



Published in final edited form as:

Science. 2021 February 26; 371(6532): . doi:10.1126/science.aay3446.

***In situ* genome sequencing resolves DNA sequence and structure in intact biological samples**

Andrew C. Payne^{1,2,*}, Zachary D. Chiang^{2,3,*}, Paul L. Reginato^{1,2,4,5,6,*}, Sarah M. Mangiameli², Evan M. Murray², Chun-Chen Yao^{2,7}, Styliani Markoulaki⁸, Andrew S. Earl³, Ajay S. Labade³, Rudolf Jaenisch^{8,9}, George M. Church^{5,6}, Edward S. Boyden^{1,4,7,10,11,12,13,†}, Jason D. Buenrostro^{2,3,†}, Fei Chen^{2,3,†}

¹Media Arts and Sciences, MIT, Cambridge, Massachusetts 02139, USA.

²Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.

³Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

⁴Department of Biological Engineering, MIT, Cambridge, MA, 02139, USA.

⁵Department of Genetics, Harvard Medical School, Boston, MA 02115, USA.

⁶Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02115, USA.

⁷Department of Brain and Cognitive Sciences, MIT, Cambridge, Massachusetts 02139, USA.

⁸Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02139, USA.

⁹Department of Biology, MIT, Cambridge, Massachusetts 02139, USA.

Correspondence: edboyden@mit.edu.

*These authors contributed equally to this work.

†These authors contributed equally to this work.

Author contributions:

A.C.P. and P.L.R. developed the protocol and performed experiments. Z.D.C. developed the computational processing pipeline. A.C.P., Z.D.C., P.L.R., S.M.M., J.D.B., and F.C. performed analyses. E.M., C.-C. Y., and A.S.L. performed supplementary experiments. S.M. performed embryo preparation under the supervision of R.J. A.S.E. designed the interactive Shiny app. A.C.P., Z.D.C., P.L.R., E.S.B., J.D.B., and F.C. wrote the manuscript with input from all authors. G.M.C., E.S.B., J.D.B., and F.C. supervised this work.

Competing interests:

The authors A.C.P., P.L.R., E.S.B., J.D.B. and F.C. are inventors on patent application 16/043,950 submitted by Harvard College and Massachusetts Institute of Technology, which covers IGS technology. J.D.B. holds patents related to ATAC-seq and is on the scientific advisory board for Camp4, Seqwell, and Celsee. F.C. is a paid consultant for Celsius Therapeutics. G.M.C. is a co-founder and SAB member of ReadCoor and is an advisor to 10x Genomics following their acquisition of ReadCoor. Conflict of interest link for G.M.C: <http://arep.med.harvard.edu/gmc/tech.html>. R.J. is an advisor/co-founder of Fate Therapeutics, Fulcrum Therapeutics, Omega Therapeutics, and Dewpoint Therapeutics.

Data and materials availability:

Raw sequencing data is available from the Sequence Read Archive under accession SRR12938556 (PGP1f) and SRR12950073 (mouse embryo). Raw images are available from the Image Data Resource (<https://idr.openmicroscopy.org>) under accession number idr0101 (86). Code for image processing and computational integration of *in situ* and *ex situ* sequencing data is available in a Zenodo repository (87), and a Github repository (https://github.com/zchiang/in_situ_genome_sequencing_processing/). Code to reproduce the analyses in this manuscript is available in a Zenodo repository (88) and a Github repository (https://github.com/zchiang/in_situ_genome_sequencing_analysis/).

Hi-C data for PGP1f is available from the Gene Expression Omnibus (GEO) under accession GSE12355. Hi-C data from early mouse embryos is available from GEO accession GSE82185. DamID data from early mouse embryos is available from GEO accession GSE112551. Genotypes for the B6C3F1 and B6D2F1 strains are available from the Mouse Genomes Project (<https://www.sanger.ac.uk/data/mouse-genomes-project/>). Mouse whole-genome sequencing data is available from SRA accession SRR7511358.

¹⁰McGovern Institute, MIT, Cambridge, Massachusetts 02139, USA.

¹¹Koch Institute, MIT, Cambridge, Massachusetts 02139, USA.

¹²Howard Hughes Medical Institute, Cambridge, MA 02139, USA.

¹³Centers for Neurobiological Engineering and Extreme Bionics, MIT, Cambridge, MA 02139, USA.

Abstract

Understanding genome organization requires integration of DNA sequence and 3D spatial context, however, existing genome-wide methods lack either base-pair sequence resolution or direct spatial localization. Here, we describe *in situ* genome sequencing (IGS), a method for simultaneously sequencing and imaging genomes within intact biological samples. We applied IGS to human fibroblasts and early mouse embryos, spatially localizing thousands of genomic loci in individual nuclei. Using these data, we characterized parent-specific changes in genome structure across embryonic stages, revealed single-cell chromatin domains in zygotes, and uncovered epigenetic memory of global chromosome positioning within individual embryos. These results demonstrate how *in situ* genome sequencing can directly connect sequence and structure across length scales from single base pairs to whole organisms.

One Sentence Summary:

In situ genome sequencing (IGS) enables simultaneous sequencing and imaging of genomes in intact samples, uncovering genome-wide organizing principles across length scales in cells and early embryos.

The genome of an organism encodes not only its genes, but also principles of spatial organization that regulate gene expression and control cellular function (1, 2). Accordingly, mapping spatial genome organization at high resolution is important for understanding its diverse regulatory roles in health, disease, and development (3, 4). Principles of genome architecture have mostly been uncovered by methods based on DNA sequencing of chromatin contacts (5), such as Hi-C (6), and methods which probe targeted genomic loci using microscopy, such as DNA fluorescence *in situ* hybridization (FISH) (7). Hi-C applied to populations of cells has revealed genome-wide organizing principles (8–12), and single-cell variations have uncovered cell-to-cell heterogeneity (13–17). DNA FISH has similarly revealed genome architecture at single-cell resolution (18, 19). More recent studies have shown how these approaches can complement each other by imaging Hi-C defined features in single cells, characterizing their heterogeneity, and validating inferred differences in chromatin conformation within and across cell types (20–26)

However, these methods cannot currently be applied jointly on the same cell, and a method to simultaneously sequence and image genomes in single cells is lacking. Efforts which combine Hi-C with microscopy (16, 27), or efforts which make FISH more like sequencing via single-nucleotide polymorphism (SNP) specific probes (23, 28), have broken important conceptual ground, but they remain limited in their imaging or sequencing throughput.

Accordingly, questions requiring both genomic and spatial analysis in single cells have been difficult to address.

***In situ* genome sequencing workflow**

Here we present a method for *in situ* genome sequencing (IGS). IGS enables DNA sequencing directly within intact biological samples, spatially localizing genome-wide paired-end sequences in their endogenous context and thus bridging sequencing and imaging modalities for mapping genomes. Our *in situ* sequencing workflow introduces innovations in three phases: *in situ* library construction, multimodal sequencing of libraries, and computational integration of spatial and genetic information.

In the first phase, we create an *in situ* sequencing library within fixed samples by amplifying an untargeted sampling of the genome in its native spatial context. To do this, we fix and treat samples using methods optimized for DNA FISH (28, 29). Next, we use Tn5 transposase to randomly incorporate DNA sequencing adaptors into fixed genomic DNA by *in situ* transposition, preserving genomic fragments in their native spatial positions (30). We circularize these fragments *in situ* by ligation of two DNA hairpins containing a unique molecular identifier (UMI) and primer sites used for subsequent multi-modal DNA sequencing (Fig. 1A ii, iii). We then clonally amplify the resulting circular templates by rolling circle amplification, yielding *in situ* DNA sequencing libraries with up to thousands of spatially-localized amplicons per nucleus (Fig. S1). We also developed a method for modulating the effective density of sequencing libraries to optimize the number of resolvable amplicons (Fig. S2). Together, this provides an approach to clonally amplify untargeted samples of a genome, creating approximately 400–500 nm sized features for *in situ* sequencing (Fig. S3).

In the second phase of our workflow, we sought to use reported (31) *in situ* sequencing protocols to determine the sequence and 3D positions of amplicons. However, current *in situ* sequencing methods yield short single-end reads (at most 30 bases), and are limited by imaging time (31). This poses a challenge for genome sequencing: the human genome encodes 3 billion bases and includes highly repetitive regions, requiring long paired-end sequencing reads to resolve many regions of the genome. To address this challenge, we combined *in situ* sequencing with high-throughput paired-end DNA sequencing. To do this, we first read amplicon-specific UMIs within fixed samples using sequential rounds of *in situ* sequencing by ligation (SBL) and fluorescence imaging (31) (Fig. 1B i, Movie S1). Immunostaining followed by additional cycles of imaging may also be performed following *in situ* sequencing. We then dissociate the *in situ* amplicons and amplify them using PCR to produce an *in vitro* sequencing library (Fig. 1B ii, iii), which we sequence on a conventional Illumina sequencer (henceforth referred to as *ex situ* sequencing) to obtain 150 bp paired-end genomic reads tagged with *in situ* sequenced, spatially-resolved UMIs. This multimodal sequencing strategy allows us to preserve spatial information while leveraging the accuracy and read-length of paired-end sequencing on the Illumina platform, which is crucial for aligning individual reads to millions of unique genomic loci.

In the third phase, we computationally match *ex situ* paired-end sequencing reads to *in situ* amplicon positions. Briefly, we deconvolve, register, and normalize fluorescence images in order to resolve the 3D centers of amplicons across multiple rounds of imaging (Fig. S4, (29)). We then quantify the fluorescence signal of each UMI-associated amplicon across four color channels over all rounds of *in situ* sequencing (Fig. 1B i, Fig. S5). The *ex situ* sequenced reads are next associated with spatial coordinates within nuclei through error-robust matching of *in situ* and *ex situ* sequenced UMIs (Fig. 1C). To do this, we implement a probabilistic matching approach using principles from single-bit error correction (32) (Fig. S5). Collectively, the integration of these methods, which include developments across library construction, sequencing, and computational analyses, enable IGS as a general strategy for spatially mapping paired-end reads (29).

