Published in final edited form as:
*Nat Biotechnol*. 2020 October 01; 38(10): 1184–1193. doi:10.1038/s41587-020-0519-y.

# 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], **Magda Bienko**[1,2,5]

Nicola Crosetto: nicola.crosetto@ki.se; Magda Bienko: magda.bienko@ki.se

[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

## Abstract

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 towards 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, gene expression, as well as A/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/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 non-randomly positioned with respect to the nuclear periphery[2–9]. The nuclear lamina is thought to be the key organizer of the radial

[4]These authors contributed equally: Gabriele Girelli, Joaquin Custodio, Tomasz Kallas.
[5]These authors jointly supervised this work: Nicola Crosetto and Magda Bienko.

**Author contributions**

*Conceptualization*: J.C., T.K., G.G., F.A., B.S., E.W., A.v.O, N.C., M.B. *Data curation*: G.G., F.A.; *Formal analysis*: G.G., F.A., E.W, B.S., J.C. *Funding acquisition*: M.B., N.C.; *Investigation*: T.K., J.C., S.K., M.B. *Methodology*: J.C., T.K., G.G., N.C., M.B. *Project administration*: M.B., N.C. *Resources:* SciLifeLab, H.B., L.X., R.M. *Software*: G.G., E.W. *Supervision*: M.B., 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., N.C. *Visualization*: G.G., F.A., J.C., M.B., N.C. *Writing*: M.B, N.C., G.G., J.C., T.K., F.A., S.K.

**Competing interests**

The Authors declare no competing interests.

arrangement of chromatin in interphase nuclei[10], by creating a large nuclear compartment where the majority of inactive chromatin clusters in the form of lamina-associated domains (LADs)[11–13]. Specialized sub-chromosomal regions such as centromeres and telomeres, as well as nucleolar organizing regions (NORs), are also non-randomly 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], GC-content[21–23], as well as 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 merely consists 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, here we develop a method that allows inferring radial locations genome-wide, all along the nuclear radius, which we name genomic loci positioning by sequencing or GPSeq. Using GPSeq, we generate 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/B compartments and subcompartments[37,38]. We develop a high-performance algorithm, *chromflock*, that dramatically improves the accuracy of whole-genome structure ensemble generation. Finally, we integrate 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 towards 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 towards 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, Online Methods, and Supplementary Methods). If the enzyme indeed gradually digests gDNA from the nuclear periphery towards 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, 45 min and 1, 2, 6 h), and used either wide-field microscopy or stimulated emission depletion microscopy followed by image deconvolution 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 following 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 is 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, Online Methods, and Supplementary Methods). To infer radial positions genome-wide, 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).

To test the reproducibility of GPSeq, we performed two replicate experiments using HindIII (Exp.1 and 2), obtaining highly correlated GPSeq scores both at 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 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 *vs.* 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 compared to regions with high DamID signal (Fig. 2h and Supplementary Fig. 3a-c). Constitutive inter-LAD regions (ciLADs)[13] were the most central, while 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 genome-wide. A step-by-step GPSeq protocol is available at **Protocol Exchange** (DOI: 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 2D 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, 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. However, the GC-content did not accurately predict radial locations genome-wide already at 1 Mb resolution (Extended Data Fig. 3a-f). Therefore, we built a multi-variable 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, 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 multivariate 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/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/B polarization were not the most central ones. In fact, the polarization was rather pronounced on chr17 and 19, which are very central, whereas A/B compartments had a similar radial arrangement on chr10 and 18, which are more peripheral (Supplementary Fig. 6d). We tested whether this would be different at the level of A/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, marks of active chromatin, such as DNA accessibility, H3K27ac and H3K4me3, as well as chromatin-bound RNA polymerase II (RNAPII) increased towards the nuclear interior in parallel with gene density and expression (Fig. 3e-f and Extended Data Fig. 4a-d). Conversely, H3K9me3, which marks heterochromatin, decreased towards the center (Fig. 3g). 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, while it increased in A1-2 and B1 (Fig. 3e). A similar trend was observed for DNA methylation (Extended Data Fig. 4e). On the other hand, H3K27ac increased towards 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). Intriguingly, H3K4me1, a mark of active and poised enhancers, decreased towards the center in all subcompartments, despite its radial profile mildly increasing towards the center when no subcompartment stratification was applied (Fig. 3h). H3K9me3 increased towards the periphery, even though it sharply increased towards the nuclear interior in the B2 subcompartment (Fig. 3g). 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).

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 up-regulated 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 compared to 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 towards the periphery (Fig. 3m, Extended Data Fig. 5a, Supplementary Table 5, and Supplementary Methods). Stratification of the Repli-seq signal by A/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, since the radial location of firing did not change throughout 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 multi-variable 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 genome-wide, 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 Online 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 Video 2–5). We first checked if 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 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. 7 a-f).

We then wondered whether the higher-order radial organization of A/B compartments is recapitulated in individual HG structures. The vast majority of the 10,000 HG structures showed a clear A/B compartment polarization at the level of individual chromosomes, which was not seen in H structures (Extended Data Fig. 8a, b). Thanks to *chromflock*, we generated 1,000 additional HG structures at 100 kb resolution, which showed the expected radial arrangement of A/B subcompartments and strongly correlated with DNA FISH (Fig. 4e, f, Extended Data Fig. 8c, d, and Online Methods). Notably, A1 and B1 subcompartments were typically the most central followed by A2 or B2, while 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/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/B subcompartment polarization in the majority of structures, which was often radially aligned in the case of larger chromosomes but much less on smaller chromosomes (Extended Data Fig. 9c, d). Importantly, 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 novel 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 (CLL) – 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[48]. Consistently, the SNV frequency progressively decreased from the nuclear periphery towards 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 towards the center, indicative of a higher burden of SNPs in active chromatin (Fig. 5a). However, when we stratified by A/B subcompartments, we found that the SNP frequency was higher in B1-2 rather than in A1-2 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). Of note, 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[51] (Supplementary Methods). Genomic loci involved in fusions localized more internally compared to 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 towards 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 compared to loci that never fuse (Extended Data Fig. 10d). The DSB frequency progressively increased towards the nuclear interior in both genic and intergenic regions (Extended Data Fig. 10e). Quantitative analysis of the radial distribution of phosphorylated histone H2A.X ($\gamma$H2A.X) – a proxy of DSBs – confirmed that endogenous breaks are more frequently detected in the inner nucleus (Extended Data Fig. 10f). Importantly, the highest DSB frequency was observed within the 5'-UTR region of protein-coding genes belonging to the most centrally located A1-2 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, in order to investigate the forces that shape the mutational landscape during evolution and in cancer.

## Discussion

We have developed a robust method to map the radial arrangement of chromatin genome-wide, which compared to the gold standard method, DNA FISH, offers orders of magnitude higher throughput. Compared to tyramide signal amplification sequencing (TSA-seq)[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 genome-wide. 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 genome-wide. However, these methods are costly and experimentally more challenging compared to GPSeq.

Together with GPSeq, we have 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 stand-alone assay to visualize chromatin accessibility in single cells, similar to 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 ChIP-seq or Hi-C to directly map chromatin occupancy and chromosome contacts along the nuclear radius.

Although in this study we have mainly used haploid cells, we show that GPSeq can also be applied to chart radiality in diploid cells. This approach, however, does not allow to distinguish the preferential radial position of loci located on homologous chromosomes. Future integration of GPSeq with whole-genome haplotyping strategies will enable to ascertain 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 over other methods, such as Hi-C or Dip-C, which require a modelling step to infer radiality.

We have 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/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/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 further 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 forty 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, while the frequency of cancer SNVs is higher at the nuclear periphery, confirming previous assumptions[61], the frequency of germline SNPs instead mildly increases towards 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 have 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 Methods

Information about antibodies, cell lines, data and code availability and statistics is available in the Life Sciences Reporting Summary.

### YFISH

A detailed step-by-step YFISH protocol is available at **Protocol Exchange** (DOI: 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 up 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 1X 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 1X 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 1X 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/1M NaCl /0.5% Triton X-100 pH 8, five times 1h each at 37 °C, while shaking. To prepare the hybridization mix, we diluted the labeled oligonucleotide to 200 nM in a hybridization buffer containing 2X SSC/25% formamide/10% dextran sulfate/1 mg/ml 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 following day, we washed the samples in washing buffer containing 2X SSC/25% formamide for 1 h at 30 °C. Finally, we incubated the samples in 2X SSC/25% formamide/0.1 ng/μl Hoechst 33342 (Thermo Fisher Scientific, cat. no. H3570) for 30 min at 30 °C, rinsed them twice in 2X 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** (DOI: 10.21203/rs.3.pex-570/v1). Briefly, we digested DNA, ligated the GPSeq adapters and washed unligated adapters using the same procedure 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 genomic DNA (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 sec ON, 90 sec 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 IVT, 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 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; GATC in Exp. 3 and 4 with MboI) associated with an abnormally high number of de-duplicated UMIs for a given digestion time (*i.e.*, condition), by identifying outliers with a chi-square 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 non-overlapping 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 the 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 log2 of these averages and rescaled it again as explained in the 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; 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 $\mu$ is the mean value of the first diagonal and $\sigma$ 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, since 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 labelled 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:

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

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, $V_q = 0.2$):

$$R_b = \sqrt[3]{V_q / N} \tag{2}$$

The calculations in *chromflock* is 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, $X = X_1, X_2, \ldots, X_N$, 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 $\theta_{i-1} \le A_{ij} < \theta_i$ 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, $S$ (*i.e.*, to create and update their contact indication matrices):

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

Initially, the assignment protocol generates the $W$ matrices, one for each structure, by including all the contacts where $A = 1$ (*i.e.*, 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, *i.e.* if bead $i$ 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 $\theta_{i-1} \le A_{i,j} < \theta_i$. Contacts are always assigned to the *most fit* structures. In other words, when $k = round(S \times A_{ij})$ contacts between bead $i$ and $j$ are being assigned to the S structures, they will be given to the $k$ structures which already have the smallest distance between bead $i$ and $j$ (*i.e.*, 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 (*i.e.*, 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 distributing 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:

$$F_b(i) = c_b s \tag{4}$$

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

6. $F_d$, a drag force defined by the viscosity $\eta$, which is proportional to the velocity of each bead and models viscosity:

$$F_d(i) = -\eta v(i) \tag{5}$$

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

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 bead $i$ 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:

