ARTICLES
https://doi.org/10.1038/s41587-020-0519-y

nature biotechnology

GPSeq reveals the radial organization of chromatin in the cell nucleus

Gabriele Girelli[1,2,4], Joaquin Custodio[1,2,4], Tomasz Kallas[1,2,4], Federico Agostini[1,2], Erik Wernersson[1,2], Bastiaan Spanjaard[3], Ana Mota[1,2], Solrun Kolbeinsdottir[1,2], Eleni Gelali[1,2], Nicola Crosetto[1,2,5] ✉ and Magda Bienko[1,2,5] ✉

With the exception of lamina-associated domains, the radial organization of chromatin in mammalian cells remains largely unexplored. Here we describe genomic loci positioning by sequencing (GPSeq), a genome-wide method for inferring distances to the nuclear lamina all along the nuclear radius. GPSeq relies on gradual restriction digestion of chromatin from the nuclear lamina toward the nucleus center, followed by sequencing of the generated cut sites. Using GPSeq, we mapped the radial organization of the human genome at 100-kb resolution, which revealed radial patterns of genomic and epigenomic features and gene expression, as well as A and B subcompartments. By combining radial information with chromosome contact frequencies measured by Hi-C, we substantially improved the accuracy of whole-genome structure modeling. Finally, we charted the radial topography of DNA double-strand breaks, germline variants and cancer mutations and found that they have distinctive radial arrangements in A and B subcompartments. We conclude that GPSeq can reveal fundamental aspects of genome architecture.

In eukaryotic cells, the genome is spatially organized and its three-dimensional (3D) architecture is vital to the proper execution of its functions[1]. One important feature of the 3D genome is that individual chromosomes are nonrandomly positioned with respect to the nuclear periphery[2–9]. The nuclear lamina is thought to be the key organizer of the radial arrangement of chromatin in interphase nuclei[10], by creating a large nuclear compartment where most of the inactive chromatin clusters in the form of lamina-associated domains (LADs)[11–13]. Specialized subchromosomal regions, such as centromeres and telomeres, as well as nucleolar organizing regions (NORs), are also nonrandomly positioned in the nucleus[14–18]. NORs contain ribosomal RNA gene clusters that coalesce to form the core of the largest nuclear body, the nucleolus, and organize chromatin within and around it[19]. Indeed, inter-chromosomal interactions around the nucleolus and nuclear speckles have been implicated in shaping the 3D genome[20].

The preferential radial location of individual genomic loci in the nucleus has been variably attributed to gene density[3,5,6], guanine–cytosine (GC) content[21–23] and chromosome size[4,7,8,24]. Additionally, transcriptional activity has also been implicated in radial nuclear organization, although it is still debated whether transcription influences radiality or vice versa[12,25–34]. Overall, the role of genomic and epigenomic features in shaping radiality remains to be quantified, despite several attempts to model the contribution of various factors[34–36]. In particular, it is unclear whether the nucleus consists merely of a peripheral transcriptionally inactive compartment as opposed to a central transcriptionally active one, or whether a finer stratification exists. In this context, a major obstacle until now has been the lack of dedicated genome-wide methods to specifically tackle this aspect of chromatin organization at high resolution. To overcome this limitation, we developed a method that allows inferring radial locations throughout the genome, all along the nuclear radius, which we named GPSeq. Using GPSeq, we generated the first high-resolution map of radial chromatin organization in human cells, which reveals a clear tendency of individual genomic regions to occupy specific radial locations, as well as gradients of chromatin modifications, transcriptional activity and replication timing and a marked polar arrangement of chromosomes with respect to A and B compartments and subcompartments[37,38]. We developed a high-performance algorithm, *chromflock*, that dramatically improves the accuracy of whole-genome structure ensemble generation. Finally, we integrated GPSeq maps with DNA breaks and mutations data, revealing radial differences in DNA damage and mutational processes.

Results

Establishment of GPSeq. We reasoned that, if we were able to gradually fragment genomic DNA (gDNA) starting from the nuclear lamina toward the nuclear center, we could then use next-generation sequencing to reconstruct the radial position of each gDNA fragment. To this end, we first identified experimental conditions that allow restriction enzymes to slowly diffuse through the nucleus of cross-linked cells and cut gDNA while progressing toward the nuclear interior. To visualize the enzyme diffusion, we developed a fluorescence in situ hybridization assay, namely YFISH, in which a Y-shaped adapter is first ligated to the cuts introduced in situ by a restriction enzyme and then detected using complementary fluorescently labeled oligos (Fig. 1a, Supplementary Table 1, Methods and Supplementary Methods). If the enzyme indeed gradually digests gDNA from the nuclear periphery toward the center, the YFISH signal should appear as a fluorescent band that progressively thickens inwards until the whole nucleus is filled (Fig. 1b). To test our hypothesis, we incubated HAP1 haploid cells for increasing times in the presence of HindIII (10, 15, 30 and 45 min and 1, 2 and 6 h) and used either wide-field microscopy or stimulated emission depletion (STED) microscopy followed by image deconvolution

[1]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. [2]Science for Life Laboratory, Stockholm, Sweden. [3]Berlin Institute of Medical Systems Biology Max Delbrück Center, Berlin, Germany. [4]These authors contributed equally: Gabriele Girelli, Joaquin Custodio, Tomasz Kallas. [5]These authors jointly supervised this work: Nicola Crosetto, Magda Bienko. ✉e-mail: nicola.crosetto@ki.se; magda.bienko@ki.se

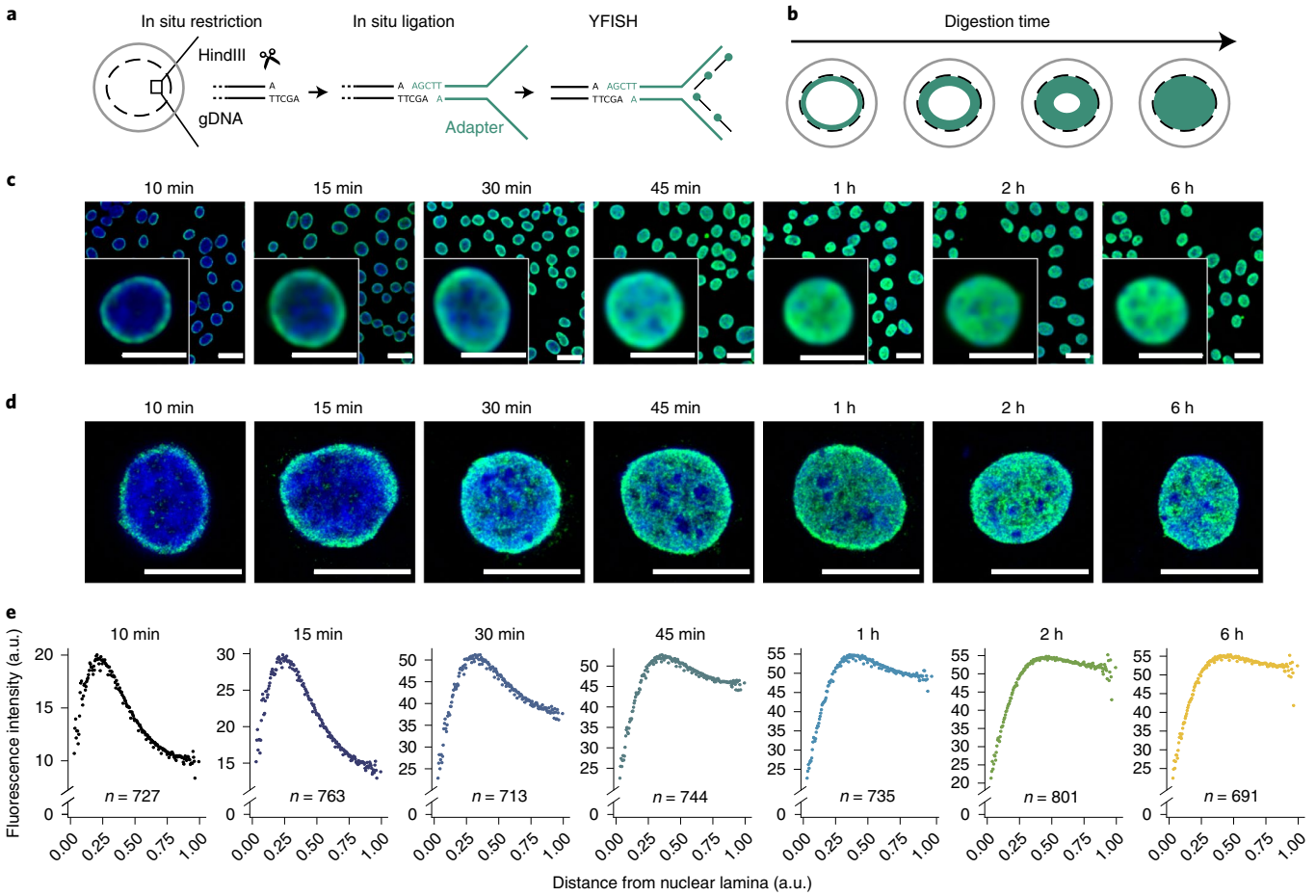


Fig. 1 | GPSeq implementation. a, Scheme of YFISH. Cross-linked permeabilized nuclei (dashed black circle) are incubated with a restriction enzyme (for example, HindIII). Digested recognition sites are ligated to a forked adapter (green), which is detected using fluorescently labeled oligos (lines with green dots) complementary to the single-stranded part of the adapter. **b**, Gradual in situ gDNA digestion. Fixed and permeabilized cells (solid gray circles) are incubated with a restriction enzyme for increasing times. The action of the enzyme is revealed by YFISH and appears as a fluorescent band (green) progressively broadening inwards starting at the nuclear periphery (dashed black circles). Each circle corresponds to a separate sample. **c**, Gradual gDNA digestion revealed by wide-field epifluorescence microscopy. Green, HindIII cut sites with ligated YFISH adapters. Blue, DNA stained with Hoechst 33342. Scale bars, 20 μm (field of view) and 10 μm (insets). Times indicate the duration of incubation with HindIII. Optical midsections are shown. A different dynamic range was used for each digestion time to highlight the pattern of digestion in individual samples. YFISH signal is not detected in Hoechst-depleted regions, which most likely represent nucleoli. **d**, Same as in **c** but using STED microscopy. Scale bars, 10 μm. Experiments shown in **c** and **d** were repeated twice with similar results. **e**, YFISH fluorescence intensity at various distances from the nuclear lamina, for each of the times shown in **c** and **d**. Each dot represents the median intensity in one of 200 radial layers. *n*, number of cells analyzed. All source data for this figure are from HAP1 cells.

to visualize the digested HindIII recognition sites (Fig. 1c,d, Extended Data Fig. 1a and Supplementary Methods). As expected, after 10 min of incubation, we detected a fluorescent band at the nuclear periphery, which expanded inwards after longer incubation times, filling the entire nucleus after 2 h (Fig. 1c–e and Extended Data Fig. 1a,b). Quantification of the YFISH signal in hundreds of single cells revealed that the enzyme diffusion was homogenous across different cells of the same sample (Extended Data Fig. 1c–f and Supplementary Methods). Within the same nucleus, the signal profile was very similar along 200 randomly drawn nuclear radii, independently of the digestion duration, suggesting that the signal expands at a relatively constant speed along all radial directions (Extended Data Fig. 1g–j and Supplementary Methods).

GPSeq reproducibility and validation. We then aimed at revealing the identity of the genomic sequences surrounding the cut sites in nuclei undergoing gradual gDNA fragmentation by ligating adapters that enable next-generation sequencing (Fig. 2a and Supplementary Table 1). We generated sequencing libraries corresponding to different HindIII incubation times and sequenced them on an Illumina platform (Supplementary Table 2, Methods and Supplementary Methods). To infer radial positions throughout the genome, we defined different radiality estimates and selected the best one by comparing their radiality scores with 3D DNA FISH measurements (Supplementary Note 1). We took advantage of our large repository of DNA FISH probes[39] and profiled 3D distances from the nuclear lamina of 68 DNA loci on 11 different chromosomes (Supplementary Fig. 1a,b, Supplementary Table 3 and Supplementary Methods). The estimate showing the highest correlation with FISH considers how the restriction probability within a given genomic window varies across consecutive digestion times (Fig. 2b,c and Supplementary Note 1). Henceforth, we refer to this estimate as the GPSeq score and employ it in all subsequent analyses. The average GPSeq score error calculated by converting the score to physical distance was 7.49% of the average nuclear radius (256.41 nm), confirming the ability of GPSeq to accurately infer radial distances (Supplementary Fig. 1c,d and Supplementary Methods).

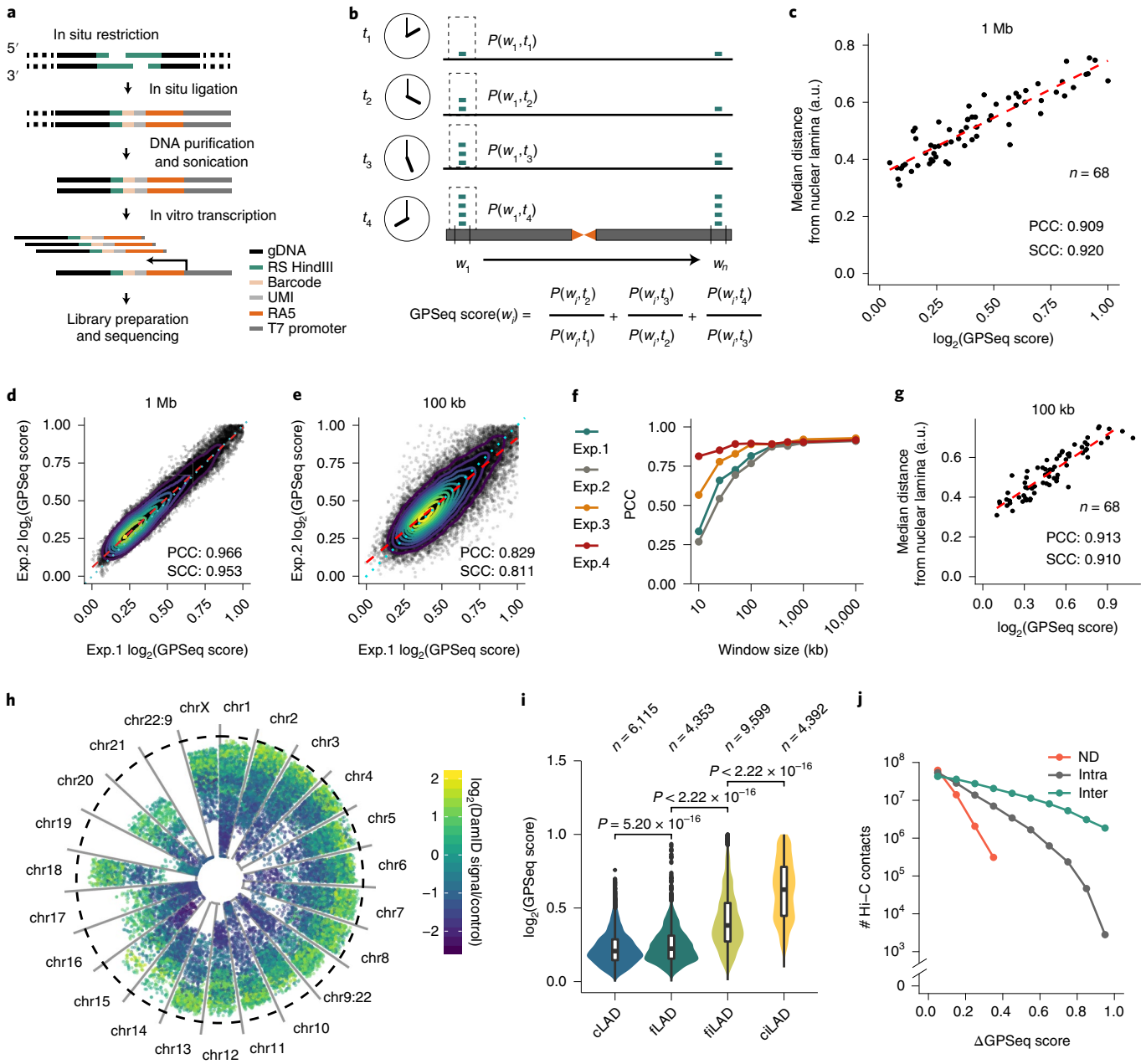


Fig. 2 | GPSeq reproducibility and validation. a, GPSeq library preparation. RS, restriction site; UMI, unique molecular identifier; RA5, Illumina adapter. **b**, GPSeq score calculation. *w*, genomic window; *P*(*w*,*t*), restriction probability in *w* after different *t* restriction times. **c**, Correlation between the $\log_2$ GPSeq score and the median 3D distance from the nuclear lamina measured by DNA FISH. Each dot represents one DNA FISH probe. **d**, Correlation between the $\log_2$ GPSeq scores (1-Mb overlapping windows, 100-kb step) in two HindIII experiments (Exp.1 and 2). Dotted cyan line, bisector; concentric curves, density contours. $n = 25{,}382$ genomic windows (points) were analyzed. **e**, Same as in **d** but at 100-kb resolution. $n = 25{,}557$ genomic windows (points) were analyzed. **f**, Correlations between the GPSeq score and the median 3D distance from nuclear lamina in two HindIII (Exp.1 and 2) and MboI (Exp.3 and 4) experiments at various resolutions. **g**, Correlation between the $\log_2$ GPSeq score averaged over the four experiments in **f** and the median 3D distance from the nuclear lamina measured by DNA FISH. Each dot represents one DNA FISH probe. **h**, $\log_2$-transformed lamin B DamID signal/control in 1-Mb genomic windows (dots) radially arranged based on their GPSeq score. The dashed line indicates nuclear lamina; chr9:22 and chr22:9 are the derivative chromosomes of the t(9;22) (q34;q11.2) translocation. $n = 26{,}350$ genomic windows (points) were analyzed. **i**, $\log_2$ GPSeq score (1-Mb nonoverlapping windows) in constitutive or facultative LADs (cLADs and fLADs, respectively) and constitutive or facultative inter-LADs (ciLADs and fiLADs, respectively). *P* values: Wilcoxon test, two sided. *n*, number of genomic windows analyzed. In all violin plots, boxes span from the 25th to the 75th percentile, and whiskers extend from −1.5 × IQR to +1.5 × IQR from the closest quartile. The dots indicate data outside the whiskers. **j**, Hi-C contacts count between pairs of 1-Mb genomic windows as a function of their GPSeq score difference (ΔGPSeq score). ND, near-diagonal Hi-C contacts (Supplementary Methods). PCC and SCC, Pearson's and Spearman's correlation coefficient, respectively. The dashed red lines indicate linear regressions. All source data for this figure are from HAP1 cells. IQR, interquartile range.