Here, we apply IGS to 106 human fibroblasts (PGP1f) and 113 cells across 57 intact early mouse embryos at the PN4 zygote (n = 24), late 2-cell (n = 20), and early 4-cell (n = 13) stages of development (Fig. 1D). Across both experiments, 66.35% of clearly resolvable amplicons (87.6% in PGP1f, 61.0% in mouse embryos) were confidently matched to an *ex situ* genomic read (Fig. S6). After cell filtering based on yield, karyotype, developmental stage, and cell cycle (29), this yielded a total of 286,335 spatially-localized genomic reads (36,602 in PGP1f, Table S1; 249,733 in mouse embryos, Table S2) with a UMI-matching false discovery rate of 0.26% (1.70% in PGP1f, 0.05% in mouse embryo, (29)). Mapped amplicons scaled with nuclear volume, spanning a median of 328 ± 114 reads per nucleus ($\pm$ SD) in the PGP1f cells, to a median of $3,909 \pm 2,116$, $2,357 \pm 1,063$, and $1,074 \pm 622$ reads per nucleus in zygote, 2-cell, and 4-cell stage embryos, respectively (Fig. S7). Sequencing coverage across the hg38 and mm10 reference genomes was comparable to whole genome sequencing (Fig. S8), and genomic reads did not show bias based on radial position (Fig. S9) or chromatin accessibility (Fig. S10). We also quantified the rate of detection for each genomic region across individual cells, as well as the distribution of genomic distance between sampled loci on the same chromosome (Fig. S11).

For downstream analyses, we annotated each read based on spatial features such as inclusion in chromosome territories and distance to nuclear landmarks (nuclear lamina, centromeres, and nucleolar precursor bodies), as well as published genomic data including A/B compartments, lamina-associated domains, and GC content (Table S3). The full embryo dataset can be interactively visualized at <https://buenrostrolab.shinyapps.io/insituseq/>.

Validation of *in situ* genome sequencing in cultured cells

To validate that our method detects features of spatial genome organization, we first examined the locations of chromosomes in interphase human male PGP1 fibroblasts (Fig. 2A,B, Movie S2). We found that autosomal reads displayed a strong tendency to spatially colocalize into two distinct spatial regions, while allosomal reads were restricted to one region, confirming the known organization of chromosomes into territories (18) (Fig. S12, Movie S3). To systematically define these territories, we used a maximum likelihood estimation approach to assign reads to homologous chromosome clusters using both the spatial and genomic positions of each read (Fig. S13). For chromosomes with two spatially-resolved homolog clusters, we found, via density-based thresholding, that 6.83% of reads

did not spatially colocalize with either cluster (Fig. S13), a larger fraction than our estimated 1.70% UMI-matching false discovery rate in PGP1f, which may be associated with long-range chromosome looping (33). Following spatial clustering, we visualized genome-wide conformations of individual chromosomes in single diploid cells by connecting the reads in each cluster according to genomic position (Fig. 2C).

We proceeded to characterize the positions of diploid chromosome territories across single cells by calculating the average spatial distance between pairwise genomic locations across the genome, at 10 Mb resolution. We found that blocks of short intrachromosomal distances were strongly delineated along the diagonal of this pairwise distance matrix (Fig. 2D), consistent with genomic organization into chromosome territories described above. The matrix also shows enrichment of shorter pairwise distances between smaller chromosomes. Additionally, we observed a positive association between chromosome size and radial distance from the nuclear center (Fig. 2E). These observations indicate that small chromosomes tend to be in closer proximity near the nuclear center, consistent with prior studies in human fibroblasts (34). These results illustrate the ability of IGS to resolve diploid chromosome territories within the nuclei of single cells and to investigate the spatial positioning of chromosomes at scale.

Repetitive DNA elements, such as transposons and endogenous retroviruses, make up approximately 50% of the human genome (35, 36), and their localization is known to play a role in normal (37) and disease-associated (38) genome organization. Although FISH-based methods can measure the localization of targeted classes of repetitive sequences (34, 39), current approaches have not simultaneously mapped the localization of many classes of repetitive sequences across the nucleus. We applied IGS to simultaneously measure the localization of repetitive sequences across the genome. We focused on the 13.9% of spatially-resolved reads that do not uniquely align to the reference genome (hg38), and aggregated them into ~250 classes of repetitive elements using Repbase (40) (Fig. S14A). We found that the number of reads associated with each element was proportional to its observed frequency in hg38 (Fig. S14B), enabling an unbiased approach to studying localization of repetitive elements.

Given our observations of radial patterns of chromosome positioning and the known radial organization of heterochromatin (41), we investigated whether repetitive elements displayed radial patterns. To do this, we compared the radial distribution of reads containing repetitive elements to permuted distributions to identify classes of repetitive elements with the strongest radial bias (Fig. 2F, (29)). We confirmed reports that Alu elements are depleted ~1 μ m from the nuclear edge (34), further validating our approach (Fig. 2G). We also found that certain types of repetitive heterochromatin, such as satellite DNA, show enrichment towards the nuclear center, while others, including AT-rich L1 elements, are overrepresented at the nuclear edge. These findings demonstrate the ability of IGS to simultaneously identify spatial localization patterns of many different repetitive sequences *de novo* in an untargeted and genome-wide manner.

Having shown that IGS confirms known features of global genome organization, we next asked whether we could characterize the structures of individual chromosomes. DNA FISH

and Hi-C studies have found that chromosome arms can be individually compartmentalized in a fashion similar to chromosome territories (44, 45). Independent localization of chromosome arms was apparent as stripes in the genome-wide distance matrix (Fig. 2D), and could be visualized in single chromosomes colored by their p and q arms (Fig. 2H). We computed pairwise distances for reads within each chromosome territory (Chrs 1–11 and Chr X, Fig. S15) and fit a power law to this relationship as described (20). Separate treatment of p and q arms resulted in an improved fit compared to fitting of all of Chr 1 (Fig. S16), in line with the expectation of differential scaling across the centromere. Indeed, intra-arm and inter-arm pairs of loci exhibited two scaling regimes when treated separately (Fig. 2I, Fig. S17). In chromosomes where we had high coverage (Chr 1–11 and Chr X), inter-arm distances were significantly larger than intra-arm distance for the range of genomic distances present in both distributions (Fig. 2J, 56–87 Mb for Chr 1; boxed region, Fig. 2I; K-S test, $p < 10^{-16}$; Fig. S18). These results extend observations of spatially polarized chromosome arms (42), and demonstrate the ability of IGS to characterize subchromosomal spatial structure. Taken together, these findings highlight the unique ability of IGS to simultaneously interrogate broad features of genome organization, including chromosome positioning, chromosome folding, and the localization patterns of repetitive sequences.

***In situ* genome sequencing in intact early mouse embryos**

The spatial organization of the genome is extensively remodeled in early embryogenesis, as the initially separate parental genomes undergo major reorganization after fertilization to prime the organism for zygotic genome activation (ZGA) (43) and, subsequently, lineage-specific cell fates (44). Studies have linked chromatin and epigenetic remodeling to various phenomena including sequence-specific localization of chromatin to nuclear landmarks (45, 46), parent specific-chromatin domain organization in single cells (47–50), and developmental specification of clonal lineages (51–53). Given the importance of spatial features, sequence-specificity, and intercellular relationships in these phenomena, we sought to apply IGS in intact early embryos to characterize genome organization in early embryogenesis across length scales. We applied IGS to intact early mouse embryos (B6C3F1 females x B6D2F1 males) spanning the PN4 zygote ($3,909 \pm 2,116$ reads/cell, median $\pm$ SD), late 2-cell ($2,357 \pm 1,063$ reads/cell), and early 4-cell ($1,074 \pm 622$ reads/cell) stages of development (Fig. 3A, Movie S4).

Collectively, imaging and sequencing methods have shown that some of the structural changes in early development are associated with nuclear landmarks, such as the centromeres (54), nuclear lamina (46) and nucleolus precursor bodies (NPBs) (55). To demonstrate that IGS is compatible with other imaging modalities and to investigate the organizational roles of these landmarks, we performed co-immunostaining for CENP-A (centromere) and Lamin-B1 (nuclear lamina), in addition to staining with DAPI (used to locate NPBs). The resulting images were segmented and registered to the *in situ* sequencing data in 3D, enabling us to directly localize genomic reads relative to these landmarks (Fig. 3B, (29)).

In order to resolve the maternal and paternal genomes within single cells, we first confirmed the presence of chromosome territories in all stages. As with the PGP1 fibroblasts, we found

that reads originating from a particular autosome could generally be separated into two distinct spatial clusters per nucleus (or one cluster per allosome in male embryos) (Movie S5). We then filtered cells based on yield, karyotype, developmental stage, and cell cycle (Fig. S19, (29). After filtering, we found a nearly equivalent rate of reads that did not colocalize with chromosome territories as in PGP1f (6.95%). Relative to reads within territories, these non-colocalizing reads were significantly depleted from regions proximal to the nuclear lamina and NPBs (Fig. S20; K-S test, $p < 10^{-51}$ and 10^{-109} respectively).

To assign parent-of origin to each territory, we identified spatially-localized reads that overlapped a genomic position with a heterozygous SNP in either of the parental strains. 1.40% and 1.64% of genomic reads were uniquely assigned to the maternal (B6C3F1) and paternal (B6D2F1) genomes respectively, resulting in an average of 67 haplotype-informative reads per cell. To validate these assignments, we visualized the positions of haplotype-informative reads in the PN4 zygote (Fig. 3C). At this stage of development, the parental genomes remain segregated in the larger paternal and smaller maternal pronuclei. Based on this known feature, we assigned each chromosome territory to either the maternal or paternal genome in a semi-supervised manner (Fig. 3D). We found that 97.1% of haplotype-informative reads were concordant with this assignment, where non-concordant reads may be attributable to genomic sequencing errors, UMI matching errors, or strain impurities. We then used our haplotype-informative reads to assign entire chromosome territories to parent-of-origin across the 2-cell and 4-cell stage embryos (29). This approach enables a strong majority of reads (82.26%), even those not overlapping a SNP, to be assigned to parent-of-origin through co-localization with haplotype-resolved reads in the same territory.

Developmental transitions in embryonic genome organization

We then sought to examine previously-described principles of global genome organization, focusing on parental haplotype (56, 57), centromere-telomere position, (54), (16), and GC content (6, 16) (Fig. 4A).

