HHS Public Access

Author manuscript

Nat Struct Mol Biol. Author manuscript; available in PMC 2023 April 11.

Published in final edited form as:

Nat Struct Mol Biol. 2022 October ; 29(10): 1011–1023. doi:10.1038/s41594-022-00839-y.

MiOS, an integrated imaging and computational strategy to model gene folding with nucleosome resolution

Maria Victoria Neguembor[1,a,*], **Juan Pablo Arcon**[2,a], **Diana Buitrago**[2,3,a], **Rafael Lema**[2], **Jürgen Walther**[2], **Ximena Garate**[1], **Laura Martin**[1], **Pablo Romero**[2], **Jumana AlHaj Abed**[4], **Marta Gut**[5,6], **Julie Blanc**[5], **Melike Lakadamyali**[7], **Chao-ting Wu**[4], **Isabelle Brun Heath**[2], **Modesto Orozco**[2,8,9,*], **Pablo D. Dans**[2,10,11,b,*], **Maria Pia Cosma**[1,6,9,12,b,*]

[1]Center for Genomic Regulation (CRG). Barcelona Institute of Science and Technology. Barcelona, Spain

[2]Institute for Research in Biomedicine (IRB Barcelona). Barcelona Institute of Science and Technology. Barcelona, Spain

[3]Departamento de Física y Matemáticas, Universidad Autónoma de Manizales, Manizales, Colombia

[4]Department of Genetics. Harvard Medical School. Boston, MA, USA

[5]CNAG-CRG, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Baldiri Reixac 4. Barcelona, Spain

[6]Universitat Pompeu Fabra (UPF). Barcelona, Spain

[7]Department of Physiology, Perelman School of Medicine. University of Pennsylvania. Philadelphia, PA, USA

[8]Faculty of Biology, University of Barcelona. Barcelona, Spain

[*]Correspondence to: Maria Victoria Neguembor (victoire.neguembor@crg.eu); Modesto Orozco (modesto.orozco@irbbarcelona.org); Pablo D. Dans (pablo.dans@unorte.edu.uy, pdans@pasteur.edu.uy); Maria Pia Cosma (pia.cosma@crg.eu).

AUTHOR CONTRIBUTIONS STATEMENT

Original idea and conceptualization by MVN, PDD, IBH, MPC and MO. MPC, MVN, PDD, JPA and MO wrote the article with contributions from all the authors. MVN produced all imaging results (iOS and MiOS) together with XG and LM. RL produced all the capture-Hi-C/MNase-seq results that were post-processed and analyzed by DB, under the supervision of IBH. JPA developed and validated the restraint-based model. JPA and DB performed Hi-C based simulations of chromosome segments. JW generated the coarse-grained chromatin structures at nucleosome level, and the deconvolution of MNase-seq signals. Fitting algorithms and fitting results were generated by PR together with JW. All modeling, simulations, and fitting results were supervised and analyzed by PDD and MO. MPC supervised the generation and analyses of all the imaging results with contribution of ML. Design of Oligopaint probes was performed by JAA and C-tW. MG and JB performed the sequencing of capture-Hi-C/MNase-seq experiments. PDD and MVN integrated all the results and were the scientific coordinators of the project.

COMPETING INTERESTS STATEMENT

C-tW holds or has patent filings pertaining to imaging, and her laboratory holds a sponsored research agreement with Bruker Inc. Although non-equity holding, C-tW is a co-founder of Acuity Spatial Genomics; through personal connections to George Church, she has equity in companies associated with him, including 10x Genomics and Twist. The remaining authors declare no competing interests.

Peer Review Information:

Nature Structural and Molecular Biology thanks Mattia Conte and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Editor recognition statement (if applicable to your journal):

Sara Osman was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

[9]ICREA. Barcelona, Spain

[10]Department of Biological Sciences, CENUR Litoral Norte, Universidad de la República (UdelaR), Salto, Uruguay

[11]Bioinformatics Unit. Institut Pasteur de Montevideo. Montevideo, Uruguay

[12]CAS Key Laboratory of Regenerative Biology, Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China

[a]These authors contributed equally.

[b]These authors jointly supervised this work.