To test the reproducibility of GPSeq, we performed two replicate experiments using HindIII (Exp.1 and 2), obtaining highly correlated GPSeq scores at both 1-Mb and 100-kb resolution. The inter-experiment variability of the GPSeq score was low even in the case of loci localized in the innermost part of the nucleus (Fig. 2d,e and Extended Data Fig. 2a–d). This suggests that there is a clear

tendency for a given genomic locus to be found at a specific radial location, all along the nuclear radius. The GPSeq scores obtained using a different enzyme, MboI, were highly correlated with those obtained using HindIII, despite the two enzymes having the opposite GC content bias (Extended Data Fig. 2e–j, Supplementary Note 2 and Supplementary Methods). All the experiments yielded GPSeq scores that strongly correlated with radiality measurements by DNA FISH (Fig. 2f and Supplementary Table 4). The correlation with DNA FISH was even higher when the GPSeq scores from the four experiments were averaged together (Fig. 2g and Supplementary Table 4). Hence, we used averaged GPSeq scores in all subsequent analyses.

To test the possible effect of DNA accessibility, we compared the restriction probability at different time points and the GPSeq score with DNA accessibility measured by assay for transposase-accessible chromatin using sequencing (ATAC-seq)[40] (Supplementary Table 5 and Supplementary Methods). The correlation between the ATAC-seq signal and the restriction probability increased with the time of digestion, reaching a moderate correlation for longer digestion times (Supplementary Fig. 2a and Supplementary Note 2). Of note, the GPSeq score showed a lower correlation with the ATAC-seq signal than the restriction probability of the longest time point (Pearson's correlation coefficient = 0.451 versus 0.72) (Supplementary Fig. 2b).

To validate GPSeq, we compared it with lamin B DamID previously performed in HAP1 cells[13] (Supplementary Table 5 and Supplementary Methods). The GPSeq score and the DamID signal were anti-correlated, and genomic regions with low DamID signal had a broader GPSeq score range than regions with high DamID signal (Fig. 2h and Supplementary Fig. 3a–c). Constitutive inter-LAD regions (ciLADs)[13] were the most central, whereas constitutive LADs (cLADs) were the most peripheral, suggesting that the nuclear mid zone is less conserved across different cell types (Fig. 2i). We also assessed whether the contact frequency measured by Hi-C[37] drops when the radial distance between two genomic loci increases. Indeed, frequently contacting loci shared very similar radial locations (Fig. 2j, Supplementary Table 5 and Supplementary Methods). Altogether, these results demonstrate that GPSeq is a reliable and reproducible method for inferring radial locations throughout the genome. A step-by-step GPSeq protocol is available at Protocol Exchange (https://doi.org/10.21203/rs.3.pex-570/v1).

Radial arrangement of chromatin in the nucleus. We then examined how chromosomes and various chromatin features are radially arranged in the nucleus. Individual chromosomes showed unique GPSeq score profiles with considerable variability along the same chromosome (Fig. 3a,b, Supplementary Fig. 4, Supplementary Fig. 5a, Supplementary Video 1 and Supplementary Methods). We used the GPSeq score to draw two-dimensional maps of the relative abundance of individual chromosomes in concentric nuclear layers, which showed that small chromosomes were depleted in the outer layers (Fig. 3c, Supplementary Fig. 5b and Supplementary Methods). Indeed, the GPSeq score and chromosome size were anti-correlated, but the relatively low strength of this anti-correlation suggested that chromosome size alone was not an accurate predictor of radiality (Fig. 3d and Supplementary Fig. 5c,d). Gene density and gene expression alone were also weak predictors of radiality at chromosomal level (Supplementary Fig. 5e,f). Notably, GC content was the only feature consistently correlated with the GPSeq score across different resolutions. However, the GC content did not accurately predict radial locations throughout the genome already at 1-Mb resolution (Extended Data Fig. 3a–f). Therefore, we built a multivariable model combining both genomic (cell-type-independent) and epigenomic (cell-type-specific) features (Supplementary Methods). A model combining chromosome size and GC content yielded the highest accuracy in predicting the radial location of individual chromosomes, with no added benefit from using information about gene density or expression (R^2 = 93.9%; prediction error = 0.073) (Extended Data Fig. 3g and Supplementary Table 6). At 1-Mb resolution, the most accurate model included GC content, gene density, gene expression and chromosome size (R^2 = 74.1%; prediction error = 0.12) (Extended Data Fig. 3h and Supplementary Table 6). An independent two-replicate experiment using the GM06990 diploid lymphoblastoid cell line showed a highly conserved radial chromatin arrangement compared to HAP1 cells (Pearson's correlation coefficient between averaged GPSeq scores of the two cell lines = 0.88) (Supplementary Table 2 and Supplementary Methods). Accordingly, the multivariable model built on HAP1 GPSeq data could accurately predict radiality in GM06990 cells at 1-Mb resolution (average prediction error = 0.1). Altogether, these results demonstrate that cell-type-invariant features of the linear genome, such as GC content, establish a radial blueprint, which is then shaped by cell-type-specific features, such as gene expression.

Higher-order radial organization of the genome. Next, we examined how A and B compartments defined by Hi-C[37] are radially arranged. As expected, A compartments were typically more central than B compartments (Supplementary Fig. 6a–c, Supplementary Table 5 and Supplementary Methods). We wondered whether this polarity is present on all chromosomes, especially those preferentially located in the inner part of the nucleus. Surprisingly, chromosomes without clear A and B polarization were not the most central ones. In fact, the polarization was rather pronounced on chr17 and chr19, which are very central, whereas A and B compartments had a similar radial arrangement on chr10 and chr18, which are more peripheral (Supplementary Fig. 6d). We tested whether this would be different at the level of A and B subcompartments[38], given that individual subcompartments showed different GPSeq score distributions (Supplementary Fig. 6e,f). Examination of individual chromosomes revealed similar subcompartment polarization patterns, with A1 being consistently more central than B2 and B3 (Supplementary Fig. 7).

We then wondered how the radial arrangement of different subcompartments affects the spatial distribution of active and inactive chromatin (Supplementary Table 5 and Supplementary Methods). Overall, features and marks of active chromatin, such as DNA accessibility, H3K27ac and H3K4me3, as well as chromatin-bound RNA polymerase II, increased globally toward the nuclear interior in parallel with gene density and expression (Fig. 3e,f and Extended Data Fig. 4a–d). Notably, we found that each feature had a rather characteristic radial profile across different subcompartments. For example, DNA accessibility remained flat along the nuclear radius in the B2 subcompartment, whereas it increased in A1 and A2 and in B1 (Fig. 3e). A similar trend was observed for DNA methylation (Extended Data Fig. 4e). H3K27ac increased toward the nuclear interior mainly in A1 and A2 subcompartments but decreased in B2, whereas H3K4me3, a mark of active promoters, increased only in A1 (Fig. 3f and Extended Data Fig. 4a). On the other hand, H3K9me3, which marks heterochromatin, decreased towards the center genome-wide, even though it sharply increased towards the nuclear interior in the B2 subcompartment (Fig. 3g). Intriguingly, H3K4me1, a mark of active and poised enhancers, decreased toward the center in all subcompartments, despite its radial profile mildly increasing toward the center when no subcompartment stratification was applied (Fig. 3h). H3K27me3, a mark associated with the Polycomb repressive complex, followed the radial distribution of the B1 subcompartment (Fig. 3i and Supplementary Fig. 6e). In turn, the pattern of H3K27me3 was reflected in the radial distribution of homeobox genes, a well-established Polycomb target[41], which were enriched in the nuclear mid zone (Fig. 3j and Supplementary Methods).

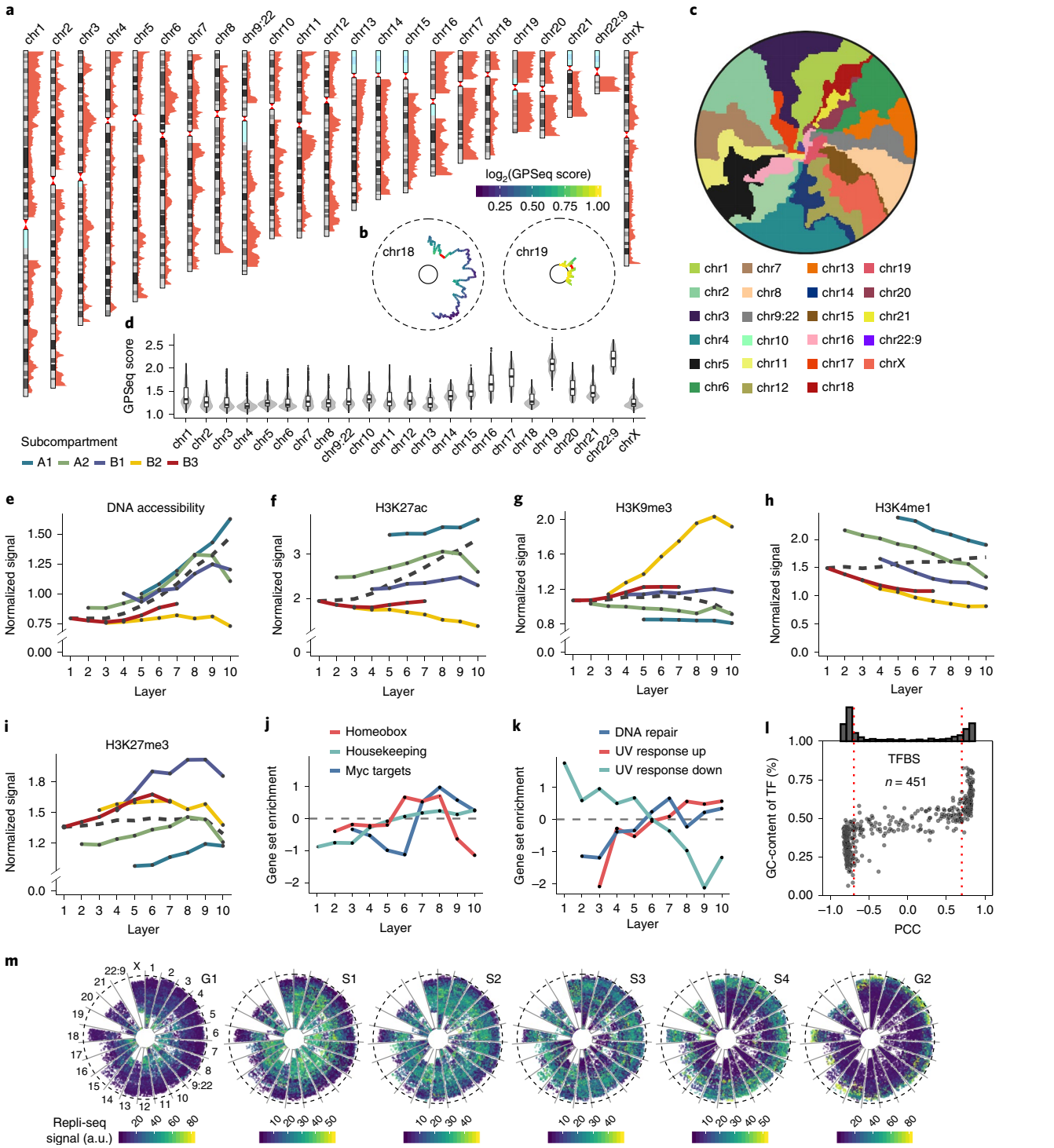


Fig. 3 | Radial organization of chromatin in human cells. a, GPSeq score profiles along individual chromosomes (1-Mb overlapping windows, 100-kb step size). **b**, Circular plots of chr18 and chr19 radial location (1-Mb nonoverlapping windows). Dashed circle, nuclear lamina; solid circle, nuclear center. Red, masked-out pericentromeric regions. **c**, Preferential radial location of individual chromosomes. For each chromosome, the number of pixels is proportional to the number of the genomic windows in that chromosome. Chromosomes are assigned to five nuclear layers of equal thickness based on their GPSeq score. The angular order of the chromosomes is arbitrary. **d**, Distribution of GPSeq score per chromosome (1-Mb overlapping windows, 100-kb steps). In all violin plots, boxes span from the 25th to the 75th percentile, and whiskers extend from −1.5 × IQR to +1.5 × IQR from the closest quartile. The dots indicate data outside the whiskers. **e**–**i**, Radial distribution of DNA accessibility and various histone marks in A and B subcompartments (100-kb resolution). The dashed lines indicate radial distribution without stratifying by subcompartment. Mean normalized signals are shown (Supplementary Methods). **j**,**k**, Radial profiles of selected gene sets. **l**, Pearson's correlation coefficient (PCC) between the $\log_2$ GPSeq score (1-Mb overlapping genomic windows, 100-kb step size) and the number of predicted TFBSs ranked based on GC content. $n = 26{,}350$ genomic windows (points) were used to calculate the PCC for $n = 451$ TFBSs. Dashed vertical red lines indicate PCC of −0.7 and 0.7, respectively. **m**, Repli-seq signal in 1-Mb genomic windows radially arranged based on their GPSeq score in six cell cycle subphases (G1, S1–S4 and G2). chr9:22 and chr22:9 are the derivative chromosomes of t(9;22) (q34;q11.2) translocation. $n = 26{,}350$ points (genomic windows) are shown in each plot. All source data for this figure are from HAP1 cells, with the exception of Repli-seq data, which are from K562 cells. IQR, interquartile range.

The observation that homeobox genes have a distinctive radial pattern prompted us to examine whether the same holds for genes involved in other pathways. In most cases, the radial distribution of genes belonging to different hallmark pathways was not significantly different from the distribution of all genes (Supplementary Table 7 and Supplementary Methods). However, some groups of genes did show a peculiar radial arrangement (Fig. 3j,k and Extended Data Fig. 4f). For example, genes downregulated in response to UV damage were enriched at the nuclear periphery, whereas genes upregulated upon UV were enriched in central nuclear layers where DNA repair genes also accumulated (Fig. 3k). Predicted transcription factor binding sites (TFBSs) were also radially distributed, with more than 70% of all TFBSs being either strongly correlated or anti-correlated with the GPSeq score (Fig. 3l, Supplementary Table 8 and Supplementary Methods). Altogether, these results suggest that the radial arrangement of chromatin defines how regulatory elements and genes are spatially distributed, which might have important functional consequences.

Radial progression of DNA replication. We then investigated the correlation between chromatin radiality and replication timing. Based on the literature, we expected that early-replicating regions would be more central than late-replicating ones[22,42]. Indeed, although replication fork firing appears to occur simultaneously at various radial locations, we found that genome-wide replication proceeds gradually, starting from the innermost part of the nucleus and progressing toward the periphery (Fig. 3m, Extended Data Fig. 5a, Supplementary Table 5 and Supplementary Methods). Stratification of the Repli-seq signal by A and B subcompartments revealed that B2 and B3 heterochromatin replicates late even in central nuclear layers (Extended Data Fig. 5b). This analysis also showed that the observed gradual radial progression of the replication wave is mainly driven by the A2 and B1 subcompartments, because the radial location of firing did not change throughout the S phase in other subcompartments (Extended Data Fig. 5c). Notably, the addition of individual epigenetic marks or replication timing did not substantially improve the predictive power of the multivariable model described above, typically increasing the R^2 of less than 1% (Supplementary Table 9).

Whole-genome reconstructions. Having demonstrated the ability of GPSeq to reliably infer radial locations throughout the genome, we sought to integrate GPSeq and Hi-C data to predict the 3D genome structure in single cells. To this end, we developed *chromflock*, a high-performance algorithm that builds on PGS[18,43] and enables direct integration of GPSeq and Hi-C information to generate ensembles of thousands of whole-genome structures based on molecular dynamics (Supplementary Software and Methods). We generated 10,000 structures at 1-Mb resolution, either using Hi-C data only (H) or combining Hi-C with GPSeq (HG) (Fig. 4a and Supplementary Videos 2–5). We first checked whether H structures are similar to those previously obtained with PGS. Indeed, the predicted structures were consistent with the distance matrix built from the original Hi-C data, showing that smaller chromosomes tend to cluster in the nuclear center (Supplementary Fig. 8a–d). Moreover, radiality profiles along individual chromosomes matched those previously obtained with PGS (Supplementary Fig. 8e). These features were not recapitulated in H structures generated using only Hi-C intra-chromosomal contacts, and even when all contacts were used, H structures did not recapitulate GPSeq radiality profiles and poorly correlated with DNA FISH (Supplementary Fig. 9a–h). In contrast, HG structures recapitulated the tendency of small chromosomes to cluster in the nuclear interior, with the exception of chr18, and were significantly more consistent with the distance matrix calculated from the original Hi-C map, compared to H structures (Fig. 4b,c and Extended Data Fig. 6a–c). Accordingly, HG structures were highly correlated with GPSeq radial profiles and DNA FISH (Fig. 4d and Extended Data Fig. 6d). Remarkably, even when trans contacts were omitted from the Hi-C input data, the structures closely resembled HG ones (Extended Data Fig. 7a–f).