We began by examining the spatial separation of parental genomes, as imaging studies have shown that maternal and paternal chromatin are spatially polarized in early embryos (56, 57). To quantify the spatial separation of parental genomes across developmental stages, we analyzed the spatial inter-chromosomal neighbors of each read, and calculated a haplotype spatial separation score (Fig. 4B). We then averaged the separation scores for all reads in each cell. We found that the mean separation score significantly decreased between the zygote and 2-cell stages and between the 2-cell and 4-cell stages (K-S test, $p < 10^{-8}$ and $p < 10^{-3}$, Fig. 4C), consistent with earlier studies (56, 57). The standard deviation of mean separation scores increased with each developmental stage (SD = 0.015 for zygotes, 0.026 for 2-cell, 0.045 for 4-cell), indicating that the degree of parental genome intermixing is heterogeneous within the embryo. Further, we observed no evidence that particular chromosomes were more likely to break this separation than others (Fig. S21). These results are concordant with the hypothesis that gradual mixing is a consequence of global chromosome repositioning following mitosis.

We next examined global spatial organization of the genome along the centromere-telomere axes of chromosomes. Mouse chromosomes are acrocentric and are known to be arranged in a Rabl-like configuration in early embryos, in which centromeres cluster toward one side of the nucleus and distal telomeres cluster toward the other (54, 56). To confirm this configuration in our data, we first measured the polarity of the CENP-A stain and found that centromeres in the 2 and 4-cell stages were significantly clustered toward one side of the nucleus (K-S test, $p < 10^{-4}$ and $p < 10^{-8}$). To analyze this configuration for all chromosome positions, we assigned each read a centromere-telomere score based on its genomic position along its chromosome. When we visualized these scores in a nucleus from a 2-cell embryo, we observed that the centromere-telomere scores were highly polarized, which was supported by co-localization of the CENP-A immunostain (Fig. 4D). To quantify this polarization across all stages, we calculated a spatial neighborhood centromere-telomere score for each read (29). We then examined the relationship between centromere-telomere scores and neighborhood scores across all reads, and observed much stronger correlation in the 2-cell and 4-cell stages (Pearson's $r = 0.519$ and 0.502) than in the zygote ($r = 0.074$, Fig. 4E). The functional consequences of this transition to a Rabl-like configuration in 2- and 4-cell embryos remains unclear. In other contexts, this configuration is thought to be an extension of anaphase chromosome positioning into interphase, perhaps without cellular function (58). On the other hand, it has also been hypothesized to restrict chromatin entanglement (59); thus, it may be involved in constraining genome structure to enable the short cell cycles of the early embryo.

Finally, we examined the role of GC content in genome structure, which is strongly associated with A/B compartmentalization (6). To study this effect, we first visualized individual homologs of Chr 12 from zygotic pronuclei (Fig. 4F). We observed that genomic reads from GC-poor regions tended to localize to the periphery of the nucleus, in line with reports describing the localization of the inactive B compartment (16). To quantify this effect, we measured the correlation between GC content and distance to nuclear lamina across Chr 12 in all zygotes and observed that these two factors were correlated in both the paternal and maternal homologs (Spearman's $\rho = 0.794, 0.649$ respectively, Fig. 4G). We applied this approach to all chromosomes and found that the paternal homologs were significantly more correlated than their maternal counterparts in the zygote, but not in the 2-cell stage (Fig 4H; K-S test, $p < 10^{-5}$, n.s.). This suggests that the degree of GC compartmentalization may be influenced by the differing biological histories of the pronuclei (47).

Intriguingly, when we examined the relationship between genomic position and distance to nuclear landmarks, we observed that Chr X seemed to localize especially far from the nuclear lamina and toward the NPBs in paternal pronuclei (Fig. S22). This finding extends models of the role of NPBs in establishing epigenetic asymmetry between parental X chromosomes rapidly after fertilization and in advance of imprinted X inactivation (45). Taken together, these results demonstrate the ability of IGS to characterize 3D genome organization across diverse developmental stages with parent-specific resolution, and with respect to nuclear landmarks.

Detection of single-cell domains chromatin domains in zygotes

Next, we used IGS to examine subchromosomal spatial organization. We focused on our data with the highest genomic resolution, the zygotic pronuclei, where large-scale parent-specific reorganization of chromatin is thought to play an important role in ZGA (60). First, we characterized the scaling relationship between mean spatial and genomic pairwise distance in the zygotic parental genomes. Consistent with previous reports (47), we found that each parental genome had distinct scaling properties (Fig. 5A, 5B).

Reports analyzing genome structure in zygotes have suggested that paternal chromatin exhibits unusually weak higher order structure (> 2 Mb) (47–49). In accordance with these reports, we found that spatial distance matrices generated from the population ensemble indeed exhibited little off-diagonal structure (Fig. 5C, S23 all chromosomes). Furthermore, evidence suggests that paternal zygotic chromatin exhibits unusually weak A/B compartmentalization (47–49), and unusually large lamina associated domains (46). Our ensemble data corroborated these reports and correlated well with Hi-C (mean Pearson's $r = 0.84$) when analyzed in terms of lamin proximity (Fig. 5D, S24 all chromosomes). However, when we examined single-cell distance matrices, we found that, unlike the ensemble, single paternal pronuclei generally exhibited large blocks of spatially associated chromatin (Fig. 5E, left). To distinguish these blocks from population-defined topological domains identified in Hi-C studies, we term them single-cell domains (SCDs).

To better understand the nature of SCDs in paternal zygotes, we systematically identified individual domains in single cells (29). The SCDs we identified corresponded well to spatially distinct clusters identified by visual inspection (Fig. 5E, right). When we examined SCDs across cells and chromosomes, we observed that they were large (median size 17.5 Mb, 10 Mb inter-quartile range) relative to canonical features defined by Hi-C, and had heterogeneous sizes and boundary positions (Fig. S25). We proceeded to assess the strength of all SCD boundaries in single cells, and found they were significantly stronger and more variable than boundaries identified by the same method in the ensemble matrices (Fig. 5F; K-S test, $p < 10^{-17}$; 95% CI for Cohen's d , (0.56, 0.77), determined by bootstrapping (29)). Together, these observations suggest that the weak ensemble structure may be explained by the variability of single-cell structures. Finally, we investigated the association of SCDs with nuclear landmarks, which may suggest organizing principles, and found that their boundaries and interiors were, respectively, significantly more lamin distal and more lamin proximal than expected (Fig. 5G (29)). We found this striking in light of previous electron microscopy studies showing discrete micron-scale lamina-associated chromatin domains following ZGA (61).

Recently, both polymer simulations (47) and direct observation of chromatin structure (22) have shown how variable domain-like structures can exist in single cells when higher order ensemble structure is lacking. We speculate that, given the weak ensemble structure, the SCDs observed here may involve a similar phenomenon. The SCDs described here are larger than canonical Hi-C defined features, and were not detected in earlier single-cell Hi-C studies in zygotes (47, 50), perhaps because they are organized on length scales which are relatively less accessible to Hi-C measurements. It may be interesting to investigate the

extent to which SCDs are governed by mechanisms related to the nuclear lamina, which perhaps modulates underlying epigenetic (41) or polymer-intrinsic (47) domain-forming behaviors of chromatin.

Epigenetic memory of global chromosome positioning within early embryos

Embryonic development is thought to involve epigenetic transmission of structural and regulatory features of chromatin organization through clonal cell lineages within individual embryos (53). These mechanisms play an important role in breaking initial symmetry (51), engaging clonal lineage-specific gene expression programs (62), and cell fate commitment (44). Intercellular asymmetries influencing the developmental fate of clonal lineages have been reported as early as the 4-cell stage within individual embryos (51, 52, 62, 63).

In order to study clonal lineage-specific features at the single-cell level, it is necessary to resolve and compare cells within the same embryo. Chromosome territories have been found to form early in interphase and subsequently maintain their relative positions until prophase (18, 64, 65), so we reasoned that comparison of chromosome positioning would be a robust way to quantify the similarity of global genome organization between interphase cells. Live-cell studies using non-specific photopatterning of the nucleus have demonstrated similarity in global genome organization between sister cells in culture (65), and indeed, visual inspection of pairs of chromosomes suggested that cells within an embryo share similar chromosomal positions (Fig. 6A). We quantified similarity by comparing single-cell autosome distance matrices of pairs of cells within and between individual embryos (Fig. 6A, B, (29)). In 2-cell embryos, we found that global chromosome positioning in sister cells was significantly more correlated than in pairs of cells from different embryos (K-S test, $p < 10^{-15}$; Fig. 6C, Fig. S26). These results suggest that cells within 2-cell embryos may share memory of their common initial chromosome positioning during zygotic metaphase, if not earlier.

Next, we asked whether the similarity shared by sister cells in 2-cell embryos might be epigenetically transmitted across the second cell division, i.e. to cousin cells. While earlier work has not found heritability in the radial positioning of individual loci (66, 67), widely varying degrees of similarity in global genome organization between mother and daughter cells have been reported (64, 65). We constructed putative clonal lineage trees within each 4-cell embryo by using the ranked correlations of autosome distance matrices for each pair of cells to classify putative sister and cousin cells. The most correlated pair of cells in each embryo was designated as one set of putative sister cells, thus implying the remainder of the tree (Fig. 6D, Fig S27, (29)). As expected by the definition of the tree, we found that global chromosome positioning was significantly more correlated between putative sister cells than between pairs of cells from different 4-cell embryos. However, we also found that the same held true for putative cousin cells, which was not expected (Fig. 6E, Fig. S26; K-S test, $p < 10^{-14}$ and $p < 10^{-3}$; 95% CI for Cohen's d , (0.55, 1.29), determined by bootstrapping, (29)). Together, these results demonstrate clonal lineage-specific similarity in global chromosome positioning in early embryos, and imply that epigenetic memory of chromosome positioning

is transmitted from mother to daughter cells during the second cleavage (Fig. 6F). While the mechanisms are not fully clear, this memory may reflect minimal relative repositioning of chromosomes during congression to and departure from the metaphase plate, perhaps due to the Rab1 configuration and short cell cycles of the early embryo. This mitotic heritability of global chromosome positioning may influence processes that affect the viability and phenotype of the developing organism, such as rates of homologous recombination (HR)-mediated double-strand break (DSB) repair (68, 69) and the distribution of translocations (70–72) in the early embryo.

