programme, The Babraham Institute, Cambridge, CB22 3AT, United Kingdom

²Department of Computer Science and Applied Mathematics, and Department of Biological Regulation, Weizmann Institute of Science, Rehovot, Israel

³Department of Biological Science, Florida State University, Tallahassee FL., 32306

Summary

Chromosomes in proliferating metazoan cells undergo dramatic structural metamorphoses every cell cycle, alternating between highly condensed mitotic structures facilitating chromosome segregation, and decondensed interphase structures accommodating transcription, gene silencing and DNA replication. Here we use single-cell Hi-C to study chromosome conformations in thousands of individual cells, and discover a continuum of *cis*-interaction profiles that finely position individual cells along the cell cycle. We show that chromosomal compartments, topological associated domains (TADs), contact insulation and long-range loops, all defined by bulk Hi-C maps, are governed by distinct cell-cycle dynamics. In particular, DNA replication correlates with build-up of compartments and reduction in TAD insulation, while loops are generally stable from G1 through S and G2. Whole-genome 3D structural models reveal a radial architecture of chromosomal compartments with distinct epigenomic signatures. Our single-cell data thereby allow for re-interpretation of chromosome conformation maps through the prism of the cell cycle.

As cells progress from mitosis to interphase and back again, their chromosomes undergo dramatic structural alterations that have intrigued microscopists for over a century¹. In preparation for mitosis and cell division, chromosomes are packaged into rod-like structures

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

*Correspondence should be addressed to C.V. (csilla.varnai@babraham.ac.uk) for the 3D structural modelling, P.F.

(peter.fraser@babraham.ac.uk) for the single-cell Hi-C method, and A.T. (amos.tanay@weizmann.ac.il) for the statistical analysis.

§Co-first authors.

#Co-second authors.

Author Contributions T.N., Y.L., P.F. and A.T. conceived the project and designed the experiments. T.N. devised new single-cell Hi-C protocol. T.N. and W.L. carried out experiments. Y.L., C.V., C.D., Y.B., N.M.C., S.W. and A.T. analysed data. T.N., Y.L., C.V., C.D., P.F. and A.T. wrote the manuscript.

Author Information The authors declare no competing financial interests.

Data Availability

All data reported in this study is available at GEO GSE94489.

Code Availability

Additional analysis files and scripts are available from our website or upon request.

that are functionally inert. Upon mitotic exit, chromosomes rapidly decondense into structures that allow regulated access to functional elements, enabling genome functions such as transcription, gene silencing and DNA replication. Though the structural and organizational principles of the interphase genome have been the subject of intense study, in particular using high-throughput chromosome conformation capture (Hi-C)^{2–10}, genome-wide, high-resolution understanding of chromosome cell-cycle dynamics¹¹ is still limited by current technologies.

Here we scale up single-cell Hi-C¹² using flow cytometry sorting, a new library preparation protocol and a new automation scheme, and apply this methodology to thousands of diploid and haploid mouse embryonic stem cells (ESCs). Having thousands of single-cell maps has allowed us to develop a strategy for in-silico cell-cycle phasing of single nuclei, and use it to study the cell-cycle dynamics of chromosomal structural features. Such features include insulation profiles and looping contacts that together define topologically associated domains (TADs), and preferential contacts between generalized chromosome pseudo-compartments. We also present whole-genome, restraint-based three-dimensional (3D) models from G1 phase haploid nuclei, following the structural dynamics during the exit from mitosis. Our results reveal a continuum of dynamic chromosomal structural features throughout the cell cycle, creating a new point of reference for interpretation of Hi-C maps.

Results

Improved multiplexed single-cell Hi-C

We redesigned our original single-cell Hi-C approach¹² to employ a 4-cutter restriction enzyme, Tn5-transposase tagmentation, flow cytometry sorting of single cells into 96-well plates and partially automated library preparation (Fig. 1a). We processed a total of 2924 single F1 hybrid 129 x Castaneus mouse ESCs grown in 2i media using 1.5 million reads per cell on average (Extended Data Fig. 1a-c). Analysis of various quality control (QC) metrics (Extended Data Fig. 1d-i) confirmed the robustness of the protocol over different batches, and in particular identified karyotypic imperfections within the single-cell sample (Extended Data Fig. 1j). Overall, 1992 cells (68.1%) passed stringent quality control filters for further analysis. This provided median coverage of 393,506 restriction fragments, and 127,233 distinct >1kb contacting pairs per cell (Fig. 1b). We found the incidence of *trans*-chromosomal contacts in the single-cell maps was low (median 5.87%, Fig. 1c), indicative of high library quality¹³ and reflecting specific and restricted interfaces of chromosome territories (Extended Data Fig. 2). This data provided us with the opportunity to study variation in chromosomal conformation across a large single-cell pool, on the basis of a high-resolution reference ensemble map (Extended Data Fig. 3)

Integrated cell-cycle phasing

Mammalian mitotic chromosomes are rod-shaped, and bulk Hi-C on sorted M phase cells has revealed a characteristic scale of *cis*-chromosomal contact distances peaking between 2–12 Mb¹¹. We examined the cells with the highest fraction of contacts at this mitotic range, and found their overall distribution of contact distances was similar to that previously observed for mitotic cells (Extended Data Fig. 4a). These single-cell mitotic maps were

enriched for striking patterns of *trans*-chromosomal alignments (Fig. 1d, Extended Data Fig. 4b), suggesting that the chromosomes interacted along their entire length (head-to-head and tail-to-tail) (1CDX4_242 in Extended Data Fig. 4b) as though organized by the mitotic spindle in telophase (see, also 1CDX_74 and 1CDX_453 in Extended Data Fig. 4b for partial or disordered chromosomal alignments). Comparison of the frequency of contacts in the mitotic range to the frequency of contacts at shorter distances (short-range contacts < 2 Mb; Fig. 1e) shows a remarkable circular pattern evocative of a gradual chromosome remodelling process as cells enter and exit mitosis. These observations suggest chromosome conformation may be used to phase single cells at various stages of the cell cycle around mitosis.

Early- and late-replicating domains are defined by their copy number dynamics during S phase, and show distinct enrichment for active and inactive epigenetic marks respectively¹⁴. Indeed, we found normalized TAD coverage across the single-cell dataset reflects a strong correlation between domains previously annotated as early- or late-replicating in mouse ESCs¹⁵ (Extended Data Fig. 4c-d). We computed a *repli-score* for each cell based on the copy-number ratio of early-replicating regions to total coverage. G1 cells would be expected to have low repli-scores, whereas early S cells would be associated with increasing scores, peaking in mid-S before declining again in late S phase and returning to a low level in G2/M. Combining repli-score with the circular pattern exhibited by the frequencies of short-range and mitotic contacts yields exactly that expected trajectory (Fig. 1e). Together, the mitotic signatures and time of replication analysis define two major anchors to support phasing of the entire single-cell Hi-C dataset along the cell cycle.

We studied the distributions of contact distance scales in single cells using non-linear dimensionality reduction (NLDR) (Extended Data Fig. 4e) and clustering (Extended Data Fig. 5). Both approaches indicate the data can be globally represented by a cyclic, gradual process of coordinated change that links the highly distinct mitotic and replicating states characterized above. To capture this variation precisely we used a supervised approach that groups cells to phases and then orders cells within phases. We first identified cells approaching mitosis (pre-M regime, orange) and coming out of mitosis (post-M regime, red) by thresholding short-range and mitotic contact frequencies (based on Fig 1e, see Methods). Cells in both regimes can be ordered by the ascending and descending frequency of mitotic contacts for the pre-M and post-M regimes, respectively. Next, we observed that cells with less than 63% short-range contacts do not include the replication group and likely represent a G1 regime (blue). The mean distance of long-range contacts in these cells is correlated to the overall frequency of short-range contacts, allowing ordering of cells within this group (Extended Data Fig. 6a). Cells with over 63% short-range contacts show a continuum of repli-score and short-range contact enrichments (Extended Data Fig. 6b). These were modelled using two additional regimes, one ordered based on a gradual increase in repli-score and short-range contacts (the early-S regime, green), and one ordered based on a gradual decrease in repli-score and further increase in short-range contacts (the late-S/G2 regime, purple). Using this phasing strategy, we observed a smooth, cyclical transition of the distribution of genomic contact distances (Fig. 1f), which is insensitive to batch effects (Extended Data Fig. 6c-d) and robust to subsampling of the data or analysis of specific chromosomes (Extended Data Fig. 6e-f). In summary, we developed an in-silico cell-cycle

phasing approach that organises single cells along a smooth trajectory of highly varied chromosome conformations (Fig 1g, Extended Data Fig. 5) leading from the mitotic state through replication and back into mitosis.

To confirm our Hi-C based cell-cycle phasing, we analysed 1169 cells sorted based on Geminin and DNA content using fluorescence-activated cell sorting (FACS) and found that the ordering of cells based on their contacts is remarkably consistent with their stage as determined by FACS (Fig. 2a-b). We further hypothesise that the refined ordering we impose within each coarse-grained stage is likely to represent cell-cycle progression at high resolution. This was initially demonstrated by correlating Hoechst/Geminin FACS indices to the inferred phasing rank of 342 single-cell Hi-C profiles (Fig. 2c, Extended Data Fig. 6g), deriving correlation even within classical sorting gates. To further support this idea we analysed the distribution of key statistics for single cells ordered according to their inferred cell-cycle stage. For reference, we projected the repli-score on the cell-cycle progression (Fig. 2d), validating that it was low during G1 and G2, showed consistent gradual increase during early S phase, and decreased during late S phase. The total number of contacts (Fig. 2e), which unlike the repli-score is a-priori independent of the inferred phasing, was shown to increase gradually during S-phase, peaking at G2 at almost a 2-fold higher level than G1, before dropping sharply after mitosis. The fraction of *trans*-chromosomal contacts exhibits an unforeseen cell-cycle dynamic, with a smooth curve from its lowest point of around 3% in G2 and mitosis, to ~8% in early S phase (Fig. 2f). Next, we quantified the *trans*-alignment of chromosomes, which is a strong characteristic of mitotic cells. Remarkably, despite being a-priori unrelated to the inferred G1 ordering scheme, we found the *trans*-alignment score declined smoothly from its peak at mitosis to its nadir at the end of the G1 phase, immediately before the start of S phase (Fig. 2g; see Extended Data Fig. 6h for examples). Collectively, these results validate our phasing strategy and open the way to analyse cell-cycle dynamics of Hi-C architectural features at high resolution.