We then wondered whether the higher-order radial organization of A and B compartments is recapitulated in individual HG structures. The vast majority of the 10,000 HG structures showed a clear A and B compartment polarization at the level of individual chromosomes, which was not seen in H structures (Extended Data Fig. 8a,b). Using *chromflock*, we generated 1,000 additional HG structures at 100-kb resolution, which showed the expected radial arrangement of A and B subcompartments and strongly correlated with DNA FISH (Fig. 4e,f, Extended Data Fig. 8c,d and Methods). Notably, A1 and B1 subcompartments were typically the most central, followed by A2 or B2, whereas B3 was typically the most peripheral across all HG structures but not in H ones (Fig. 4g and Supplementary Fig. 10). To further investigate the spatial arrangement of A and B subcompartments in individual chromosomes in single structures, we devised a metric of polarization and orientation (Extended Data Fig. 9a,b and Supplementary Methods). Most chromosomes showed a strong A and B subcompartment polarization in most structures, which was often radially aligned in the case of larger chromosomes but much less for smaller chromosomes (Extended Data Fig. 9c,d). Notably, such radial arrangement of subcompartments was not recapitulated in H structures (Extended Data Fig. 9e,f). Altogether, these results demonstrate that integration of GPSeq and Hi-C data allows generating ensembles of genome structure predictions that can provide new insights into how the genome is radially organized at the single-cell level.

GPSeq reveals radial patterns of mutations and DNA breaks. It has long been speculated that heterochromatin acts like a shield to protect euchromatin from DNA damage[44]. In support of this 'bodyguard hypothesis', several studies have reported that the frequency of single-nucleotide polymorphisms (SNPs) and cancer-associated single-nucleotide variants (SNVs) is higher in heterochromatic and late-replicating genomic regions[45–48], which are conventionally associated with the nuclear periphery. On the other hand, different studies have shown that other mutation types, such as gene fusions, are more frequent in open chromatin[49], which is more abundant in the nuclear interior. To shed light on how different mutational processes relate to chromatin radiality, we integrated our GPSeq data with publically available SNP, SNV and gene fusion data (Supplementary Methods). We first assessed the radial pattern of SNVs previously identified in four cancer types, including chronic lymphocytic leukemia, a tumor that shares the hematopoietic origin with the HAP1 cell line used in this study. These mutations have been previously associated with various heterochromatin marks, in particular H3K9me3 (ref. [48]). Consistently, the SNV frequency progressively decreased from the nuclear periphery toward the center, as expected based on the 'bodyguard hypothesis', especially in the case of lung cancer and melanoma mutations (Fig. 5a). A similar analysis of SNPs identified in the 1000 Genomes Project[50] revealed a small increase toward the center, indicative of a higher burden of SNPs in active chromatin (Fig. 5a). However, when we stratified by A and B subcompartments, we found that the SNP frequency was higher in B1 and B2 rather than in A1 and A2 subcompartments (Fig. 5b). Interestingly, centrally located genomic regions belonging to the B2 subcompartment carried the highest burden of SNPs, although the differences were small (Fig. 5b). Notably, these regions were also strongly enriched in H3K9me3 (Fig. 3g). We speculate that different mutational processes and/or DNA repair mechanisms might underlie the observed differences in the radial distribution of germline SNPs and cancer SNVs.

We then examined gene fusions in The Cancer Genome Atlas (TCGA)[51] (Supplementary Methods). Genomic loci involved in

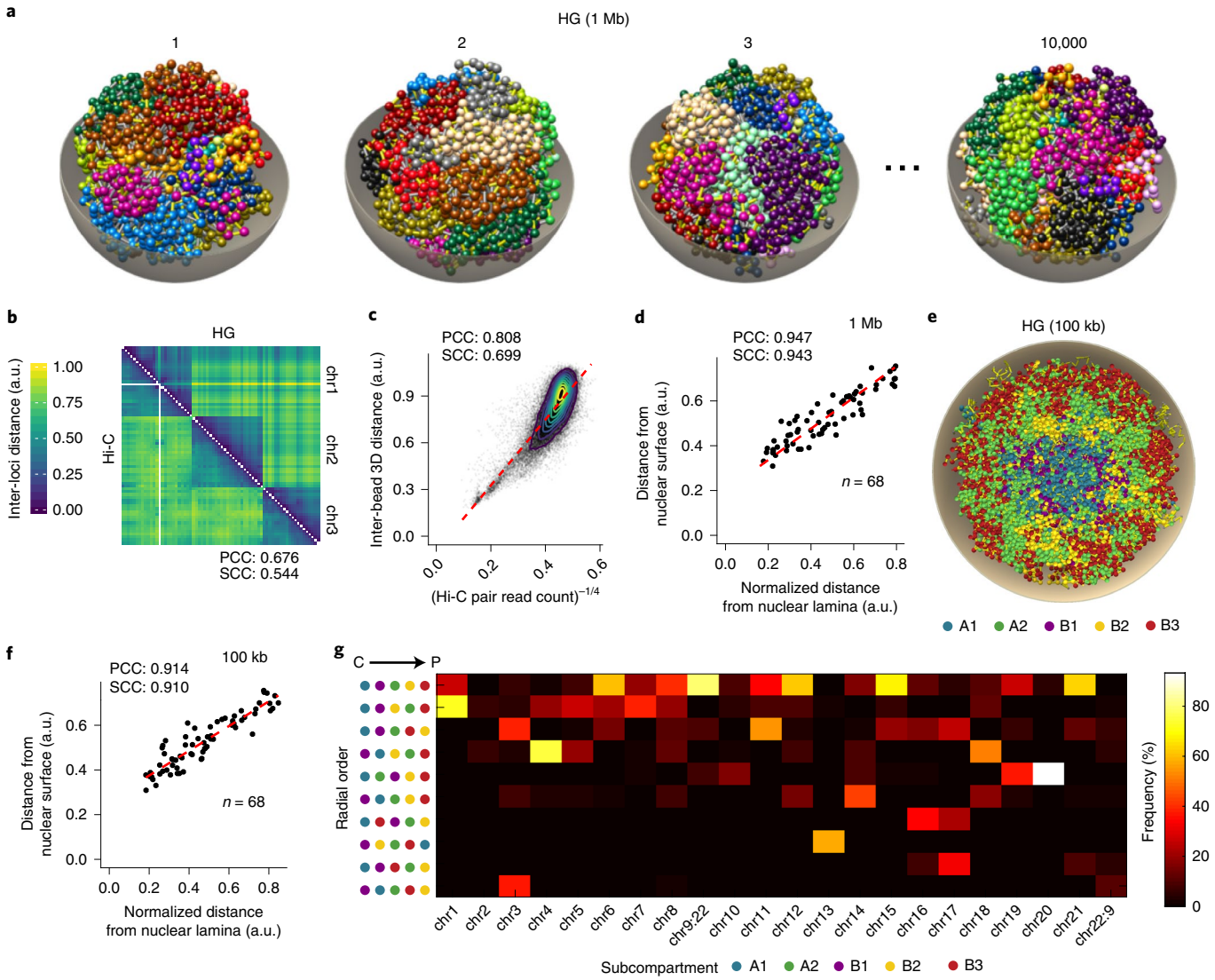


Fig. 4 | Generation of 3D genome structures by GPSeq and Hi-C integration. a, Examples of 4 out of 10,000 *chromflock* structures generated by integrating GPSeq and Hi-C data (HG structures). Each bead corresponds to a 1-Mb genomic window. Chromosomes are shown with distinct colors. Gray, modeled nuclear surface. **b**, Comparison between Hi-C and HG structures for three representative chromosomes. Upper triangle, inter-bead 3D distances in 10,000 HG structures. Bottom triangle, KR-normalized Hi-C contact frequency matrix, with each element raised to the power of −0.25. The reported correlation coefficients are for 1-Mb resolution, whereas, for simplicity, the plot shows averaged values over 10-Mb genomic windows (points). **c**, Correlation between average inter-bead 3D distance in HG structures and KR-normalized Hi-C contact frequency, with each element raised to the power of −0.25. Each dot represents a pair of 10-Mb nonoverlapping genomic windows obtained by averaging 1-Mb nonoverlapping windows. $n = 47{,}531$ pairs of genomic windows (points) were analyzed. Concentric curves indicate density contours. **d**, Correlation between radial position in HG structures and median 3D distance to nuclear lamina measured by DNA FISH. Each dot represents one DNA FISH probe. **e**, Example of one out of 1,000 HG *chromflock* structures. Each bead corresponds to a 100-kb genomic window. A and B subcompartments are shown in different colors. The modeled nuclear surface is shown in gray. **f**, Same as in **d** but for 1,000 HG structures at 100-kb resolution. **g**, Frequency of the ten most frequent A and B subcompartment radial arrangements from center (C) to periphery (P) in 1,000 HG structures, separately for each chromosome. PCC and SCC, Pearson's and Spearman's correlation coefficient, respectively. Dashed red lines indicate linear regressions.

fusions localized more internally than loci that have not been found to fuse (Fig. 5c). Notably, an analysis of the chromosome mingling frequency in the 100-kb resolution *chromflock* structures showed that the most frequently mingling loci were moderately enriched in gene fusions but only in HG structures (Fig. 5d, Extended Data Fig. 10a and Supplementary Methods). Accordingly, the number and density of unique Hi-C trans contacts increased toward the nuclear center (Extended Data Fig. 10b,c).

We then investigated whether the radial distribution of gene fusions corresponds to the one of DNA double-strand breaks (DSBs), a major DNA lesion implicated in the pathogenesis of gene fusions in cancer[52]. To this end, we took advantage of a genome-wide map of endogenous DSBs, which we previously obtained from a HAP1-related cell line[53] using our BLISS method[54] (Supplementary Methods). As expected, genomic loci frequently fused in human cancers had a higher DSB frequency than loci that have not been found to fuse (Extended Data Fig. 10d). The DSB frequency progressively increased toward the nuclear interior in both genic and intergenic regions (Extended Data Fig. 10e). Quantitative analysis of the radial distribution of phosphorylated histone H2A.X (γH2A.X)—a proxy of DSBs—confirmed that endogenous breaks are more frequently detected in the inner nucleus (Extended Data

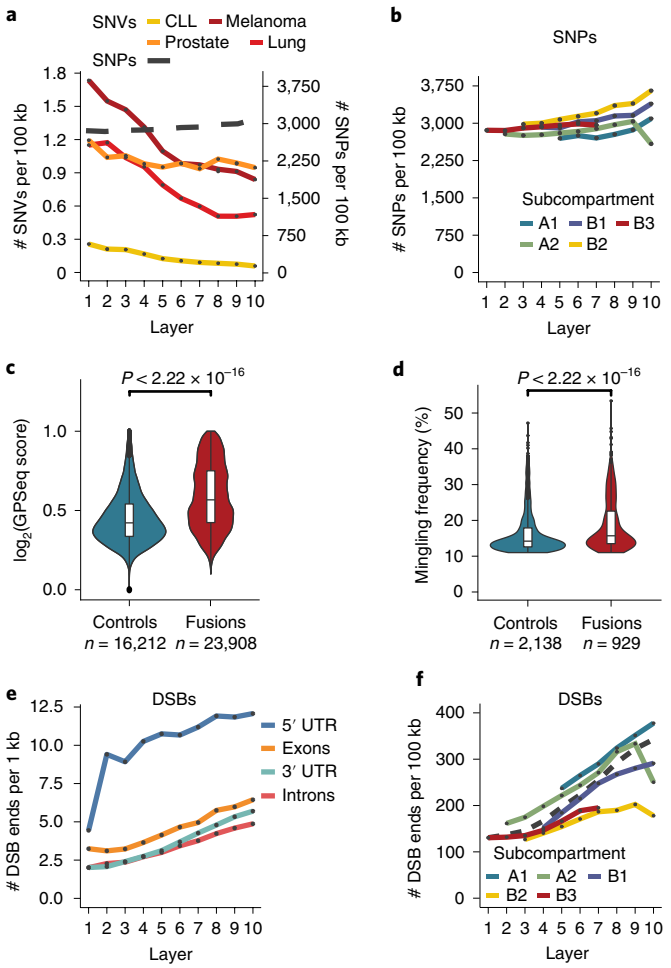


Fig. 5 | Radial distribution of mutations and DNA breaks. a, Radial distribution of SNVs in four cancer types (left axis) and of SNPs from the 1000 Genomes Project (right axis). Mean normalized signals are shown at 100-kb resolution (Supplementary Methods). **b**, Radial distribution of SNPs in A and B subcompartments. Mean normalized signals are shown. **c**, Distribution of the GPSeq score of 100-kb genomic windows overlapping (Fusions) or not (Controls) with cancer-associated gene fusions annotated in TCGA. *n*, number of genomic windows analyzed. *P* values: Wilcoxon test, two sided. **d**, Distribution of the inter-chromosome mingling frequency of the 10% most frequently mingling beads in 100-kb-resolution HG *chromflock* structures, separately for beads overlapping (Fusions) or not (Controls) with cancer-associated gene fusions annotated in TCGA. *P* value: Wilcoxon test, two sided. *n*, number of beads analyzed. **e**, Radial distribution of endogenous DSBs stratified by different parts of human protein-coding genes. **f**, Radial distribution of DSBs in A and B subcompartments. The dashed line indicates DSB radial distribution without stratifying by subcompartment. Mean BLISS signals at 100-kb resolution are shown. In all violin plots, boxes span from the 25th to the 75th percentile, and whiskers extend from −1.5 × IQR to +1.5 × IQR from the closest quartile. The dots indicate data outside the whiskers. All source data for this figure are from HAP1 cells, except for BLISS data, which are from K562 cells. CLL, chronic lymphocytic leukemia; IQR, interquartile range; UTR, untranslated region.

Fig. 10f). Notably, the highest DSB frequency was observed within the 5′ untranslated region of protein-coding genes belonging to the most centrally located A1 and A2 regions, in agreement with prior observations that DSBs tend to accumulate around the transcription start site of actively transcribed genes[54,55] where gene fusions also form preferentially[49] (Fig. 5e,f). Altogether, these results highlight the advantage of having GPSeq radial maps, to investigate the forces that shape the mutational landscape during evolution and in cancer.

Discussion

We developed a robust method to map the radial arrangement of chromatin throughout the genome, which, compared to the gold standard method, DNA FISH, offers orders of magnitude higher throughput. Compared to tyramide signal amplification sequencing[56] and lamin DamID[11], GPSeq can accurately estimate radial positions all along the nuclear radius, not only close to the nuclear lamina. In principle, genome architecture mapping (GAM)[57] could be adopted to assess radiality throughout the genome. However, given the fact that, in GAM, the total number of reads per library from a given nuclear profile is used as a proxy of radiality, it remains unclear whether this method can accurately probe for radiality at high resolution. Lastly, single-cell Hi-C[58] and diploid chromatin conformation capture (Dip-C)[59] can also be used, in principle, to infer radial positions throughout the genome. However, these methods are costly and experimentally more challenging than GPSeq.

Together with GPSeq, we developed a new FISH assay, YFISH, which allows monitoring the pattern of in situ digestion before sequencing GPSeq samples. YFISH could also serve as a standalone assay to visualize chromatin accessibility in single cells, similar to ATAC with visualization (ATAC-see)[60]. Notably, the same protocol for gradual diffusion of restriction enzymes can be adapted to other proteins, such as antibodies (Supplementary Fig. 11a–c and Supplementary Methods), opening up the possibility to develop 'radial' versions of existing assays—for instance, radial chromatin immunoprecipitation sequencing or Hi-C to directly map chromatin occupancy and chromosome contacts along the nuclear radius.

Although in this study we mainly used haploid cells, we show that GPSeq can also be applied to chart radiality in diploid cells. This approach, however, does not allow distinguishing of the preferential radial position of loci located on homologous chromosomes. Future integration of GPSeq with whole-genome haplotyping strategies will enable determination of whether homologous loci occupy similar or different radial positions in the nucleus. Irrespective of that, GPSeq can already be applied to investigate the role of different factors in shaping chromatin radiality in different cell types, including aneuploid and polyploid cells, as this does not require haplotyping. This makes GPSeq superior to other methods, such as Hi-C or Dip-C, which require a modelling step to infer radiality.

We also developed a new algorithm, *chromflock*, which extends the PGS software previously used to make 3D genome reconstructions[18]. We show that *chromflock* is able to generate ensembles of thousands of 3D genome structures that are highly consistent with radial distances measured by DNA FISH. Remarkably, at high resolution (100-kb), *chromflock* structures generated by integrating GPSeq and Hi-C data fully recapitulate the radial organization of A and B subcompartments revealed by bulk GPSeq.

Although it has been known for a long time that chromatin is radially organized, here we provide the first high-resolution radial map of the human nucleus, revealing many previously unappreciated features. We show that, even in the more central parts of the nucleus, there is a clear tendency for certain genomic loci to occupy specific radial positions. Notably, our A and B subcompartment analysis revealed that the radial distribution of chromatin features follows unique patterns. For example, DNA accessibility is higher in the repressed chromatin located in the inner portion of the nucleus in comparison to the more transcriptionally active chromatin in the A1 subcompartment, which is located farther away from the center. Intriguingly, the levels of the heterochromatin mark H3K9me3 are highest in central B2 regions, which might be needed to counteract the highly active chromatin surrounding them.