Discussion:

In this work we present *in situ* genome sequencing, unifying sequencing and imaging of genomes in intact samples. This unified approach enables *de novo* discovery of spatial organization of genomes across length scales, from single-cell subchromosomal domains to intercellular relationships. Because IGS is both sequencing and imaging-based, it can be extended in either modality based on the needs of specific experiments. We demonstrate this in early mouse embryos through integration of genotype information to spatially resolve the maternal and paternal genomes, through integration of immunofluorescence to localize genomic loci relative to nuclear landmarks, and by using whole-embryo spatial information to infer clonal lineages. This contextual information enabled us to uncover single-cell chromatin domains in zygotes with lamin-distal boundary positions and lamin-proximal interiors, as well as heritable correlations in global chromosome positioning within single early embryos.

While extant methods such as multiplexed DNA FISH and single-cell Hi-C are well-suited to measuring aspects of spatial genome organization, they cannot currently be combined in the same cell. With IGS, we spatially localize hundreds to thousands of genomic loci in single cells, achieving genomic resolutions comparable to recent genome-wide approaches based on targeted DNA FISH in fibroblasts (73, 74). Unlike these targeted methods, we further show how IGS can perform untargeted sequence localization, resolve genome structure of maternal and paternal alleles, and be applied in 3D nuclei and thick intact samples. However, due to its genome-wide sampling frequency (at most ~1 Mb in this report), IGS is currently limited in its ability to systematically examine specific genetic loci in specific cells. Targeted DNA FISH or single-cell Hi-C are thus currently more appropriate for applications requiring high-resolution interrogation of genomic features such as TAD boundaries (22, 47, 75) or enhancer-promoter loops (16, 26), and chromosome painting methods (76) may be preferred when high-throughput visualization of chromosomes territories is required. IGS therefore joins an ecosystem of conceptually new approaches (77–79) complementary to these more well-established methods.

We expect that improvements to IGS will further enable the study of genome sequence, structure, and function. In addition to the cultured cells and early embryos presented here, we anticipate extension of IGS to a broader range of cell types and intact tissues. Outside of development, IGS may be well-suited to study cancers, in particular how copy number instability and translocations contribute to tumor heterogeneity and alter nuclear morphology (80). The transposase-based library construction used in IGS may also be extended to

measure the spatial localization of the accessible genome (30, 81). Further, because nuclear volume is the primary constraint on the amplicon yield of IGS, we anticipate many-fold improvements in yield and resolution, either through smaller amplicons (82), or preferably through integration of IGS with Expansion Microscopy (83, 84), which simultaneously increases nuclear volume by 50–100-fold and enables super-resolved imaging by physical expansion of samples. Finally, following our proof-of-concept integration with immunostaining, we expect increasingly multiplexed multi-omic (85) variations of IGS will be possible. We anticipate *in situ* genome sequencing will be instrumental in unifying genomics and microscopy, and therefore sequence and structure.

Methods summary

Library construction

Cells were grown and fixed on a glass coverslip, and embryos were fixed in solution and immobilized in polyacrylamide gel in a 6-well plate. Phosphorylated DNA adapters were inserted into fixed genomic DNA *in situ* by incubating samples 1 hr (cells) or overnight (embryos) with transposase. Hairpins were hybridized to the adapters on either side of the insert, and the complex was circularized by gap-fill ligation. Hairpins contained either a UMI and primers for *in situ* and *ex situ* sequencing, or an RCA primer hybridization site. RCA primers were hybridized, and RCA was performed overnight, with aminoallyl-dUTP spiked into the reaction. Amplicons were crosslinked by reacting with BSPEG(9).

In situ sequencing

For cells, the coverslip was mounted in a flow cell and *in situ* sequencing reactions were performed using automated fluidics. For embryos, reactions were performed manually in a 6-well plate. Samples were treated with calf intestinal phosphatase before the first primer hybridization and before each cleavage reaction. *In situ* sequencing was performed using sequencing-by-ligation chemistry. Samples were exchanged into an imaging buffer following each round of sequencing. Images were acquired using confocal microscopy. Immunostaining and immunofluorescence imaging were performed after *in situ* sequencing.

Ex situ sequencing

Samples were transferred into solution and used as input to a PCR reaction. The resulting library was sequenced using high-throughput paired-end sequencing and then aligned to hg38 for PGP1f and mm10 for the early mouse embryos. Aligned reads that overlapped alleles that differ between the B6C3F1 and B6D2F1 strains were annotated as haplotype-informative.

Computational integration of *in situ* and *ex situ* sequencing data

In situ and *ex situ* sequencing data were computationally integrated using a probabilistic UMI matching approach that borrows principles from single-bit error correction to account for signal intermingling between densely-packed amplicons, as well as the decay of sequencing quality over successive rounds of *in situ* sequencing.

Detailed methods for sample preparation, library construction, multimodal sequencing, imaging, image analysis, and all analyses of data are described in the Materials and Methods section of the Supplementary Material. Oligonucleotide sequences are found in Table S4. A description of methods for kit-free synthesis of *in situ* sequencing reagents is given in the Supplementary Text and Tables S5 and S6. A description of cost, complexity and throughput of the method is found in the Supplementary Text and Table S7.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We thank the Church, Boyden, Buenrostro, and Chen labs for useful discussions. We thank S. Alon, R. Kalhor, D. Lee, M. Villanueva, M. Aryee, T. Zhao, P. Wiggins, W. Xie, B. Zhang, T. Wu, and E. Lieberman-Aiden for useful suggestions. We thank J. Strecker for purified Tn5 transposase.

Funding:

J.D.B. and F.C. acknowledge funding from the Allen Institute Distinguished Investigator award and funding from the NIH R21HG009749. F.C. also acknowledges funding from NIH DP5OD024583. E.S.B. acknowledges, for funding, John Doerr, Lisa Yang, the Open Philanthropy Project, NIH 1R01NS087950, NIH 1RM1HG008525, NIH 1R01DA045549, NIH 2R01DA029639, NIH 1R01NS102727, NIH 1U19MH114821, NIH 1R01EB024261, NIH 1R01MH110932, HHMI, the HHMI–Simons Faculty Scholars Program, IARPA D16PC00008, U.S. Army Research Laboratory and the U. S. Army Research Office under contract/grant W911NF1510548, U.S.–Israel Binational Science Foundation Grant 2014509, and NIH Director's Pioneer Award 1DP1NS087724. RJ was supported by grants from the NIH (NIH grant R37 CA084198 and R01 MH104610–20). G.M.C acknowledges funding from NIH RM1HG008525. A.C.P. and P.R. acknowledge funding from NSERC PGS-D awards. Z.D.C. acknowledges funding from NHGRI training grant T32HG002295 and the Harvard Quantitative Biology Initiative.

References and Notes

1. Bonev B, Cavalli G, Organization and function of the 3D genome. *Nature Reviews Genetics* 17 (2016), pp. 661–678.
2. Rowley MJ, Corces VG, Organizational principles of 3D genome architecture. *Nat. Rev. Genet* 19, 789–800 (2018). [PubMed: 30367165]
3. Zheng H, Xie W, The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol* 20, 535–550 (2019). [PubMed: 31197269]
4. Spielmann M, Lupiáñez DG, Mundlos S, Structural variation in the 3D genome. *Nature Reviews Genetics* 19 (2018), pp. 453–467.
5. Dekker J, Capturing Chromosome Conformation. *Science* 295 (2002), pp. 1306–1311. [PubMed: 11847345]
6. Lieberman-Aiden E et al., Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293 (2009). [PubMed: 19815776]
7. Levsky JM, Singer RH, Fluorescence in situ hybridization: past, present and future. *J. Cell Sci* 116, 2833–2838 (2003). [PubMed: 12808017]
8. Dixon JR et al., Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380 (2012). [PubMed: 22495300]
9. Sexton T et al., Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* 148, 458–472 (2012). [PubMed: 22265598]
10. Rao SSP et al., A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680 (2014). [PubMed: 25497547]
11. Nora EP et al., Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385 (2012). [PubMed: 22495304]

12. Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L, Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol* 30, 90–98 (2011). [PubMed: 22198700]
13. Nagano T et al., Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59–64 (2013). [PubMed: 24067610]
14. Nagano T et al., Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* 547, 61–67 (2017). [PubMed: 28682332]
15. Ramani V et al., Massively multiplex single-cell Hi-C. *Nat. Methods* 14, 263–266 (2017). [PubMed: 28135255]
16. Stevens TJ et al., 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* 544, 59–64 (2017). [PubMed: 28289288]
17. Tan L, Xing D, Chang C-H, Li H, Xie XS, Three-dimensional genome structures of single diploid human cells. *Science* 361, 924–928 (2018). [PubMed: 30166492]
18. Cremer T, Cremer M, Chromosome territories. *Cold Spring Harb. Perspect. Biol* 2, a003889 (2010). [PubMed: 20300217]
19. Fraser J, Williamson I, Bickmore WA, Dostie J, An Overview of Genome Organization and How We Got There: from FISH to Hi-C. *Microbiology and Molecular Biology Reviews* 79 (2015), pp. 347–372. [PubMed: 26223848]
20. Wang S et al., Spatial organization of chromatin domains and compartments in single chromosomes. *Science* 353, 598–602 (2016). [PubMed: 27445307]
21. Cattoni DI et al., Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. *Nat. Commun* 8, 1–10 (2017). [PubMed: 28232747]
22. Bintu B et al., Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* 362 (2018), doi:10.1126/science.aau1783.
23. Nir G et al., Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genet* 14, e1007872 (2018). [PubMed: 30586358]
24. Szabo Q et al., TADs are 3D structural units of higher-order chromosome organization in *Drosophila*. *Science Advances* 4 (2018), doi:10.1126/sciadv.aar8082.
25. Finn EH et al., Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* 176, 1502–1515.e10 (2019). [PubMed: 30799036]
26. Mateo LJ et al., Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* 568, 49–54 (2019). [PubMed: 30886393]
27. Cardozo Gizzi AM et al., Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Mol. Cell* 74, 212–222.e5 (2019). [PubMed: 30795893]
28. Beliveau BJ et al., Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nat. Commun* 6, 7147 (2015). [PubMed: 25962338]
29. Materials and methods are available as supplementary materials.
30. Chen X et al., ATAC-se reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat. Methods* 13, 1013–1020 (2016). [PubMed: 27749837]
31. Lee JH et al., Highly multiplexed subcellular RNA sequencing in situ. *Science* 343, 1360–1363 (2014). [PubMed: 24578530]
32. Moffitt JR et al., High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. U. S. A* 113, 11046–11051 (2016). [PubMed: 27625426]
33. Volpi EV et al., Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J. Cell Sci* 113 (Pt 9), 1565–1576 (2000). [PubMed: 10751148]
34. Bolzer A et al. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol* 3, e157 (2005). [PubMed: 15839726]
35. de Koning APJ, Gu W, Castoe TA, Batzer MA, Pollock DD, Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* 7, e1002384 (2011). [PubMed: 22144907]