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

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

$$E = \sum_{i}^{N}(E_s(i) + E_c(i) + E_r(i)) + \sum_{i}^{N}\sum_{j}^{N}(E_v(i,j) + E_a(i,j)) \tag{7}$$

and the total forces are:

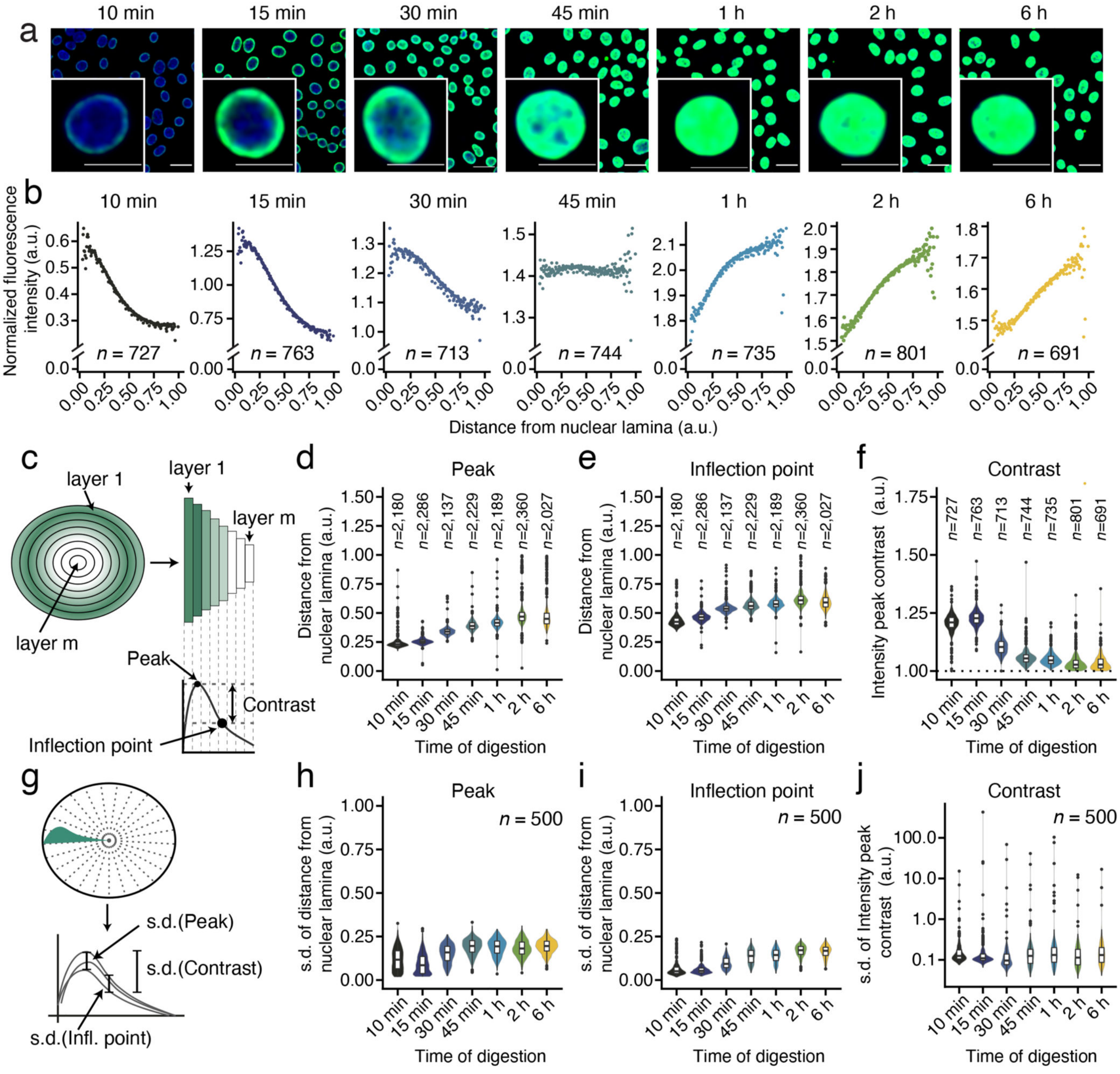
$$F = \nabla E + F_b + F_d \tag{8}$$

We have derived an analytical expression for $\nabla E$ which has been verified against the numerical gradient. We have 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 non-overlapping 1 Mb bins. We have excluded chromosome Y from the analysis as done in ref.[65]. The list of theta values we used is: (1,0.2,0.1,0.5,0.02,0.01,0.001), *i.e.*, we used 7 epochs. The theta values are the same as used in PGS, however we included 0.001 in order to use a larger proportion of the inter contacts. Furthermore, we used $c_v = 1$, $c_s = 1$, $c_a = 1$, $R_s = 1$, $V_q = 0.2$, $R_c = (2.1 + 0.9p)R_b$, 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 re-assignments. 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 run 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).