More than 40 years ago, it was proposed that constitutive heterochromatin at the nuclear periphery protects the more central active

chromatin from DNA damage[44]. Our results suggest that this 'bodyguard hypothesis' might explain the spatial distribution of certain mutation types, but not all. For example, whereas the frequency of cancer SNVs is higher at the nuclear periphery, confirming previous assumptions[61], the frequency of germline SNPs instead mildly increases toward the nuclear interior. This observation is not in disagreement with previous studies, which showed a correlation between SNPs and late-replicating chromatin[45,62]. In fact, our results show that the highest burden of SNPs is found in H3K9me3 heterochromatin, which is indeed late replicating. However, this fraction of heterochromatin tends to be located in the nuclear interior, unlike the majority of heterochromatin. It is important to note that, despite being preferentially localized in the nuclear interior, smaller chromosomes do contain heterochromatin, which is thus embedded in a highly transcriptionally active environment. This might explain the different propensity of heterochromatin located at various radial positions to undergo different mutational processes. One limitation of this analysis, however, is the fact that our radial maps were not obtained in the same cell type from which the mutations are likely to arise.

In conclusion, we developed a 'user-friendly' and versatile assay that significantly expands the existing toolkit for studying the 3D genome. GPSeq can be readily applied to explore the conservation, dynamics and functional relevance of genome radiality in different cell types and conditions as well as the influence of nuclear shape on radiality.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-020-0519-y.

Received: 15 March 2019; Accepted: 9 April 2020;
Published online: 25 May 2020

References

1. Sleeman, J. E. & Trinkle-Mulcahy, L. Nuclear bodies: new insights into assembly/dynamics and disease relevance. *Curr. Opin. Cell Biol.* **28**, 76–83 (2014).
2. Croft, J. A. et al. Differences in the localization and morphology of chromosomes in the human nucleus. *J. Cell Biol.* **145**, 1119–1131 (1999).
3. Bridger, J. M., Boyle, S., Kill, I. R. & Bickmore, W. A. Re-modelling of nuclear architecture in quiescent and senescent human fibroblasts. *Curr. Biol. CB* **10**, 149–152 (2000).
4. Cremer, M. et al. Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosome Res.* **9**, 541–567 (2001).
5. Boyle, S. et al. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum. Mol. Genet.* **10**, 211–219 (2001).
6. Mayer, R. et al. Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. *BMC Cell Biol.* **6**, 44 (2005).
7. Bolzer, A. et al. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* **3**, e157 (2005).
8. Sun, H. B., Shen, J. & Yokota, H. Size-dependent positioning of human chromosomes in interphase nuclei. *Biophys. J.* **79**, 184–190 (2000).
9. Tanabe, H. et al. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc. Natl Acad. Sci. USA* **99**, 4424–4429 (2002).
10. van Steensel, B. & Belmont, A. S. Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell* **169**, 780–791 (2017).
11. Guelen, L. et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
12. Peric-Hupkes, D. et al. Molecular maps of the reorganization of genome–nuclear lamina interactions during differentiation. *Mol. Cell* **38**, 603–613 (2010).
13. Kind, J. et al. Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134–147 (2015).
14. Alcobia, I., Dilão, R. & Parreira, L. Spatial associations of centromeres in the nuclei of hematopoietic cells: evidence for cell-type-specific organizational patterns. *Blood* **95**, 1608–1615 (2000).
15. Alcobia, I., Quina, A. S., Neves, H., Clode, N. & Parreira, L. The spatial organization of centromeric heterochromatin during normal human lymphopoiesis: evidence for ontogenically determined spatial patterns. *Exp. Cell Res.* **290**, 358–369 (2003).
16. Molenaar, C. et al. Visualizing telomere dynamics in living mammalian cells using PNA probes. *EMBO J.* **22**, 6631–6641 (2003).
17. Weierich, C. et al. Three-dimensional arrangements of centromeres and telomeres in nuclei of human and murine lymphocytes. *Chromosome Res.* **11**, 485–502 (2003).
18. Tjong, H. et al. Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. *Proc. Natl Acad. Sci. USA* **113**, E1663–E1672 (2016).
19. Németh, A. & Längst, G. Genome organization in and around the nucleolus. *Trends Genet.* **27**, 149–156 (2011).
20. Quinodoz, S. A. et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* **174**, 744–757 (2018).
21. Federico, C. et al. Gene-rich and gene-poor chromosomal regions have different locations in the interphase nuclei of cold-blooded vertebrates. *Chromosoma* **115**, 123–128 (2006).
22. Grasser, F. et al. Replication-timing-correlated spatial chromatin arrangements in cancer and in primate interphase nuclei. *J. Cell Sci.* **121**, 1876–1886 (2008).
23. Hepperger, C., Mannes, A., Merz, J., Peters, J. & Dietzel, S. Three-dimensional positioning of genes in mouse cell nuclei. *Chromosoma* **117**, 535–551 (2008).
24. Kreth, G., Finsterle, J., von Hase, J., Cremer, M. & Cremer, C. Radial arrangement of chromosome territories in human cell nuclei: a computer model approach based on gene density indicates a probabilistic global positioning code. *Biophys. J.* **86**, 2803–2812 (2004).
25. Andrulis, E. D., Neiman, A. M., Zappulla, D. C. & Sternglanz, R. Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature* **394**, 592–595 (1998).
26. Sadoni, N. et al. Nuclear organization of mammalian genomes. Polar chromosome territories build up functionally distinct higher order compartments. *J. Cell Biol.* **146**, 1211–1226 (1999).
27. Kosak, S. T. et al. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* **296**, 158–162 (2002).
28. Kosak, S. T. et al. Coordinate gene regulation during hematopoiesis is related to genomic organization. *PLoS Biol.* **5**, e309 (2007).
29. Finlan, L. E. et al. Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet.* **4**, e1000039 (2008).
30. Reddy, K. L., Zullo, J. M., Bertolino, E. & Singh, H. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* **452**, 243–247 (2008).
31. Takizawa, T., Meaburn, K. J. & Misteli, T. The meaning of gene positioning. *Cell* **135**, 9–13 (2008).
32. Therizols, P. et al. Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science* **346**, 1238–1242 (2014).
33. Shachar, S. & Misteli, T. Causes and consequences of nuclear gene positioning. *J. Cell Sci.* **130**, 1501–1508 (2017).
34. Cook, P. R. & Marenduzzo, D. Transcription-driven genome organization: a model for chromosome structure and the regulation of gene expression tested through simulations. *Nucleic Acids Res.* **46**, 9895–9906 (2018).
35. Ganai, N., Sengupta, S. & Menon, G. I. Chromosome positioning from activity-based segregation. *Nucleic Acids Res.* **42**, 4145–4159 (2014).
36. Küpper, K. et al. Radial chromatin positioning is shaped by local gene density, not by gene expression. *Chromosoma* **116**, 285–306 (2007).
37. Lieberman-Aiden, E. et al. Comprehensive mapping of long range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
38. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
39. Gelali, E. et al. iFISH is a publically available resource enabling versatile DNA FISH to study genome architecture. *Nat. Commun.* **10**, 1636 (2019).
40. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
41. Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H. & Helin, K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* **20**, 1123–1136 (2006).
42. Schermelleh, L., Solovei, I., Zink, D. & Cremer, T. Two-color fluorescence labeling of early and mid-to-late replicating chromatin in living cells. *Chromosome Res.* **9**, 77–80 (2001).
43. Hua, N. et al. Producing genome structure populations with the dynamic and automated PGS software. *Nat. Protoc.* **13**, 915–926 (2018).
44. Hsu, T. C. A possible function of constitutive heterochromatin: the bodyguard hypothesis. *Genetics* **79**, 137–150 (1975).

45. Stamatoyannopoulos, J. A. et al. Human mutation rate associated with DNA replication timing. *Nat. Genet.* **41**, 393–395 (2009).
46. Liu, L., De, S. & Michor, F. DNA replication timing and higher-order nuclear organization determine single-nucleotide substitution patterns in cancer genomes. *Nat. Commun.* **4**, 1502 (2013).
47. Morganella, S. et al. The topography of mutational processes in breast cancer genomes. *Nat. Commun.* **7**, 11383 (2016).
48. Schuster-Böckler, B. & Lehner, B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**, 504–507 (2012).
49. Chiarle, R. et al. Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* **147**, 107–119 (2011).
50. Clarke, L. et al. The International Genome Sample Resource (IGSR): a worldwide collection of genome variation incorporating the 1000 Genomes Project data. *Nucleic Acids Res.* **45**, D854–D859 (2017).
51. Hu, X. et al. TumorFusions: an integrative resource for cancer-associated transcript fusions. *Nucleic Acids Res.* **46**, D1144–D1149 (2018).
52. Mertens, F., Johansson, B., Fioretos, T. & Mitelman, F. The emerging complexity of gene fusions in cancer. *Nat. Rev. Cancer* **15**, 371–381 (2015).
53. Gothe, H. J. et al. Spatial chromosome folding and active transcription drive DNA fragility and formation of oncogenic MLL translocations. *Mol. Cell* **75**, 267–283.e12 (2019).
54. Yan, W. X. et al. BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat. Commun.* **8**, 15058 (2017).
55. Lensing, S. V. et al. DSBCapture: in situ capture and sequencing of DNA breaks. *Nat. Methods* **13**, 855–857 (2016).
56. Chen, Y. et al. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol.* **217**, 4025–4048 (2018).
57. Beagrie, R. A. et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).
58. Nagano, T. et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
59. Tan, L., Xing, D., Chang, C.-H., Li, H. & Xie, X. S. Three-dimensional genome structures of single diploid human cells. *Science* **361**, 924–928 (2018).
60. Chen, X. et al. ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat. Methods* **13**, 1013–1020 (2016).
61. Gonzalez-Perez, A., Sabarinathan, R. & Lopez-Bigas, N. Local determinants of the mutational landscape of the human genome. *Cell* **177**, 101–114 (2019).
62. Koren, A. et al. Differential relationship of DNA replication timing to different forms of human mutation and variation. *Am. J. Hum. Genet.* **91**, 1033–1040 (2012).

Methods

YFISH. A detailed step-by-step YFISH protocol is available at Protocol Exchange (https://doi.org/10.21203/rs.3.pex-570/v1). Briefly, we performed in situ restriction using either 10 µl of HindIII-HF (NEB, cat. no. R3104S) or 8 µl of MboI (NEB, cat. no. R0147M) in 400 µl at 37 °C for different durations, ranging from 1 min to 30 min, in the case of MboI, and 6 h in the case of HindIII. We stopped the reaction by placing the samples in ice-cold 1 × PBS/50 mM EDTA/0.01% Triton X-100 and washing them multiple times on ice. Afterwards, we dephosphorylated the samples by incubating them in 400 µl of 1 × calf intestinal alkaline phosphatase buffer containing 6 µl of calf intestinal alkaline phosphatase (Promega, cat. no. M1821) for 2 h at 37 °C. Next, we ligated YFISH adapters at a final concentration of 0.2 µM in 300 µl of 1 × T4 DNA ligase buffer containing 36 µl of T4 DNA ligase (Thermo Fisher Scientific, cat. no. EL0014), by incubating the samples for 18 h at 16 °C. The next day, we washed unligated adapters by incubating the samples in 10 mM Tris-HCl/1 M NaCl/0.5% Triton X-100, pH 8, five times for 1 h each at 37 °C while shaking. To prepare the hybridization mix, we diluted the labeled oligonucleotide to 200 nM in a hybridization buffer containing 2 × SSC/25% formamide/10% dextran sulfate/1 mg ml^{-1} *E. coli* tRNA/0.02% bovine serum albumin (BSA). We placed the coverslips onto a piece of Parafilm, with cells facing a 300-µl droplet of hybridization mix, and incubated the samples in a humidity chamber for 18 h at 30 °C. The next day, we washed the samples in washing buffer containing 2 × SSC/25% formamide for 1 h at 30 °C. Finally, we incubated the samples in 2 × SSC/25% formamide/0.1 ng µl^{-1} Hoechst 33342 (Thermo Fisher Scientific, cat. no. H3570) for 30 min at 30 °C, rinsed them twice in 2 × SSC and mounted them in ProLong Gold Antifade Mountant (Thermo Fisher Scientific, cat. no. P36930) before imaging. We imaged all the samples using either wide-field epifluorescence microscopy or STED microscopy, as described in the Supplementary Methods.

GPSeq. A detailed step-by-step GPSeq protocol is available at Protocol Exchange (https://doi.org/10.21203/rs.3.pex-570/v1). Briefly, we digested DNA, ligated the GPSeq adapters and washed unligated adapters using the same procedure as described above for YFISH. We then scraped the cells off the coverslips and digested them in 110 µl of 10 mM Tris-HCl/100 mM NaCl/50 mM EDTA/1% SDS, pH 8, containing 10 µl of Proteinase K (NEB, cat. no. P8107S), for 18 h at 56 °C. The next day, we inactivated the enzyme by increasing the temperature to 96 °C for 10 min. We purified gDNA using phenol-chloroform extraction and precipitated gDNA using glycogen (Sigma, cat. no. 10901393001) and sodium acetate, pH 5.5 (Life Technologies, cat. no. AM9740) in ice-cold ethanol (VWR, cat. no. 20816.367) for 18 h at −80 °C. We resuspended the DNA pellets in 100 µl of TE buffer and sonicated them in a Bioruptor Plus machine with the following settings: 30 s ON, 90 s OFF, high mode, 16 cycles. Afterwards, we concentrated gDNA down to a final volume of 8 µl in nuclease-free water, using AMPure XP (Beckman Coulter, cat. no. A63881). We performed in vitro transcription on each sample separately, with the MEGAscript T7 Transcription Kit (Thermo Fisher Scientific, cat. no. AM1334-5), using the same amount of gDNA (between 50 and 300 ng; see Supplementary Table 2) for each sample in a final volume of 20 µl and incubating the samples for 14 h at 37 °C. After in vitro transcription, we added 1 µl of DNAse I (Thermo Fisher Scientific, cat. no. AM2222) to the sample and incubated it for 15 min at 37 °C. We then purified the RNA with Agencourt RNAClean XP beads (Beckman Coulter, cat. no. A63987). Lastly, we prepared sequencing libraries using the TruSeq Small RNA Library Preparation Kit (Illumina, cat. no. RS-200-0012), following the manufacturer's instructions with some modifications, as described in the step-by-step protocol. We sequenced all the libraries on the NextSeq 500 system (Illumina) using the NextSeq 500/550 High Output v2 kit (75 cycles) (Illumina, cat. no. 20024906).

GPSeq score calculation. First, we pre-processed the sequencing data using a custom pipeline (*gpseq-seq-gg*) featuring quality control, read filtering based on the expected adapter sequence, adapter trimming, mapping, filtering of the mapping output, filtering of reads mapped away from restriction sites and unique molecular identifier (UMI)-based read de-duplication (Supplementary Methods). Summary statistics of the pipeline output are available in Supplementary Table 2. We discarded restriction sites (AAGCTT in Exp.1 and 2 with HindIII and GATC in Exp. 3 and 4 with MboI) associated with an abnormally high number of de-duplicated UMIs for a given digestion time (that is, condition), by identifying outliers with a chi-squared method and a significance of 0.01. We then binned the genome using either 1-Mb overlapping windows sliding in steps of 100 kb (1-Mb resolution) or nonoverlapping 100-kb windows (100-kb resolution). For each condition, we considered all the restriction sites that had been cut, to calculate a digestion probability, based on which we calculated the GPSeq score. We generated a BED-like file containing the GPSeq score per window and masked it based on a manually curated mask of repetitive and low-complexity regions (Supplementary Table 7). To be able to compare different experiments, we rescaled the calculated GPSeq score. More details on the actual GPSeq score calculation and rescaling are available in Supplementary Note 1. The algorithm is implemented in the *gpseqc_estimate* script, which is part of the *gpseqc* Python3 package, available at http://github.com/ggirelli/gpseqc. This analysis was implemented as a snakemake flow[63], available at http://github.com/ggirelli/gpseqc-snakemake. To average the GPSeq score across different experiments, we first averaged the score of each window across the experiments and then calculated the $\log_2$ of these averages and rescaled it again, as explained in Supplementary Note 1.