Abstract

The linear sequence of DNA provides invaluable information about genes and their regulatory elements along chromosomes. However, to fully understand gene function and regulation, we need to dissect how genes physically fold in the three-dimensional (3D) nuclear space. Here we describe immuno-OligoSTORM (iOS), an imaging strategy that reveals the distribution of nucleosomes within specific genes in super-resolution, through the simultaneous visualization of DNA and histones. We combine iOS with restraint-based and coarse-grained modeling approaches to integrate super-resolution imaging data with Hi-C contact frequencies and deconvoluted MNase-sequencing information. The resulting method, called Modeling immuno-OligoSTORM (MiOS), allows quantitative modeling of genes with nucleosome resolution and provides information about chromatin accessibility for regulatory factors, such as RNA polymerase II. With MiOS, we explore intercellular variability, transcriptional-dependent gene conformation and folding of housekeeping and pluripotency-related genes in human pluripotent and differentiated cells, thereby obtaining the highest degree of data integration achieved so far.

Editor summary:

The authors present Modeling immuno-OligoSTORM (MiOS), a super-resolution imaging and computational strategy to model 3D gene folding at multiple genomic scales, reaching nucleosome resolution at the single-gene level.

INTRODUCTION

Unraveling the 3D genome organization of chromatin fiber is paramount to understanding gene function. In recent years, significant advances have been made through the use of genomic approaches, mainly chromosome conformation capture (3C)–based techniques, such as Hi-C, ChIA-PET and Micro-C, among others [1–3]. Obtaining 3D spatial organization of chromosomes and identifying proximal chromatin interactions between different genomic regions has greatly advanced our understanding of gene regulation. Chromosomes have been shown to segregate into specific nuclear regions, called chromosome territories, as well as into heterochromatin and euchromatin, which are partitioned in A and B compartments [4]. Topologically-associated domains (TADs) [5–8] and loops [9] further subdivide the chromatin fiber [10].

In parallel, a variety of imaging approaches, such as stochastic optical reconstruction microscopy (STORM), OligoSTORM, OligoFISSEQ, Optical Reconstruction of Chromatin Architecture (ORCA), Hi-M, DNA-MERFISH and DNA seqFISH+ [11–21] allowed specific chromosome regions, genes and nucleosome clutches to be visualized at high resolution, opening the possibility of determining their spatial conformation within a single nucleus.

In addition to novel genomic and imaging techniques, a number of computational models have been developed to estimate physical genomic distances from Hi-C matrices and imaging [16–19,22–28]. Results using these models have confirmed the presence at the single-cell level of compartments, TADs and chromatin nanodomains, as well as a high degree of cell–to–cell variability for the retrieved structures [16–18,22].

Despite the large amount of work dedicated to unravelling the 3D structure of chromatin regions, the methods used to integrate high-resolution genomics with imaging approaches are still limited. For instance, we still lack approaches that incorporate epigenetic information with nucleosome positioning and that reach nucleosome-resolution modeling of genes. Achieving this degree of integration and resolution is essential for obtaining a realistic view of chromatin complexity and folding in single cells—and thus for understanding gene function from a global perspective.

Here, we introduce MiOS (Modeling immuno-OligoSTORM), a combined imaging and computational strategy to model 3D gene folding at multiple genomic scales, down to nucleosome resolution. To achieve this, we developed immuno-OligoSTORM (iOS), combining OligoSTORM and DNA-PAINT imaging to simultaneously acquire super-resolved images of DNA at specific gene loci, as well as the nucleosomes within these regions. MiOS integrates modeling approaches, iOS super-resolution (SR) images, Hi-C and micrococcal nuclease–sequencing (MNase-seq) to reach a high degree of imaging, genomic and epigenomic data integration that is close to single-nucleosome resolution. The modeled structures reproduce average experimental data from cell populations while including information about single-cell variability. We used MiOS to analyse a conserved region comprising the key pluripotency-associated genes *NANOG* and *DPPA3* (hereafter *STELLA*) and the housekeeping genes *GAPDH* and *IFFO1*. The results unveiled the conformational changes between human induced pluripotent stem cells (hiPSCs) and human fibroblasts (hFibs) at nucleosome resolution and inferred the physical properties of these genes. We report differences in RNA polymerase II accessibility to these genes and observed TAD repositioning, with an increased spatial proximity between *NANOG* and *STELLA* in hiPSCs as compared to hFibs. Thus, MiOS is a powerful multiscale tool for studying gene folding and function, as it integrates the analysis of the chromatin fiber structure in super-resolution with not only the short-range 3D contacts but also the nucleosome position data at the single-gene level. The use of MiOS is expected to expand the current limits of integrative and structural biology.