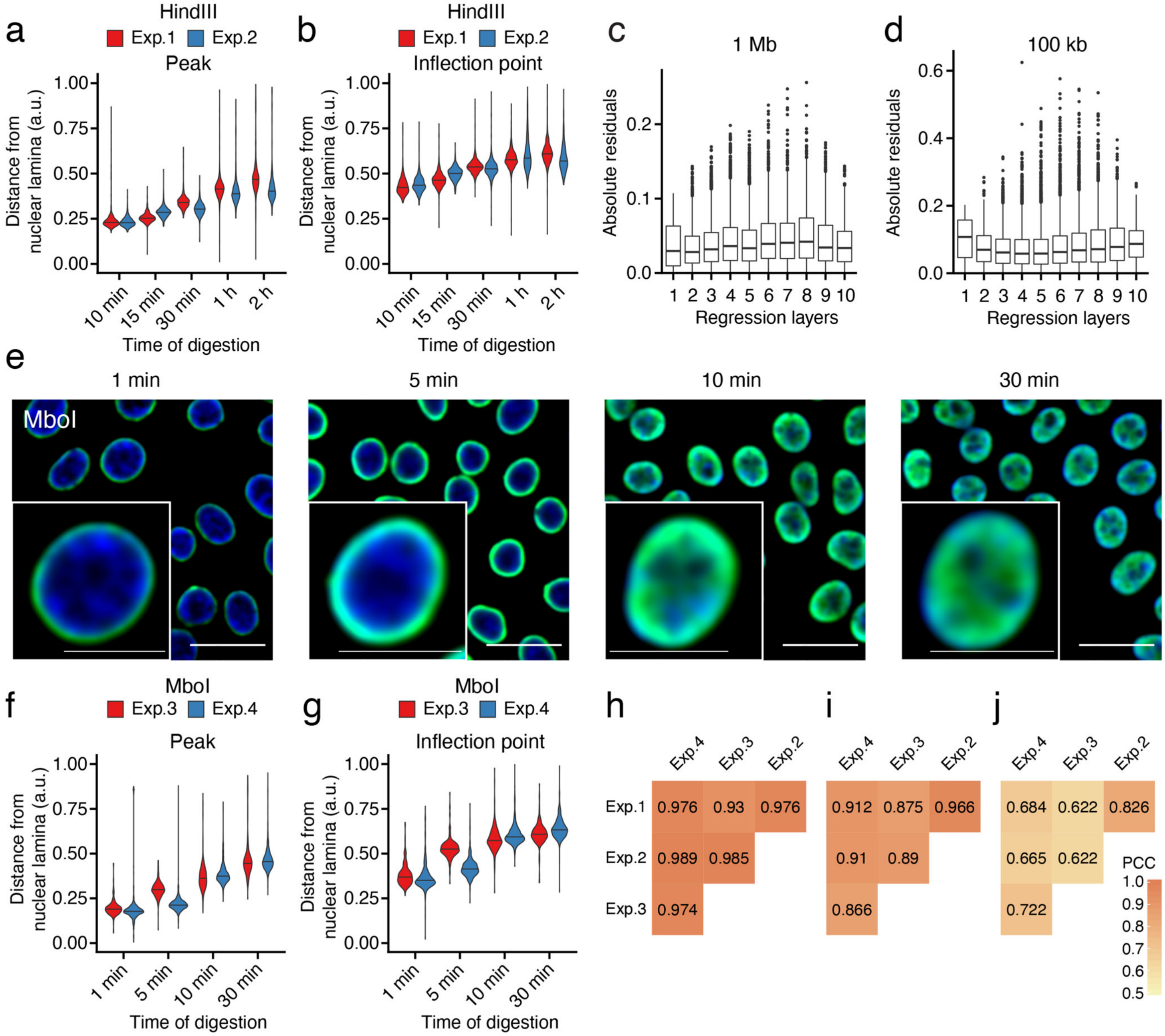
## Extended Data



**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

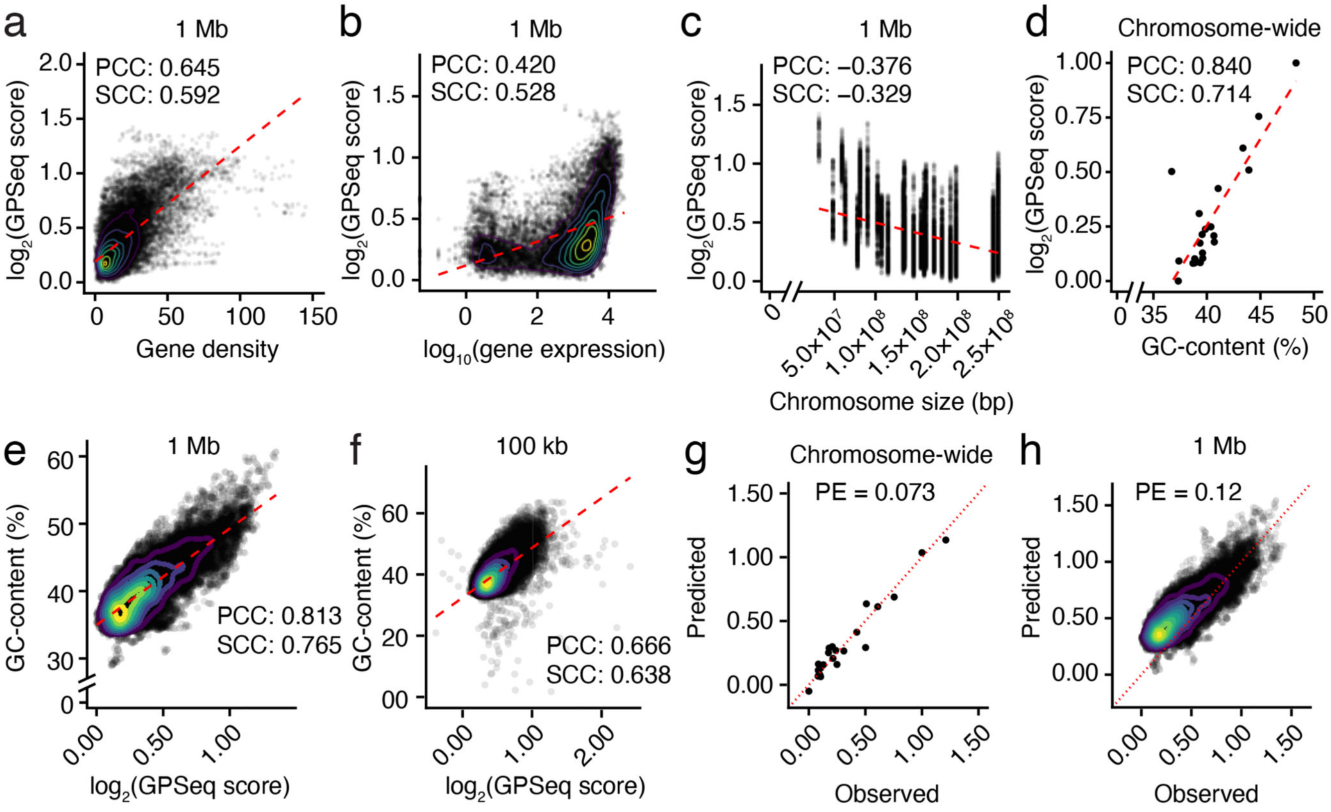
DNA stained with Hoechst 33342. Each dot represents the median intensity in one of 200 radial layers. *n*, number of cells analyzed. (**c**) Calculation of YFISH signal inter-cellular variability. Top: each nucleus is divided in *m* concentric layers of equal thickness and the mean fluorescence intensity 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 (dashed lines) are randomly drawn inside each 3D segmented nucleus and the YFISH intensity profile (green) is evaluated at 100 points (dashed 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**

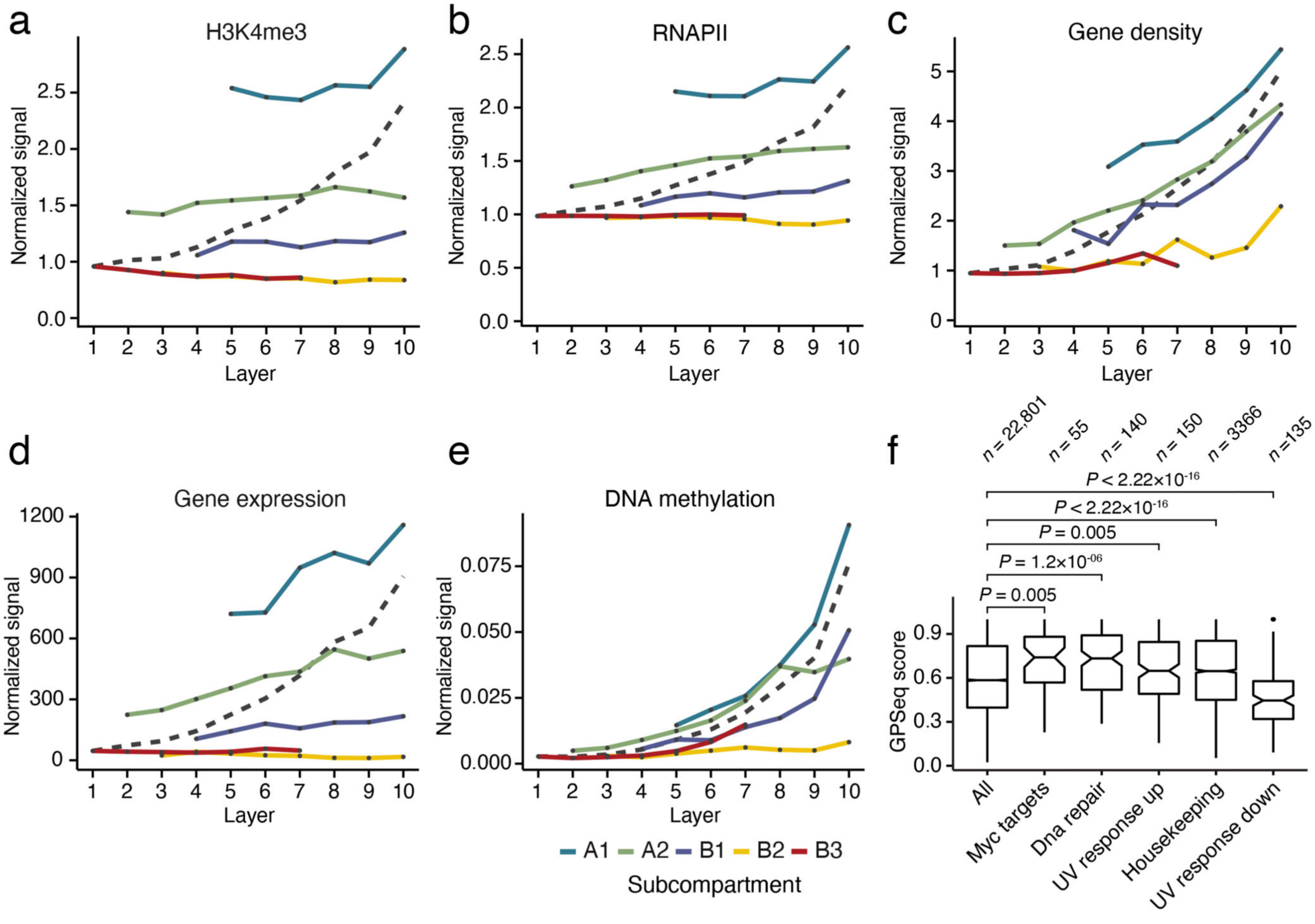
(**a**) Distribution of the position of the peak in the YFISH fluorescence intensity radial profile (see Extended Data Fig. 1c) at different restriction times, in two HindIII experiments (Exp.1 and 2). (**b**) Same as in (a), but for the position of the inflection point. (**c**) Distribution of the absolute 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 (*i.e.*, 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.



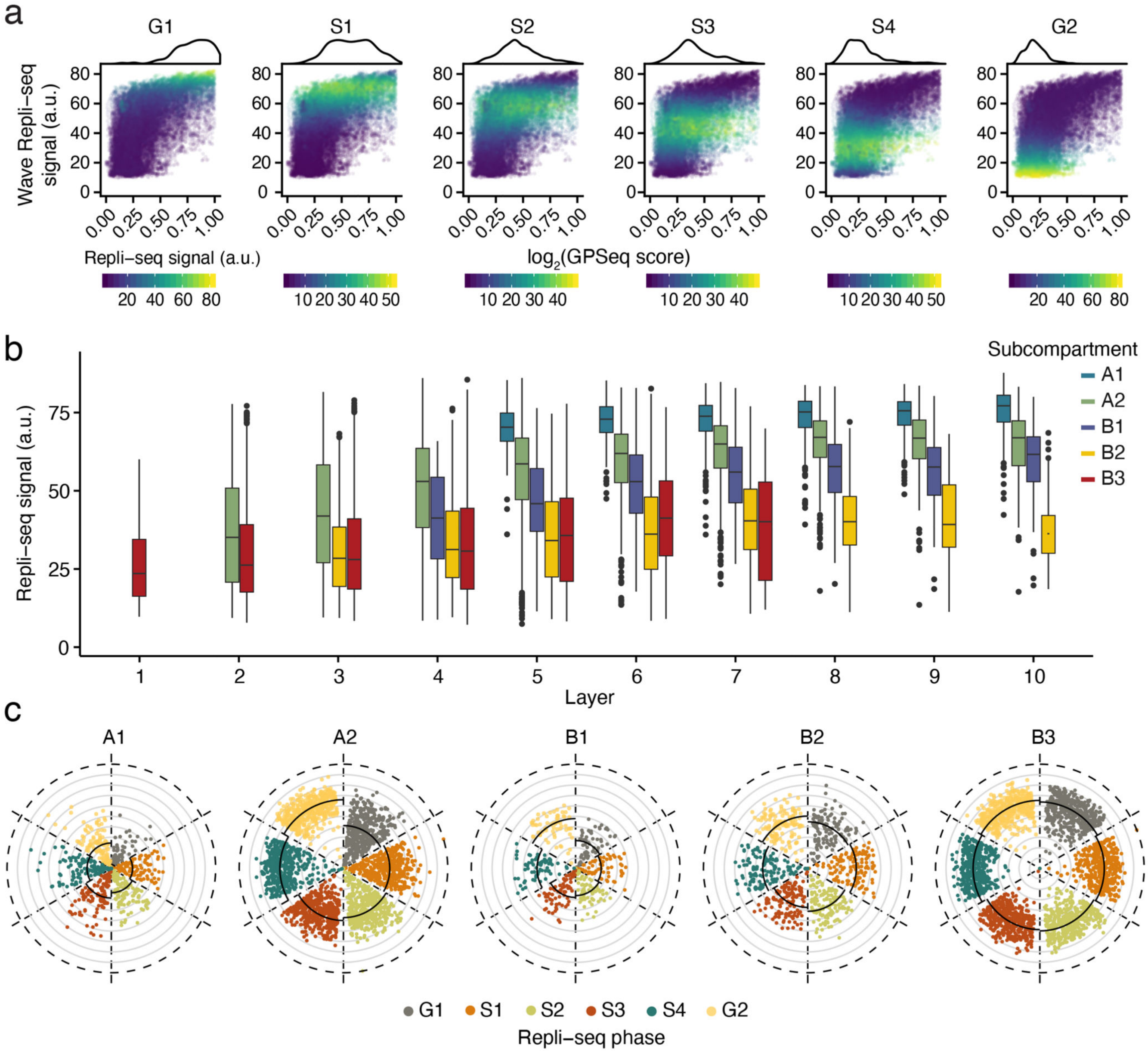
**Extended Data Fig. 3. Predictors of chromatin radiality**

(**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 multi-variable model including both chromosome size and GC-content as described in the Online Methods. PE, prediction error. Dashed red line: linear regression. 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 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.