Generation of 3D genome structures. We started by generating a contact probability matrix A using Hi-C data previously obtained using HAP1 cells (experiment 4DNFI1E6NJQJ from ref. [64]), following the procedure described in ref. [65] with the following exceptions: (1) we did not use any low-pass filtering of the input data; (2) we corrected for the presence of the t(9;22) (q34;q11.2) translocation in HAP1 cells; and (3) after KR normalization, we handled the outliers on the first-off diagonal by shifting back values outside the interval $(\mu \pm 2\sigma)$, where μ is the mean value of the first diagonal and σ is the standard deviation of the first diagonal (per chromosome). This heuristic removed some of the streaks (strong horizontal and vertical lines) that otherwise were introduced by the pre-processing described in ref. [65]. We then generated populations of putative single-cell 3D genome structures using a custom software, namely *chromflock* (https://github.com/elgw/chromflock/), which we designed to emulate the state-of-the-art PGS package[43] as much as possible. PGS features a deconvolution step in which the input Hi-C data is deconvolved into individual (one per structure) binary contact-indication matrices, which resemble single-cell Hi-C contact maps. However, we could not apply PGS to our GPSeq data from haploid HAP1 cells, because this method was designed for diploid cell lines only. Moreover, the PGS package does not directly allow integration of data obtained with complementary assays, such as Hi-C and GPSeq, into the simulations. We implemented *chromflock* in the C99 programming language and executed from bash script using GNU Parallel. We created the 3D renderings for this paper using Chimera[66] unless otherwise stated. The main input to *chromflock* is a $N \times N$ contact probability matrix A, where N is the number of beads and each element A_{ij} specifies the probability of bead i being in contact with bead j. A label vector L has to be supplied, where the value of L_i specifies to which chromosome the bead i belongs. The label vector is necessary for the compression heuristics described below (also employed in PGS) and also allows *chromflock* to output Chimera (cmm) files, where chromosomes are labeled with individual colors. We denote the number of structures to be generated by S. For simulations, we converted the GPSeq score into radius g or distance from the nucleus center:

$$g = 1 - \log_2(\text{GPSeq score}) \quad (1)$$

Finally, we shifted the values falling outside of the [0, 1] interval to the closest boundary. The geometry of the simulations, corresponding to the nucleus interior, is the unit sphere. We set the radius of the N beads, R_b, so that the beads occupy 20% of the volume of the sphere (volume quotient, $V_q = 0.2$):

$$R_b = \sqrt[3]{V_q / N} \quad (2)$$

The calculations in *chromflock* are divided into epochs, which are assignment steps followed by molecular dynamics simulations. Initially, each structure, s, has an empty contact-indication matrix $W^{(s)}$. At the beginning of each epoch, contacts are assigned to structures in the population, and then the beads coordinates are updated using molecular dynamics. To determine in which epoch a contact should be introduced to the structures, we use a list, $\theta = (\theta_1 = 1, \theta_2, \theta_3, \ldots)$. In the i-th epoch, the contacts for which $\theta_{i-1} \leq A_{ij} < \theta_i$ are assigned to round$(S \times A_{ij})$ structures. During the first epoch, the contacts where $A_{ij} = 1$ are used. The assignment step is responsible for enforcing restraints to the individual structures, S (that is, to create and update their contact-indication matrices):

$$W^{(s)}, s = 1, \ldots, S \quad (3)$$

Initially, the assignment protocol generates the W matrices, one for each structure, by including all the contacts where $A = 1$ (that is, contacts that bind adjacent beads physically together and which should be present in all structures). At each subsequent epoch, new contacts are introduced in the structures as described above. Typically, each epoch iterates several times to allow constraints that cannot be satisfied to move to other structures (that is, if beads i and j are set to be close in structure s ($W^{(s)}_{i,j} = 1$) but they are not physically close in structure s, that constraint is removed and assigned to the most fit structure). Each time an epoch is re-iterated, the contacts $W_{i,j}$ are reset, where $\theta_{i-1} \leq A_{i,j} < \theta_i$. Contacts are always assigned to the most fit structures. In other words, when $k = \text{round}(S \times A_{ij})$ contacts between beads i and j are being assigned to the S structures, they will be given to the k structures, which already have the smallest distance between beads, i and j (that is, the k structures where $||X_i - X_j||$ is minimal). The molecular dynamics step of each epoch uses the Verlet integration scheme to solve the Langevin equation. When a structure is initialized, the positions of the beads are taken randomly from a uniform distribution over the simulation domain. In subsequent runs, the simulation continues from the last coordinates. The forces field consists of:

1. F_v, which enforces steric hindrance (that is, volume exclusion) to preclude beads from occupying the same volume or overlap;
2. F_s, which keeps the beads inside the simulation domain (unit sphere);
3. F_a, which makes the beads attract one another (if that is specified by W);
4. F_c, which models a chromosome compression force used at the first epoch. This heuristic is suggested in ref. [43] and helps distribute the contact constraints more evenly between the structures;

5. F_b, which encodes a Brownian force, simulating the net effect of smaller molecules, which are not modeled explicitly:

$$F_b(i) = c_b s \quad (4)$$

where s is drawn from an isotropic 3D Gaussian with $\sigma = 1$ using the highly efficient method by McFarland[67];

1. F_d, a drag force defined by the viscosity η, which is proportional to the velocity of each bead and models viscosity:

$$F_d(i) = -\eta v(i) \quad (5)$$

F_v, F_s, F_a and F_c are defined in terms of their potential function or error as follows:

1. $E_v(i, j)$ is the volume exclusion potential that keeps the beads from overlapping and is set equal to $c_v(d_{i,j} - 2R_b)^2$ if $d_{i,j} < 2R_b$ or otherwise equal to 0. We set the distance between beads i and j, $d_{ij} = ||X_i - X_j||$, and the radius of bead i, $r_i = ||X_i||$;
2. $E_s(i)$ is the potential that keeps the beads inside the nuclei and is set equal to $c_s(r_i + R_b - R_s)^2$ if $r_i > R_s - R_b$ or otherwise equal to 0;
3. $E_a(i, j)$ is the potential that keeps beads attracted to each other and is set equal to $c_a(d_{i,j} - R_c)^2$ if $W_{i,j} = 1$ and $d_{ij} > R_c$ or otherwise equal to 0;
4. $E_c(i)$ is the compression potential and is set equal to $c_c||X_i - m_k||^2$ when bead i belongs to chromosome k, where m_k is the center of mass of chromosome k;
5. $E_r(i, j)$ is the potential for radial preference and is set equal to $c_r(r_i - g_i)^2$ if g_i is finite or otherwise equal to 0. We use non-finite values to indicate that no radial preference is set.

We let the volume exclusion force vary with time as:

$$F_v(x) = \frac{1}{2}(1 + erf(\beta(p - 0.5))) \quad (6)$$

where we set $\beta = 5$ and p is the proportion of iterations taken (that is, $p \in [0, 1]$). Hence, the total error is:

$$E = \sum_i^N (E_s(i) + E_c(i) + E_r(i)) + \sum_i^N \sum_j^N (E_v(i, j) + E_a(i, j)) \quad (7)$$

and the total forces are:

$$F = \nabla E + F_b + F_d \quad (8)$$

We derived an analytical expression for ∇E, which has been verified against the numerical gradient. We used cell lists to speed up the calculation of E_v, which otherwise would be $O(N^2)$. Unless otherwise stated, we used 10,000 structures (S = 10,000) and binned the genome in nonoverlapping 1-Mb bins. We excluded chromosome Y from the analysis as done in ref. [65]. The list of theta values we used is: 1, 0.2, 0.1, 0.5, 0.02, 0.01 and 0.001 (that is, we used seven epochs). The theta values are the same as used in PGS; however, we included 0.001 to use a larger proportion of the inter contacts. Furthermore, we used $c_v = 1$, $c_v = 1c_s = 1$, $c_a = 1$, $R_s = 1$, $V_q = 0.2$, $R_c = (2.1 + 0.9p)R_b$ and $\eta = 0.5$. When GPSeq data are used, we set $c_r = 0.005$; otherwise we set $c_r = 0$. We ran each epoch for three cycles of reassignments. We used 7,000 time steps in the molecular dynamics. To generate structures at 100-kb resolution, the parameters that we used for the 1-Mb-resolution structures did not yield distinct chromosome territories. Hence, we added a compression stage of the chromosomes at each epoch, instead of just the first one. We then ran 8,000 iterations without any compression to relax the structures.

Statistical analyses. We conducted all statistical analyses in the R software environment (v3.5.1, https://www.r-project.org).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Source data for Figures, Extended Data Figures, Supplementary Figures, Supplementary Tables and Supplementary Notes are available at https://github.com/ggirelli/GPSeq-source-data. The following GPSeq data have been deposited in the GEO Repository GSE135882:

1. Raw and pre-processed GPSeq sequencing data
2. Bead coordinates for *chromflock*-generated whole-genome structures
3. Genome-wide GPSeq scores at chromosome-wide, 1-Mb and 100-kb resolution
4. GPSeq scores in genomic windows centered on the midpoint of the DNA FISH probes shown in Supplementary Fig. 1a, at 1-Mb and 100-kb resolution

Previously published data sets used in the analyses, for which accession numbers are available, are described in Supplementary Table 5. For SNPs, tumor SNVs and gene fusions, we used the following data sets:

1. Chronic lymphocytic leukemia, lung cancer, prostate cancer and melanoma SNVs were obtained from the supplementary tables of the corresponding papers described in ref. [48]
2. SNPs from the 1000 Genomes Project Phase 3 were downloaded from https://www.internationalgenome.org/
3. TCGA gene fusions were downloaded from https://www.tumorfusions.org/

Code availability

The following code was used and is available at the indicated links:

1. *pygpseq*: https://github.com/ggirelli/pygpseq/releases/tag/v3.3.4
2. *pygpseq-scripts*: https://github.com/ggirelli/pygpseq-scripts/releases/tag/v0.0.1
3. *iFISH-singleLocus-analysis*: https://github.com/ggirelli/iFISH-singleLocus-analysis/releases/tag/v1.0
4. *gpseq-seq-gg*: https://github.com/ggirelli/gpseq-seq-gg/releases/tag/v2.0.3
5. *bed-fix-chrom-rearrangement*: https://github.com/ggirelli/bed-fix-chrom-rearrangement/releases/tag/v0.0.1
6. *gpseqc*: https://github.com/ggirelli/gpseqc/releases/tag/v2.3.6.post1
7. *gpseqc-snakemake*: https://github.com/ggirelli/gpseqc-snakemake/releases/tag/v1.0
8. *bioTrackBinner*: https://github.com/ggirelli/bioTrackBinner/releases/tag/v0.0.1
9. *ggkaryo2*: https://github.com/ggirelli/ggkaryo2/releases/tag/v0.0.3
10. *chromflock*: https://github.com/elgw/chromflock/releases/tag/0.1

References

63. Köster, J. & Rahmann, S. Snakemake-a scalable bioinformatics workflow engine. *Bioinformatics* **34**, 3600 (2018).
64. Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci. USA* **112**, E6456–E6465 (2015).
65. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* **30**, 90–98 (2011).
66. Pettersen, E. F. et al. UCSF Chimera-a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
67. McFarland, C. D. A modified ziggurat algorithm for generating exponentially- and normally-distributed pseudorandom numbers. *J. Stat. Comput. Simul.* **86**, 1281–1294 (2016).

Acknowledgements

We thank A. van Oudenaarden (Hubrecht Institute) for initial discussions on GPSeq data analysis and I. Solovei (UMC Munich), M.A. Marti-Renom (CRG Barcelona), S.L. Klemm (Stanford) and B. Bouwman (Bienko-Crosetto lab) for critically reading the manuscript and providing ideas. We thank L. Xu and R. Mirzazadeh (Bienko-Crosetto lab) for helping with FISH probe production. We acknowledge H. Blom at the Advanced Light Microscopy facility at the Science for Life Laboratory (SciLifeLab) for acquiring and processing STED images and for providing computing resources. We acknowledge the van Steensel laboratory for providing HAP1 lamin DamID data generated in the frame of the 4D Nucleome project. This work was supported by a postdoctoral scholarship from the Swedish Society for Medical Research to E.W.; by funding from the Swedish Research Council (2018-02950), the Swedish Cancer Research Foundation (CAN 2018/728), the Ragnar Söderberg Foundation (Fellows in Medicine 2016) and the Strategic Research Programme in Cancer (StratCan) at the Karolinska Institutet to N.C.; and by funding from the Science for Life Laboratory, the Karolinska Institutet KID Funding Program, the Swedish Research Council (621-2014-5503), the Human Frontier Science Program (CDA-00033/2016-C), the Ragnar Söderberg Foundation (Fellows in Medicine 2016) and the European Research Council under the European Union's Horizon 2020 research and innovation program (StG-2016_GENOMIS_715727) to M.B.

Author contributions

Conceptualization: J.C., T.K., G.G., F.A., B.S., E.W., A.v.O, N.C. and M.B.; data curation: G.G. and F.A.; formal analysis: G.G., F.A., E.W., B.S. and J.C.; funding acquisition: M.B. and N.C.; investigation: T.K., J.C., S.K. and M.B.; methodology: J.C., T.K., G.G., N.C. and M.B.; project administration: M.B. and N.C.; resources: SciLifeLab, H.B., L.X. and R.M.; software: G.G. and E.W.; supervision: M.B. and N.C.; validation: J.C., T.K., A.M., S.K., E.G., L.X., R.M., G.G., F.A., E.W., M.B. and N.C.; visualization: G.G., F.A., J.C., M.B. and N.C.; writing: M.B, N.C., G.G., J.C., T.K., F.A. and S.K.

Competing interests

The authors declare no competing interests.

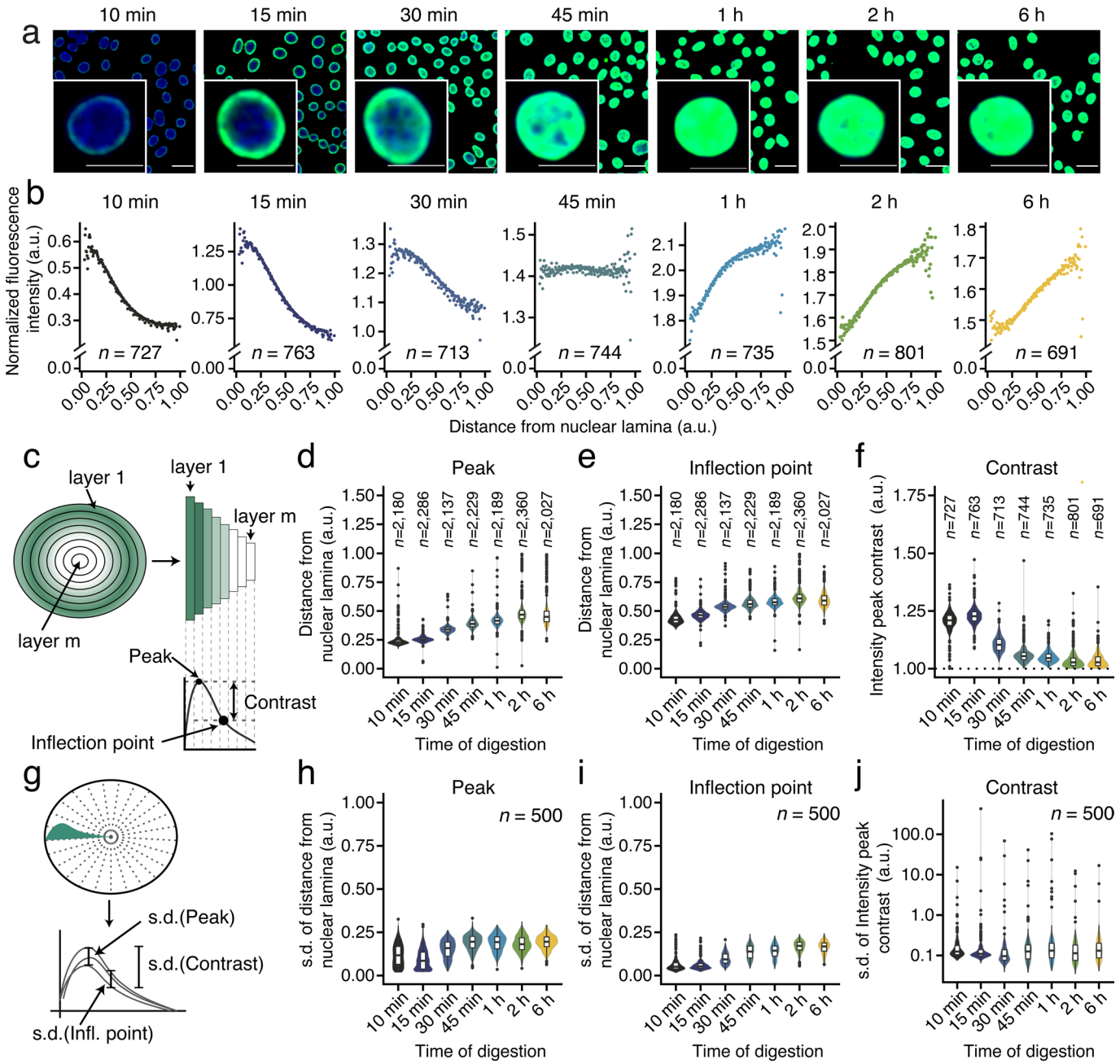
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41587-020-0519-y.

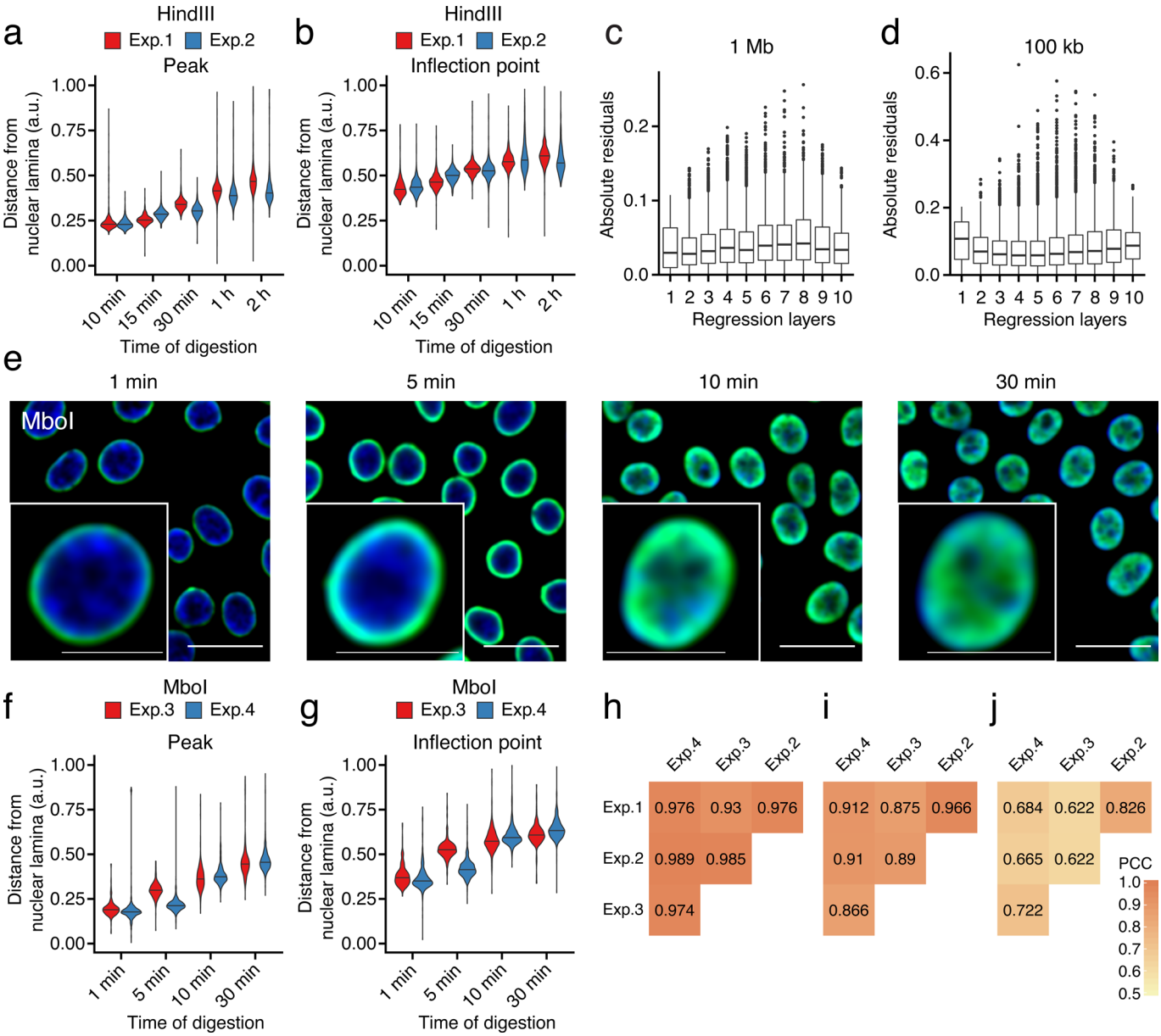
Supplementary information is available for this paper at https://doi.org/10.1038/s41587-020-0519-y.