Cell-cycle chromosomal dynamics

We identified 3434 TAD borders in the pooled Hi-C map using analysis of contact insulation signatures⁶ (Extended Data Fig. 3). We created sub-pools of G1, early-S and late-S/G2 cells *in silico* to compare insulation patterns between the three major cell-cycle phases. We found that the locations of TAD borders are generally unchanged in these pools, but that the average intensity of the borders is dynamic (Extended Data Fig. 7a-c). We calculated the mean insulation at borders for each cell (Fig. 3a) and found that the intensity of insulation varies markedly over the cell cycle. Insulation is not detectable in mitotic maps (Extended Data Fig. 7d-e), and emerges rapidly upon exit from mitosis in early G1. Insulation increases to its maximum during G1, but declines when replication begins, plateauing at mid S-phase at its lowest interphase level and remaining fixed through G2 until mitosis. We note that since insulation is defined by the depletion of short-range contacts crossing TAD borders, it may be affected by changes in the frequency of contacts in the 25 kb-2 Mb distance range. However, this overall frequency only partially predicts single-cell insulation (Extended Data Fig. 7f).

Next, we studied the dynamics of compartmentalisation of the genome into active (A) and non-active (B) groups of TADs by analysing the *cis*-chromosomal compartment linkage. For each cell we estimated the depletion of A-B contacts given the observed A-A and B-B counts (Methods, Fig. 3b). Similar to insulation, compartmentalisation is weakest during mitosis, as expected from the condensed uniform structure at that stage¹¹. However, in contrast to TAD insulation, compartmentalisation is weak during G1 phase and increases upon entering S phase, with a further increase in late S phase, reaching a maximum in G2, just before a rapid and complete loss approaching mitosis. Compartmentalisation calculated from *trans*-chromosomal contacts shows exactly the same trend (Extended Data Fig. 7g). The distinct cell-cycle dynamics of TAD insulation and compartment structure suggest that independent mechanisms contribute to these phenomena.

To study the cell-cycle dynamics of insulation at individual TAD borders and compare it to the global trend above, we grouped single cells into 20 time-slots according to their inferred cell-cycle phase, and pooled data from each time-slot to compute an insulation time series for each border. We then clustered the borders based on their dynamics in these time-slots, and examined the mean insulation in each cluster per cell, along the inferred phasing. Borders demonstrate similar qualitative dynamics in all clusters. However, each cluster varies in timing and intensity of deterioration in insulation during S-phase (Fig. 3c). Cluster C4 includes borders showing the strongest G1 insulation and C2 borders show the earliest loss of insulation in S. Interestingly, borders in both clusters tend to be located between early-replicating TADs (Fig. 3d). Borders in C3 show a gradual increase in insulation well into S-phase, followed by abrupt loss of insulation in mid-S. This cluster is enriched for borders separating early-replicating from mid/late-replicating TADs (Fig. 3c). C1, which has the weakest G1 insulation, shows gradual and delayed loss of insulation, and is enriched for borders between late-replicating domains. These data suggest that loss of insulation in S-phase is temporally coupled to the replication process. Specifically, replication through a border or in both adjacent TADs is linked to loss of insulation (C1, C2, C4), whereas replication of only one adjacent TAD is linked to a temporary increase in insulation (C3). In all cases, insulation remains low post-replication, not re-emerging until the next cell cycle.

We identified pairs of looping loci by analysis of pooled matrices and shuffled pooled matrices (Extended Data Fig. 7h). We focused on 2036 top-scoring loops that were associated with convergent CTCF motifs and evidence for CTCF/Cohesin binding (Methods). We then grouped loops according to the replication time of their anchors (early-early vs. late-late) and computed for each group and each single cell the total number of looping contacts in 20x20 kb bins around the loop anchors, normalized by the number of contacts in the surrounding 60x60 kb bin (Fig. 3e). We found that loop enrichment increases after mitosis, is generally stable through G1, but then decreases slightly in early S phase for early-replicating loops and in late S phase for later-replicating loops. Spatial enrichment analysis shows conserved localized loop enrichment (Fig. 3f). We do not detect loops that are specific to a certain cell-cycle phase (Extended Data Fig. 7i), suggesting that even when decreased in average intensity during replication, loops are significantly observed in the data. However, the assay's resolution is not sufficient to allow detection of individual loop turnover at single-cell resolution, leaving unresolved regimes for rapid loop dynamics. Interestingly, loops are still apparent in cells approaching mitosis despite the fast

chromosomal condensation – unlike cells after mitosis, which do not have any contact enrichment around loops (Fig. 3f). The notable lack of loop contact enrichment in mitotic maps does not necessarily imply loops are disassembled at this stage, since mitotic condensation may increase the local contact frequency around loop anchors and decrease Hi-C sensitivity.

Given the relative stability of looping contacts that demarcate TAD structures and the dynamic cell-cycle rise and fall in TAD insulation, we next asked if chromatin structure within TADs changes during the cell cycle, or if TADs constitute stable chromatin blocks that are re-organised dynamically. We computed average intra-TAD contact distances and used these as a measure of TAD condensation per cell, reasoning that condensed TADs will be characterised by higher mean contact distance than de-condensed TADs^{3,16}. We compared condensation for groups of TADs while controlling for size and time of replication (Fig. 3g). In G1, early-replicating TADs are de-condensed compared to mid- or late-replicating TADs, consistent with open chromatin in active regions. Early-replicating TADs are further de-condensed in early S phase, before showing a remarkable increase in condensation at mid-S, likely after their replication is complete. Following that, and during late S and G2, early-replicating TADs increase their mean condensation to levels comparable to those of late-replicating TADs in G1. Condensation of mid-early replicating TADs follows a similar trend, with some delay and damped amplitude. Mid-late and late-replicating TADs, on the other hand, show constitutively high condensation in G1 and throughout S, with further marked increases observed for mid-late TADs in G2. We note that the potential involvement of *trans* sister chromatid contacts in driving the observed increase in mean contact distance cannot be determined unambiguously. However, the delay between replication of early domains at the start of S phase and their increased compaction in mid/late S suggests sister-chromatid contacts cannot fully account for the dynamics observed.

We next generated 1247 QC positive single-cell Hi-C maps from haploid mouse ESCs, to allow systematic analysis of conformation in single-copy chromosomes. We confirmed haploid ESC maps show comparable QC metrics to diploid cells, with less than half of the contacts per cell and some specific patterns of chromosome loss that we filtered in-silico (Extended Data Fig. 8a-c). We derived a pooled ensemble haploid Hi-C map and computed TAD structures, insulation and A/B compartments (compared to diploid in Extended Data Fig. 3f). We then used the same approach for cell-cycle phasing that was developed for diploid cells (Fig. 4a). Remarkably, we observe a similar cyclical trend, and analogous dynamics of insulation, loop enrichment and TAD condensation (Extended Data Fig. 8d-g). The universality of these processes shows that haploid ESCs are comparable to normal diploid cells, and a good model for studying chromosome conformation and dynamics, while providing the opportunity for unambiguous 3D modelling of single-copy chromosomes for G1 cells.

A spectrum of epigenetic compartments

We defined A-scores for each TAD as the fraction of *trans*-chromosomal contacts made with the A-compartment, and found a spectrum that generalizes the original discrete partitioning of the genome into two compartments (Fig. 4b). A-scores correlate with time of replication;

high A-score TADs are early-replicating and low A-score TADs are late-replicating (Fig. 4c). These scores are conserved between diploid and haploid cells and growth conditions (Extended Data Fig. 9a), suggesting that a continuum of compartment affinities is a robust property of chromosomal domains. Such a continuum may be created if individual TADs switch compartments in individual cells, thereby creating a gradient of A-compartment affinities. Alternatively, it is possible that the A-score gradient is conserved at the single-cell level, representing a structural continuum, rather than a two-compartment structure that is heterogeneous across a cell population. To explore these questions at single-cell resolution we needed sufficient TAD coverage in each cell, so we computed a single-cell A-compartment association score for each TAD, using the mean A-score of its long-range (> 2 Mb) *cis*-contacting TADs. *Cis* A-association scores and *trans* A-scores are highly consistent on average (Fig. 4d), but the *cis* A-association score is defined by more contacts, while being constrained by the chromosomal larger-scale domain architecture. The standard deviation of A-association scores for individual TADs across the entire haploid single-cell cohort was less than 0.06 (Fig. 4e), suggesting limited variability of A-score between cells. Further analysis (Fig. 4f) showed that A-score variation decreases in early and late S-phase, consistent with the overall increase in A/B compartment strength above. Both sorting TADs by their mean A-association scores (Fig. 4g) and grouping these distributions by unsupervised clustering (Extended Data Fig. 9b) showed that TADs vary quantitatively in their A-association, but no significant group of TADs shows a bimodal compartment association distribution. The same analysis indicated that compartment association polarises in late S/G2, with many TADs shifting from broadly spread association in G1 or early S, to preferential contacts with TADs in a narrower range of A-scores following replication.

To further explore this effect we discretised the A-association score continuum, deriving 20 groups of TADs that corresponded to putative *pseudo-compartments*. We tested how such pseudo-compartments contact one another *trans*- and *cis*-chromosomally, and observed significant *cis*-chromosomal interaction between TADs within the same and neighbouring pseudo-compartments (Fig. 4h and Extended Data Fig. 9e, top). In *trans* we observed a more binary contact structure in G1 cells, transitioning toward a gradient of pseudo-compartment preferences following replication (Fig. 4i). A-association analysis of diploid cells shows similar behaviour (Extended Data Fig. 9c,d,f). Analysis of RNA-seq, histone marks¹⁵ and Lamin association¹⁷ suggested pseudo-compartments are characterized by a gradient of epigenetic signatures (Extended Data Fig. 9g-h). Overall, the ordering of TADs according to A-association score is significantly correlated with a linear model weighting H3K4me1 and H3K4me3 scores ($r=0.79$, $p<<0.001$). Together, these data suggest pseudo-compartments capture large-scale *trans*- and *cis*-chromosomal conformation at the population and single-cell level, generalising the binary A/B compartment scheme.

3D whole genome modelling and analysis

Using G1 haploid cell data we developed a structure sampling approach for inferring whole genome 3D conformation. We applied stringent contact filtering to eliminate contacts that are more likely to be technical artifacts (Methods), removing a median of 0.89% of the contacts from each cell (Extended Data Fig. 10a). We applied a simulated annealing protocol¹² to sample over 30,000 whole genome 3D models for 190 single-cell datasets,

using contacts as distance restraints at 0.1, 0.5 and 1Mb resolutions. The resulting 3D models satisfied nearly all observed *cis*- and *trans*-chromosomal contacts, with a median of <0.012% violated constraints per cell (Extended Data Fig. 10b-c). Subsampling 10,000 contacts per cell produced consistent models (Extended Data Fig. 10d-e), while cross-validation tests showed that structural modelling is sufficiently robust to re-discover *trans*-chromosomal distances between a pair of contacting chromosomes, even when all contacts between those chromosomes are discarded (Extended Data Fig. 10f).