MAIN TEXT

We devised a strategy to generate 3D models of the chromatin fiber at high spatial resolution by integrating multiple sources of genomic, epigenetic, and imaging data. Specifically,

we combined super-resolution STORM and DNA-PAINT imaging, capture Hi-C, capture MNase-seq and restraint-based and coarse-grained modeling, to identify how genes are folded in the nuclear space. Our overall aim was to obtain 3D structures at multiple size scales, reaching nucleosome resolution. To develop the MiOS strategy, we studied a region spanning 2.3 Mb of human chromosome 12 (chr12: 6,140,000–8,460,000) (Extended Data Fig. 1a). This region comprises broadly expressed genes, including *GAPDH* and *IFFO1*, in the proximal subregion, and a conserved cluster of early developmental pluripotency genes, including *STELLA* and *NANOG*, in a distal subregion [29] (Extended Data Fig. 1a–f). Accordingly, previous Hi-C [9] and ChIP-seq [30,31] analyses in human IMR90 fibroblasts showed that the proximal subregion belongs to the active A compartment and is enriched for active chromatin marks, while the distal subregion containing pluripotency-related genes falls into the repressed B compartment (Extended Data Fig. 1a). Thus, we investigated the differential 3D conformations of these genes in human IMR90 fibroblasts (hFibs) and fibroblast-derived human induced pluripotent stem cells (hiPSCs) [32].

iOS allows simultaneous SR imaging of genes and histones

To visualize the nuclear position and conformation of individual genes, we designed STORM-compatible Oligopaint probes spanning coding and regulatory regions in *GAPDH-IFFO1*, *STELLA* and *NANOG* (Extended Data Fig. 1a) to perform OligoSTORM [12,15]. Additionally, we carried out dual color Oligopaint, to measure pairwise distances between the *GAPDH–IFFO1* locus and the *STELLA* or *NANOG* locus with diffraction-limited imaging (Fig. 1a). The measured mean distance between *GAPDH* and *NANOG* was 0.90 ± 0.41 μm in hFibs, and slightly shorter (0.86 ± 0.37 μm) in hiPSCs (Fig. 1b). Interestingly, although *GAPDH* and *STELLA* are closer in the genomic sequence than *GAPDH* and *NANOG* (Extended Data Fig. 1a), they appeared further apart in space (1.20 ± 0.55 μm) in hFibs (Fig. 1c), suggesting that genomic contacts could influence the physical arrangement and thus the 3D distance of these genes in nuclear space.

To incorporate information on nucleosome positioning into our models, we developed immuno-OligoSTORM (iOS), combining the strategies of OligoSTORM and DNA-PAINT [33] to sequentially image DNA and histone H3 at super-resolution (see Methods). Briefly, we performed Oligopaints FISH labeling followed by an immunostaining protocol compatible with DNA-PAINT. We then imaged the samples in the presence of a hybrid imaging buffer containing STORM-compatible oxygen scavenging conditions as well as DNA-PAINT imager strands. We first acquired diffraction-limited images of the Oligopaint-labeled loci and then OligoSTORM images at a high frame rate. Finally, using DNA-PAINT imaging, we acquired images of histones at a lower frame rate. We used fiduciary beads across all imaging steps to spatially overlay OligoSTORM and DNA-PAINT images, with a localization precision below 20 nm in x-y for both techniques (Extended Data Fig. 1g) [11,34]. iOS enabled the detection of core histone H3 located within and around the genes of interest (Fig. 1d, e). The technique could potentially be used to image any nuclear protein to assess its interactions with any given genomic region.