Correspondence and requests for materials should be addressed to N.C. or M.B.

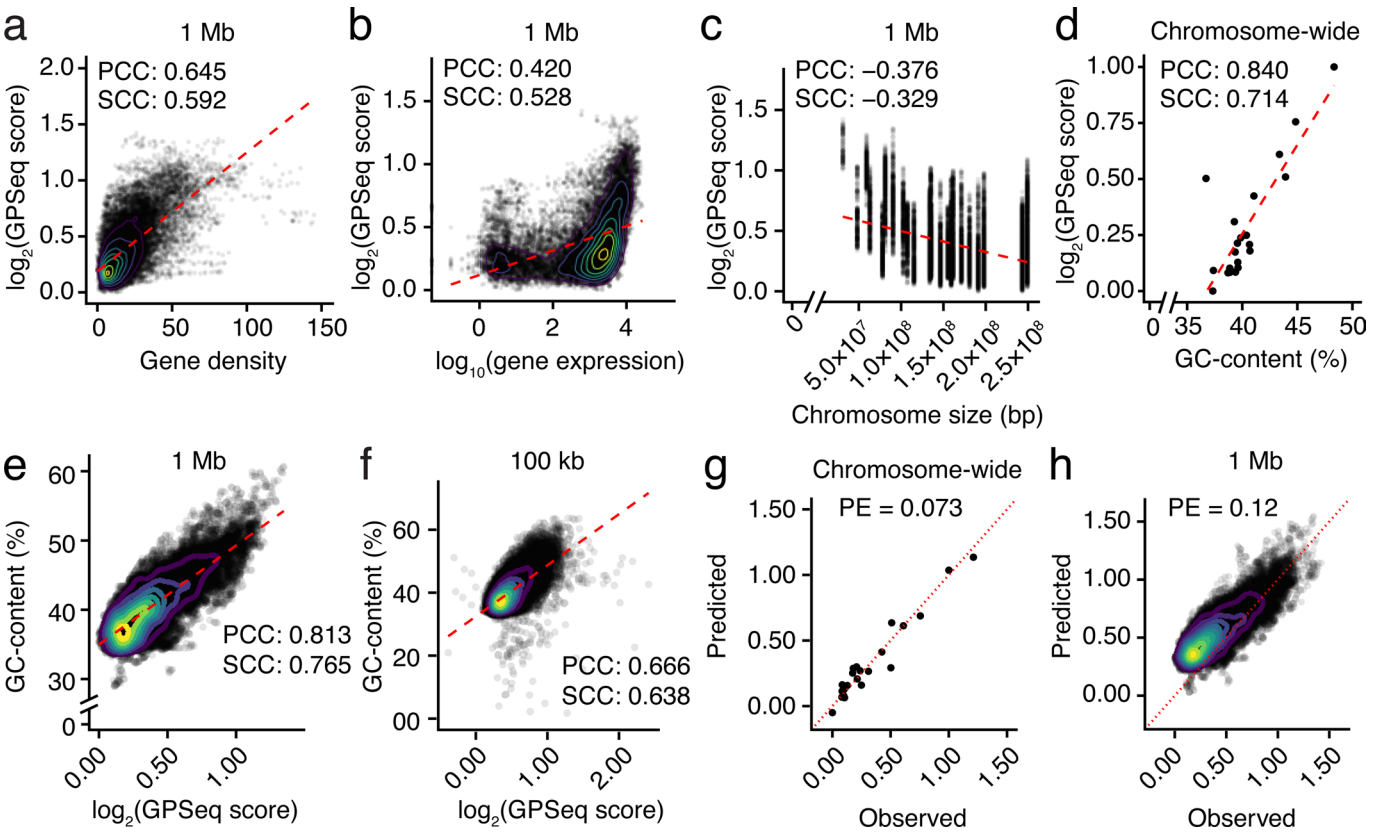
Reprints and permissions information is available at www.nature.com/reprints.



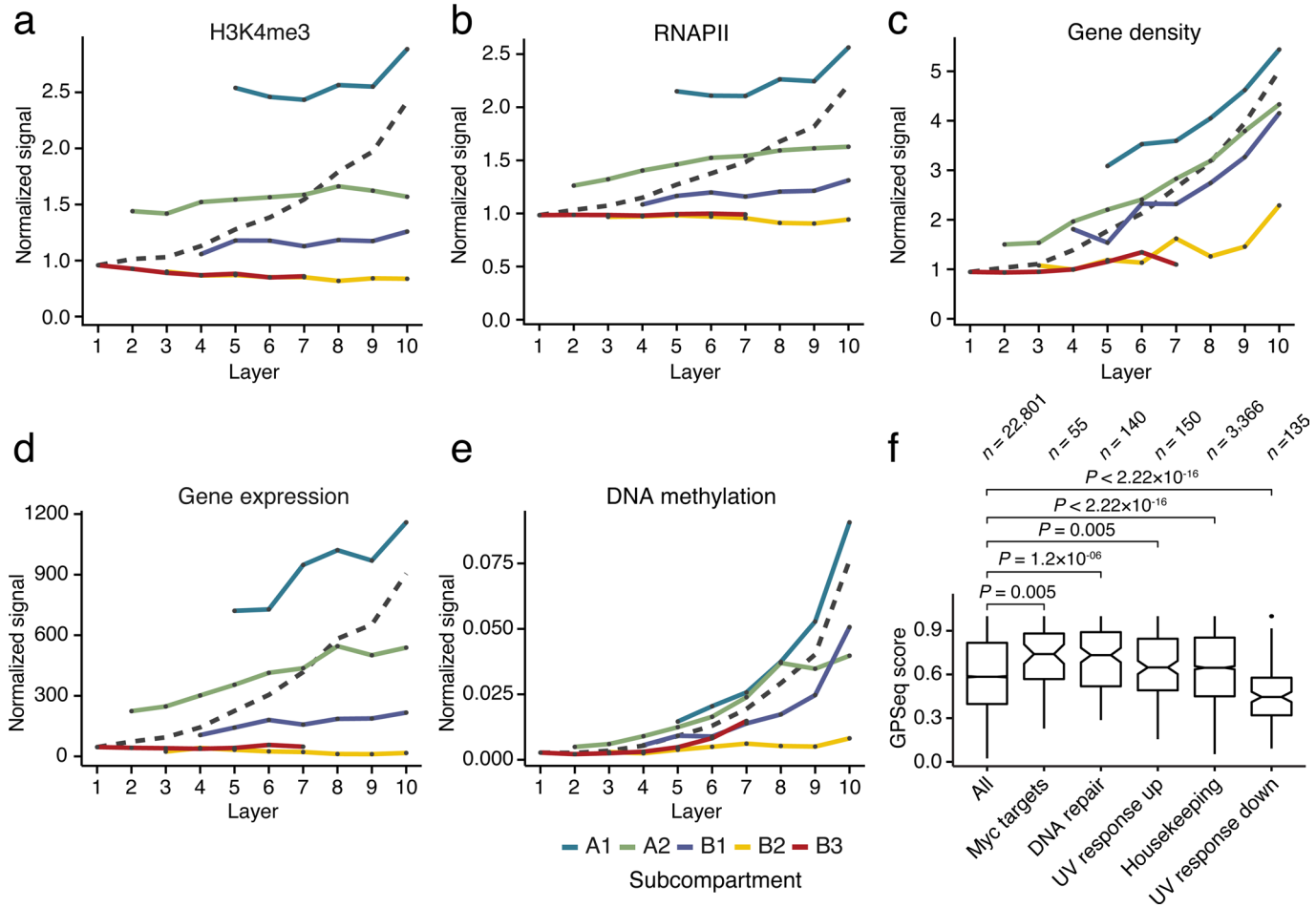
Extended Data Fig. 1 | Monitoring gradual gDNA restriction by YFISH. (**a**) Gradual gDNA digestion with HindIII revealed by wide-field epifluorescence microscopy. Green: HindIII cut sites. Blue: DNA stained with Hoechst 33342. Scale bars: 20 μm (field-of-view) and 10 μm (insets). Times indicate the duration of incubation with HindIII. Mid optical sections are shown. The same dynamic range was used for each digestion time. The experiment was repeated twice with similar results. (**b**) Normalized YFISH fluorescence intensity at various distances from the nuclear lamina, for each of the times shown in (a). The YFISH signal was normalized over the fluorescence intensity of DNA stained with Hoechst 33342. Each dot represents the median intensity in one of 200 radial layers. *n*, number of cells analyzed. (**c**) Calculation of YFISH signal inter-cellular variability. Top: each nucleus is divided in *m* concentric layers of equal thickness and the mean fluorescence intensity per layer is calculated. Bottom: for each restriction time, the peak, inflection point, and contrast are calculated from the distribution of the mean fluorescence intensity in all the nuclei. (**d-f**) Distributions of the peak position (d), inflection point position (e), and peak contrast (f) at various digestion times, for the samples of which (a) are representative images. *n*, number of nuclei analyzed as described in (c). (**g**) Calculation of YFISH signal intra-cellular variability. Top: 200 radii (as exemplified by the dotted lines) are randomly drawn inside each 3D segmented nucleus and the YFISH intensity profile (green) is evaluated at 100 points (as exemplified by the dotted lines) evenly spaced along each radius. Bottom: the standard deviation (s.d.) of the positions of the peak and inflection point and of the peak contrast are calculated from all the YFISH signal profiles from the same nucleus. (**h–j**) Distributions of the standard deviation (s.d.) of the peak position (h), inflection point position (i), and peak contrast (j) at various digestion times, for the samples of which (a) are representative images. *n*, number of nuclei analyzed as described in (g). In all the violin plots in the figure, each box spans from the 25th to the 75th percentile and the whiskers extend from -1.5×IQR to +1.5×IQR from the closest quartile, where IQR is the inter-quartile range. Dots: outliers (data falling outside whiskers). All the source data for this figure are from HAP1 cells.



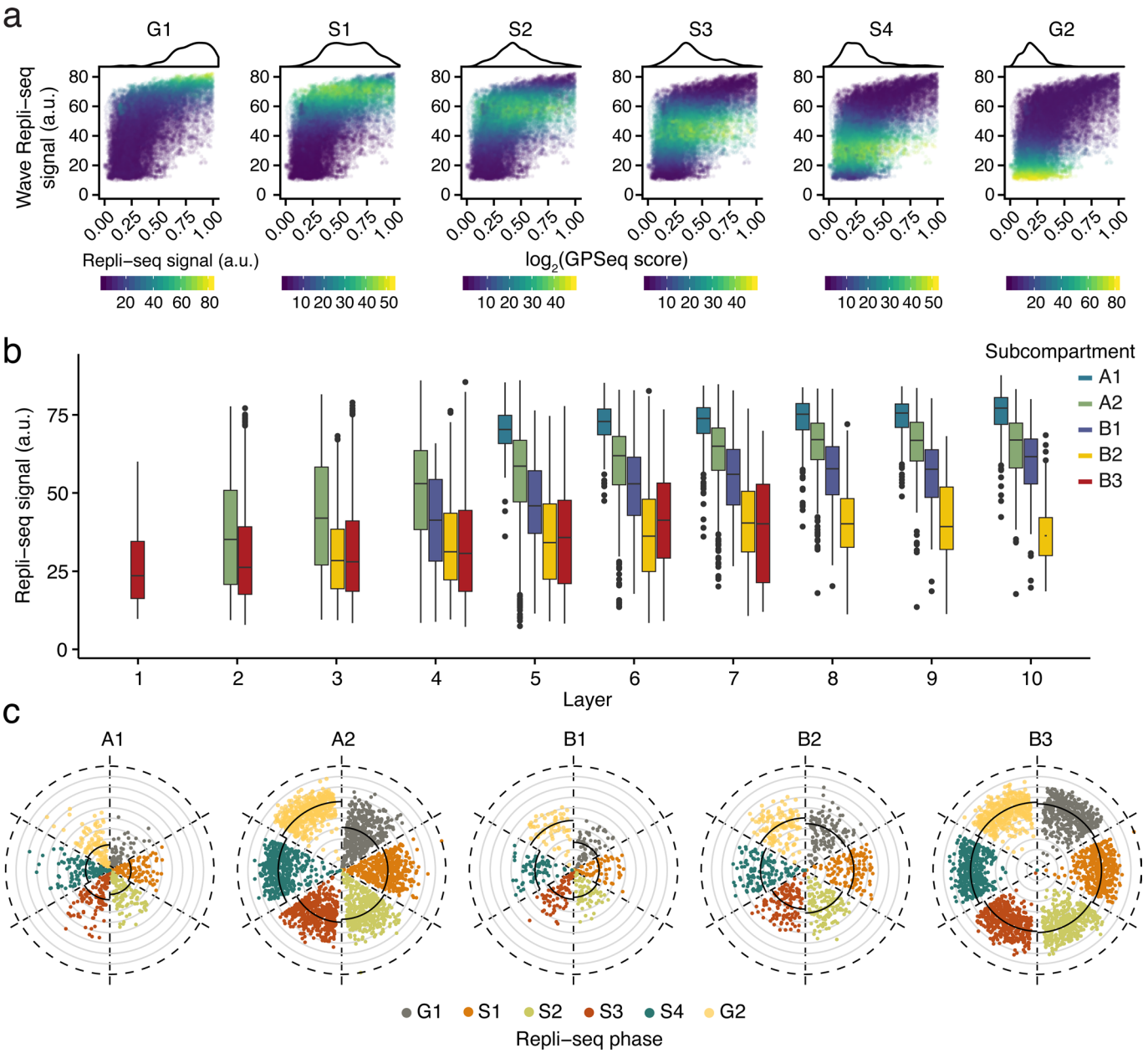
Extended Data Fig. 2 | Quantification of gradual gDNA restriction and GPSeq reproducibility. (**a**) Distribution of the position of the peak in the YFISH fluorescence intensity radial profile (see Extended Data Fig. 1c) at different restriction times, in two HindIII experiments (Exp.1 and 2). (**b**) Same as in (a), but for the position of the inflection point. (**c**) Distribution of the absolute residuals of the linear regression fitting between the log2 GPSeq score (1 Mb resolution, overlapping windows with 100 kb step size) in two HindIII experiments (Exp.1 and 2). The regression layers were generated by dividing the linear regression line into 10 bins of equal size. (**d**) Same as in (c) but correlating the GPSeq score at 100 kb resolution. All box plots in (c, d) span from the 25th to the 75th percentile and whiskers extend from -1.5×IQR to +1.5×IQR from the closest quartile, where IQR is the inter-quartile range. Dots: data falling outside whiskers. (**e**) Gradual gDNA digestion with MboI revealed by wide-field epifluorescence microscopy. Green: MboI cut sites. Blue: DNA stained with Hoechst 33342. Scale bars: 20 μm (field-of-view) and 10 μm (insets). Times indicate the duration of incubation with MboI. Mid optical sections are shown. The same dynamic range was used for all the digestion times. The experiment was repeated twice with similar results. (**f, g**) Same as in (a, b), but for MboI experiments (Exp.3 and 4). (**h**) Correlation between the GPSeq score in four GPSeq experiments at chromosome resolution (*that is*, using genomic windows of the size of each chromosome). (**i**) Same as in (h) but at 1 Mb resolution (overlapping windows, 100 kb step size). (**j**) Same as in (h) but at 100 kb resolution (non-overlapping windows). In all the violin plots in the figure, the median is shown as a black line and the violins extend from the min to the max value. Sample size information for (a-d), (f, g) and (i, j) is available in Supplementary Table 11. All the source data for this figure are from HAP1 cells.