36. Miga KH et al., Telomere-to-telomere assembly of a complete human X chromosome. *Nature* (2020), 585, 79–84 (2020). [PubMed: 32663838]
37. Cournac A, Koszul R, Mozziconacci J, The 3D folding of metazoan genomes correlates with the association of similar repetitive elements. *Nucleic Acids Res* 44, 245–255 (2016). [PubMed: 26609133]
38. Casa V, Gabellini D, A repetitive elements perspective in Polycomb epigenetics. *Front. Genet* 3, 199 (2012). [PubMed: 23060903]
39. Hausmann M, Lee J-H, Sievers A, Krufczik M, Hildenbrand G, COMBINatorial Oligonucleotide FISH (COMBO-FISH) with Uniquely Binding Repetitive DNA Probes. *Methods Mol. Biol* 2175, 65–77 (2020). [PubMed: 32681484]
40. Bao W, Kojima KK, Kohany O, Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 6, 11 (2015). [PubMed: 26045719]
41. Falk M et al., Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* (2019), 570, 395–399 (2019). [PubMed: 31168090]
42. Muller H, Gil J Jr, Drinnenberg IA, The Impact of Centromeres on Spatial Genome Architecture. *Trends Genet* 35, 565–578 (2019). [PubMed: 31200946]
43. Hug CB, Vaquerizas JM, The Birth of the 3D Genome during Early Embryonic Development. *Trends Genet* 34, 903–914 (2018). [PubMed: 30292539]
44. Chazaud C, Yamanaka Y, Lineage specification in the mouse preimplantation embryo. *Development* 143, 1063–1074 (2016). [PubMed: 27048685]
45. Namekawa SH, Payer B, Huynh KD, Jaenisch R, Lee JT, Two-step imprinted X inactivation: repeat versus genic silencing in the mouse. *Mol. Cell. Biol* 30, 3187–3205 (2010). [PubMed: 20404085]
46. Borsos M et al., Genome–lamina interactions are established de novo in the early mouse embryo. *Nature* 569, 729–733 (2019). [PubMed: 31118510]
47. Flyamer IM et al., Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* 544, 110–114 (2017). [PubMed: 28355183]
48. Du Z et al., Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* 547, 232–235 (2017). [PubMed: 28703188]
49. Ke Y et al., 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. *Cell* 170, 367–381.e20 (2017). [PubMed: 28709003]
50. Collombet S et al., Parental-to-embryo switch of chromosome organization in early embryogenesis. *Nature* 580, 142–146 (2020). [PubMed: 32238933]
51. Torres-Padilla M-E, Parfitt D-E, Kouzarides T, Zernicka-Goetz M, Histone arginine methylation regulates pluripotency in the early mouse embryo. *Nature* 445, 214–218 (2007). [PubMed: 17215844]
52. Burton A et al., Single-cell profiling of epigenetic modifiers identifies PRDM14 as an inducer of cell fate in the mammalian embryo. *Cell Rep* 5, 687–701 (2013). [PubMed: 24183668]
53. Burton A, Torres-Padilla M-E, Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nat. Rev. Mol. Cell Biol* 15, 723–734 (2014). [PubMed: 25303116]
54. Aguirre-Lavin T et al., 3D-FISH analysis of embryonic nuclei in mouse highlights several abrupt changes of nuclear organization during preimplantation development. *BMC Dev. Biol* 12, 30 (2012). [PubMed: 23095683]
55. Fulka H, Aoki F, Nucleolus Precursor Bodies and Ribosome Biogenesis in Early Mammalian Embryos: Old Theories and New Discoveries. *Biol. Reprod* 94 (2016), doi:10.1095/biolreprod.115.136093.
56. Mayer W, Smith A, Fundele R, Haaf T, Spatial separation of parental genomes in preimplantation mouse embryos. *J. Cell Biol* 148, 629–634 (2000). [PubMed: 10684246]
57. Reichmann J et al., Dual-spindle formation in zygotes keeps parental genomes apart in early mammalian embryos. *Science* 361, 189–193 (2018). [PubMed: 30002254]
58. Cowan CR, Carlton PM, Cande WZ, The polar arrangement of telomeres in interphase and meiosis. Rabl organization and the bouquet. *Plant Physiol* 125, 532–538 (2001). [PubMed: 11161011]

59. Pouokam M et al., The Rab1 configuration limits topological entanglement of chromosomes in budding yeast. *Sci. Rep* 9, 6795 (2019). [PubMed: 31043625]
60. Lee MT, Bonneau AR, Giraldez AJ, Zygotic Genome Activation During the Maternal-to-Zygotic Transition. *Annual Review of Cell and Developmental Biology* 30 (2014), pp. 581–613.
61. Ahmed K et al., Global Chromatin Architecture Reflects Pluripotency and Lineage Commitment in the Early Mouse Embryo. *PLoS One* 5, e10531 (2010). [PubMed: 20479880]
62. Plachta N, Bollenbach T, Pease S, Fraser SE, Pantazis P, Oct4 kinetics predict cell lineage patterning in the early mammalian embryo. *Nat. Cell Biol* 13, 117–123 (2011). [PubMed: 21258368]
63. Piotrowska-Nitsche K, Perea-Gomez A, Haraguchi S, Zernicka-Goetz M, Four-cell stage mouse blastomeres have different developmental properties. *Development* 132, 479–490 (2005). [PubMed: 15634695]
64. Walter J, Schermelleh L, Cremer M, Tashiro S, Cremer T, Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during subsequent interphase stages. *J. Cell Biol* 160, 685–697 (2003). [PubMed: 12604593]
65. Gerlich D et al., Global chromosome positions are transmitted through mitosis in mammalian cells. *Cell* 112, 751–764 (2003). [PubMed: 12654243]
66. Kind J et al., Single-cell dynamics of genome-nuclear lamina interactions. *Cell* 153, 178–192 (2013). [PubMed: 23523135]
67. Thomson I, Gilchrist S, Bickmore WA, Chubb JR, The Radial Positioning of Chromatin Is Not Inherited through Mitosis but Is Established De Novo in Early G1. *Curr. Biol* 14, 166–172 (2004). [PubMed: 14738741]
68. Lee C-S et al., Chromosome position determines the success of double-strand break repair. *Proc. Natl. Acad. Sci. U. S. A* 113, E146–54 (2016). [PubMed: 26715752]
69. Phillips ER, McKinnon PJ, DNA double-strand break repair and development. *Oncogene* 26, 7799–7808 (2007). [PubMed: 18066093]
70. Vanneste E et al., Chromosome instability is common in human cleavage-stage embryos. *Nat. Med* 15, 577–583 (2009). [PubMed: 19396175]
71. Zhang Y et al., Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* 148, 908–921 (2012). [PubMed: 22341456]
72. Engreitz JM, Agarwala V, Mirny LA, Three-dimensional genome architecture influences partner selection for chromosomal translocations in human disease. *PLoS One* 7, e44196 (2012). [PubMed: 23028501]
73. Nguyen HQ et al., 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. *Nat. Methods* 17, 822–832 (2020). [PubMed: 32719531]
74. Su J-H, Zheng P, Kinrot SS, Bintu B, Zhuang X, Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* 182, 1641–1659.e26 (2020). [PubMed: 32822575]
75. Luppino JM et al., Cohesin promotes stochastic domain intermingling to ensure proper regulation of boundary-proximal genes. *Nat. Genet* 52, 840–848 (2020). [PubMed: 32572210]
76. Fields BD, Nguyen SC, Nir G, Kennedy S, A multiplexed DNA FISH strategy for assessing genome architecture in *Caenorhabditis elegans*. *Elife* 8 (2019), doi:10.7554/eLife.42823.
77. Beagrie RA et al., Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* 543, 519–524 (2017). [PubMed: 28273065]
78. Quinodoz SA et al., Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* 174 (2018), doi:10.1016/j.cell.2018.05.024.
79. Girelli G et al., GPSeq reveals the radial organization of chromatin in the cell nucleus. *Nat. Biotechnol*, 1–10 (2020). [PubMed: 31919444]
80. Zink D, Fischer AH, Nickerson JA, Nuclear structure in cancer cells. *Nat. Rev. Cancer* 4, 677–687 (2004). [PubMed: 15343274]
81. 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* 10, 1213–1218 (2013). [PubMed: 24097267]