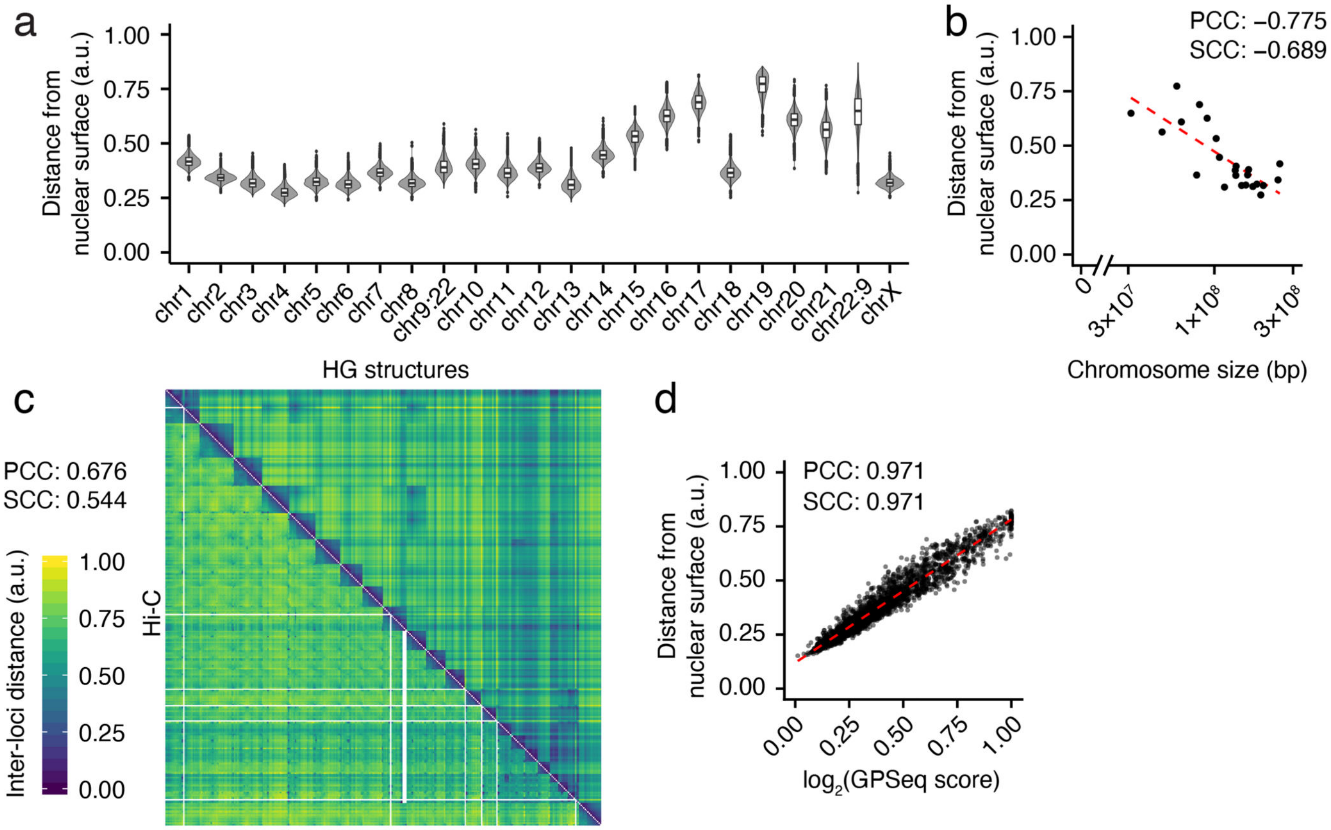
**Extended Data Fig. 4. Radial distribution of chromatin marks and gene expression**

(**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.



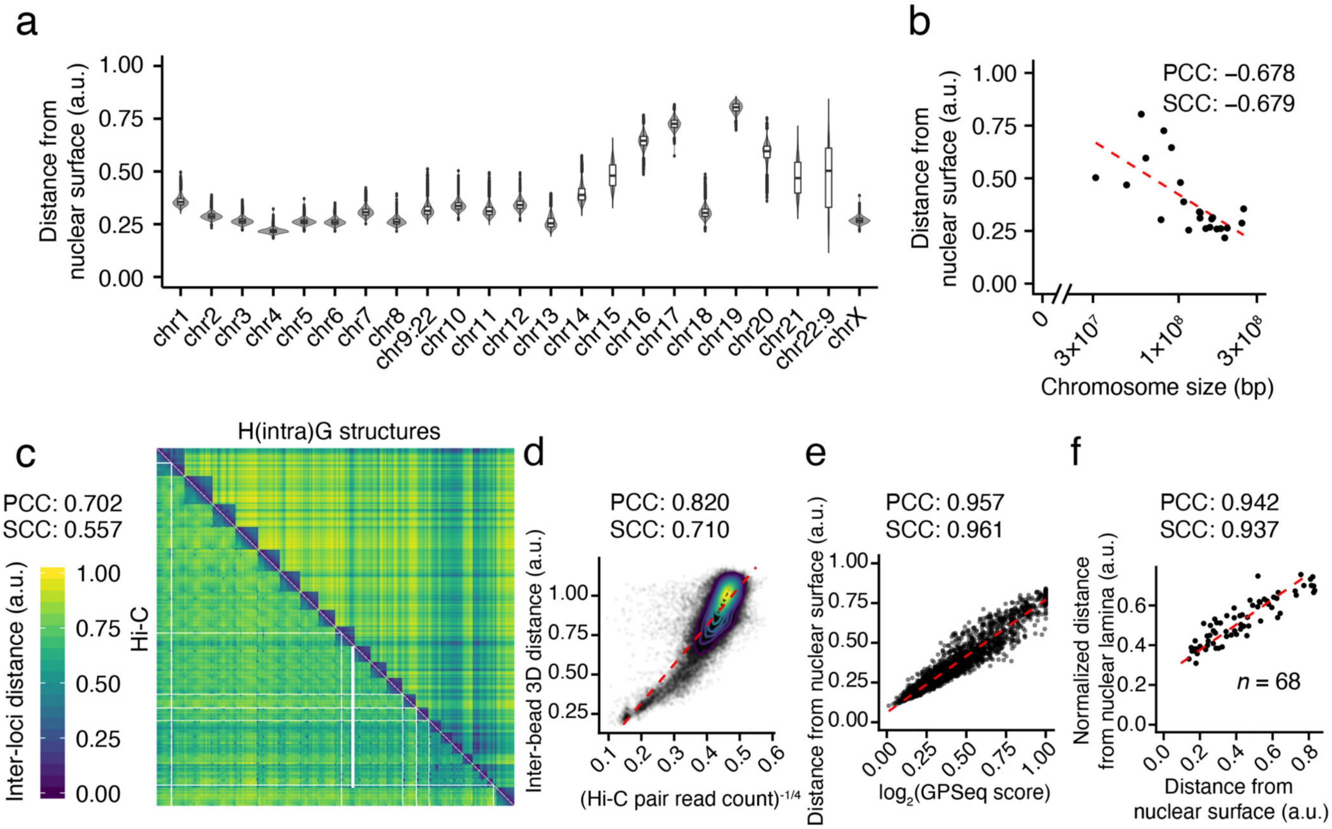
**Extended Data Fig. 5. Radial progression of DNA replication**

(**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 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.



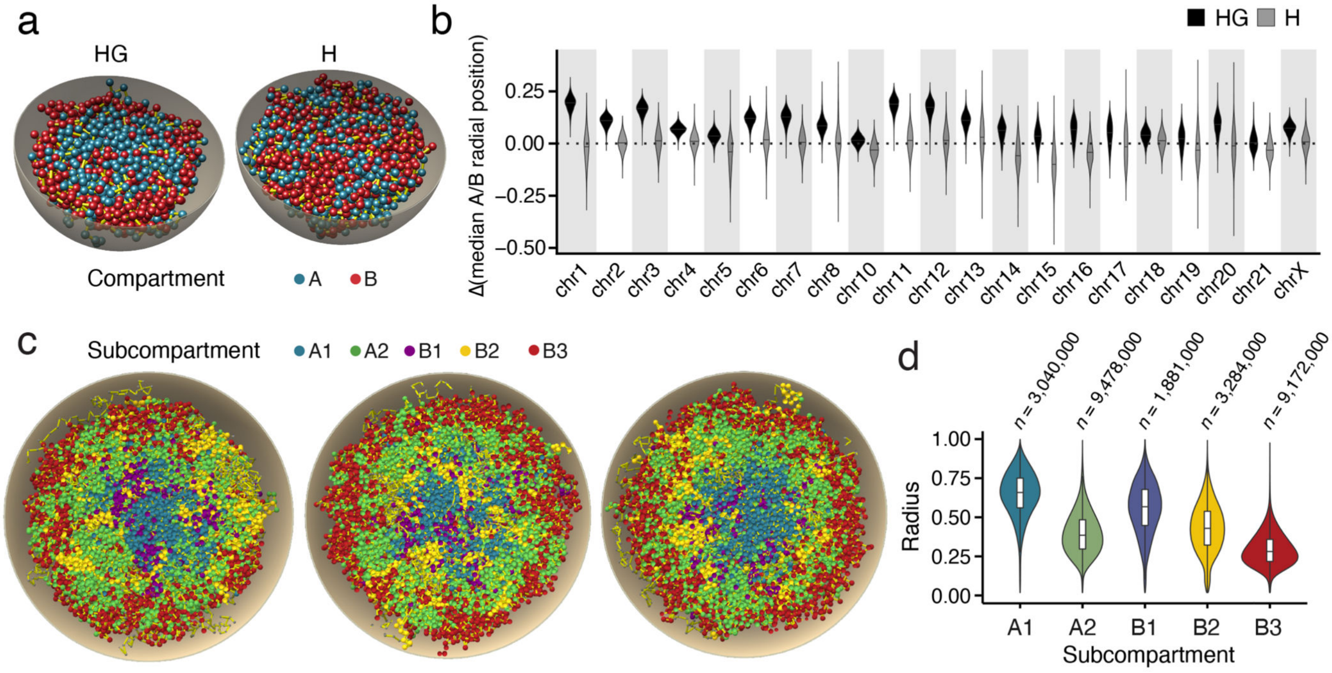
**Extended Data Fig. 6. Analysis of *chromflock* structures generated using both GPSeq and Hi-C data (HG structures)**

(**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 –0.25. The reported correlation coefficients are for 1 Mb resolution, while the plot shows averaged values over 10 Mb genomic windows (points) 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.