Restraint-based chromatin modeling preserves cell variability

Before integrating iOS data with Hi-C conformational information, we calibrated our restraint-based chromatin model using single-cell microscopy data from Bintu *et al.* [18]. The input median distance matrix for the model was obtained by processing approximately 3,500 single cells through multiplexed, FISH-based 3D diffraction-limited imaging over a 2 Mb region of chromosome 21, at 30-kb resolution (see Supplementary Note). The refined modeling method generated a reduced ensemble of 70 structures that efficiently reproduced the experimental distance matrix ($r_{Spearman}$ = 0.98, r_{HiCRep} = 0.96, Fig. 2a), and that fulfilled virtually all the input distance restraints (99.7%; Extended Data Fig. 2a). The predicted modeled structures exhibited a conformational variability compatible with that obtained by single-cell microscopy (see Fig. 2b and Extended Data Fig. 2b–f). In fact, about 80% of the experimental structures that showed discernible TAD-like organization can be mapped to one of the 70 modeled structures of the reduced ensemble. Of note, different TAD-like patterns were evident among the different single structures, which are not necessarily equivalent to the TADs revealed in the ensemble matrix (e.g. compare Figs. 2a and 2b). To test the similarity between the experimental and modeled structures, as well as their degree of conformational variability, we compared a common structural parameter that characterizes polymeric entities, such as the end-to-end distance (Extended Data Fig. 2b). Average values for both sets were close to each other (817.2 nm and 831.5 nm for the experimental and modeled ensembles, respectively), and the distributions showed no significant difference (p = 0.16, two-sided Mann-Whitney test; Extended Data Fig. 2b). We also directly compared the cartesian coordinates of the different structures to further demonstrate that the model is an accurate representation of the experimental ensemble. Each experimental structure was superposed onto each modeled structure, and the optimal roto-translational fit was obtained by minimizing the root mean square deviation (rmsd) of probe/bead positions. The similarity between the best-fitted structures (optimally aligned) was quantified by computing their rmsd and comparing the resulting distribution to a null distribution consisting of the rmsd between all fitted (aligned) experimental structures (Extended Data Fig. 2c). The experimental model distribution showed a statistically significant decrease in rmsd with respect to the null distribution ($p < 10^{-16}$, two-sided Mann-Whitney test), indicating that structures from the modeled ensemble resemble the experimental dataset. Finally, principal component analysis (PCA) of cartesian coordinates revealed a high degree of overlap in the sampled conformational space between the experimental and modeled ensembles. The variance in the essential deformation space (i.e. the first 10 modes) was comparable between ensembles, and the projection plot over the first two principal components coincided (Extended Data Fig. 2d,e). Furthermore, the root mean square inner product (RMSIP) between the first 10 eigenvectors reached a value of 0.61 (Z_{score} = 19.6), indicating that the main deformation movements sampled by the experimental and theoretical ensembles are highly similar, and that the reduced set of modeled structures fits well the structural diversity sampled experimentally.

We then tested the validity of our distance restraint–based method on another genomic system, using data from different species and with different resolution. We modeled the 3D conformation of the α-globin locus using Capture-C data obtained at 4-kb resolution for a 300-kb region that included α-globin genes (chr11: 32,000,000–32,300,000) in erythroid

and embryonic stem (ES) cells [35,36]. After converting contact data to average distances (see Supplementary Note), theoretical structural ensembles reproduced more than 99.9% of the experimental restraints, and the ensemble-derived contact matrices showed high correlation for both erythroid and ES cells when compared to the experimental ones (Fig. 2c). For instance, it is known that the α-globin locus undergoes a major conformational change upon erythroid differentiation [37]. In line with this, a self-interacting domain encompassing the α-globin genes and enhancers was formed only in erythroid cells (Fig. 2c), with additional anti-diagonal contacts that resulted in an almost symmetric hairpin-like structure (Fig. 2d), as previously suggested [35]. Finally, we explored the ability of these predicted structures to reproduce previously published FISH data [38] (see probe localization; Fig. 2e). The model showed a significant decrease in the 3D distance between the extremes of the α-globin domain (A-Ex) as compared to a control, non-interacting region (A-Cx) in erythroid cells, but not in ES cells (Fig. 2f), in agreement with experimental results [38]. In conclusion, the modeling strategy was proven to be robust using data from different organisms (human and mouse) and at different resolutions (30 kb and 4 kb); of note, it was able to reproduce population-derived average data while preserving cell–to–cell variability in the 3D genomic structure.