82. Clausson C-M et al., Compaction of rolling circle amplification products increases signal integrity and signal-to-noise ratio. *Sci. Rep* 5, 12317 (2015). [PubMed: 26202090]
83. Chen F, Tillberg PW, Boyden ES, Expansion microscopy. *Science* 347, 543–548 (2015). [PubMed: 25592419]
84. Alon S et al., Expansion Sequencing: Spatially Precise In Situ Transcriptomics in Intact Biological Systems (2020), p. 2020.05.13.094268.
85. Zhu C, Preissl S, Ren B, Single-cell multimodal omics: the power of many. *Nat. Methods* 17, 11–14 (2020). [PubMed: 31907462]
86. Williams E et al., Image Data Resource: a bioimage data integration and publication platform. *Nat. Methods* 14, 775–781 (2017). [PubMed: 28775673]
87. Payne AC, Chiang ZD, Reginato PL, Mangiameli SM, zchiang/in_situ_genome_sequencing_processing: In situ genome sequencing processing code, Version 1.02, Zenodo (2020); 10.5281/zenodo.4299227.
88. Payne AC, Chiang ZD, Reginato PL, zchiang/in_situ_genome_sequencing_analysis: In situ genome sequencing analysis code, Version 1.0, Zenodo (2020); 10.5281/zenodo.4299125.
89. Solovei I et al., Spatial preservation of nuclear chromatin architecture during three-dimensional fluorescence in situ hybridization (3D-FISH). *Exp. Cell Res* 276, 10–23 (2002). [PubMed: 11978004]
90. Lareau CA et al., Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat. Biotechnol* 37, 916–924 (2019). [PubMed: 31235917]
91. Picelli S et al., Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* 24, 2033–2040 (2014). [PubMed: 25079858]
92. Lee JH et al., Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc* 10, 442–458 (2015). [PubMed: 25675209]
93. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ, ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol* 109, 21.29.1–21.29.9 (2015).
94. Langmead B, Salzberg SL, Langmead. 2013. Bowtie2. *Nat. Methods* 9, 357–359 (2013).
95. Smith T, Heger A, Sudbery I, UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* 27, 491–499 (2017). [PubMed: 28100584]
96. Massingham T, Goldman N, Error-correcting properties of the SOLiD Exact Call Chemistry. *BMC Bioinformatics* 13, 145 (2012). [PubMed: 22726842]
97. Berg S et al., ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* 16, 1226–1232 (2019). [PubMed: 31570887]
98. Ueda HR et al., Tissue clearing and its applications in neuroscience. *Nat. Rev. Neurosci* 21, 61–79 (2020). [PubMed: 31896771]
99. Quinlan AR, Hall IM, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842 (2010). [PubMed: 20110278]
100. Servant N et al., HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* 16, 259 (2015). [PubMed: 26619908]
101. Durand NC et al., Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* 3, 95–98 (2016). [PubMed: 27467249]

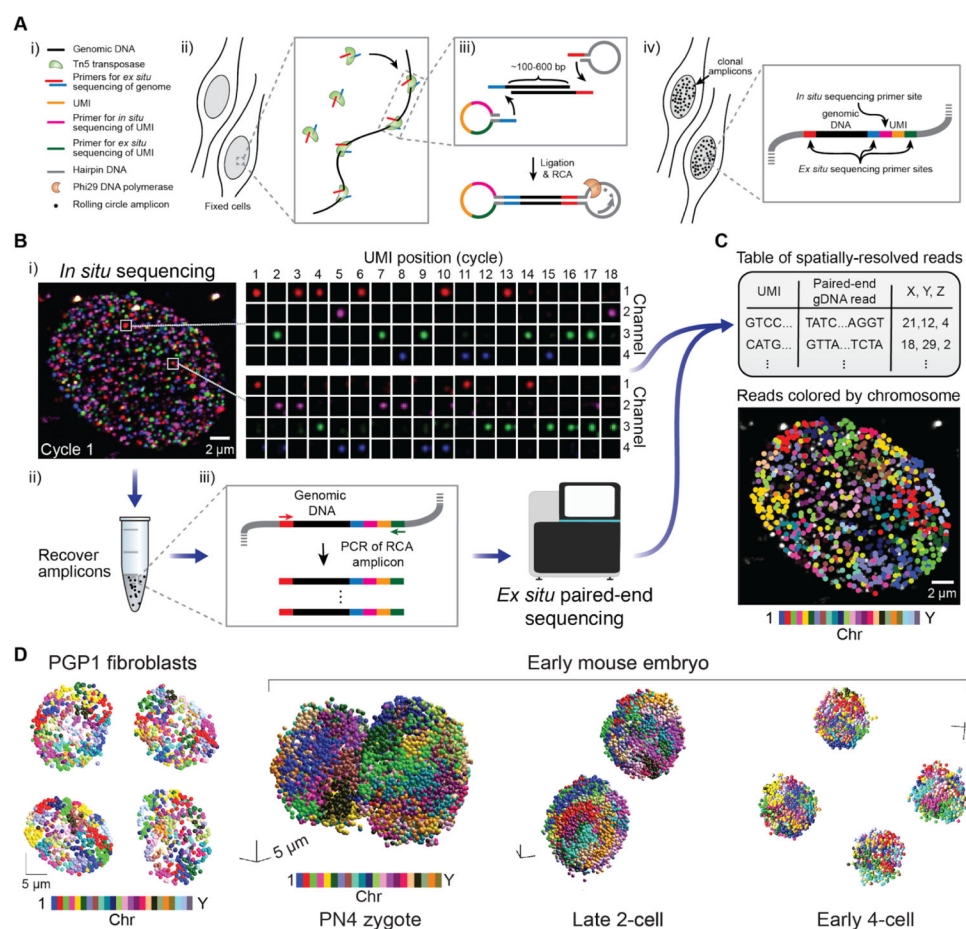


Figure 1: Method for *in situ* genome sequencing.

(A) *In situ* genomic DNA library construction. i) Legend. ii) Adaptor insertion. iii) Insert circularization by hairpin ligation, followed by *in situ* rolling circle amplification (RCA). iv) Clonal amplicons contain primers for *in situ* and *ex situ* sequencing. (B) Workflow for *in situ* genome sequencing. i) *In situ* sequencing localizes unique molecular identifiers (UMIs). 4-channel imaging of two representative amplicons over 18 rounds of *in situ* sequencing. ii) Amplicon dissociation following *in situ* sequencing. iii) PCR and *ex situ* sequencing of amplicons associates genomic sequences with UMIs. (C) Top: paired-end sequences are spatially localized by integrating *in situ* and *ex situ* sequencing data. Bottom: matched reads, colored by chromosome, are overlaid on their imaged amplicon library (below). (D) *In situ* sequenced nuclei from cultured fibroblasts and intact embryos at the PN4 zygote, late 2-cell, and early 4-cell stages, with spatially-localized reads colored by chromosome.

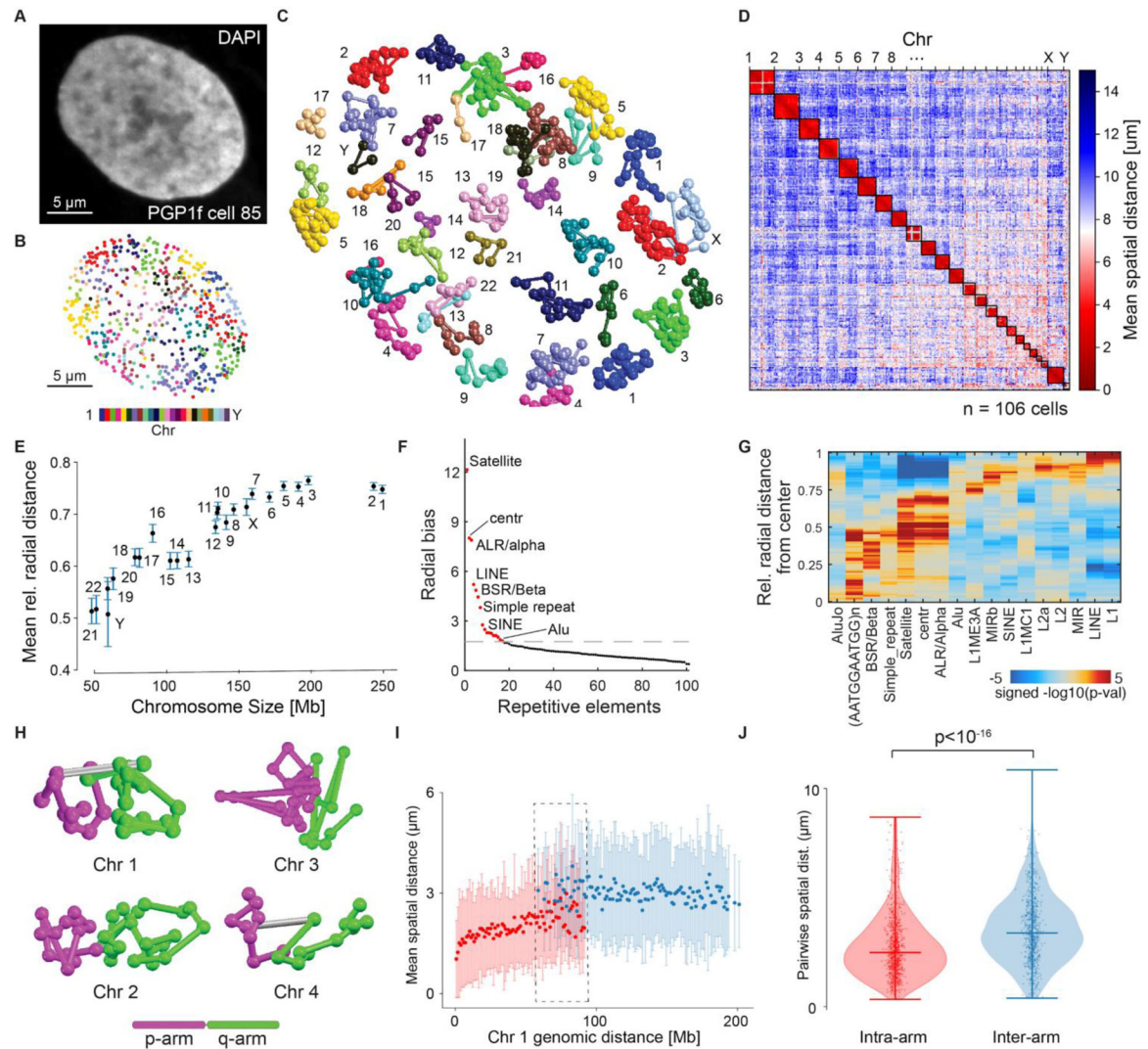


Figure 2: IGS characterizes spatial features of the human genome.