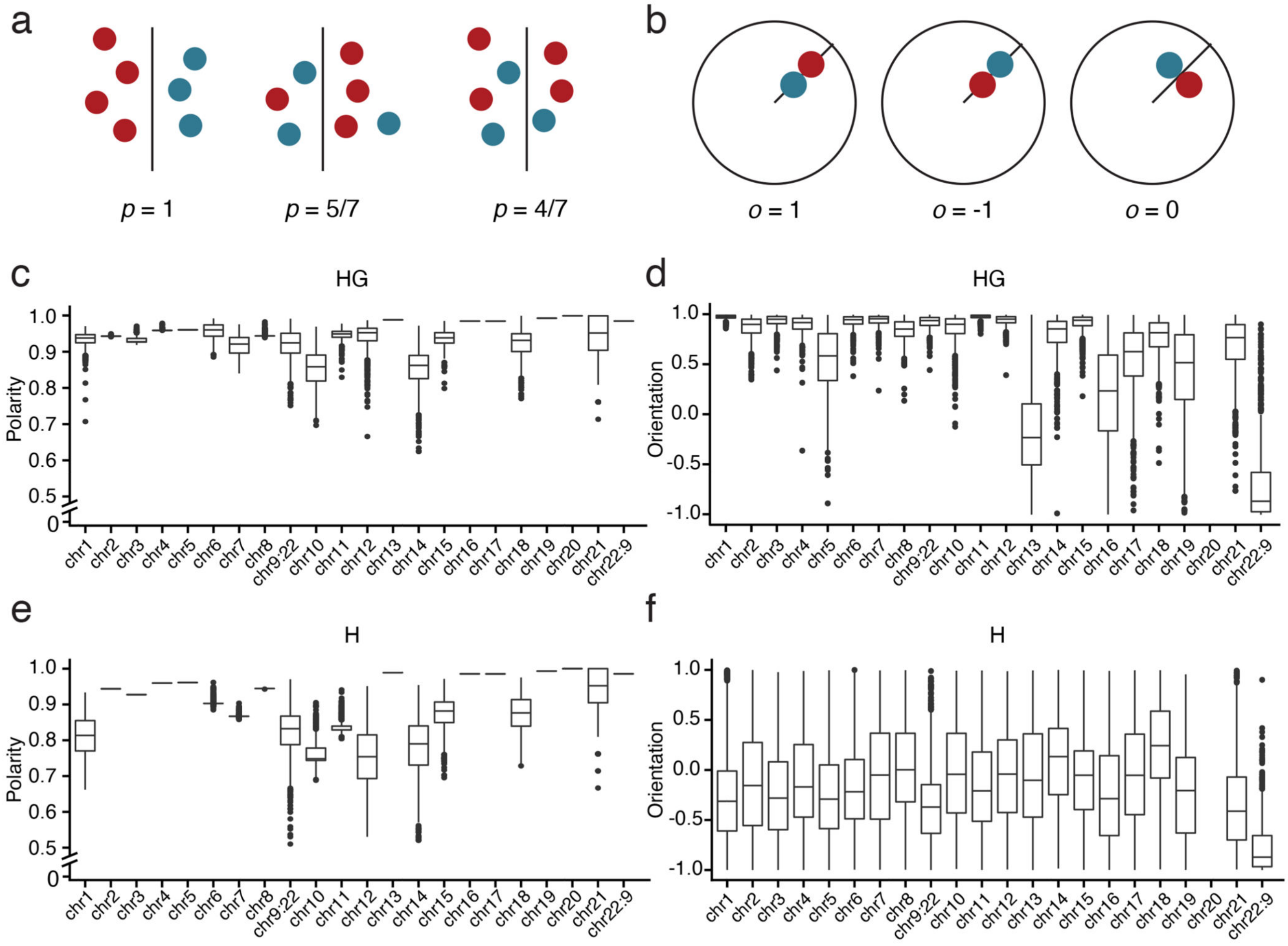
**Extended Data Fig. 7. Analysis of *chromflock* structures generated using GPSeq and Hi-C intra-chromosomal contacts only (H(intra)G)**

(**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 –0.25. The reported correlation coefficients are for 1 Mb resolution, while the plot shows averaged values over 10 Mb genomic windows (points) 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.



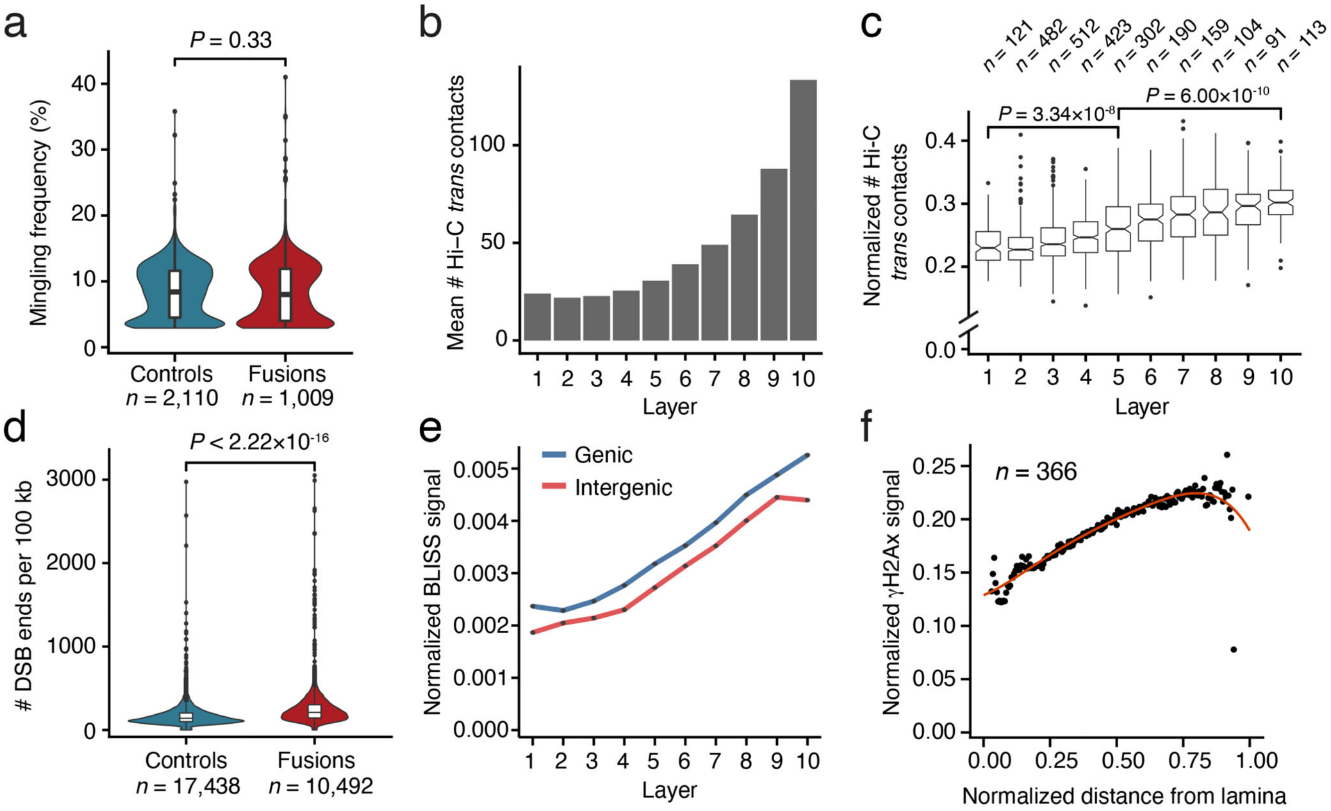
**Extended Data Fig. 8. Radial organization of A/B compartments and subcompartments in *chromflock* structures**

(**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. Dots: outliers (data falling outside whiskers).



**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 (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.



**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 (*i.e*., 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 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).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank Alexander van Oudenaarden (Hubrecht Institute) for initial discussions on GPSeq data analysis, and Irina Solovei (UMC Munich), Marc A. Marti-Renom (CRG Barcelona), Sandy L. Klemm (Stanford) and Britta Bouwman (Bienko-Crosetto lab) for critically reading the manuscript and providing ideas. We thank Lei Xu and Reza Mirzazadeh (Bienko-Crosetto lab) for helping with FISH probe production. We acknowledge Hans 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 (SSMF) 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 Karolinska Institutet to N.C.; 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 programme (StG-2016_GENOMIS_715727) to M.B.

## Data Availability Statement

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*-generate whole-genome structures
3. Genome-wide GPSeq scores at chromosome-wide, 1 Mb, and 100 kb resolution
4. GPSeq score 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 datasets 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 datasets:

1. Chronic Lymphoid 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 Statement

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

1. Sleeman JE, Trinkle-Mulcahy L. Nuclear bodies: new insights into assembly/dynamics and disease relevance. Curr Opin Cell Biol. 2014; 28:76–83. [PubMed: 24704702]
2. Croft JA, et al. Differences in the localization and morphology of chromosomes in the human nucleus. J Cell Biol. 1999; 145:1119–1131. [PubMed: 10366586]
3. Bridger JM, Boyle S, Kill IR, Bickmore WA. Re-modelling of nuclear architecture in quiescent and senescent human fibroblasts. Curr Biol CB. 2000; 10:149–152. [PubMed: 10679329]
4. Cremer M, et al. Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. Chromosome Res Int J Mol Supramol Evol Asp Chromosome Biol. 2001; 9:541–567.
5. Boyle S, et al. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. Hum Mol Genet. 2001; 10:211–219. [PubMed: 11159939]
6. Mayer R, et al. Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. BMC Cell Biol. 2005; 6:44. [PubMed: 16336643]
7. Bolzer A, et al. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. PLoS Biol. 2005; 3:e157. [PubMed: 15839726]
8. Sun HB, Shen J, Yokota H. Size-dependent positioning of human chromosomes in interphase nuclei. Biophys J. 2000; 79:184–190. [PubMed: 10866946]
9. Tanabe H, et al. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. Proc Natl Acad Sci U S A. 2002; 99:4424–4429. [PubMed: 11930003]
10. van Steensel B, Belmont AS. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell. 2017; 169:780–791. [PubMed: 28525751]
11. Guelen L, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature. 2008; 453:948–951. [PubMed: 18463634]
12. Peric-Hupkes D, et al. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell. 2010; 38:603–613. [PubMed: 20513434]
13. Kind J, et al. Genome-wide maps of nuclear lamina interactions in single human cells. Cell. 2015; 163:134–147. [PubMed: 26365489]
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. 2000; 95:1608–1615. [PubMed: 10688815]

15. Alcobia I, Quina AS, 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. 2003; 290:358–369. [PubMed: 14567993]

16. Molenaar C, et al. Visualizing telomere dynamics in living mammalian cells using PNA probes. EMBO J. 2003; 22:6631–6641. [PubMed: 14657034]

17. Weierich C, et al. Three-dimensional arrangements of centromeres and telomeres in nuclei of human and murine lymphocytes. Chromosome Res Int J Mol Supramol Evol Asp Chromosome Biol. 2003; 11:485–502.