To study changes in chromosomal structure as cells emerge from mitosis into G1, we analysed over 20,000 models from 126 cells with > 10,000 contacts per cell and < 0.5% constraint violations. Analysis of large-scale *cis*-chromosomal shapes shows that individual chromosomes rapidly de-condense from a rod-like mitotic conformation to a more spherical structure as G1 progresses (Fig. 5a-b, Extended Data Fig. 10g-h). We monitored the expansion of chromosomes by measuring the distance between adjacent 500 kb chromatin segments in A or B compartments in each cell (Fig. 5c). A and B compartments are similarly compact upon mitotic exit ($P > 0.05$, two-sided MW test), but as cells progress through G1, A compartments expand more rapidly and to a larger extent than B compartments ($P < 0.005$, one-sided MW test), consistent with decondensed euchromatin and compact heterochromatin. Remarkably, the twenty high-resolution pseudo-compartments form a gradient that generalises this observation, with the strongest B-like pseudo-compartments most compact and the strongest A-like pseudo-compartments the most expanded (Fig. 5d).

We then computed the mean nuclear radial positioning of A and B compartments as cells exit mitosis. We observed B compartments occupying more peripheral locations and A compartments tending toward the nuclear interior for cells 9-16 and onward ($P < 0.001$, one-sided MW test, Fig. 5e). Pseudo-compartments generalise this trend, showing progressively more differentiated radial positioning through G1 phase (Fig. 5f). The structural models also confirm and refine the preferential contacts among pseudo-compartments (as shown in Fig. 4h-i), and show that long-range *cis*-contacts are re-established very early in G1 phase, while *trans* compartmentalisation is not re-established until late G1 (Fig. 5g-h), consistent with 4C observations¹⁸. Finally, distributions of chromosomal conformations also suggest preferential contacts of Nucleolar Organiser Regions¹⁹ (NOR, $P < 0.01$ one-sided MW test, Extended Data Fig. 10i).

Discussion

Using new high-throughput single-cell Hi-C technology, we generated and analysed single-cell contact maps from thousands of cells, and established a strategy for cell-cycle phasing that decouples TADs, compartments and loops that were previously superimposed in bulk maps. We confirmed that genome folding is highly heterogeneous at the single-cell level, but variation is far from random, reflecting a combination of deterministic dynamics and stochastic effects. Our data suggest that the most dramatic source of deterministic dynamics in the mouse ESC is the cell cycle. Consistent with previous reports at lower temporal resolution¹¹, we show chromosome condensation in preparation for mitosis, and rapid expansion in early G1, while also finding extensive remodelling during replication.

Importantly, the effects of these major transitions extend throughout the entire cell cycle, such that the genome is not stably folded at any point in interphase.

Our methodology and other techniques for high-resolution or high-throughput single-cell Hi-C20–22 now allow high-resolution analysis of other factors driving chromosome folding and genome regulation. It will be extremely important to interpret future single-cell studies and bulk Hi-C experiments through the cell-cycle prism introduced here. An emerging challenge is therefore the derivation of comprehensive models, combining cell-cycle chromosomal dynamics with gene regulatory processes during differentiation22 or response to stimuli. Using such models, it may become possible to understand how nuclei are capable of supporting the remarkably deterministic and pronounced chromosomal dynamics of mitosis and replication while implementing robust and flexible gene regulatory programs – a question that is rapidly becoming one of the key challenges in epigenetics.

Methods

Cell culture, fixation and sorting

Two mouse ESC lines were used. One diploid F1 hybrid ESC from an intercross of *Mus musculus* 129/SV-Jae and *Mus musculus castaneus* (129 x Castaneus, a gift from Joost Gribnau) and one haploid ESC23 (H129-1; ECACC 14040203, gift from Martin Leeb and Anton Wutz) were grown either in 2i medium without feeder cells or ES-DMEM with fetal bovine serum (FBS) on feeder cells24. Both cell lines were tested for mycoplasma contamination. For harvesting ESCs from cultures on feeder cells, the feeder cells were depleted using Feeder Removal MicroBeads (Miltenyi Biotec), followed by fixation and permeabilisation of the enriched ESCs and staining for Oct-3/4-immunoreactivity as described below. Oct-3/4-positive cells were collected by Fluorescence-activated cell sorting (FACS), ready for subsequent Hi-C processing.

To fix the cells, 50 to 200 million ESCs were suspended in relevant medium and fixed for 10 min by adding formaldehyde at a final concentration of 2% at room temperature before quenching with 127 mM glycine for 5 min on ice. The cells were washed with PBS and permeabilised in 10 mM Tris-HCl (pH 8), 10 mM NaCl, 0.2% IGEPAL CA-630 with cOmplete EDTA-free protease inhibitor cocktail (Roche) for 30 min on ice with intermittent agitation, and spun to collect the nuclei pellet, which was ready for Hi-C processing.

To enrich for specific cell types before the Hi-C process, cells were blocked with PBS-FT (5% FBS, 0.1% Tween-20 in PBS) at room temperature for 1 hr, incubated with a primary antibody in PBS-FT as required (anti-Oct3/4, Santa-Cruz sc-5279, at 1:100 dilution; or anti-Geminin, Abcam 175799 at 1:400 dilution) for 1 hr on ice. Samples were then washed with PBS-FT, and incubated with the secondary antibodies (Alexa Fluor 647-anti-mouse IgG, Thermo Fisher A-31571, or Alexa Fluor 555-anti-rabbit IgG, Thermo Fisher A-21428) for 30 min on ice, washed with 2% FBS in PBS, stained with Hoechst 33342 at the final concentration of 15 µg/ml and subjected to FACS by Influx (BD Biosciences) to collect nuclei cells of interest.

For some diploid cells (batch id = 27 to 35 in Extended Data Fig. 1i), the same labelling procedure with anti-Geminin antibody and Hoechst 33342 was applied after Hi-C processing. Then Geminin-immunoreactivity and DNA content of individual single cells were recorded during FACS into 96-well plates.

Hi-C processing

The cells (approximately half million to several million) were washed with 1.24 x NEBuffer 3 (New England Biolabs; 62 mM Tris-HCl [pH 7.9], 124 mM NaCl, 12.4 mM MgCl₂, 1.24 mM DTT) and suspended in 400 µl of 1.24 x NEBuffer 3. Six µl of 20% SDS was added and incubated at 37°C for 60 min with constant agitation, then 40 µl of 20% Triton X-100 was added and incubated at 37°C for 60 min with constant agitation. Next, 50 µl of 25 U/µl Mbo I (New England Biolabs) was added and incubated at 37°C overnight with constant agitation. To label the digested DNA ends, 1.56 µl of 10 mM dCTP, 1.56 µl of 10 mM dGTP, 1.56 µl of 10 mM dTTP, 39 µl of 0.4 mM biotin-14-dATP (Thermo Fisher) and 10.4 µl of 5 U/µl DNA polymerase I, large (Klenow) fragment (New England Biolabs) were added and incubated at 37°C for 45 min with occasional mixing. The sample was then spun and supernatant partially removed leaving 50 µl with cells, followed by addition of 100 µl of 10x T4 DNA ligase reaction buffer (New England Biolabs), 10 µl of 100x BSA (New England Biolabs), water and 10 µl of 1 U/µl T4 DNA ligase (Thermo Fisher) were added to make the total volume 1 ml, and incubated at 16°C overnight. Then the nuclei were passed through a 30 µm cell strainer and single nuclei were sorted into individual empty wells in 96 well plates using an Influx cell sorter. The plates were sealed and stored at -80°C until further processing.

Single-cell Hi-C library preparation

To prepare single-cell Hi-C libraries from single nuclei in a 96 well plate, 5 µl of PBS was added to each well, the plate was sealed and crosslinks reversed by incubating at 65°C overnight. Hi-C concatemer DNA was fragmented and linked with sequencing adapters using the Nextera XT DNA Library Preparation Kit (Illumina), by adding 10 µl of Tagment DNA Buffer and 5 µl of Amplicon Tagment Mix, incubating at 55°C for 5 min, then cooling down to 10°C, followed by addition 5 µl of Neutralize Tagment Buffer and incubation for 5 min at room temperature. Hi-C ligation junctions were then captured by Dynabeads M-280 streptavidin beads (Thermo Fisher; 20 µl of original suspension per single-cell sample). Beads were prepared by washing with 1 x BW buffer (5 mM Tris-Cl pH 7.5, 0.5 mM EDTA, 1 M NaCl), resuspended in 4 x BW buffer (20 mM Tris-Cl pH 7.5, 2 mM EDTA, 4 M NaCl; 8 µl per sample), and then mixed with the 25 µl sample and incubated at room temperature overnight with gentle agitation. Using a Bravo automated liquid handling system (Agilent Technologies), the beads were then washed four times with 200 µl of 1 x BW buffer, twice with 200 µl of 10 mM Tris-Cl pH 7.5 at room temperature, and resuspended in 25 µl of 10 mM Tris-Cl pH 7.5. Single-cell Hi-C libraries were amplified from the beads by adding 15 µl of Nextera PCR Master Mix, 5 µl of Index 1 primer of choice and 5 µl of Index 2 primer of choice. Samples were then incubated at 72°C for 3 min, 95°C for 30 sec followed by the thermal cycling at 95°C for 10 sec, 55°C for 30 sec and 72°C for 30 sec for 12 or 18 cycles, then incubated at 72°C for 5 min. The supernatant was separated from the beads, and the 96 supernatants from a 96 well plate that had 12 cycles of amplification were combined

together whereas the supernatants from 18 cycles of amplification were processed uncombined. The combined or uncombined supernatant was purified with AMPure XP beads (Beckman Coulter; 0.6 times volume of the supernatant) according to manufacturer's instructions and eluted with 10 mM Tris-Cl pH 8.5 (100 µl when 96 samples were combined; 30 µl when sample was uncombined). The eluate was purified once more with AMPure XP beads (equal volume to the previous eluate) and eluted with 11 µl of 10 mM Tris-Cl pH 8.5.

Strand-specific nuclear RNA-seq library preparation

Ten to 20 million unfixed ESCs (grown in 2i medium without feeder cells) were washed with cold PBS, resuspended in 0.5 ml of cold buffer RLN (50 mM Tris pH 7.5, 140 mM NaCl, 1.5 mM MgCl₂, 1 mM DTT, 0.4% IGEPAL CA-630) and incubated for 5 min on ice. The samples were spun and the resultant pellets (nuclei) were washed with buffer RLN and subject to RNA isolation using RNeasy Mini Kit (Qiagen) and QIAshredder spin columns (Qiagen) according to the instructions from the manufacturer. The strand-specific RNA-seq libraries were prepared from 200 ng each of nuclear RNA samples using TruSeq Stranded Total RNA LT Sample Prep Kit (Illumina) according to manufacturer's instructions.

library QC and sequencing

Before sequencing, the libraries were quantified by qPCR (Kapa Biosystems) and the size distribution was assessed with Agilent 2100 Bioanalyzer (Agilent Technologies). They were sequenced by either 2 x 50 bp, 2 x 75 bp or 2 x 150 bp paired-end run by HiSeq 1500, HiSeq 2500 or NextSeq 500 (Illumina).