(A) DAPI stain of a PGP1f nucleus after *in situ* library construction. (B) 601 spatially-localized reads in the same PGP1f nucleus, colored by chromosome. (C) Exploded view reveals conformations of chromosome territories, shown as *in situ* reads (balls) connected according to sequential genomic position (sticks). (D) Genome-wide population mean pairwise distance matrix of 106 PGP1f cells binned at 10 Mb. (E) Chromosome size vs. normalized mean radial distance from the nuclear center for 106 diploid-resolved PGP1f cells. Error bars denote 95% confidence interval of the mean determined by bootstrapping. (F) The 103 most abundant repetitive elements ordered by radial bias, defined as the variability of binned distances relative to a permuted background from the nuclear center for 106 PGP1f cells. The dashed grey line represents the threshold for elements shown in (G). (G) Radial enrichment/depletion by binned distance from the nuclear center for the repetitive elements with the strongest radial bias from (F). (H) Ball-and-stick models for Chr 1–4 in the same single-cell, demonstrating spatial polarization between the p and q arms of each chromosome. (I) Genomic distance vs. spatial distance for Chr 1, distinguishing intra-arm and inter-arm measurements. Error bars: standard deviation. Dashed: range in which

both measurements can be compared at reasonable sampling depth ($n > 20$ per 1 Mb bin). (J) Intra-arm and inter-arm distance distributions in the dashed range in (I) are distributed differently ($n = 819$ intra-arm, 766 inter-arm, 144 Chr 1 territories, K-S test, $p < 10^{-16}$). Violin plot indicates median and range.

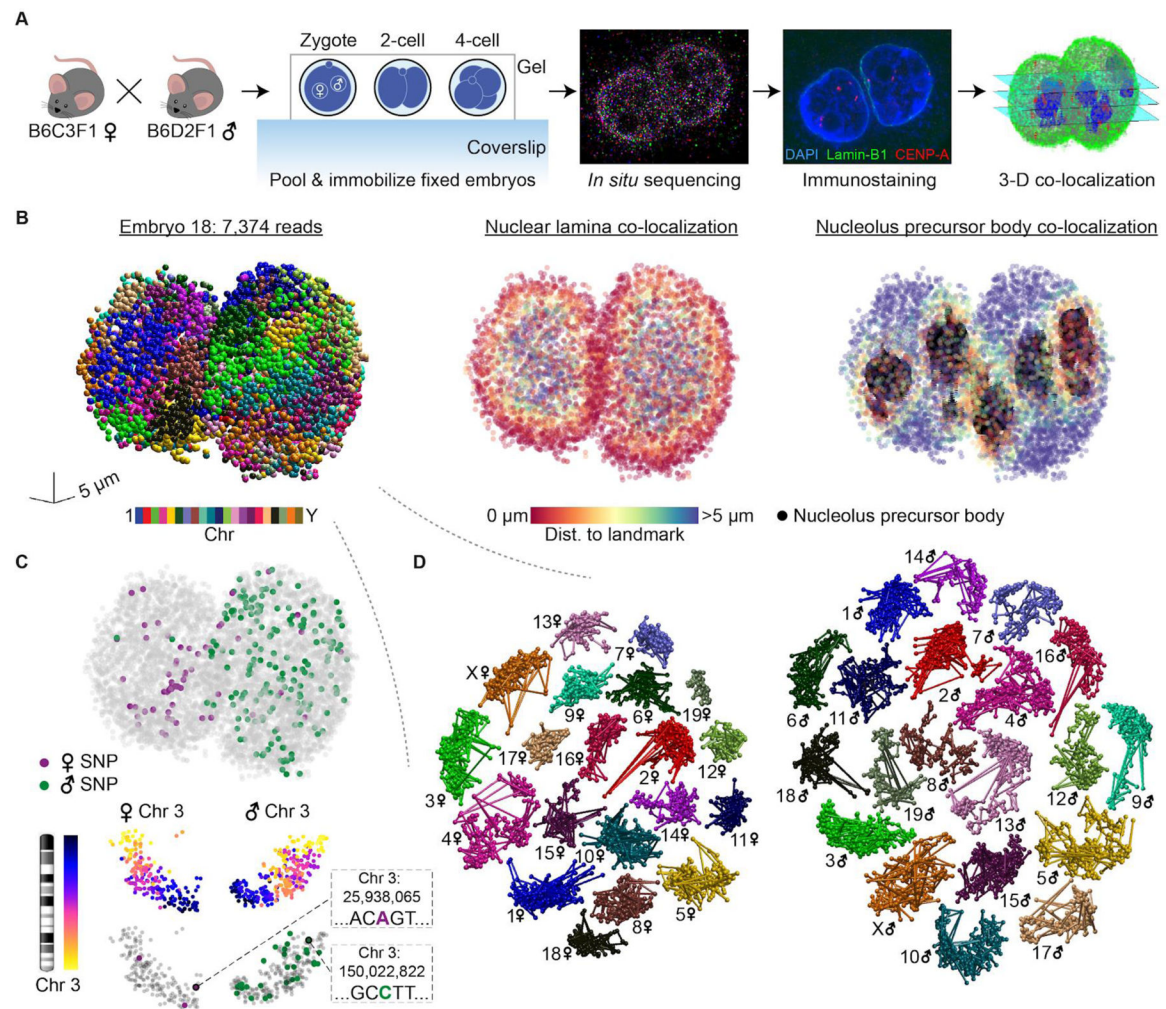


Figure 3: IGS enables high-resolution genomic and spatial profiling of intact early mouse embryos

(A) Workflow. B6C3F1 x B6D2F1 embryos at the zygote, 2-cell, and 4-cell stages are pooled, fixed and immobilized in a polyacrylamide gel. Following *in situ* sequencing, DAPI and immunofluorescence staining of CENP-A and Lamin-B1 are performed. (B) Representative zygote with 7,374 spatially-localized reads colored by chromosome (left), distance to the nuclear lamina (middle), and distance to nearest nucleolus precursor body (right). (C) Amplicons from (B), with reads colored by parental haplotype assignment for the intact embryo (top), reads colored by genomic position for Chr 3 homologs (middle), and reads colored by parental haplotype assignment for Chr 3 homologs (bottom). Boxes show two haplotype-informative Chr 3 SNPs. (D) An exploded view of chromosome territories from (B) for the maternal (left) and paternal pronuclei (right).

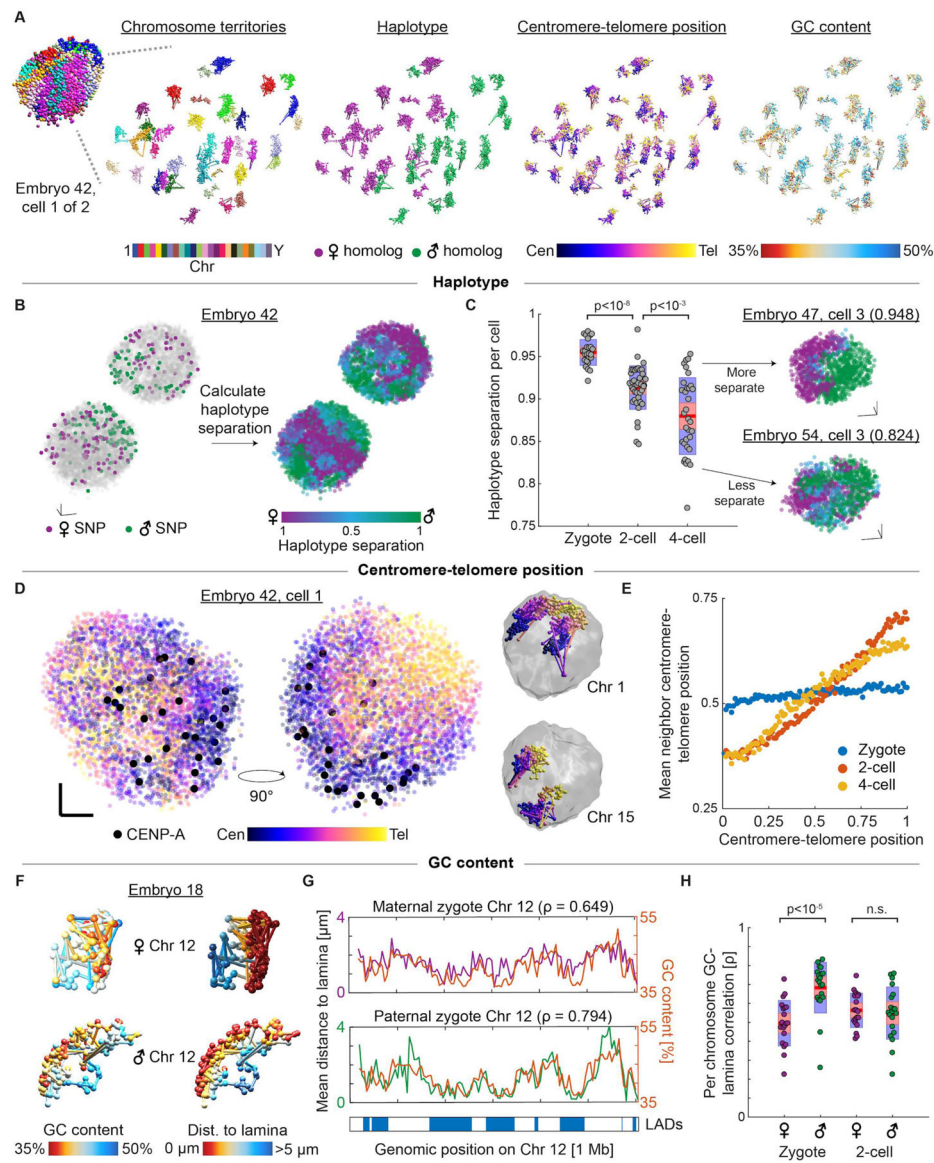


Figure 4: IGS characterizes developmental transitions in embryonic genome organization. (A) Exploded view of a single nucleus from a 2-cell embryo colored by chromosome territories, haplotype, centromere-telomere position, and GC content. (B) 2-cell embryo with spatially-localized reads colored by parental haplotype assignment (left) and haplotype separation score (right). (C) Boxplots showing mean haplotype separation score per cell across developmental stages (left; K-S test, $p < 10^{-4}$ and $p < 10^{-8}$). Grey dots represent mean scores of single cells. Distribution mean (red line), 95% confidence interval, (red box), and 1 standard deviation (blue box) are indicated. Two cells representing extreme scores (> 1 SD) are shown (right). (D) Nucleus from (A) with spatially-localized reads colored by centromere-telomere position, shown from two angles 90 degrees apart (left). Black dots indicate the position of CENP-A as identified from immunostains. Chr 1 and Chr 15 homologs from this cell are shown (right) to illustrate the Rabl-like configuration. (E) Mean centromere-telomere position of spatial neighbors as a function of centromere-telomere

position for each stage. **(F)** Chr 12 homologs from a representative zygote with spatially-localized reads colored by GC content (left) and distance to lamina (right). **(G)** Plots showing the relationship between GC content and average distance to the nuclear lamina for 1 Mb bins in Chr 12 of the maternal and paternal zygotic pronuclei. Zygotic lamina-associated domains (LADs) defined by DamID are displayed below. **(H)** Boxplots showing Spearman's ρ between GC content and distance to lamina for 1 Mb bins, partitioned by haplotype and developmental stage (K-S test, $p < 10^{-5}$ and n.s.). Dots represent single chromosomes. Distribution mean (red line), 95% confidence interval, (red box), and 1 standard deviation (blue box) are indicated. $n = 24$ zygotes, 40 2-cell, 49 4-cell nuclei for all panels. Scale bars: 5 μm in all directions.