18. Tjong H, et al. Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. Proc Natl Acad Sci U S A. 2016; 113:E1663–1672. [PubMed: 26951677]

19. Németh A, Längst G. Genome organization in and around the nucleolus. Trends Genet TIG. 2011; 27:149–156. [PubMed: 21295884]

20. Quinodoz SA, et al. Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell. 2018; 174:744–757.e24 [PubMed: 29887377]

21. Federico C, et al. Gene-rich and gene-poor chromosomal regions have different locations in the interphase nuclei of cold-blooded vertebrates. Chromosoma. 2006; 115:123–128. [PubMed: 16404627]

22. Grasser F, et al. Replication-timing-correlated spatial chromatin arrangements in cancer and in primate interphase nuclei. J Cell Sci. 2008; 121:1876–1886. [PubMed: 18477608]

23. Hepperger C, Mannes A, Merz J, Peters J, Dietzel S. Three-dimensional positioning of genes in mouse cell nuclei. Chromosoma. 2008; 117:535–551. [PubMed: 18597102]

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. 2004; 86:2803–2812. [PubMed: 15111398]

25. Andrulis ED, Neiman AM, Zappulla DC, Sternglanz R. Perinuclear localization of chromatin facilitates transcriptional silencing. Nature. 1998; 394:592–595. [PubMed: 9707122]

26. Sadoni N, et al. Nuclear organization of mammalian genomes. Polar chromosome territories build up functionally distinct higher order compartments. J Cell Biol. 1999; 146:1211–1226. [PubMed: 10491386]

27. Kosak ST, et al. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. Science. 2002; 296:158–162. [PubMed: 11935030]

28. Kosak ST, et al. Coordinate gene regulation during hematopoiesis is related to genomic organization. PLoS Biol. 2007; 5:e309. [PubMed: 18031200]

29. Finlan LE, et al. Recruitment to the nuclear periphery can alter expression of genes in human cells. PLoS Genet. 2008; 4:e1000039. [PubMed: 18369458]

30. Reddy KL, Zullo JM, Bertolino E, Singh H. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. Nature. 2008; 452:243–247. [PubMed: 18272965]

31. Takizawa T, Meaburn KJ, Misteli T. The meaning of gene positioning. Cell. 2008; 135:9–13. [PubMed: 18854147]

32. Therizols P, et al. Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. Science. 2014; 346:1238–1242. [PubMed: 25477464]

33. Shachar S, Misteli T. Causes and consequences of nuclear gene positioning. J Cell Sci. 2017; 130:1501–1508. [PubMed: 28404786]

34. Cook PR, Marenduzzo D. Transcription-driven genome organization: a model for chromosome structure and the regulation of gene expression tested through simulations. Nucleic Acids Res. 2018; 46:9895–9906. [PubMed: 30239812]

35. Ganai N, Sengupta S, Menon GI. Chromosome positioning from activity-based segregation. Nucleic Acids Res. 2014; 42:4145–4159. [PubMed: 24459132]

36. Küpper K, et al. Radial chromatin positioning is shaped by local gene density, not by gene expression. Chromosoma. 2007; 116:285–306. [PubMed: 17333233]

37. Lieberman-Aiden E, et al. Comprehensive mapping of long range interactions reveals folding principles of the human genome. Science. 2009; 326:289–293. [PubMed: 19815776]

38. Rao SSP, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014; 159:1665–1680. [PubMed: 25497547]

39. Gelali E, et al. iFISH is a publically available resource enabling versatile DNA FISH to study genome architecture. Nat Commun. 2019; 10:1636. [PubMed: 30967549]

40. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013; 10:1213–1218. [PubMed: 24097267]

41. Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. 2006; 20:1123–1136. [PubMed: 16618801]

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 Int J Mol Supramol Evol Asp Chromosome Biol. 2001; 9:77–80.

43. Hua N, et al. Producing genome structure populations with the dynamic and automated PGS software. Nat Protoc. 2018; 13:915–926. [PubMed: 29622804]

44. Hsu TC. A possible function of constitutive heterochromatin: the bodyguard hypothesis. Genetics. 1975; 79 Suppl:137–150. [PubMed: 1150080]

45. Stamatoyannopoulos JA, et al. Human mutation rate associated with DNA replication timing. Nat Genet. 2009; 41:393–395. [PubMed: 19287383]

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. 2013; 4:1502. [PubMed: 23422670]

47. Morganella S, et al. The topography of mutational processes in breast cancer genomes. Nat Commun. 2016; 7:11383. [PubMed: 27136393]

48. Schuster-Böckler B, Lehner B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. Nature. 2012; 488:504–507. [PubMed: 22820252]

49. Chiarle R, et al. Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell. 2011; 147:107–119. [PubMed: 21962511]

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. 2017; 45:D854–D859. [PubMed: 27638885]

51. Hu X, et al. TumorFusions: an integrative resource for cancer-associated transcript fusions. Nucleic Acids Res. 2018; 46:D1144–D1149. [PubMed: 29099951]

52. Mertens F, Johansson B, Fioretos T, Mitelman F. The emerging complexity of gene fusions in cancer. Nat Rev Cancer. 2015; 15:371–381. [PubMed: 25998716]

53. Gothe HJ, et al. Spatial Chromosome Folding and Active Transcription Drive DNA Fragility and Formation of Oncogenic MLL Translocations. Mol Cell. 2019; 75:267–283.e12 [PubMed: 31202576]

54. Yan WX, et al. BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. Nat Commun. 2017; 8:15058. [PubMed: 28497783]

55. Lensing SV, et al. DSBCapture: in situ capture and sequencing of DNA breaks. Nat Methods. 2016; 13:855–857. [PubMed: 27525976]

56. Chen Y, et al. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. J Cell Biol. 2018; 217:4025–4048. [PubMed: 30154186]

57. Beagrie RA, et al. Complex multi-enhancer contacts captured by genome architecture mapping. Nature. 2017; 543:519–524. [PubMed: 28273065]

58. Nagano T, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature. 2013; 502:59–64. [PubMed: 24067610]

59. Tan L, Xing D, Chang C-H, Li H, Xie XS. Three-dimensional genome structures of single diploid human cells. Science. 2018; 361:924–928. [PubMed: 30166492]

60. Chen X, et al. ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing. Nat Methods. 2016; 13:1013–1020. [PubMed: 27749837]

61. Gonzalez-Perez A, Sabarinathan R, Lopez-Bigas N. Local Determinants of the Mutational Landscape of the Human Genome. Cell. 2019; 177:101–114. [PubMed: 30901533]

62. Koren A, et al. Differential relationship of DNA replication timing to different forms of human mutation and variation. Am J Hum Genet. 2012; 91:1033–1040. [PubMed: 23176822]

63. Köster J, Rahmann S. Snakemake-a scalable bioinformatics workflow engine. Bioinforma Oxf Engl. 2018; 34:3600.

64. Sanborn AL, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A. 2015; 112:E6456–6465. [PubMed: 26499245]

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. 2011; 30:90–98. [PubMed: 22198700]

66. Pettersen EF, et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004; 25:1605–1612. [PubMed: 15264254]

67. McFarland CD. A modified ziggurat algorithm for generating exponentially- and normally-distributed pseudorandom numbers. J Stat Comput Simul. 2016; 86:1281–1294. [PubMed: 27041780]