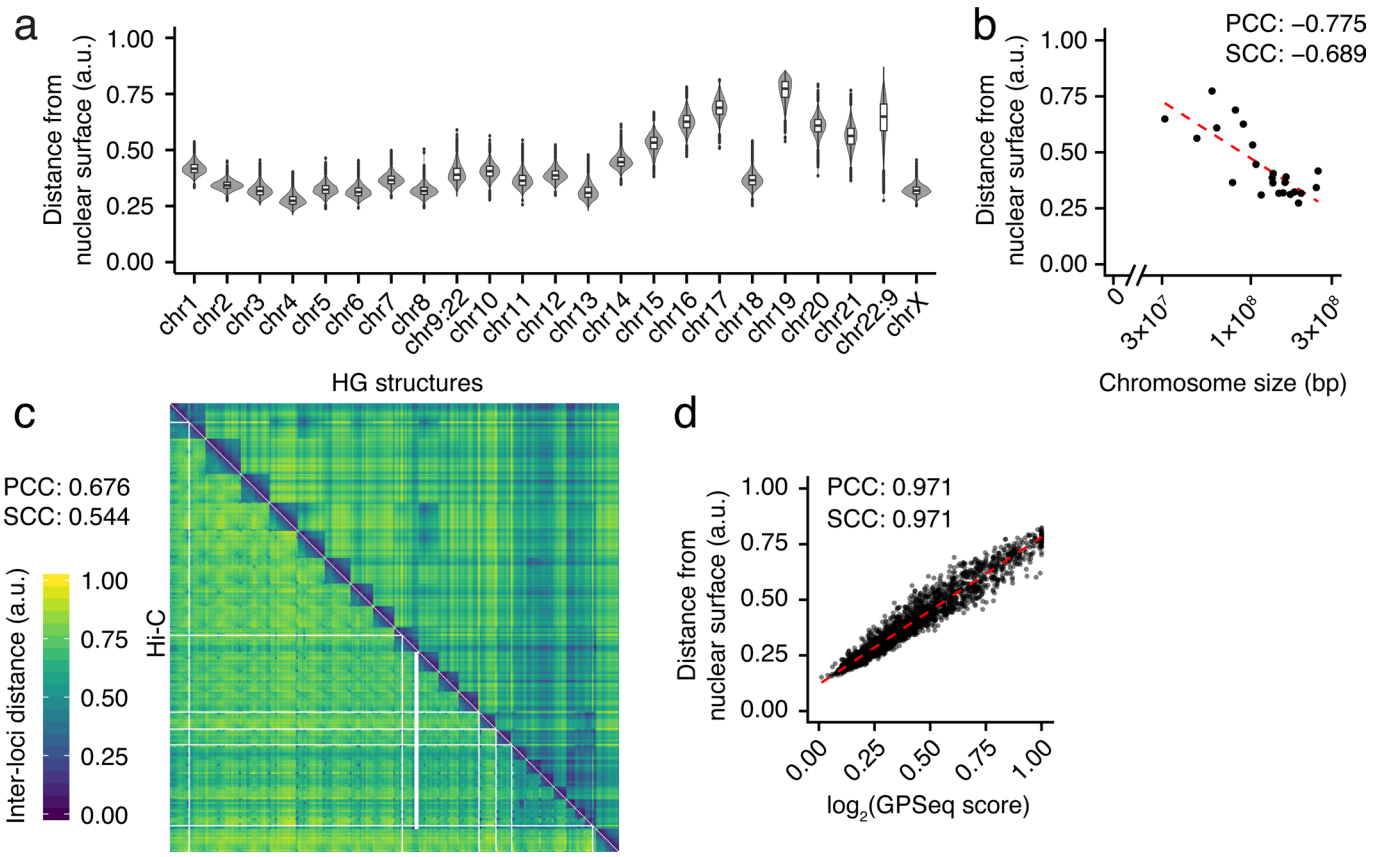
Extended Data Fig. 3 | Predictors of chromatin radiality. (**a**) Correlation between the log2 GPSeq score and the mean number of transcription start sites (TSS, one TSS per gene) at 1 Mb resolution (overlapping genomic windows, 100 kb step size). Each dot represents one out of 26,330 genomic windows analyzed. (**b**) Correlation between the log2 GPSeq score and the average RNA-seq reads count at 1 Mb resolution (overlapping genomic windows, 100 kb step size). Each dot represents one out of 26,330 genomic windows analyzed. (**c**) Correlation between the log2 GPSeq score (1 Mb resolution, overlapping genomic windows with 100 kb step size) and chromosome size in base-pairs (bp). Each dot represents a single 1 Mb genomic window. (**d**) Correlation between the log2 GPSeq score (chromosome resolution) and the median GC-content per Mb per chromosome. Each dot represents one chromosome. (**e**) Correlation between the log2 GPSeq score (1 Mb resolution, overlapping genomic windows with 100 kb step size) and the median GC-content per Mb per chromosome. Each dot represents a single 1 Mb window. $n = 25{,}026$ genomic windows (points) were analyzed. (**f**) Same as in (e) but at 100 kb resolution (non-overlapping windows). $n = 25{,}342$ genomic windows (points) were analyzed. (**g**) Predicted over observed chromosome-wide GPSeq score. The prediction is based on a multivariable model including both chromosome size and GC-content as described in the Methods. PE, prediction error. Dotted red line: bisector. Each dot represents one chromosome. (**h**) Same as in (g) but using 1 Mb overlapping genomic windows with 100 kb step and using GC-content, chromosome size, gene expression and gene density to model the GPSeq score. $n = 26{,}293$ genomic windows (points) were analyzed. In all the plots in the figure, PCC and SCC are the Pearson's and Spearman's correlation coefficient, respectively. Dashed red lines: linear regressions. All the source data for this figure are from HAP1 cells.



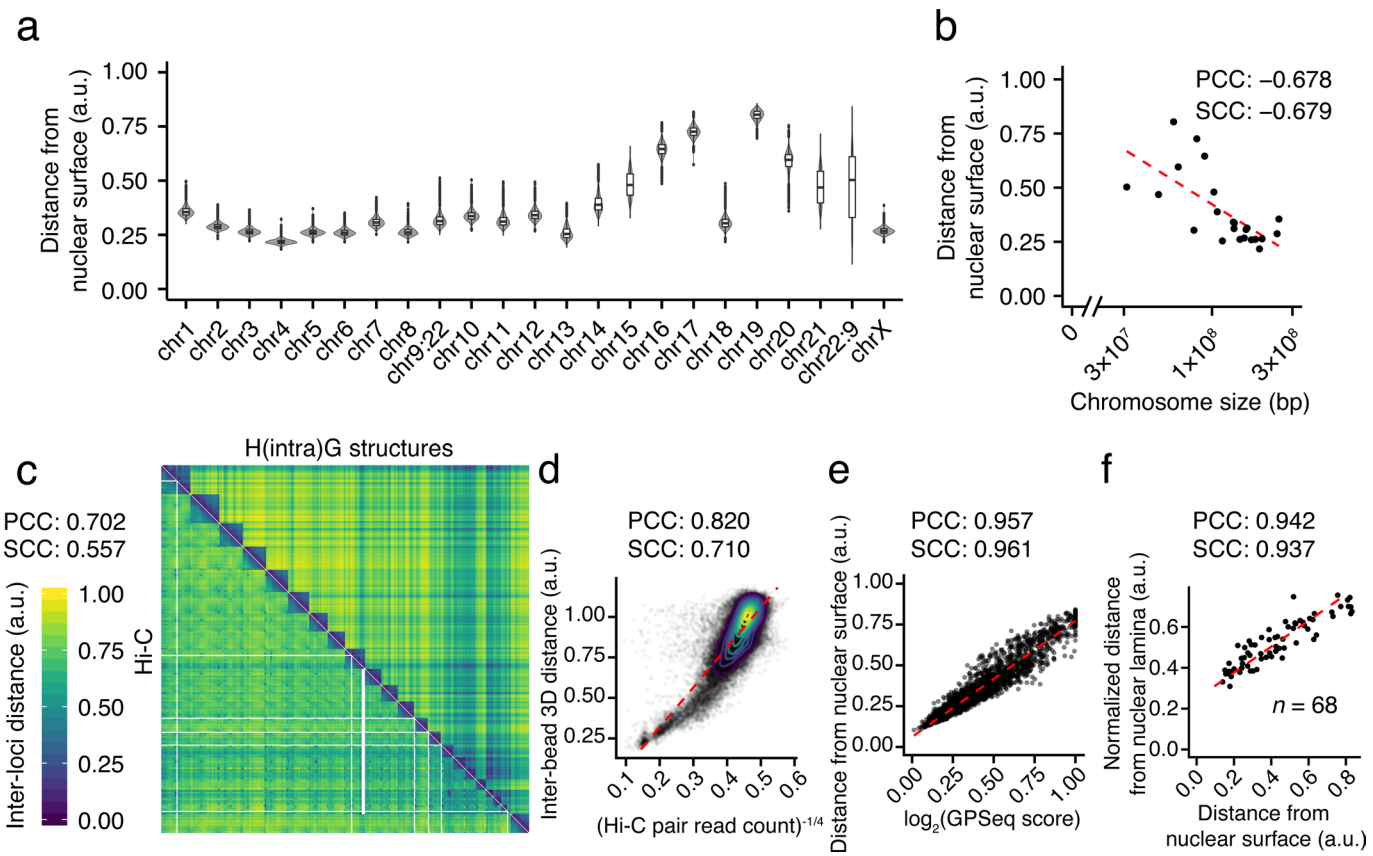
Extended Data Fig. 4 | Radial distribution of chromatin marks and features as well as gene expression. (**a–e**) Mean normalized signal of various chromatin features in ten concentric nuclear layers, divided by A/B subcompartments. Gene density was calculated as the mean number of transcription start sites (TSS, one TSS per gene) per 100 kb, and gene expression was calculated as the average RNA-seq reads count per 100 kb (Supplementary Methods). The dashed grey lines show the radial distribution of the features without dividing by subcompartment. (**f**) Distribution of the log2 GPSeq scores of all the genes and of each gene set pathway. *P*-values: Wilcoxon test, two-sided. *n*, number of genes. Box plots span from the 25th to the 75th percentile and whiskers extend from −1.5×IQR to +1.5×IQR from the closest quartile, where IQR is the inter-quartile range. All the source data for this figure are from HAP1 cells, except for DNA methylation data, which are from K562 cells.



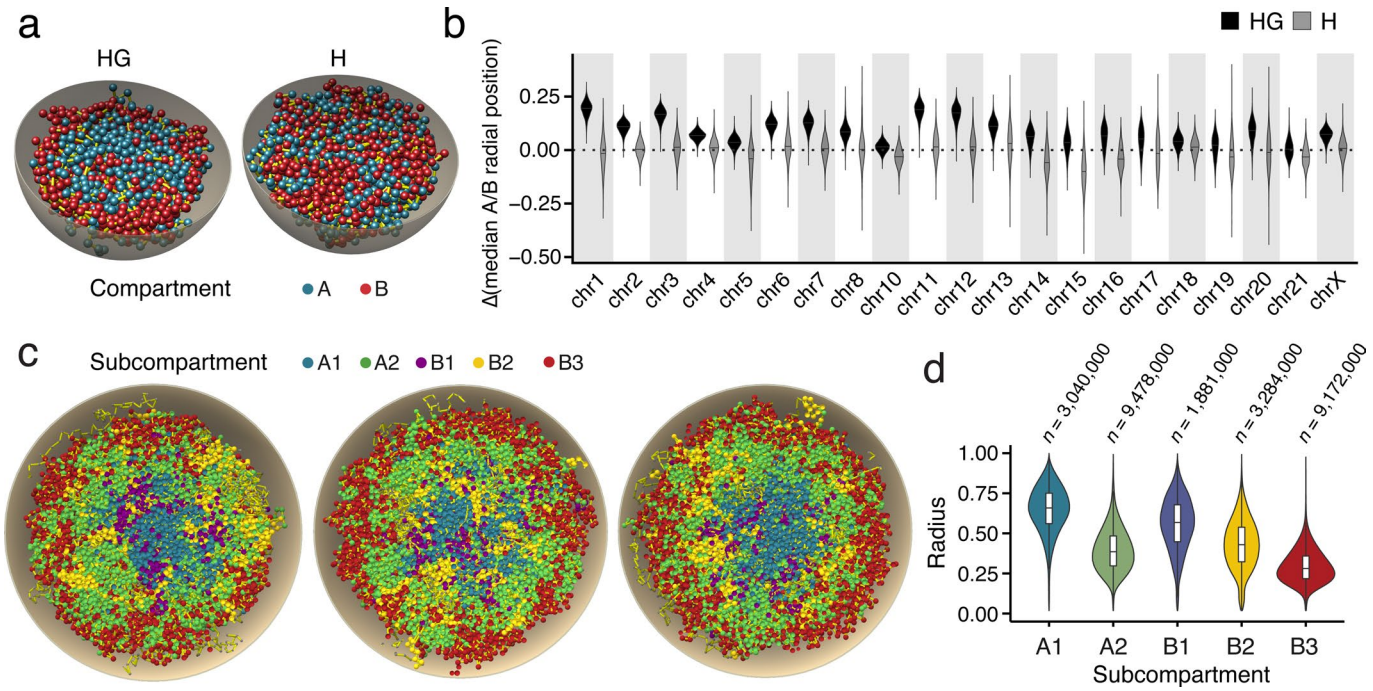
Extended Data Fig. 5 | Radial progression of DNA replication. (**a**) Correlation between the log2 GPSeq score and the Repli-seq signal after wavelet transformation, at 1 Mb resolution (overlapping genomic windows, 100 kb step size). Each dot represents a single 1 Mb genomic window out of 26,330 genomic windows (dots) analyzed. The dots are colored based on the cell cycle sub-phase (G1, S1-4, G2). The density distribution on top of each scatterplot corresponds to the density of the log2 GPSeq score of the 5% bins with the highest Repli-seq signal in the indicated sub-phase. (**b**) Distribution of the Repli-seq signal by A/B subcompartment type in ten concentric nuclear layers. In all the boxplots, each box spans from the 25th to the 75th percentile and whiskers extend from -1.5×IQR to +1.5×IQR from the closest quartile, where IQR is the inter-quartile range. Dots: outliers (data falling outside whiskers). (**c**) Repli-seq signal in 100 kb genomic windows (dots) radially arranged based on their GPSeq score, separately for each sub-phase and A/B subcompartment. Only the 5% bins with the highest Repli-seq signal in the indicated sub-phase are reported. Solid black lines indicate the mean in each sector. Dashed circles: nuclear lamina. Grey circles separate ten concentric nuclear layers. Sample size information is available in Supplementary Fig. 6f (b) and in Supplementary Table 11 (c). GPSeq source data for this figure are from HAP1 cells, while the Repli-seq data are from K562 cells.



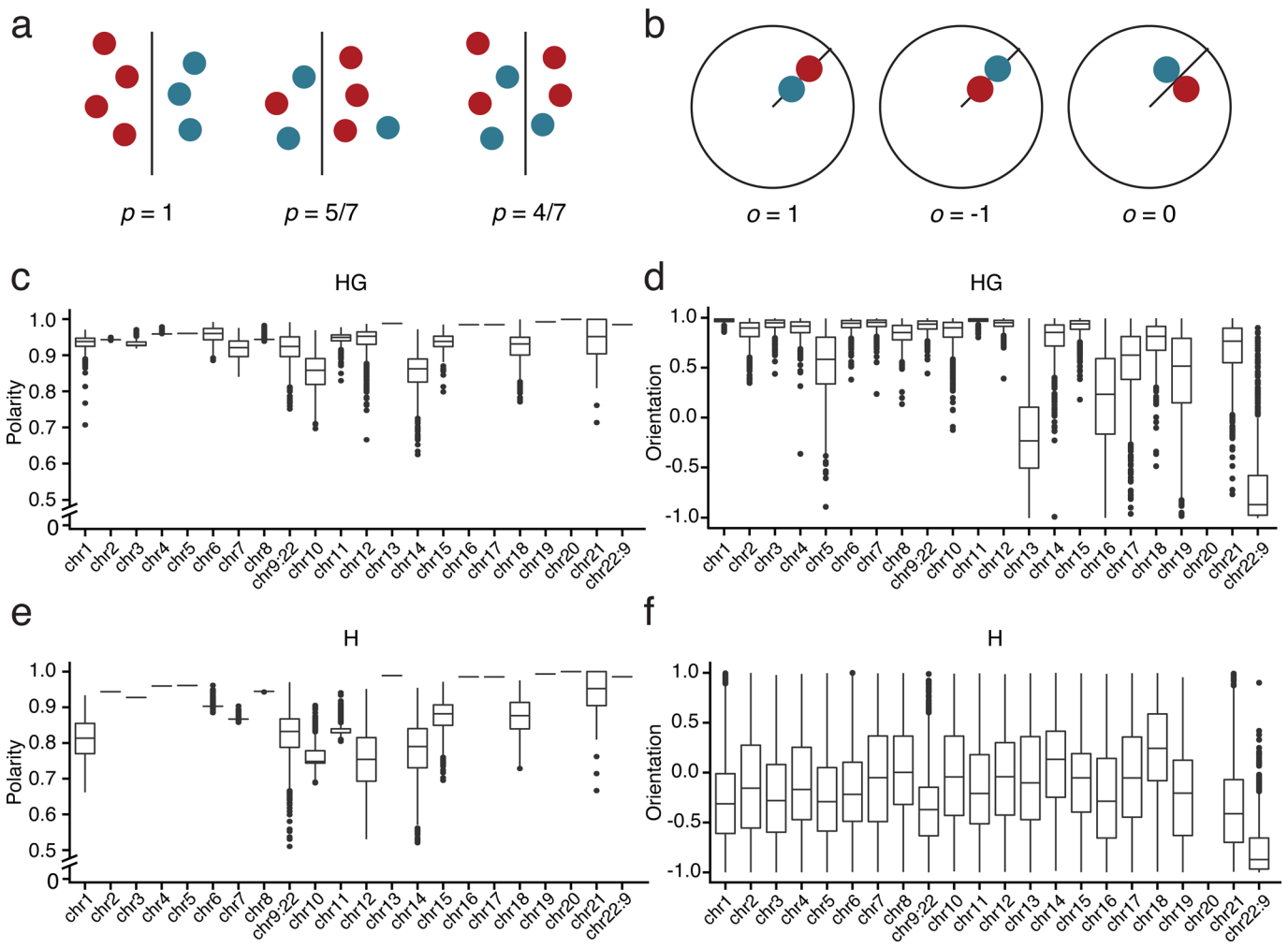
Extended Data Fig. 6 | Analysis of *chromflock* structures generated using both GPSeq and Hi-C data (HG structures). (**a**) Distribution of the average distance from the modeled nuclear surface of 1 Mb beads in 10,000 HG structures per chromosome. chr9:22 and chr22:9 are the derivative chromosomes of the t(9;22)(q34;q11.2) translocation. (**b**) Correlation between the average chromosome distance from the modeled nuclear surface in HG structures and chromosome size in base-pairs (bp). Each dot corresponds to one chromosome. (**c**) Distance matrix heatmap. The upper triangle shows the inter-bead 3D distances in HG structures. The bottom triangle shows the KR-normalized Hi-C contact frequency matrix, with each element raised to the power of −0.25. The reported correlation coefficients are for 1 Mb resolution, while the plot shows averaged values over 10 Mb genomic windows for simplicity. (**d**) Correlation between the distance from the modeled nuclear surface position of 1 Mb beads in HG structures, and the log2 GPSeq score of the corresponding windows. $n = 2{,}627$ genomic windows (points) were analyzed.