Sequence processing pipeline

Paired-end reads of sequencing batches were de-multiplexed to single-cell datasets based on the two 8 bp unique identification tags attached to each cell sample during library amplification. Reads lacking a perfect match were discarded. Reads were broken down to segments on their matches to the MboI recognition site (GATC), concatenating segments shorter than 16 bp with their adjacent segment. Each segment was independently mapped to the *Mus Musculus* genome (assembly mm9) Using Bowtie225 in *end-to-end* alignment mode. The uniquely mapped (MAPQ > 36) segments from each paired-end read were consolidated into a chain of segments by merging segments mapped to overlapping genomic regions. Next, the segments chain was translated to a chain of fragment-ends (fends) by associating each segment with its downstream fend. Finally, adjacent fend pairs from all chains were used to comprise the contact map of the cell. For the single-cell analysis, we used the distinct contacts in each cell, ignoring the number of times each contact appeared in the map. When analysing pooled contact maps we also considered the number of times each contact appeared.

Quality controls and selection of high quality contact maps

We calculated several metrics to evaluate the quality of each contact map:

1. *Coverage*: Total number of contacts

2. *Trans fraction*: Fraction of inter-chromosomal contacts out of the total number of contacts
3. *Non-digested fraction*: fraction of contacts between fends distanced less than 1 kb from each other out of the total number of contacts. The vast majority of these ultra-close cis contacts originate from non-digested DNA and therefore this metric quantifies the digestion efficiency of the experiment.
4. *Max chromosomal coverage aberration*: We calculated coverage enrichment per chromosome and cell by comparing the number of contacts a chromosome forms with the expected number given the mean fraction of contacts per chromosome across the entire single cell pool. The *max chromosomal coverage aberration* of a cell is the maximal enrichment (or depletion) of this log₂-ratio across all its chromosomes.
5. *Strongest contact decay bin*: The contact distance bin with the most contacts (see below for definition of the logarithmic decay bins)

The following filters were applied on the contact maps to select high quality ones:

Metric	Diploid condition	Haploid condition
Coverage	20K-700K	10K-1M
<i>Trans</i> fraction	< 15%	< 20%
Non-digested fraction	< 55%	< 50%
Max chromosomal coverage aberration	½ to 2	½ to 2
Strongest contact decay bin	> 45 (46 kb)	> 45 (46 kb)

We found no systematic technical differences between 2i- and serum- cultured cells (Extended Data Figs. 1h-j and 8a-c), so we applied the same quality control filters for both. For the diploid analysis we only used 2i-cultured cells.

In our previous version of the single-cell Hi-C protocol¹² we found that molecules supported by a single read (singleton contacts) usually represented spurious ligations and were therefore discarded. In contrast, the improved single cell Hi-C approach shows no significant difference in QC metrics between singly and doubly covered molecules (represented as segment chains in the processing pipeline). For example, we found that singleton chains have the same cis to trans ratio as multiply covered chains (mean of 93% versus 92.9%), suggesting low level of spurious contacts is represented in both classes. We note that in addition to overall improved processing robustness and quality, the sequencing depth of our current libraries, in contrast to the previous version, is usually far from saturation (Extended Data Fig. 1b-c) so the fraction of singleton chains is much higher (mean 21%, Extended Data Fig. 1a).

Repli-score and time of replication analysis

Repli-score quantifies the coverage enrichment of early replicating regions over late replicating ones. To calculate it we characterized each fend by the mean Repli-chip value

around it (10Kb upstream and downstream), using replicate 1 of the *Mus Musculus* 129 ES-D3 Repli-chip dataset¹⁵. Repli-chip values are bimodal around zero, with early replicating loci having positive values and late loci negative ones. The repli-score of a cell is the frequency of early replicating fends in its contact map, divided by the minimal frequency found across all QC passed cells (to scale the score to be relative to one).

The generation of ordered correlation matrix of normalized domain copy-number (Extended Data Fig. 4c) is very sensitive to chromosomal copy-number variations. Therefore we selected cells with strictly normal chromosomal coverage (cells with *max chromosomal coverage aberration* between 0.82-1.23), ending up using 46% of the cells. We next counted per cell the total number of contacts formed by each domain (*trans* and *cis* above 23.17Kb). Domains without time of replication information and domains with extremely low (< 0.04) or high (> 0.16) mean number of contacts per Kb (4.8% of the domains) were discarded. The 50 domains with the earliest mean time of replication and the 50 domains with the latest one were marked as top-early and top-late, respectively. Contact numbers were transformed to contact frequencies. The Pearson correlation matrix between all pairs of domains was then calculated, and domains were ordered by their coverage profile mean Pearson correlation with the top-early domains coverage profile minus the mean Pearson correlation with the top-late profiles. The same steps were repeated to order the domains by their A-score instead of by their time of replication (Extended Data Fig. 10g) only with ordering the domains by their difference of mean correlation with the 50 domains with highest A-score to their mean correlation with the 50 domains with lowest A-score.

In-silico cell phasing over the cell cycle

We counted the number of *cis* (intra-chromosomal) contacts per cell, binning contacts by distance into logarithmic bins (143 bins, first one for contacts distanced < 1Kb, then each bin covers an exponent step of 0.125, using base 2). Contacts in bins 1-37 were found to be noisy and were discarded, making bins 38-143 the valid bins. The following metrics were used to phase the cells:

- *% near* - percentage of contacts in bins 38-89 out of all valid bins
- *% mitotic* – percentage of contacts in bins 90-109 out of all valid bins
- *farAvgDist* – mean contact distance considering bins ≥ 98
- *rawRepliScore* – fraction of early-replicating fends out of all fends in the contact map (see above for details)

Each cell was assigned to a group and cells within each group were ordered by these criteria (scale means subtracting the mean and dividing by the standard deviation):

Group	Typical phase	Assignment Criteria		Ordering by
		Diploid cells	Haploid cells	
1	Post-M	<i>% mitotic</i> 30 Δ <i>% near</i> 50	<i>% mitotic</i> 30 Δ <i>% near</i> 42	– <i>% mitotic</i>

Group	Typical phase	Assignment Criteria		Ordering by
		Diploid cells	Haploid cells	
2	G1	% near 63	% near 61.1	$scale(\% \text{ near}) + scale(farAvgDist)$
3	Early to mid-S	$63 < \% \text{ near} \leq 78.5$	$61.1 < \% \text{ near} \leq 77$	$\frac{\% \text{ near}}{VAR(\% \text{ near})} + \frac{rawRepliScore}{VAR(rawRepliScore)}$
4	Mid-S to G2	% near > 78.5	% near > 77	$\frac{\% \text{ near}}{VAR(\% \text{ near})} - \frac{rawRepliScore}{VAR(rawRepliScore)}$
5	Pre-M	% near > 50 $\wedge$ % near + 1.8 $\times$ % mitotic > 100	% near > 42 $\wedge$ % near + 1.8 $\times$ % mitotic > 100	% mitotic

Cells are first assigned to groups 1 and 5 and those remaining are assigned to groups 2-4. To compensate for the arbitrary thresholds (which may introduce discontinuity on the interfaces between the three main phases), the initial assignment to groups is then fine-tuned by clustering the contact distance distribution profiles of the cells to small clusters (k-means, $k = \text{number of cells} / 10$) and reassigning cells in each cluster to the majority vote of that cluster. This resulted in reassigning 5.7% and 8.2% of the cells in the diploid and haploid datasets, respectively, with almost all (>99.8%) re-assignments involving adjacent groups.

To test the stability of our phasing procedure among the different chromosomes we compared the positions of cells by phasing using disjoint sets of chromosomes with roughly the equivalent length (Extended Data Fig. 7c, 1st set: chromosomes 2, 3, 5, 6, 9, 10, 12, 15, 17 and 19, 2nd set: chromosomes X, 1, 4, 7, 8, 11, 13, 14, 16, and 18). Phasing positions determined by both sets largely concur, with most of the differences concentrated on the interfaces between G1 and early S, or early S late S/G2 groups. The positions of only 38 out of the 1942 cells participated in both sets (<2) differ by more than 10%.

To perform the non-linear dimensionality reduction on the 2i diploids' contact distance profiles (Extended Data Fig. 4e), spectral embedding based on the above logarithmic bins (discarding the first <1Kb distance bin) was used to project the cells onto a two-dimensional plane while preserving local similarities. Specifically, for each two cells we computed the distance between their discretized decay distributions as:

$$d_{i,j} = D_{KL}(Cell_i, Cell_j) + D_{KL}(Cell_j, Cell_i)$$

We then generated the adjacency matrix of the k-NN graph ($k=7$), and used the 2nd and 3rd-smallest eigenvectors of the graph Laplacian as the projection axes.

Pooled contact maps

Contacts from single cell maps were pooled together to generate aggregated maps. We used the total number of contacts when creating ensemble-like maps, e.g. for the pool of all cells and the phase groups, and distinct contacts when creating contact maps comprising a few dozen cells (e.g. Fig. 1h, Fig. 3b). To normalize the pooled maps (Extended Data Figs. 3a and 8h), to compare pooled contacts around epigenetic landmarks (Fig. 3e), and to compare *cis* contact enrichment over a background contact distribution (Fig. 4h, Extended Data Fig.

10e-f), we sampled a randomized ensemble Hi-C map²⁶ and compared contact distributions in the observed and randomized data using the Shaman tool (<https://bitbucket.org/tanaylab/shaman>). Briefly, we shuffled contacts using a Markov Chain Monte Carlo-like approach, first within each chromosome and then genome-wide, such that the primary factors that define contact distributions are preserved - the marginal coverage and contact distance distribution - but any compartment or TAD structure that may be present is not retained. The resulting expected ensemble map has an identical number of contacts for each locus, and identical probability for a contact at a given genomic distance, as in the observed map. We generated shuffled maps from the pools of 2i diploids single-cells and the pool of 2i and serum haploids single-cells.

Insulation, domain and border calling and domains initial A/B classification

The insulation of a locus (Supplementary Fig. 1) corresponds to depletion of contacts in the “a” region. We calculate an insulation score by considering a distance around the locus (scale) and comparing the total number of contacts that violate insulation (in region *a*) with the total number of contacts around the locus (in regions *a*, *b₁* and *b₂*). In both counts contacts distanced less than 1Kb from the diagonal are ignored (these are mostly non-digested contacts). The negative log-ratio of these counts correlates positively with strong insulation. We calculated insulation on the pooled contact maps in 1Kb resolution, using a scale of 300Kb. Domains were then identified as contiguous regions in which insulation score is below the 90% quantile of the genome-wide distribution. Domains shorter than 20Kb or longer than 4Mb and domains containing a non-mappable region longer than 25kb were discarded, resulting in 2894 domains (or 2649 domains for the haploid pool of cells).