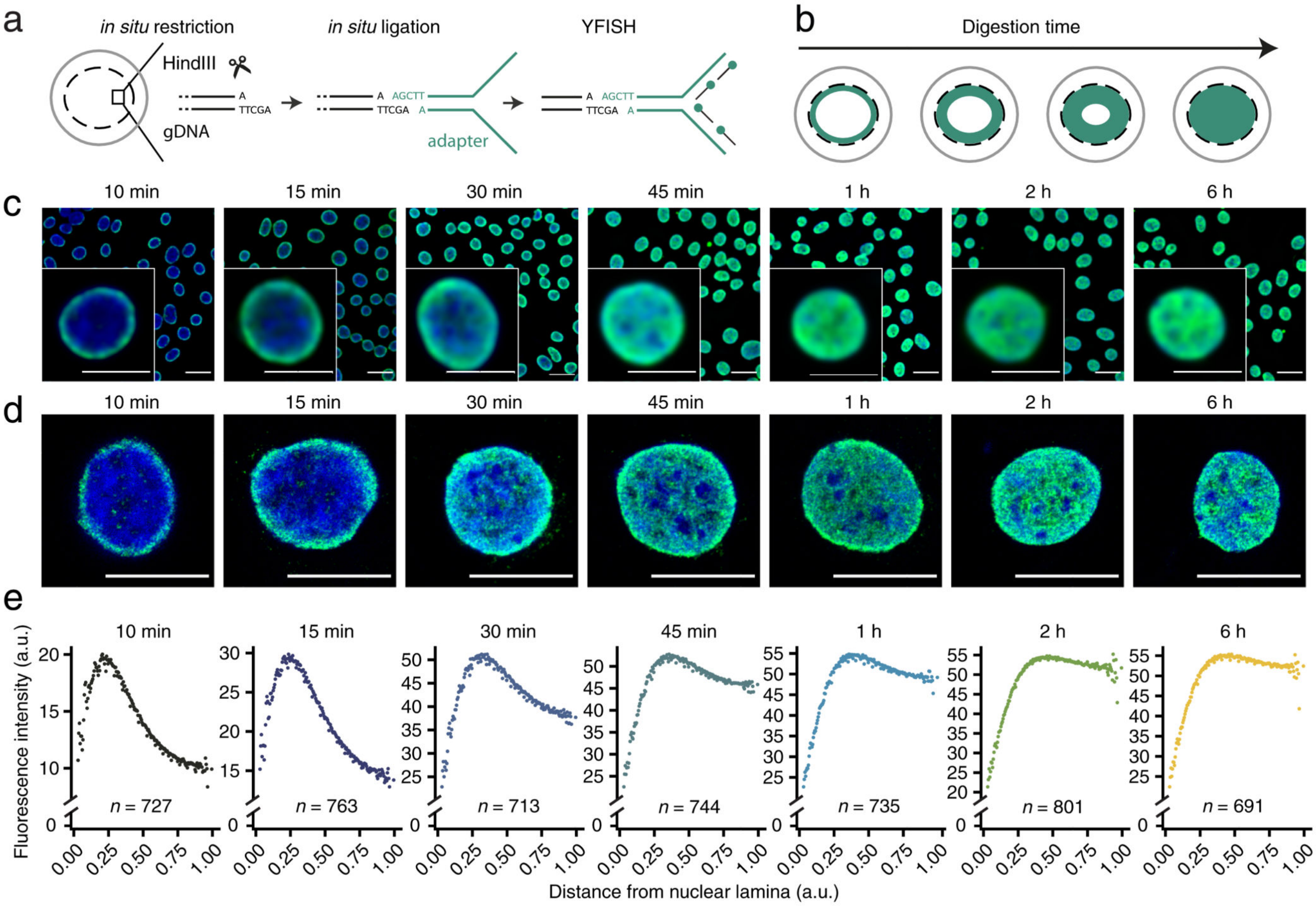


**Fig. 1. GPSeq implementation.**

(**a**) Scheme of YFISH. Cross-linked permeabilized nuclei (dashed circle) are incubated with a restriction enzyme (*e.g.*, HindIII). Digested recognition sites are ligated to a forked adapter (green), which is detected using fluorescently labeled oligos (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 in order 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, 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, 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.

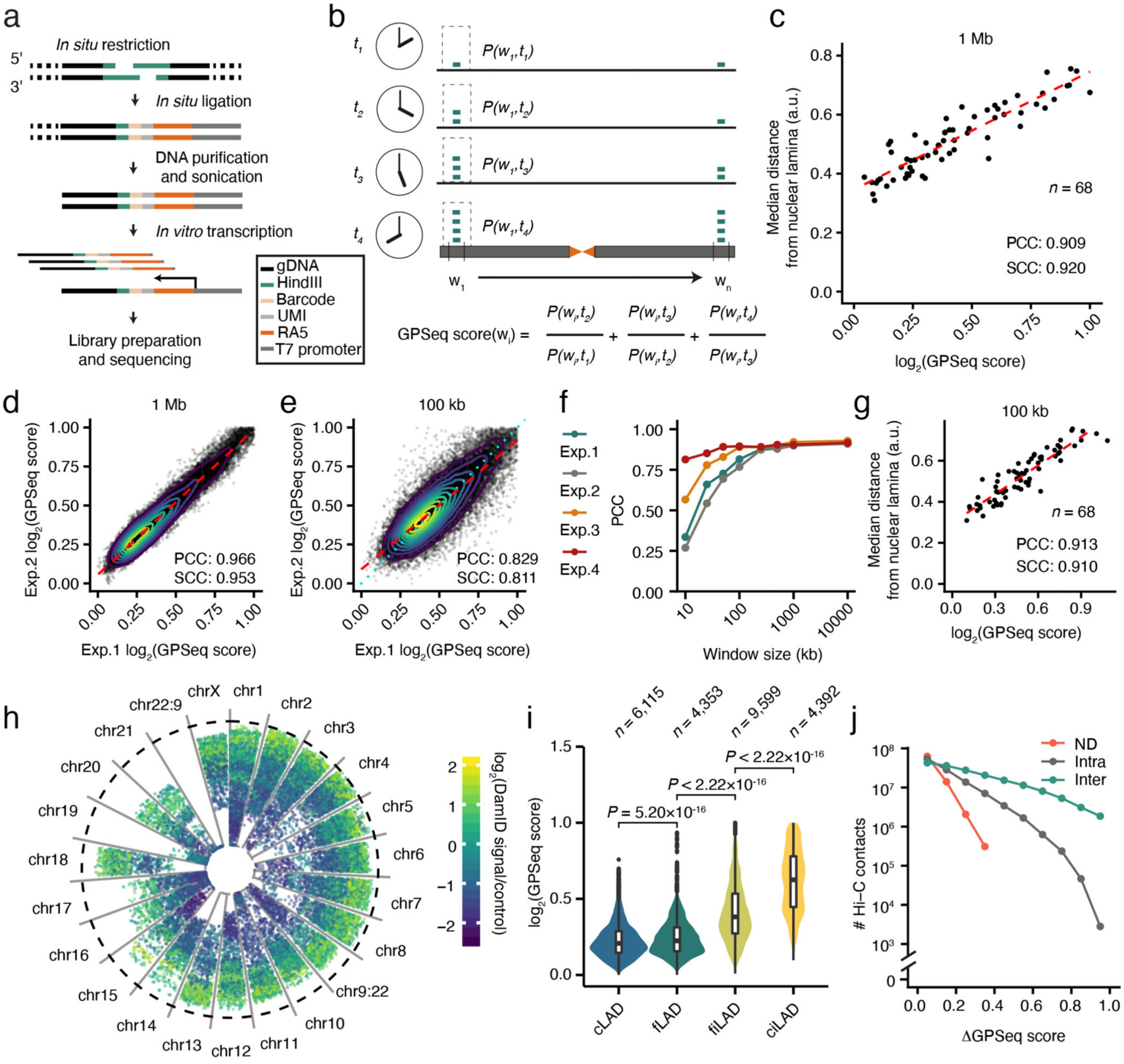


**Fig. 2. GPSeq reproducibility and validation.**

(**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* following different *t* restriction times. (**c**) Correlation between the log2 GPSeq score and the median 3D distance from the nuclear lamina. Each dot represents one DNA FISH probe. (**d**) Correlation between the log2 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 log2 GPSeq score averaged over the four experiments in (f) and the median 3D distance from nuclear lamina. Each dot represents one DNA FISH probe. (**h**) Log2 Lamin B DamID signal/control in 1 Mb genomic windows (dots) radially arranged based on their GPSeq score. Dashed line: 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**) Log2 GPSeq score (1 Mb non-overlapping 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. IQR: inter-quartile range. Dots: data outside 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. Dashed red lines: linear regressions. All source data for this figure are from HAP1 cells.

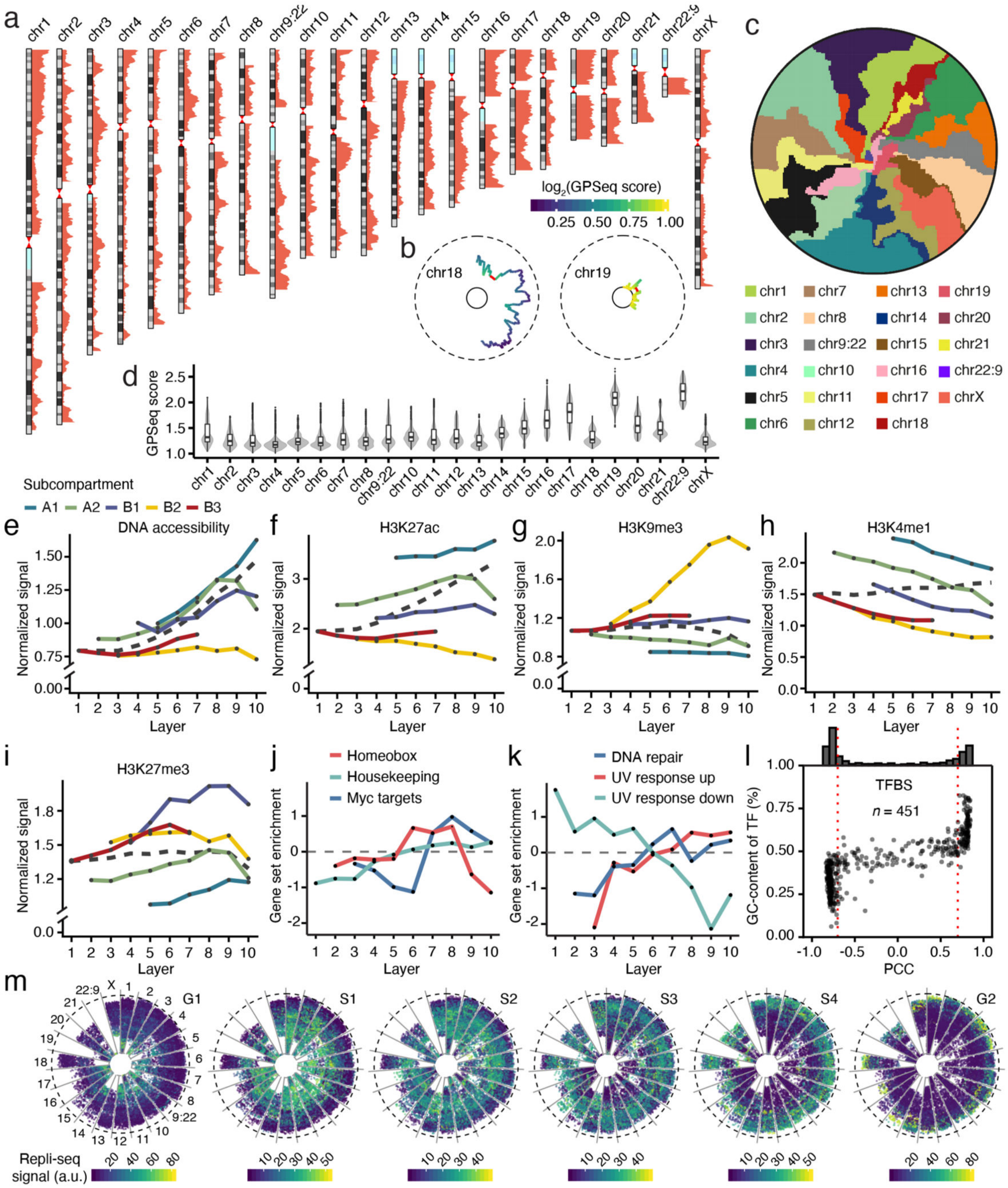


**Fig. 3. Radial organization of chromatin in human cells.**

(**a**) GPSeq score profiles along individual chromosomes (1 Mb overlapping windows, 100 kb step size). (**b**) Circular plots of chr18 and 19 radial location (1 Mb non-overlapping windows). Dashed circle: nuclear lamina; solid circle: nuclear center. Red: masked-out peri-centromeric 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. IQR: inter-quartile range. Dots: data outside whiskers. (**e-i**). Radial distribution of DNA accessibility and various chromatin marks in A/B subcompartments (100 kb resolution). Dashed lines: 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 log2 GPSeq score (1 Mb overlapping genomic windows, 100 kb step size) and the number of predicted transcription factor binding sites (TFBSs) ranked based on GC-content. 26,350 genomic windows (points) were used to calculate the PCC for $n = 451$ TFBSs. (**m**) Repli-seq signal in 1 Mb genomic windows radially arranged based on their GPSeq score in six cell cycle sub-phases (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.