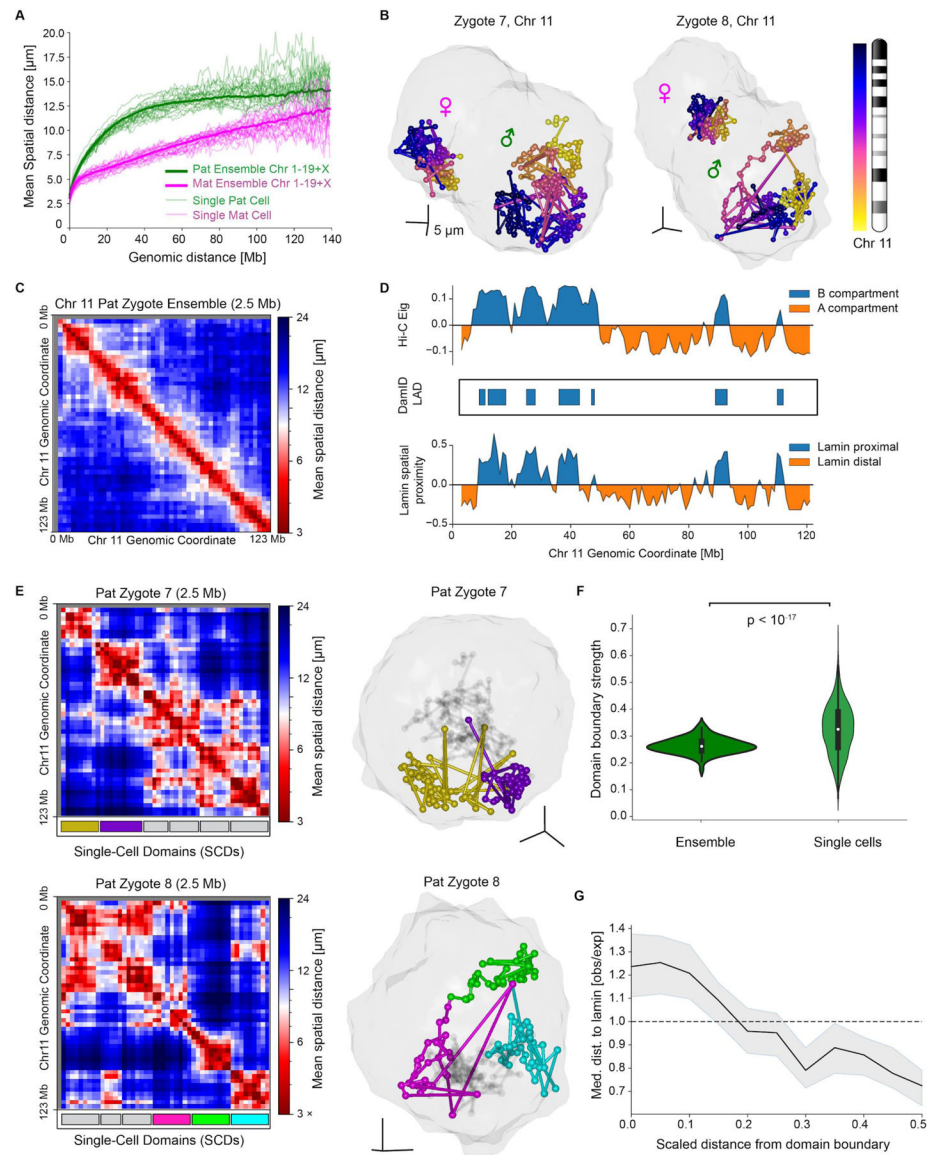


Figure 5: IGS reveals single-cell domains in zygotes.

(A) Global relationship between genomic and spatial distance in zygotes for all chromosomes, distinguishing the parental genomes. (B) Visualization of Chr 11 homologs in two zygotes according to parent-of-origin. (C) Population ensemble mean spatial distance matrix for paternal Chr 11, constructed at 2.5 Mb resolution (24 zygotic pronuclei, 2317 reads). (D) Comparison across measurement modalities for the population of paternal zygotic Chr 11. Top row: Hi-C defined eigenvalues and compartment calls. Middle row: DamID-defined population lamina associated domains. Bottom row: lamin-proximal and lamin-distal regions defined with IGS (24 zygotic pronuclei, 2317 reads). (E) Top left: single-cell mean distance matrix for paternal Chr 11 in a representative zygote, with single-cell domain boundaries (SCDs) marked below (263 reads). Top right: visualization of individual paternal SCDs in the same zygote. To assist visualization, two SCDs are shown in color (purple, gold), while the remaining SCDs are shown in grey. Bottom left: single-cell

mean distance matrix for paternal Chr 11 in a second representative zygote, with single-cell domain boundaries marked below (213 reads). Bottom right: visualization of three paternal SCDs in the second zygote. To assist visualization, three SCDs are shown in color (magenta, lime, cyan) while the remaining SCDs are shown in grey. **(F)** Comparison of single-cell and ensemble domain boundary strengths spanning all detectable boundaries in Chr 1–19+X (74 ensemble boundaries, 1057 single-cell boundaries, K-S test, $p < 10^{-17}$). **(G)** Scaled distance from SCD boundary versus observed/expected median distance to nuclear lamina, measured genome-wide (Chr 1–19+X, $N = 1262$ SCDs). Envelope indicates 95% confidence interval determined by bootstrapping.

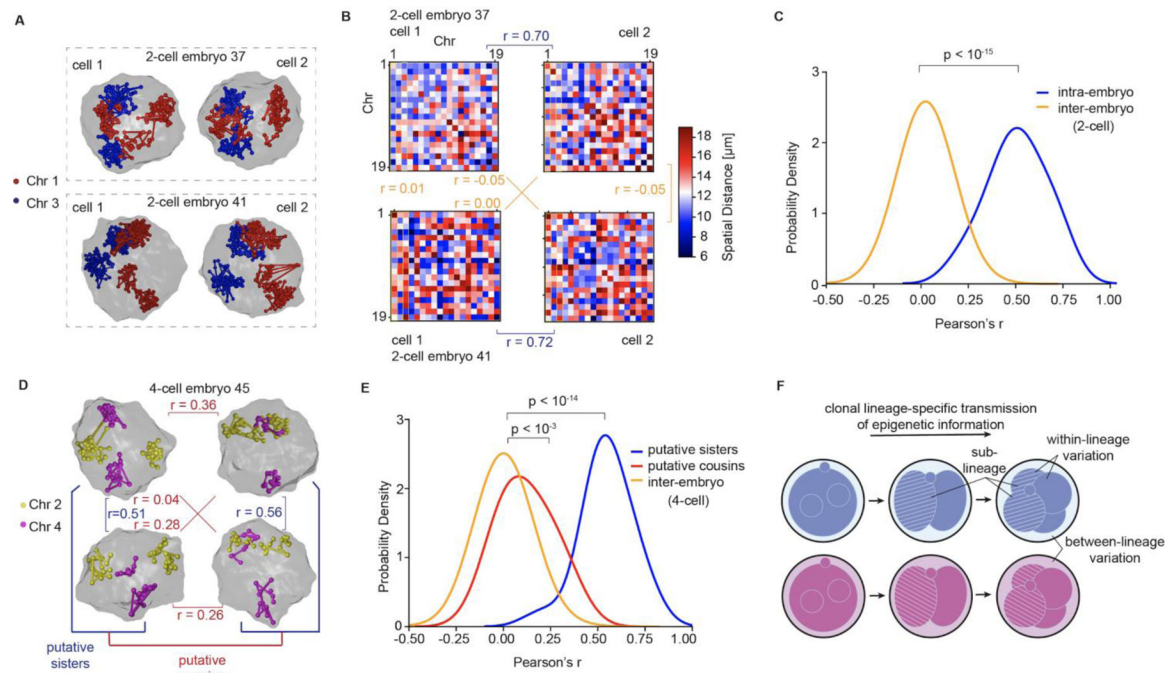


Figure 6. IGS uncovers epigenetic memory of global chromosome positioning within single embryos.

(A) Positioning of Chr 1 and 3 in the cells of 2-cell embryos 37 (top) and 41 (bottom). (B) Pairwise correlations between autosome distance matrices for the cells in (A). Intra-embryo and inter-embryo correlations are shown in blue and orange, respectively. (C) Probability distributions of correlations between autosome distance matrices for intra-embryo and inter-embryo pairs of cells among 2-cell embryos. K-S test, $p < 10^{-15}$; $n = 20$ intra-embryo pairs and $n = 760$ inter-embryo pairs, among 20 2-cell embryos. (D) Positioning of Chr 2 and 4 in the cells of 4-cell embryo 45. Pairs of cells are putatively classified as sister and cousin cells based on correlation of global chromosome positioning, with the most correlated pair classified as sisters. Correlations between sister and cousin cells are shown in blue and red, respectively. (E) Probability distributions of correlations between autosome distance matrices for pairs of putative sister cells, cousin cells, and inter-embryo pairs of cells among 4-cell embryos. K-S test, $p < 10^{-14}$ for sisters vs. inter-embryo and $p < 10^{-3}$ for cousins vs. inter-embryo; $n = 18$ sister pairs, $n = 36$ cousin pairs, and $n = 933$ inter-embryo pairs, among 13 4-cell embryos. (F) Model of epigenetic memory transmission within clonal lineages.