The initial assignment of domains to the A and B compartments was done by k-means clustering (K=2) the inter-domain contact profile, using only inter-chromosomal (*trans*) contacts, and computing log₂ ratio of observed and expected (based on genomic length) inter-TAD contacts²⁰. Domains in each cluster exhibit distinct epigenetic and time of replication signatures (Extended Data Fig. 3d-e) pointing us to assign the 1221 domains in cluster 1 to the B-, inactive-compartment and 1673 domains in cluster 2 to the A-, active compartment (haploids had 1353 A- and 1296 B-domains).

Domain borders were called from the insulation profile of the pooled map by identifying highly insulating regions between domains (insulation above the 90% quantile) and selecting in each element the 1KB with highest insulation score. To calculate the cell mean insulation over a set of borders *B* (either all borders as in Fig. 3a or a subset of borders as in Fig. 3b) the total number of border violating contacts is compared to the total number of contacts around the border. Denoting *A*[*b*] as the number of contacts in the *a* region of border *b* (Supplementary Fig. 1) and similarly *B₁*[*b*] and *B₂*[*b*]:

$$INS_B = \log_2 \frac{\sum_{b \in B} A[b] + B_1[b] + B_2[b]}{\sum_{b \in B} A[b]}$$

To cluster borders by their insulation profile across the inferred cells phasing (Fig. 3b) cells were divided to 20 slices (100 cells in each slice grouped according to phasing), insulation was calculated per border and slice and the resulting 20 length border insulation profiles were clustered ($k=4$). Fig. 3b shows per cell (and not per slice) the mean insulation over each of the resulting clusters.

Inter-chromosomal alignment

The head-to-head alignment of chromosomes in single-cell mitotic Hi-C maps, likely represents late telophase or very early G1 nuclei (Extended Data Fig. 4b). To quantify the degree of head-to-head alignment of the chromosomes of each cell, we extracted trans-chromosomal contacts and scaled the coordinates of the contacting fends by the length of their respective chromosomes. Alignment was then approximated using the Pearson correlation between the two scaled fend coordinate vectors (Fig. 2g).

A/B compartment score

Intra-chromosomal, inter-domain contacts were classified according to the compartment association of the contacting domains into A-A, A-B and B-B. Contacts within domains less than 2Mb apart were discarded. We summarized the statistics of observed total compartment contact per cell as (O_{AA} , O_{AB} , O_{BB}). The compartment score of a cell is the depletion of A-B contacts over the count expected by the marginal distribution of A or B contacts:

$$T = O_{AA} + O_{AB} + O_{BB} \quad P(A) = \frac{2O_{AA} + O_{AB}}{T} \quad P(B) = \frac{2O_{BB} + O_{AB}}{T}$$

$$CompScore = \log_2 \frac{2 \cdot P(A) \cdot P(B) \cdot T}{O_{AB}}$$

Analyzing CTCF loops

We followed these steps to generate a set of loops:

- Genomic regions with strong CTCF ChIP-seq¹⁵ signal were screened (50 bp regions on which either of the two ChIP-seq replicates genome-wide quantile is larger than 99%)
- Strong CTCF motifs were screened (Motif strength > 99.988% of the genome-wide 50 bp regions quantiles distribution)
- Each 50bp window was classified as *C* (only ChIP), *F* (ChIP + forward motif), *R* (ChIP + reverse motif) or *B* (ChIP + forward and reverse motif).
- Each ChIP-supported motif is classified by its upstream and downstream (200Kb) mean time of replication (ToR) as early (if mean ToR is positive) or late (if mean ToR is negative), resulting in the following anchor counts (discarding 116 loops with missing ToR information):

Upstream ToR	Downstream ToR	Denoted	Count
Early	Early	EE	55,778

Upstream ToR	Downstream ToR	Denoted	Count
Early	Late	EL	2,459
Late	Early	LE	2,598
Late	Late	LL	10,278

- Convergently oriented ChIP-supported motifs (B/F anchor upstream to a B/R one) distanced 200Kb-1Mb from each other were considered as loop candidates (total of 760,712)
- Loop candidates were further filtered by coverage enrichment in the normalized pool of 2i diploids contact map, requiring contact enrichment score higher than 60 within a 20x20Kb window centered on the loop, resulting with 61,796 loop candidates
- Loop foci enrichment (number of contacts in a 20x20Kb vs. 60x60Kb window centers on the loop) was calculated and the top quintile (20%, 12,361 loops) was selected.
- Overlapping loops (with overlapping 60x60Kb window center on the loop) were filtered, resulting in a final set of 2036 loops, with the following anchors classification:

1 st anchor	2 nd anchor	Denoted	Count
EE	EE	Early-Early	762
EE	LL	Early-Late	161
LL	EE	Late-Early	123
LL	LL	Late-Late	456
Mixed (EL/LE)	Mixed (EL/LE)	Mixed	534

Loop foci enrichment (Fig. 3e) quantifies how concentrated contacts are around a loop. We calculate the ratio between contacts in a small window (20x20kb) centered on the loop and contacts in a larger window (60x60kb), normalizing by the expected ratio if contacts were uniformly distributed (1/9). To get a mean loop foci enrichment over a group of loops, the sum of contacts in all small windows is compared with the sum of contacts in the larger ones.

Spatial analysis of contact distribution around loops (Fig. 3f) was performed by grouping observed and shuffled contacts in a 50x50kb window around loop anchors into 3Kb or 14Kb square bins (higher resolution of 3Kb bins is only possible for the 3 large group of cells: G1, Early-S and Late-S/G2), and color coding according to the per-bin log fold ratios over shuffled-data statistics.

Domain condensation

The condensation level of a domain is quantified by the mean contact distance of contacts within the domain. To ensure domain coherence for condensation analysis (Fig. 3g) we used a more fine-grained domain definition with a higher threshold on the insulation (85

percentile for calling domains, instead of 90) resulting in 3,683 domains, out of which 3,622 had time of replication information. Each domain was classified by its mean time of replication and its length, resulting in the following groups of domains:

		Domains length range			
	mean ToR (mT)	1-244K	244-490K	490-980K	980-1500K
Early	mT > 0.84	399	439	223	24
Mid-Early	0 < mT < 0.84	401	361	267	47
Mid-Late	-0.66 < mT < 0	142	174	218	116
Late	mT < -0.66	113	128	223	99
Total		1055	1102	931	286

As expected from earlier reports, domains length was correlated with time of replication, but sufficient statistics was available for analysis in less populated bins:

	1-244K	244-490K	490-980K	980-1500K
Early	36.77%	40.46%	20.55%	2.21%
Mid-Early	37.27%	33.55%	24.81%	4.37%
Mid-Late	21.85%	26.77%	33.54%	17.85%
Late	20.07%	22.74%	39.61%	17.58%

Computing distributions of single-cell compartment association scores

In order to analyze the compartmentalization behavior of domains in a population of single cells, we defined two scores for each domain:

A-score: quantifies the tendency of a domain to interact with A domains. Calculated solely from *trans*-chromosomal contacts of an ensemble-like contact map (pool of singles or bulk Hi-C).

A-association score: quantifies the interaction with A domains of a given domain in a specific cell. Calculated from *cis*-chromosomal contacts (> 2Mb) of a single-cell.

To compute A-score we first calculate the fraction of *trans* contacts each domain has with A domains, which we term *A-fraction* (see above for details how domains are clustered to the A and B groups). We then calculate for each domain the mean A-fraction of the domains it interacts with in *trans* (considering the number of contacts between the domains as weights), which results in an A-score for each domain. (see comparisons of A-scores from different maps in Extended Data Fig. 9a).

Because of the sparsity of *trans* contacts in a single-cell contact map, we used *cis* contacts to explore the dynamics of compartments behavior at a single-cell level. A-association scores per single-cell and domain, are computed using *cis* contacts at distances above 2 Mb. We first filtered domains whose A-score was calculated by at least 20 *trans* contacts (discarding

90 out of 2649 haploid domains and 5 out of 2894 diploid domains). Next, for each single-cell, we only used domains with more than three contacts with at least two other domains. We then sampled two of the interacting domains and took their mean A-score as the A-association score of the domain in that cell. By sampling two domains we control for possible differences in A-score variance that are due to differences in marginal TAD coverage.

To study haploid A-scores, while maximizing consistency with the diploid system, we computed for each haploid domain, the overlapping diploid domains and calculated their weighted mean (diploid) A-score. We then defined a critical value of weighted A-score means to re-define A/B labels in haploids, using a threshold that maximizes the similarity between these A/B labels and the labels derived by standard compartment clustering as performed for diploids. Given these revised A/B labels for haploid TADs, we repeated the procedure described above (computing haploid A-score and haploid A-association) using haploid single cell maps. Note that in this way we used the diploid compartments solely for bootstrapping the haploid process.

Ranking domains by A-association score distributions and assigning to pseudo-compartments

For unsupervised analysis, domains were clustered by the distribution of their A-association scores in each group of cells (k-means, $k = 20$, Extended Data Fig. 9b). Given the absence of clusters showing bimodal distributions, domains were ordered by their mean A-association score at the late S to G2 group, where compartmentalization is peaking (Fig. 3b). Pseudo-compartments were then defined by discretizing the ranking into 20 equal size groups.

Creating normalized pseudo-compartment interaction maps

We derived an observed pseudo-compartment interaction map from the contact map pooling cells in each cell-cycle phase by summing the number of contacts between domains in each pseudo-compartment, separately for *cis*- (> 2Mb) and for *trans*-chromosomal contacts. To derive the *cis*- and *trans*-chromosomal contact enrichments (Fig. 4h-I and Extended Data Fig. 9f) we took the log ratio of the observed values and the expected contingency table generated from the observed marginal pseudo-compartment coverage across all three groups. For the *cis* pseudo-compartment maps in Extended Data Fig. 9e, we took the log of the element-wise ratio between the observed pseudo-compartment map and an expected pseudo-compartment map summed from the shuffled pool contact map.