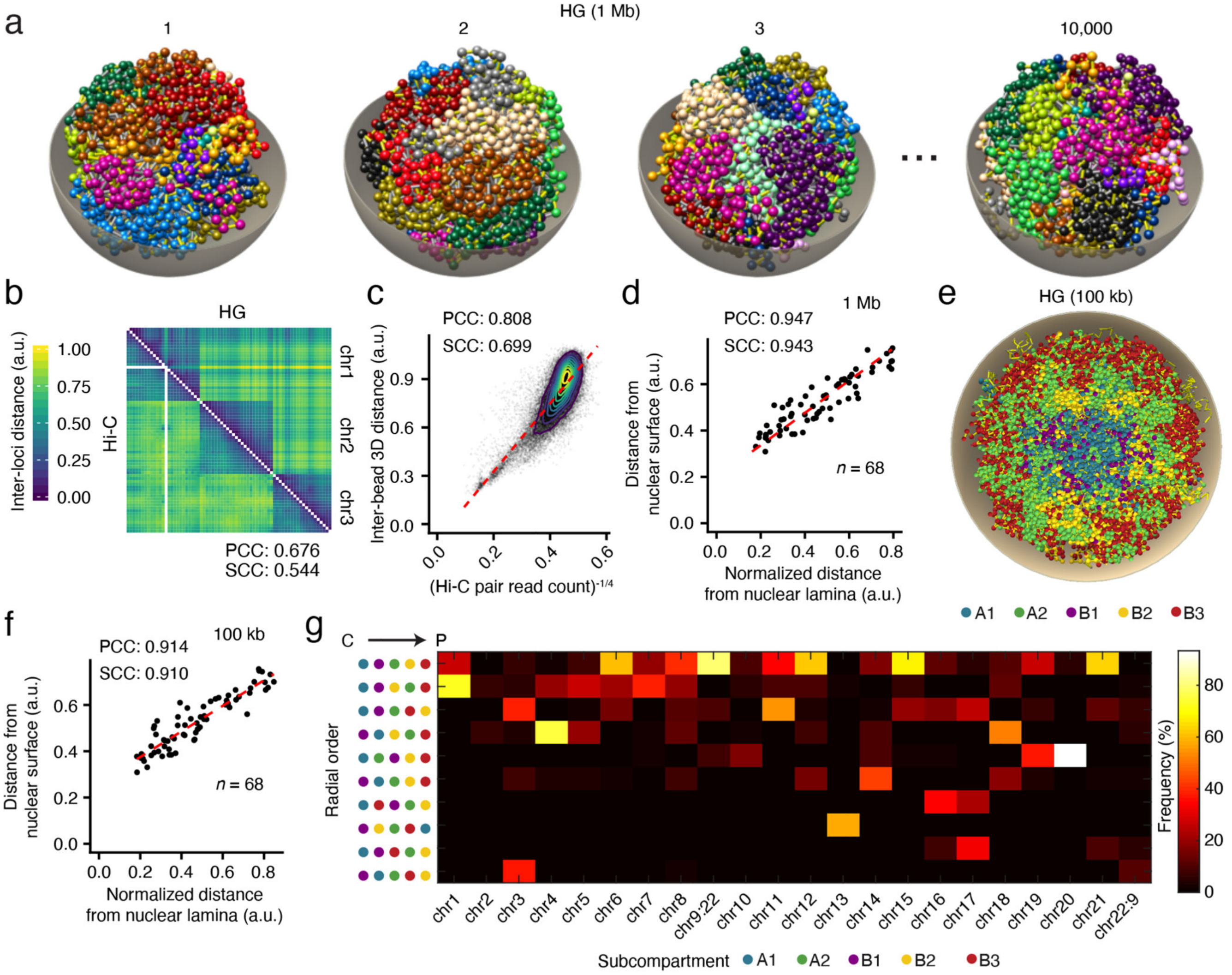


**Fig. 4. Generation of 3D genome structures by GPSeq and Hi-C integration.**

(**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, while 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 non-overlapping genomic windows obtained by averaging 1 Mb non-overlapping windows. *n* = 47,531 pairs of genomic windows (points) were analyzed. Concentric curves: 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/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 10 most frequent A/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: linear regressions.

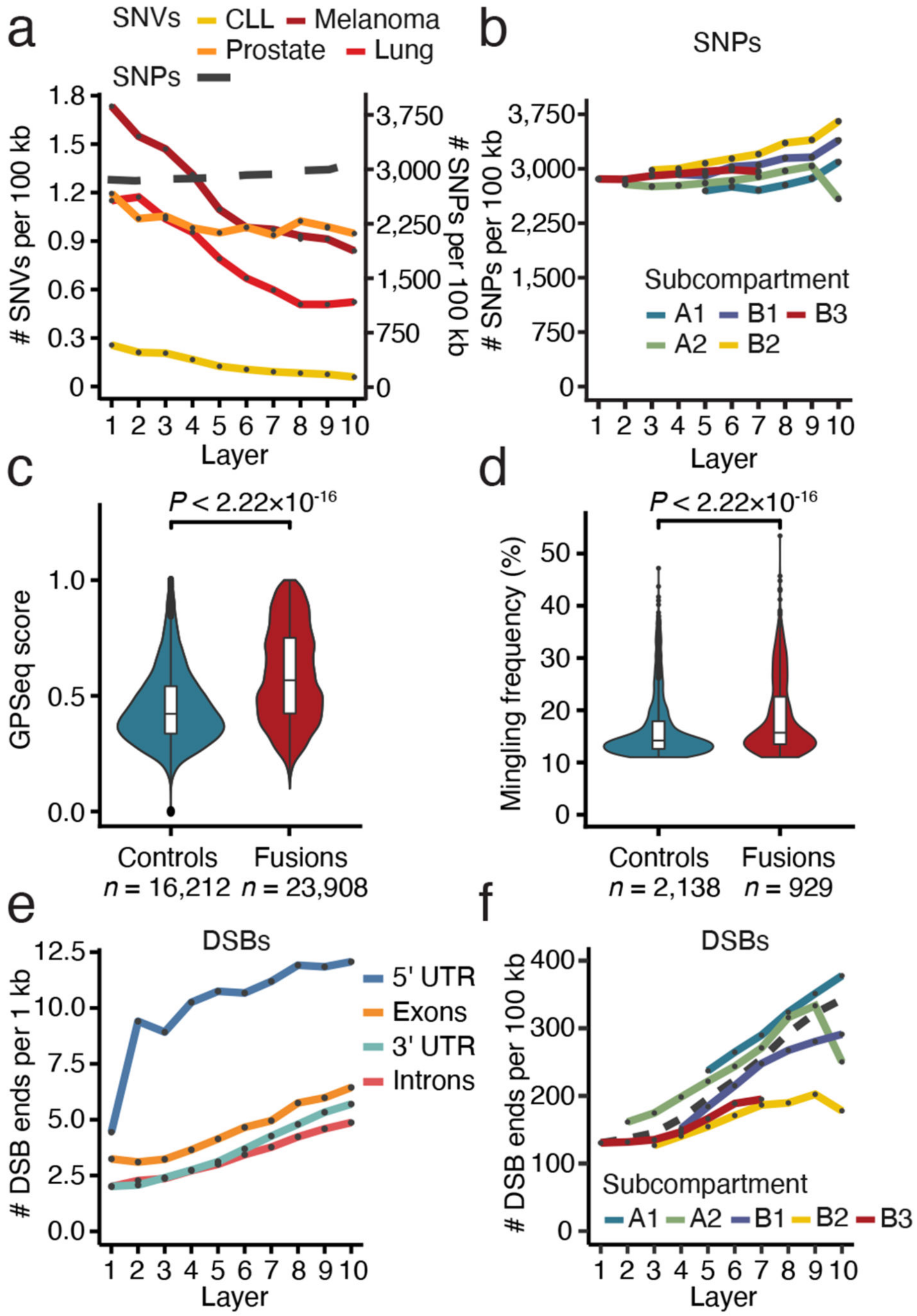


**Fig. 5. Radial distribution of mutations and DNA breaks.**

(**a**) Radial distribution of single-nucleotide variants (SNVs) in four cancer types (left axis) and of single-nucleotide polymorphisms (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/B subcompartments. Mean normalized signals are shown. (**c**) Distribution of the GPSeq score of 100 kb genomic windows overlapping (Fusions) or not (Controls) with genomic fusions from The Cancer Genome Atlas. *n*, number of genomic windows analyzed. *P*-values: Wilcoxon test, two-sided. (**d**) Distributions 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 in K562 cells. (**f**) Radial distribution of DSBs in A/B subcompartments. Dashed line: 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. IQR: inter-quartile range. Dots: data outside whiskers. All source data for this figure are from HAP1 cells, except for BLISS data, which are from K562 cells.