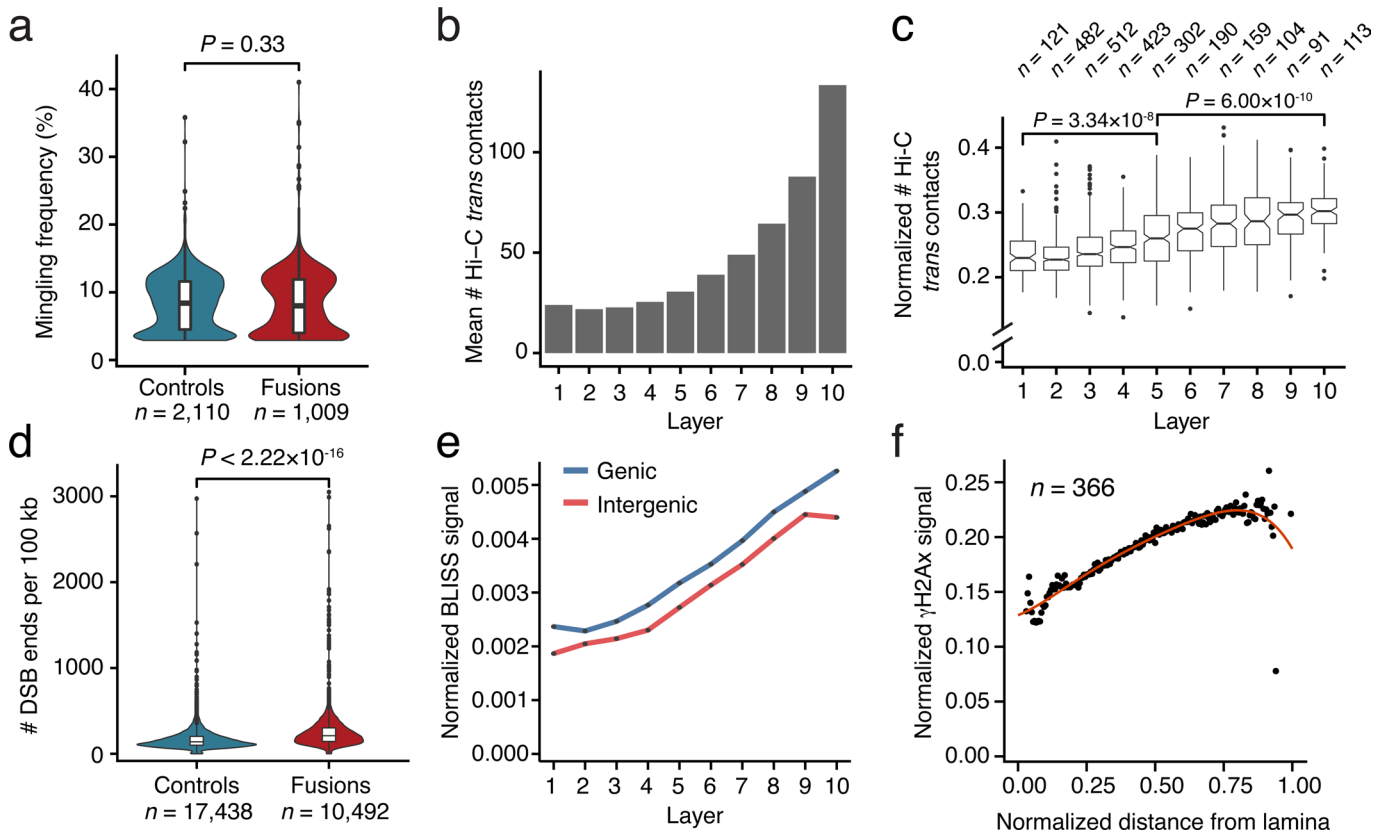
Extended Data Fig. 7 | Analysis of *chromflock* structures generated using GPSeq and Hi-C intra-chromosomal contacts only (H(intra)G). (**a**) Distribution of the average distance from the modeled nuclear surface of 1 Mb beads in 10,000 H(intra)G structures. chr9:22 and chr22:9 are the derivative chromosomes of the t(9;22)(q34;q11.2) translocation. (**b**) Correlation between the average chromosome distance from the modeled nuclear surface in H(intra)G structures and chromosome size in base-pairs (bp). Each dot corresponds to one chromosome. (**c**) Distance matrix heatmap. The upper triangle shows the inter-bead 3D distances in H(intra)G structures. The bottom triangle shows the KR-normalized Hi-C contact frequency matrix, with each element raised to the power of −0.25. The reported correlation coefficients are for 1 Mb resolution, while the plot shows averaged values over 10 Mb genomic windows for simplicity. (**d**) Correlation between the average inter-bead 3D distance in H(intra)G structures and the KR-normalized Hi-C contact frequency. Each dot represents a pair of 10 Mb non-overlapping genomic windows, each obtained by averaging 1 Mb non-overlapping bins. n = 47,531 genomic window pairs (points) were analyzed. Density contours are shown as concentric curves. (**e**) Correlation between the distance from the modeled nuclear surface position of 1 Mb beads in H(intra)G structures and the log2 GPSeq score of the corresponding windows. n = 2,627 genomic windows (points) were analyzed. (**f**) Correlation between the radial position in H(intra)G structures and the median 3D distance to the nuclear lamina measured by DNA FISH. Each dot represents one of the FISH probes (n = 68) shown in Supplementary Fig. 1a. In all the violin plots in the figure, each box spans from the 25th to the 75th percentile, whiskers extend from −1.5×IQR to +1.5×IQR from the closest quartile, where IQR is the inter-quartile range. Dots: outliers (data falling outside whiskers). In all the figure, PCC and SCC are the Pearson's and Spearman's correlation coefficient, respectively. Dashed red lines: linear regressions.



Extended Data Fig. 8 | Radial organization of A/B compartments and subcompartments in *chromflock* structures. (**a**) Examples of A/B arrangement in *chromflock* structures (1 Mb resolution) built using both GPSeq and Hi-C (HG) or only Hi-C (H) data. In all the structures, each bead represents a single 1 Mb genomic window. Elements connecting the beads are shown in yellow. The modeled nuclear surface is shown in grey. (**b**) Distribution of the difference in the median distance from the modeled nuclear surface of 1 Mb A-compartment beads *vs.* B-compartment beads per structure ($n = 10{,}000$) per chromosome (either for the HG or the H structures). Grey shades are used to visually distinguish different chromosomes. Sample size information is available in Source Data. (**c**) Examples of subcompartment arrangement in three out of 1,000 HG structures at 100 kb resolution. In all the structures, each bead represents a single 100 kb genomic window. The modeled nuclear surface is shown in grey. (**d**) Distribution of the distance to the modeled nuclear surface of the 100 kb beads belonging to different A/B subcompartments in 1,000 HG structures. *n*, number of beads belonging to each A/B subcompartment pooled from all the 1,000 structures. In all the violin plots in the figure, each box spans from the 25th to the 75th percentile, whiskers extend from –1.5×IQR to +1.5×IQR from the closest quartile, where IQR is the inter-quartile range.



Extended Data Fig. 9 | Polarity and orientation of A1 and B3 subcompartments in 100 kb-resolution *chromflock* structures. (**a**) Examples of possible arrangements of two subcompartments (red and blue) and their corresponding polarity score, *p* (see Supplementary Methods for how *p* is calculated). (**b**) Same as in (a), but for the orientation score, *o*. (**c**) Distributions of polarity scores in structures built using GPSeq and Hi-C data (HG), separately for each chromosome. (**d**) Same as in (c), but for orientation scores. (**e, f**) Same as in (c, d), respectively, but for structures built using only Hi-C data (H). Each boxplot in (c-f) corresponds to $n = 1{,}000$ structures. chr9:22 and chr22:9 are the derivative chromosomes of the t(9;22)(q34;q11.2) translocation. Box plots span from the 25th to the 75th percentile and whiskers extend from −1.5 × IQR to +1.5 × IQR from the closest quartile, where IQR is the inter-quartile range.



Extended Data Fig. 10 | Relationship between chromosome mingling, cancer-associated gene fusions and DSBs. (**a**) Distribution of the inter-chromosome mingling frequency of the 10% most frequently mingling beads in 100 kb-resolution *chromflock* structures, separately for beads overlapping (Fusions) or not (Controls) with cancer-associated gene fusions annotated in TCGA. Structures were generated using Hi-C data only (*that is*, without GPSeq integration). *P*-value: Wilcoxon test, two-sided. *n*, number of beads analyzed. (**b**) Average number of Hi-C trans-chromosomal contacts per 1 Mb genomic window in ten concentric layers defined based on the GPSeq score. (**c**) Distribution of the normalized number of trans-chromosomal Hi-C contacts (trans/all) per 1 Mb genomic window in the same layers as in (b). *P*-values: Wilcoxon test, two-sided. *n*, number of genomic windows analyzed. (**d**) Distributions of the total BLISS read count per 100 kb genomic windows, separately for windows overlapping (Fusions) or not (Controls) with cancer-associated gene fusions annotated in TCGA. *P*-value: Wilcoxon test, two-sided. *n*, number of genomic windows analyzed. (**e**) Radial distribution of DSBs in genic *vs*. intergenic genomic regions in ten concentric nuclear layers defined based on the GPSeq score. (**f**) Radial profile of γH2A.X along the nuclear radius. The intensity of γH2A.X immunofluorescence was normalized by the intensity of DNA staining using Hoechst 33342 using the same approach as for quantifying YFISH signal radial profiles (Supplementary Methods). Each point represents the median γH2A.X signal intensity in one of 200 radial layers. *n*, number of cells analyzed. The red line is a polynomial fit to the points. In all the violin plots and boxplots in the figure, boxes extend from the 25th to the 75th percentile, the midline represents the median, and whiskers extend from –1.5×IQR to +1.5×IQR from the closest quartile, where IQR is the inter-quartile range. Dots: outliers (data falling outside whiskers).

Corresponding author(s): Dr. Magda Bienko (magda.bienko@ki.se)
Dr. Nicola Crosetto (nicola.crosetto@ki.se)

Last updated by author(s): Apr 3, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	NIS Elements AR software (Nikon, v5.11.20) was used for image acquisition.

Data analysis	SEQUENCING DATA. We demultiplexed raw data using the online BaseSpace platform from Illumina. We performed quality check and processed the FASTQ files with a custom pipeline to generate de-duplicated read counts per restriction enzyme recognition site as BED files. All the scripts used are available here: https://github.com/ggirelli/gpseq-seq-gg/releases/tag/v2.0.3. We used FastQC v0.11.5 for quality control, scan_for_matches (http://blog.theseed.org/servers/2010/07/scan-for-matches.html) for pattern filtering, and bwa (0.7.17-r1188) for alignment. We calculated GPSeq scores using our "GPSeqC" Python3 package (https://github.com/ggirelli/gpseqc/releases/tag/v2.3.6). We corrected the genomic coordinates in the BED files for the presence of the t(9;22)(q34;q11.2) translocation using bed-fix-chrom-rearrangement (v0.0.1), a Python3 script available at https://github.com/ggirelli/bed-fix-chrom-rearrangement/releases/tag/v0.0.1. We performed binning of biological data tracks using bioTrackBinner (v0.0.1), a suite of R scripts available at https://github.com/ggirelli/bioTrackBinner/releases/tag/v0.0.1. We removed ChIP-seq adapter sequences using TrimGalore (v0.4.4_dev), perform alignment using bwa-mem (v0.7.17-r1188), and removed PCR duplicates using Picard MarkDuplicates (v2.18.11). We generated genomic coverage tracks in bigWig format using the bamCoverage module from deeptools (v3.2.1) with --binSize 50 option. IMAGE ANALYSIS. We deconvolved microscopy images, when needed, using Huygens Professional (v17.04). We analyzed immunofluorescence and FISH images radial profiles using our "pygpseq" Python3 package available at https://github.com/ggirelli/pygpseq/releases/tag/v3.3.4 (an extensive list of dependencies and corresponding versions is available at the provided link). Further analyses were performed with a suite of Python3 and R scripts: pygpseq-scripts (v0.0.1) available at https://github.com/ggirelli/pygpseq-scripts/releases/tag/v0.0.1. For FISH probe signal analysis, we used our in-house suite DOTTER written in MATLAB (MATLAB and Statistics Toolbox Release R2018a) and C99 with GSL (https://www.gnu.org/software/gsl/). We generated 3D genome structures using our chromflock (v0.1) software available at https://github.com/elgw/chromflock/tree/0.1. We performed further downstream analyzes using custom R scripts and used the ggplot2 package to generate the plots. We generated chromosome ideogram plots using ggkaryo2 (v0.0.3), a prototype R package available at https://github.com/ggirelli/ggkaryo2/releases/tag/v0.0.3. When possible, we implemented pipelines as snakemake flows and made them available as GitHub repositories (links in manuscript).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All HAP1 sequencing data were deposited in the Gene Expression Omnibus (GEO) under accession code GSE135882.
All GM06990 sequencing data were deposited in the GEO under accession code GSE135882.
Source data for all figure panels have been deposited on GitHub at: https://github.com/ggirelli/GPSeq-source-data

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical tool was used to choose sample size a priori . All experiments were performed in replicates (coming from different cell cultures), as indicated in Main, Figure legends, and Supplementary Table 2, 11 and 12.
Data exclusions	We excluded chromosomes 9 and 22 from Hi-C related analyses due to the presence of the t(9;22)(q34;q11.2) translocation.
Replication	We performed GPSeq experiments on HAP1 cells in two replicates, either using HindIII or MboI enzyme, for a total of 4 experiments. The experiments were all in very good agreement, with MboI-based experiments reaching a higher resolution, as expected (see Supplementary Notes for a detailed explanation). We then performed an additional GPSeq experiment on GM06990 cells, again in two replicates, which showed very good agreement between them and with HAP1 experiments, supporting our claim about the conservation of radial arrangement in different cell types. All libraries showed at least 25% of the recognition sites being cut and a minimum of 2.5 million unique reads. To validate our sequencing results, we undertook a major DNA FISH effort, probing a total of 68 genomic loci distributed along 11 chromosomes to assure as proper coverage of the genome. We only analyzed datasets with at least 100 cells showing a proper FISH signal and used more than 20,000 FISH signals in total for the analysis (min: 101, max: 742, average: 314 nuclei analyzed per dataset).
Randomization	A comparison of different cell lines or different treatments was not the aim of the current study. As such, randomization is not relevant.
Blinding	Blinding is not relevant to our study, as samples were processed identically through standard experimental and computational procedures, that should not bias outcomes.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Mouse Anti-phospho-Histone H2A.X (Ser139) (Millipore, cat. no. 05-636, lot no. 2250524, monoclonal clone JBW301, dilution 1:200); ; Rabbit Anti-Histone H2A (Cell Signaling Technology, cat. no. 12349S, lot no. 1, monoclonal clone D6O3A, dilution 1:500); Goat Anti-rabbit IgG ATTO 488 conjugate (Abcam, cat. no. ab150077, lot no. GR322463-1, polyclonal, dilution 1:500)

Validation

Mouse Anti-phospho-Histone H2A.X (Ser139) (Millipore, cat. no. 05-636), references:
- Cell Rep. 2015 Oct 20;13(3):451-459. doi: 10.1016/j.celrep.2015.09.017. Epub 2015 Oct 8.
- Science. 2015 Jan 9;347(6218):185-188. doi: 10.1126/science.1261971.
- Nat Commun. 2015 Apr 29;6:7035. doi: 10.1038/ncomms8035

Rabbit Anti-Histone H2A (Cell Signaling Technology, cat. no. 12349S), was validated using SimpleChIP® Enzymatic Chromatin IP Kits (www.cellsignal.com/products/primary-antibodies/histone-h2a-d6o3a-rabbit-mab/12349).
References:
- Cancer Cell. 2018 Feb 12;33(2):322-336.e8. doi: 10.1016/j.ccell.2018.01.002.
- Nat Commun. 2018 Nov 7;9(1):4654. doi: 10.1038/s41467-018-07016-0.
- Nat Med. 2018 Jun;24(6):758-769. doi: 10.1038/s41591-018-0034-6. Epub 2018 May 21.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

HAP1 human chronic myeloid leukemia cells were purchased from Horizon Discovery (cat. No. C859).
GM06990 lymphoblastoid cells were purchased from Coriell Institute (cat. No. GM06990).

Authentication

No additional authentication was performed.

Mycoplasma contamination

All cell lines tested negative for Mycoplasma contamination.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used in this study.