Epigenetic annotations

Track	Source	Usage	Processing
Time of replication	Repli-chip of ES-D316 (rep 1)	Calculating repli-score and classifying TADs	Mean value on regions (TADs, upstream/downstream to landmarks, 10kb around a fend for repli-score)
Haploid RNA-seq	In-house (experimental details above)	4j	Log2(mean of 2 replicates over TAD)

Track	Source	Usage	Processing
Diploid RNA-seq	In-house (experimental details above)	Ext 3e, Ext 10h	$-\log_2(1 - \max(\text{reps max percentile}))$
H3K4Me1	ChIP-seq of ES-Bruce416	4j	Mean of 2 replicates over TAD
H3K4me3	ChIP-seq of ES-Bruce416	4j	Mean of 2 replicates over TAD
H3K4me3	ChIP-seq of ES-Bruce416 (rep 1)	Ext 3e	Fraction of TAD with peaks (peak = score > 99% percentile)
H3K27me3	ChIP-seq of ES-Bruce416	Ext 3e	Mean of 2 replicates over TAD
LaminB1	mESC LaminB1 DamID18	4j, Ext 3e	mean LaminB1 value over TAD
CTCF	ChIP-seq of ES-Bruce416	Fig 3d-e	Details on loops section above

Mapping and filtering before 3D modelling

Since structural modelling is sensitive to low levels of noise, haploid G1 cells were processed using stringent contact filtering to remove contacts that are more likely to be technical artefacts. We first used HiCUP27, applying a di-tag size selection from 50 bps to 850bps, for mapping di-tags and filtering out common Hi-C artefacts. Putative PCR duplicates were not removed by HiCUP, instead the filtered data was then passed a new tool (SiCUP) for further single-cell Hi-C specific filtering. We removed reads mapping to the Y chromosome, to short restriction fragment (less than 21bps) and to regions defined as problematic by ENCODE. We also filtered reads mapping to fragment ends forming multiple interactions in one percent or more of the datasets. To avoid potential artefacts we removed singleton di-tags. In haploid G1 cells there is only one copy of the genome, hence after removal of PCR-duplicates each observed fragment end should be in contact with at most one other fragment end. Consequently, multiple contacts from the same fragment were removed entirely. An exception to this was when a fragment end (A) interacted with two other fragments ends (B and C) which were close together (defined here as when B and C were within 20 MboI fragments). In such instances the strand orientation of the reads mapping to B and C were typically the same, to a degree not expected by chance (as defined by a chi-squared test when evaluating the whole dataset). We reasoned that in such instances these apparently distinct interactions were in fact derived from one initial Hi-C interaction. Consequently, when this was observed, not all the di-tags were discarded. Instead, if the Hi-C interaction was in trans, a random di-tag was discarded. Alternatively, when the Hi-C interaction was in cis, the di-tag representing the shortest Hi-C interaction was retained.

We also filtered out unsupported contacts. For each cell, using the filtered contacts, we first derived a connectivity graph of the genome. Nodes of the graph represented 1Mb segments of the genome, and each edge represented a single contact mapped onto 1 Mb resolution, so any two nodes of the graph might be connected by more than one edge. We defined a contact as unsupported if upon deletion of that contact, the shortest path connecting its two end nodes would be longer than 3 edges. These unsupported contacts (median 1.06% of contacts, Extended Data Fig. 10a) were removed from the sc-HiC libraries before 3D modelling.

Polymer model for whole-genome modelling

We modeled chromatin as a simple non-overlapping beads-on-a-string polymer¹² subject to data derived constraints. The energy function consists of a bond potential that ensures the connectivity of the polymer chain within chromosomes, a second term that prevents overlaps of chromatin and a third term that enforces the observed contacts as distance constraints with different contact points for each cell. Note that no other physical parameter, e.g. the overall shape or size of the nuclear envelope, is constrained by our modelling approach, hence the overall three-dimensional shape of the entire genome is purely determined by the filtered and supported chromosome contacts in the corresponding single-cell Hi-C library. We implemented this coarse-grained polymer model to be used with the GROMACS 5.0.6 molecular dynamics package²⁸. The energy function and the exact values of the model parameters are listed in the Supplementary Material. Briefly, the exclusion energy term is a harmonic function for distances below d_0 with a force constant k^{excl} , and the bond and constraint energy terms are defined by a flat-bottomed harmonic function with no energy penalty for distances between d_0 to d_1 , and a harmonic potential with a force constant k outside this range which turns into a linear function for distances greater than d_2 . 3D models were optimised using a simulated annealing protocol, using a piecewise linear temperature profile, and different initial conditions to assess convergence robustness.

Quality control of 3D models

To avoid underdetermined models, we only considered models that were computed using at least 10,000 filtered contacts. The root mean square distance (RMSD, the smallest possible Euclidean distance) of models, indicating the reproducibility of models, as a function of the number of contacts is shown in Extended Data Fig. 10d. We also discarded models with high numbers of violated constraints. For this, we considered a constraint violated when its distance was greater than d_2 (constrained distances are expected to be greater than d_1 by construction, due to thermal fluctuations). At 1 Mb resolution there were no violated constraints for 171 of the 190 cells (Extended Data Fig. 10b), and the fraction of violated constraints strongly correlated at the 500 kb and 100 kb resolutions (Pearson $R = 0.96$). For further analysis, we selected 127 cells that on median had no more than 0.5% violated constraints at the 500 kb resolution and no more than 0.1% violated constraints at the 100 kb resolution (shown as black dots in Extended Data Fig. 10c). We further discarded one cell that had more than 10% of a chromosome missing, resulting in 126 ordered G1 cells with high-quality models for downstream analysis.

We tested robustness to the initial conditions by using permuted chromosomes along a spiral as well as folded structures of other nuclei as the initial chromatin conformation. Results were robust to changes to the initial conditions, indicating the convergence of the simulations, which was measured by computing the distribution of the root-mean-square-distance (RMSD) between aligned models of the same cell compared to different cells (Extended Data Fig. 10e). Models at different resolutions were compared to one another by downsampling the higher resolution structures to 1 Mb, and scaling the nuclear radius of the model to the nuclear radius of the 1 Mb model. The downsampled models computed at different resolutions were consistent at 1 Mb resolution. Models were also robust when scaling the model parameters by a factor of half or two.

Cross validation

To test the robustness of our 3D modelling approach to missing contacts, we performed a cross validation test. For each cell, we identified the 5 most strongly interacting chromosome pairs, measured by the number of contacts between the two chromosomes. For each of these 5 chromosome pairs, we removed all *trans* contacts between that chromosome pair and recomputed a structure model for that cell (we only delete interactions between one chromosome pair at a time). We computed the distance distribution of the removed constraints in the models computed with those constraints removed (Extended Data Fig. 10f blue line), and compared it to their distance distribution in randomly selected cells (control, Extended Data Fig. 10f green line). The distances of these removed contacts are noticeably shorter in their corresponding model than they are in randomly selected other cell, indicating that the proximity of the chromosome pair is preserved. This suggests that these sparse datasets contain sufficient information about chromosome level interactions for robust structural modelling. This is probably due to their indirect coupling through their contacts with their neighbouring chromosomes. In comparison, the distances of the unsupported contacts filtered out before modelling (Extended Data Fig. 10f light blue line) were much more similar to the control, indicating that most of those interactions are likely noise.

Chromosome decondensation

We grouped the 126 cells into 7 time groups in G1 (Extended Data Fig. 10h). As the chromatin structure varied most in early G1, we used 4 time groups in the first quarter of the 126 G1 ordered cells, and 1 group for each of the remaining three quarters. We computed the nuclear radius from its radius of gyration. As an indicator of nucleus sphericity, we also computed the inertia ellipsoid of the nucleus. The detailed description of how these quantities are computed can be found in the Supplementary Material. We noted that although the nuclear lamina is not modelled by the force field, the models turn out to be near spherical ($a \approx b \approx c$) and roughly the same size (Extended Data Fig. 10g, nucleus a/c ratio). We also noted that low mappability regions tend to be more flexible and uncertain in their position in the computed models, illustrated by the unconstrained region around chrX:30Mb (the white chromatin region looping out of the models in Extended Data Fig. 10e). To illustrate changes in the overall chromatin structure, we selected an M-phase and a late (118th of the 126 ordered cells) G1 cell. We plotted the whole genome structure and a karyotyped image at 100 kb resolution (Fig. 5a). The structure models were visualised and aligned using VMD29 1.9.1, the images were rendered using PovRay 3.7 (<http://www.povray.org>). The colour scheme used for the chromosomes was chosen to provide maximum contrast³⁰.

To quantify the decondensation of chromosomes during G1, we fitted an ellipsoid to each chromosome of every model and averaged the values over all models of the same cell. As a measure of sphericity, we used the longest-to-shortest semiaxis ratio (Extended Data Fig. 10g, chromosome a/c ratio) together with the middle-to-shortest semiaxis ratio (Extended Data Fig. 10g, chromosome b/c ratio). Larger a/c ratios together with $b/c \approx 1$ in early G1 indicate rod-like structures that decondense into a more spherical structure as a/c decreases to $a \approx b \approx c$. We plotted the distribution of all chromosome a/c ratios over cells in the 7 time groups with illustrating examples of chromosome 1 from each time group in Fig. 5b of the Main text.

3D organisation of compartments

To investigate the spatial organisation of compartmentalisation, we assigned compartment labels to the beads using a majority vote. Beads with no compartments were excluded from further analysis. As an indicator of local decompaction of chromatin, we computed the average distance of neighbouring chromatin segments of the same compartment, normalised by the d_1^{bond} bond length parameter. We computed the average compartment decompaction values for each model, and averaged them across all 3D models of the same cell. The distribution over all cells in the 7 time groups is shown in Fig. 5c-d. We tested if the mean A and B compartment decompaction values came from the same distribution using Mann–Whitney statistics.

We also computed the average radial positioning of the compartments. For each bead, we computed its radial position as a volume fraction, increasing linearly between the nucleus centre-of-mass (0) and the nuclear radius (1) in the volume enclosed by a sphere of that distance radius from the nucleus centre. We averaged the radial position values of all beads of the same compartment, and across all 3D models of the same cell. We plotted the distribution of the radial positions across the cells in the 7 time groups in Fig. 5e-f. We tested if the mean A and B compartment radial position values came from the same distribution by performing a Mann–Whitney test on each time group.

Testing the enrichment of compartment co-localisation between pairs of pseudo-compartments in long *cis* (> 2 Mb genomic distance) or *trans* was done separately. For this analysis, we defined two beads in contact if their distance was within 10% of the nucleus diameter. For any two pseudo-compartments, we counted the contacting bead pairs in long *cis* or *trans*, for which one bead belonged to one pseudo-compartment, while the other bead belonged to the other pseudo-compartment. We averaged these counts over all models of all cells in the 7 time groups of G1, and rounded them down to the nearest integer, resulting the observed counts. We shuffled the bead contacts while keeping the number of contact beads of each compartment constant, resulting the expected counts. We plotted the $\log_2(\text{observed}/\text{expected})$ ratios between any two pseudo-compartments as a matrix (Fig. 5g-h).

Nuclear Organiser Region clustering

For each chromosome, we used the first constrained bead as the pericentromeric region closest to the centromere. Chromosomes were divided into two groups, Nucleolar Organiser Region (NOR) chromosomes (chromosomes 11, 12, 15, 16, 18 and 19), and other chromosomes. We computed all pairwise distances between the pericentromeric regions within these two groups normalised by the nucleus diameter, and plotted the distribution of the top one third of these distances across all models of cells for each time group (External Data Fig. 10i). We averaged these distances for each model and each cell, and tested if the mean of the top third shortest centromere distances came from the same distribution by performing a Mann–Whitney test on each time group.

Extended Data

Extended Data Fig. 1. Technical QCs.

Panels in this figure show data for all the diploid cells analysed, as specified in panel **i**. **a**) Testing library saturation. Showing the fraction of segment chains supported by a single read in each batch, batch colours match their sorting criteria, see panel **i** for details. **b**) Number of unique molecules captured per cell against the number of sequenced reads. **c**) Number of reads per unique molecule against number of sequenced reads. **d**) Observed (red) and

expected (by binomial reshuffling of the observed contacts) number of cells each trans contact appears in. **e)** Correlation of the fraction of trans contacts and close cis (< 1 kb, non-digested) with the fraction of contacts in different distance bins. **f)** Stratification of contacts by the orientation of contacting fragments against contact distance. **g)** Distribution of the logarithmic decay bin with the most contacts per cell, dashed line at 15.5 mark QC threshold below which cells are discarded. **h)** QC metrics of single cells coloured by FACS sorting criteria: “2n DNA” and “2n<DNA 4n” by Hoechst staining (H), “G1”, “early-S”, “mid-S” and “late-S/G2” by Hoechst (H) and Geminin (G) staining. Vertical lines mark experimental batches. Shown from top to bottom: total number of contacts (coverage); fraction of inter-chromosomal contacts (%trans); fraction of very close (< 1 Kb) intra-chromosomal contacts (%no dig); early replicating coverage enrichment (repli-score); fraction of fragment ends (fends) covered more than once (%dup fend). Horizontal dashed lines mark thresholds used to filter good cells. **i)** Details on the experimental batches, showing the number of cells in each batch, number of cells passed QC, mean number of reads per cell (MRPC, in million reads), FACS sorting criteria (H = Hoechst, G = Geminin) and the medium used to grow the cells. Batches are colored by FACS sorting criteria. **j)** Coverage enrichment per cell (column) and chromosome (rows), where expected coverage is calculated from the frequency in the pooled cells and the total number of contacts in each cell. We discard cells that have any aberrantly covered chromosome (at least 2 fold enrichment or depletion, shown on the right “bad” panel). Left panel shows all cells ordered by batch ID, with lower panel colored by FACS sorting criteria (as in panels **h-i**).

Extended Data Fig. 2. Trans-chromosomal contacts.

a) Examples of inter-chromosomal contact maps for several pairs of chromosomes. Showing contact maps of single cells (blue, each point is a contact) and the pooled contact map on the same chromosomes, using 500 Kb square bins. **b)** Distribution of the number of contacts between selected pairs of chromosomes. Showing the number of contacts per square Mb; red dashed line marks the cutoff for interacting chromosomes. **c)** Fraction of cells in which each pair of chromosomes was interacting (the pair of chromosomes had normalised interaction above the cutoff shown in panel **b**).

Extended Data Fig. 3. Pooled contact map and population data.

a) Normalised chromosomal contact maps of the pooled diploid 2i single cells (showing chromosomes 10 and 11). **b)** A chromosome ideogram showing the division of domains to the A and B compartments by k-means clustering of domains trans-chromosomal contact profiles. **c)** A chromosome ideogram showing the inferred A-score of each domain. **d)** Comparing A and B domains by their lengths. **e)** Comparing A and B domains by their mean time of replication (ToR); percentage of H3K4me3 peaks; mean H3K27me3; mean LaminB1 values; mean RNA-Seq. **f)** Genome-wide comparison of insulation scores across

reference bulk Hi-C dataset (from Olivares-Chauvet et al, 2016) and different pools of single cells. Insulation is calculated every 10 kb using a scale of 300 kb, pearson correlation is shown at the legend.

Extended Data Fig. 4. Mitotic cells and dynamics of coverage by replication time.

a) Contact decay profile of our pool of diploid 2i singles (pool), a reference mouse ESC ensemble Hi-C dataset (Ens mES, from Olivares-Chauvet et al., 2016, supplementary reference 26), human K562 mitotic population Hi-C (K562 M, from Naumova et al, 2013,

reference 11) and our pooled 8 most mitotic-like cells (top M). **b)** Genome-wide contact maps of 8 putative mitotic diploid cells (cells with the highest fraction of contacts at 2-12 Mb distances). **c)** Correlation matrix of domain coverage across cells. Ordered by the mean correlation to the top 50 earliest replicating domains minus the mean correlation with the latest 50 domains. Time of replication of each domain is shown at the bottom. **d)** Comparing the fraction of contacts associated with the latest- and earliest-replicating fends ($\text{ToR} < -0.5$ and $\text{ToR} > 0.5$, respectively). Cells are colored as in Fig. 1e. **e)** Projection of diploid 2i cells onto a two-dimensional plane using their contact decay profile (>1 kb) by a non-linear dimensionality reduction algorithm (see methods). Cells are colored by their repli-score, which was not used to position them on the plane.

Extended Data Fig. 5. Clustering cells by contact decay and phasing cells over the cell cycle.

a) Chromosomal contact maps (chromosome 6) of pooled cells per cluster defined in Fig. 1f and the mean decay trend of each cluster (red) compared to the rest of clusters (grey). **b)** K-means ($k = 12$) clusters of single-cell contact decay profiles. **c)** Mean far contact distances and repli-score against the fraction of short-range contacts per clusters in panel **a** (red – cells in cluster, grey – all cells), ordering clusters by their mean %short-range.

Extended Data Fig. 6. Phasing cells over the cell cycle and quality controls.

a) Mean contact distance among the far (4.5-225 Mb) range against the fraction of short-range (23 kb-2 Mb) contacts. Dashed red lines mark the cutoffs used to divide cells into the main 3 groups (G1, early-S, late-S/G2). **b)** Similar to panel **a** but showing repli-score against the fraction of short-range contacts. **c)** Batch of origin of each phased cell using all diploid 2i and serum cells, colored by the batch FACS sorting criteria (see Extended Data Fig. 1h for batch information) **d)** Similar to panel **c** but for haploid 2i and serum cells (see Extended Data Fig. 8b for batch information) **e)** Testing the stability of our phasing, we compare the

positions of cells by phasing using only half the chromosomes (1st half: 2, 3, 5, 6, 9, 10, 12, 15, 17, 19, 2nd half: X, 1, 4, 7, 8, 11, 13, 14, 16, 18). Position of only 38 cells (< 2%) differ in more than 10% of the number of cells (outliers marked in red, 10% margin in dashed grey). Phasing groups marked in black lines. **f**) Chromosomal comparison of the contact decay metrics used to phase the cells in each group: mean far contact distance for the G1 cells (left) and %short-range for G1, S and G2 cells (right). Showing smoothed ($n = 51$) trends per chromosome on top, chromosomes colored by length (light grey – chromosome 19, black – chromosome 1). Data for specific chromosomes: chromosome 1 (middle) and chromosome 11 (bottom) as red dots. **g**) Hoechst and Geminin FACS indices of cells in batches for cells in batches 27-35 against the cells inferred phasing position. **h**) Genome-wide contact maps of representative cells along the inferred G1 phase, their position marked at the bottom of Fig 2g.

Extended Data Fig. 7. Insulation of sub-groups and loops.

a) Insulation profiles of the main three phased groups (G1, early-S, late-S/G2) over 6 Mb regions in chromosomes 1-5. **b)** Comparing border insulation of phased groups. Dashed blue lines show the mean insulation value per group. **c)** Showing mean insulation per cell on mouse ESC TAD borders taken from Dixon et al. 2013 (reference 5; using the center point of borders smaller than 80 kb) compared to the mean cell insulation over our inferred borders. **d)** Insulation profile over a 6 Mb region in chromosome 1 (same as panel **a**), showing insulation of pooled cells (black), pooled mitotic cells (orange) and a shuffled pool

of mitotic cells (red). **e**) Genome-wide distribution of insulation values for pooled-, pooled-mitotic- and shuffled-pooled-mitotic cells, same colours as in panel **c**. **f**) Mean border insulation per cell against the fraction of short-range (< 2 Mb) contacts. **g**) A/B compartment score on trans-chromosomal contacts, single-cell mean values in dots coloured by FACS sorting, mean trend (black) and mean trend of *cis*-chromosomal A/B compartment in dashed black (same trend as in Fig. 3b). **h**) Normalized contact maps of the regions in panel **a**, circling the loops detected in distances 200 kb to 1 Mb (marked in dashed black line) **i**) Comparing loop foci enrichment calculated per phased group, showing the Pearson correlation in the legend.

Extended Data Fig. 8. Haploid cells technical QCs.

a QC metrics of single cells colored by FACS sorting criteria: G1 (H) and G1/S (H) by Hoechst staining (see color legend at bottom). Vertical lines mark experimental batches. Shown from top to bottom: total number of contacts (coverage); fraction of inter-chromosomal contacts (%trans); fraction of very close (< 1 kb) intra-chromosomal contacts (%no dig); early replicating coverage enrichment (repli-score); fraction of fragment ends (fends) covered more than once (%dup fend). Horizontal dashed lines mark thresholds used to filter good cells. **b**) Details on the experimental batches, showing the number of cells in

each batch, the number of cells that passed QC, mean number of reads per cell (MRPC, in million reads), FACS sorting criteria (H = Hoechst,) and the medium used to grow the cells. **c)** Coverage enrichment per cell (column) and chromosome (rows), with expected coverage calculated from the pooled cells. We discard cells that have any aberrantly covered chromosome (at least 2 fold enrichment or depletion, shown on the left “bad” panel). The right panel shows all cells. **d)** Similar to Fig. 2d-g but produced from the haploid cells, outliers coloured in black. **e)** Similar to Fig. 3a-b but produced from the haploid cells, outliers coloured in black. **f)** Similar to Fig. 3e but produced from the haploid cells. **g)** Similar to Fig. 3g but produced from the haploid cells, outliers coloured in black.

Extended Data Fig. 9. Haploid cells clustering, diploid cells pseudo-compartment analysis.

a) Comparing trans A-score of a reference bulk Hi-C dataset (from Olivares-Chauvet et al., 2016) and different pools of single cells, using the set of domains inferred from the pool of diploid 2i cells (A domains in red and B in black). Showing on top of each panel the correlation of the domains' A-score in each dataset. **b)** Per-domain distributions of A-association score across haploid cells in each of the three main cell cycle inferred phases (from left to right: G1, early S, late S to G2). Domains were clustered (k-means, $k = 20$) by their distribution in all groups. Inter- and intra- cluster ordering by mean A-association score

at late-S/G2. **c)** Similar to Fig. 4g but for diploid 2i cells. **d)** Contingency table of haploid and diploid inferred pseudo-compartments, considering all overlapping intervals between the haploid and diploid sets of domains. **e)** Long-range (> 2 Mb) *cis*-chromosomal contact enrichments between TAD groups according to pseudo-compartments for haploid (top) and 2i diploid (bottom) cells. Showing the expected contact frequency by contacts in the shuffled pooled contact maps on the right. **f)** Long-range (> 2 Mb) *cis*-chromosomal (top) and *trans*-chromosomal (bottom) 2i diploid contact enrichments between TAD groups according to pseudo-compartments. Showing the expected contact frequency by the marginal pseudo-compartments coverage on the right. **g)** TADs are sorted by their mean A-association in late S – G2 (x-axis, left – strongly B, right – strongly A), and their mean RNA, H3K4me1, Lamin-B1 and H3K4me3 levels are shown (y-axis).

Extended Data Fig. 10. Whole genome structure modelling - quality control and clustering of peri-centromeric regions.

a) The distribution of the fraction of unsupported contacts in the pre-filtered contact set in all 190 haploid cells in 2i inferred as being in G1. **b)** The distribution of the fraction of violated constraints in the 190 haploid G1 cells. **c)** The fraction of violated constraints in the 3D models of 190 haploid G1 cells at 500 kb versus 100 kb resolutions (Pearson $R = 0.96$). Black dots represent cells that have at least 10,000 contacts and no more than 0 violations at 1 Mb, 0.5% violations at 500 kb and 0.1% violations at 100 kb. Cells with no violations at

either 500 kb or 100 kb resolution are not shown. **d)** Mean RMSD between all models of the same cell with at least 10,000 (black) and fewer than 10,000 (red) filtered contacts, for cells at 1 Mb (dots), 500 kb (open diamonds) and 100 kb (triangles) resolutions. **e)** Model reproducibility across cells. The RMSD distribution between all models for the same cell (red) compared to the RMSD distribution between models for different cells for the 126 cells with the highest quality models. The images show aligned structures of 106 x 1 Mb models (top), 80 x 500 kb models (centre) and 5 x 100 kb models (bottom) for NXT-1091 (38th of 126 ordered G1 cells) with a mean RMSD at the peak of the red curves. **f)** Cross-validation test. Average trans-chromosomal distance distributions, computed using a set of 200 random trans-contacts between any two chromosomes, with the contact points distributed uniformly on the two chromosomes. Red curve: trans-distances between the 5 most strongly contacting chromosome pairs in models using all supported contacts. Blue curve: same as red curve, but using the models computed with the contacts between that chromosome pair removed. Cyan curve: trans-distances of unsupported contacts in the models computed with all supported contacts present. Green curve: trans-distances of chromosome pairs in models of all cells using all supported contacts. **g)** Nucleus and chromosome shape properties in the 126 cells with high-quality models: the nuclear radius (in arbitrary units), the longest-to-shortest (a/c) semiaxis ratio of the ellipsoid fitted to the whole nucleus model, and the longest-to-shortest (a/c) and middle-to-shortest (b/c) semiaxis ratios of the ellipsoids fitted to each chromosome of the 3D models. For the nucleus size and a/c ratio, values in each model are shown, for the chromosome fitted ellipsoid semiaxis ratios the model-averaged value is shown for each cell. **h)** The 7 cell groups in the inferred G1 time order for the 126 cells with the highest-quality models. **i)** Distances of the top third shortest distances between NOR chromosomes or non-NOR chromosomes in 7 cell groups along the inferred G1 time order. Distances are normalised by the nucleus diameter.

Acknowledgements

This work was supported by the NIH 4D Nucleome 1U01HL129971-01 and the European Research Council 340152. We thank Drs M. Leeb, A Wutz and J. Gribnau for haploid and diploid ESCs.

References

1. Paweletz N, Walther Flemming: pioneer of mitosis research. *Nat Rev Mol Cell Biol.* 2001; 2:72–75. [PubMed: 11413469]
2. Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 2009; 326:289–93. [PubMed: 19815776]
3. Sexton T, et al. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell.* 2012; 148:458–72. [PubMed: 22265598]
4. Nora EP, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature.* 2012; 485:381–5. [PubMed: 22495304]
5. Dixon JR, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature.* 2012; 485:376–80. [PubMed: 22495300]
6. Sofueva S, et al. Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* 2013; 32:3119–29. [PubMed: 24185899]
7. Rao SSP, et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell.* 2014; 159:1665–1680. [PubMed: 25497547]
8. Zuin J, et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A.* 2014; 111:996–1001. [PubMed: 24335803]

9. de Wit E, et al. CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell*. 2015; 60:676–684. [PubMed: 26527277]
10. Flavahan WA, et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*. 2016; 529:110–114. [PubMed: 26700815]
11. Naumova N, et al. Organization of the mitotic chromosome. *Science*. 2013; 342:948–53. [PubMed: 24200812]
12. Nagano T, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*. 2013; 502:59–64. [PubMed: 24067610]
13. Nagano T, et al. Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome Biology*. 2015; 16:175. [PubMed: 26306623]
14. Pope BD, et al. Topologically associating domains are stable units of replication-timing regulation. *Nature*. 2014; 515:402–405. [PubMed: 25409831]
15. Stamatoiyannopoulos JA, et al. An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol*. 2012; 13:418. [PubMed: 22889292]
16. Boettiger AN, et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states - nature16496.pdf. *Nature*. 2016; 529:418–422. [PubMed: 26760202]
17. Peric-Hupkes D, et al. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell*. 2010; 38:603–13. [PubMed: 20513434]
18. Dileep V, et al. Topologically-associating domains and their long-range contacts are established during early G1 coincident with the establishment of the replication timing program Authors. *Genome Res*. 2015; 25:1104–1113. [PubMed: 25995270]
19. Britton-Davidian J, Cazaux B, Catalan J. Chromosomal dynamics of nucleolar organizer regions (NORs) in the house mouse: micro-evolutionary insights. *Heredity*. 2012; 108:68–74. [PubMed: 22086078]
20. Ramani V, et al. Massively multiplex single-cell Hi-C. *Nat Methods*. 2017; 14:263–266. [PubMed: 28135255]
21. Stevens TJ, et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature*. 2017; 544:59–64. [PubMed: 28289288]
22. Flyamer IM, et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*. 2017; 544:110–114. [PubMed: 28355183]
23. Leeb M, Wutz A. Derivation of haploid embryonic stem cells from mouse embryos. *Nature*. 2011; 13:1–5.
24. Leeb M, Perry ACF, Wutz A. Establishment and Use of Mouse Haploid ES Cells. *Curr Protoc Mouse Biol*. 2015; 5:155–85. [PubMed: 26069082]
25. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9:357–359. [PubMed: 22388286]
26. Olivares-Chauvet P, et al. Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature*. 2016; 540:296–300. [PubMed: 27919068]
27. Wingett S, et al. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Research*. 2015; 4:1310. [PubMed: 26835000]
28. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun*. 1995; 91:43–56.
29. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph*. 1996; 14:33–38. [PubMed: 8744570]
30. Green-Armytage P. A Colour Alphabet and the limits of colour coding. *JAIC-Journal Int Colour Assoc*. 2010; 10:1–23.

Figure 1. Multiplexed single-cell Hi-C reveals heterogeneous, cell-cycle dependent chromosomal architectures.

a) Single-cell Hi-C schematic. **b**) Number of informative contacts retrieved per QC passed cell. Median 127,233 (dashed line). **c**) Fraction *trans*-chromosomal contacts per QC passed cell. Median 5.87% (dashed line). **d**) Genome-wide contact map of a representative mitotic cell (1CDX4_242). **e**) Percentage cells with short-range contacts (< 2 Mb) versus contacts at the mitotic band (2-12 Mb, left), and repli-score (right). Cells are grouped by % short-range and % mitotic contacts and coloured by group. **f**) Single-cell contact decay profiles ordered

by *in-silico* inferred cell-cycle phasing, with approximate cell-cycle phases shown on top. Each column represents a single cell. **g**) Selected phased and pooled contact maps.

Figure 2. Validation of inferred cell-cycle phasing.

a) FACS sorting scheme **b)** Positions of FACS sorted cells along a circular layout of their *in-silico* phasing (clockwise, M is on top). **c)** Comparing in-silico phasing rank to combined Hoechst and Geminin FACS indices for 324 indexed-sorting cells. Black lines demarcate approximate cell-cycle phases. **d)** Repli-score per cell ordered by cell-cycle phasing (left to right). FACS-sorted cells coloured as in **a** and **b**, and remaining cells are grey. Outliers (0.5%) are red. Cyclic average window ($n = 100$) trend shown in black. **e-g)** Similar to **d**, showing additional single cell statistics.

Figure 3. Distinct and context specific cell-cycle dynamics of insulation, compartmentalisation, loops, and domain condensation.

a) Mean contact insulation at TAD borders, phased cells coloured as in Fig. 2. **b)** A/B compartmentalisation score **c)** Clustering borders (k-means, $k = 4$) according to insulation profile over inferred phasing. Mean insulation of the borders in a cluster (blue dots), mean cluster insulation trend (blue line) and global trend (black line). **d)** Mean time of replication ± 200 kb of the Fig. 3b clustered borders. Early-replicating positive, late-replicating negative. **e)** Convergent CTCF loops contact enrichment. Showing data for 762 early-early loops and 456 late-late loops. **f)** Aggregated contacts around convergent CTCF loops distanced 0.2-1 Mb from each other, pooling cells by their *in-silico* phasing group and

normalising by the expected profile using the shuffled pooled contact map. **g)** Mean intra-TAD contact distance stratifying domains by length, and mean time of replication.

Figure 4. Haploid single-cell Hi-C reveals a spectrum of TAD pseudo-compartments.

a) Similar to **1g**, but analysing haploid cells. **b)** Chromosome idiogram depicting inferred A compartment score of each TAD. **c)** TADs' mean time of replication versus A-score. **d)** TADs' mean *cis* A-association score versus *trans* A-score. **e)** Standard deviation versus mean A-association scores per TAD across cells. Colours mark groups of TADs stratified by mean A-association. **f)** Standard deviation of A-association score across TADs stratified as in **e** and averaged separately over cells in each of the three main inferred cell-cycle phases: G1, early S and late S-G2. (* = < 0.05, ** = < 0.01, *** = < 0.001, single side MW-test). **g)** Per-TAD distributions of A-association scores in the three phased cell-cycle populations. **h)** Long-range (> 2 Mb) *cis*-chromosomal and **i)** *trans*-chromosomal contact frequency between TAD groups according to pseudo-compartments..

Figure 5. Whole genome structure modelling.

a) Whole genome conformations, and separated individual chromosomes. Models computed at 100 kb resolution. **b)** Top: Chromosome 1 examples along inferred G1 phase with fitted ellipsoid of inertia. Chromosomes coloured by genomic position pericentromeric (red), telomeric (blue). Bottom: Chromosome de-condensation illustrated by distribution of a/c semiaxis ratios of all chromosomes. **c)** Decompaction of chromatin in A and B compartments, defined as average distance between neighbouring 500 kb segments of same compartment. **d)** Same as **c** for the 20 pseudo-compartments. **e)** Average radial position of A

and B compartments within the nucleus, defined as the volume fraction enclosed by a sphere of that radius. Zero = nuclear centre, 1 = periphery. **f)** Same as **e** for the 20 pseudo-compartments. **g)** Long-range (> 2 Mb) *cis*-contact enrichment between 20 pseudo-compartments. **h)** *Trans*-contact enrichment between the 20 pseudo-compartments.

Supplementary Figure 1. Locus insulation