Combining iOS and Hi-C reveals gene spatial rearrangements

We next aimed to integrate iOS data obtained in hFibs and hiPSCs with conformational Hi-C data, by applying the restraint-based model. Thus, we first performed *ad hoc* capture Hi-C in hFibs and hiPSCs at a 5-kb resolution on the 2.3 Mb region of chromosome 12, including the genes of interest. Our data reproduced and was highly correlated with previously published results [9], showing clearly identifiable TADs in hFibs (Extended Data Fig. 3a–c). Conformational gene rearrangements, including different genomic contacts and TAD reorganization, were evident in hiPSCs as compared to hFibs (Extended Data Fig. 3b–f). Long-range contacts that were clearly present in hFibs were lost in hiPSCs, which also showed formation of new TADs (Fig. 3a, b). For instance, *NANOG* is located near the boundary of two TADs, and *STELLA* is at the edge of a separated TAD, in hFibs; in sharp contrast, both genes fall into a single TAD in hiPSCs (Fig. 3c, d). The insulation analysis [39] confirmed a clear change in TAD size and boundaries in genomic regions, including both *STELLA* and *NANOG* loci (Fig. 3e). In hiPSCs, increased interaction counts between *STELLA* and *NANOG* were observed when comparing bin pairs at a 5-kb resolution that covered the genomic regions of the two genes in both cell types (Supplementary Table 1). Interestingly, these results agree with previous observations reported in mouse ES cells [29,40], suggesting that the contacts of *STELLA* and *NANOG* in pluripotent cells are conserved across species.

As a first step to combine Hi-C with iOS, we converted capture Hi-C contact maps for hFibs and hiPSCs into average distance matrices that served as input restraints between genomic loci (beads) for our restraint-based model (see Supplementary Note) (Fig. 4a, b). The molecular dynamics simulation resulted in highly diverse ensembles of structures (158 and 174 structures for hFibs and hiPSCs, respectively) that fulfilled, as a group, around 80% of these experimental restraints (Fig. 4c, d). Representative structures within the ensemble satisfied nearly half of the restraints (Fig. 4e,f). To test the accuracy of the model,

we compared the contact matrices derived from the predicted 3D conformations with the experimental ones from capture Hi-C data (Fig. 4a, b), obtaining high resemblance regarding contact domains and high correlation coefficients ($r_{Spearman}$ = 0.83, r_{HiCRep} = 0.86 for hFibs and $r_{Spearman}$ = 0.80, r_{HiCRep} = 0.70 for hiPSCs). By measuring distances between genes in the ensemble of structures, clear spatial rearrangements emerged in hiPSCs with respect to hFibs: the predicted 3D distance between *NANOG* and *STELLA* of 0.16 ± 0.13 μm in hFibs decreased in hiPSCs, to 0.05 ± 0.05 μm (Fig. 4c, d; $p < 1e\text{-}16$, two-tailed Mann-Whitney test). This result is compatible with both genes belonging to the same TAD in the average population analysis. On the other hand, the distance between genes in hFibs as compared to hiPSCs was 1.06 ± 0.11 μm vs. 0.94 ± 0.10 μm for *GAPDH–STELLA*, and 0.99 ± 0.08 μm vs. 0.92 ± 0.10 μm for *GAPDH–NANOG* (Fig. 4c, d). Thus, although *GAPDH* and *STELLA* are closer in genomic sequence than *GAPDH* and *NANOG* (Extended Data Fig. 1a), the predicted distances inferred by the simulated structures appeared larger for *GAPDH–STELLA* than for *GAPDH–NANOG* (Fig. 4c, d); this situation was also observed with dual color Oligopaint imaging in hFibs (Fig. 1c). This suggests that the 3D organization of the region maintains these two loci at a shorter distance than the one expected based on their linear genomic distance, further reinforcing the relevance of genomic contacts in the spatial arrangement of these genes. Importantly, the consistency between the Hi-C–based structures (generated with no added input from imaging into the model) and the actual measurements in imaged nuclei provides an independent validation that supported the accuracy of the model.

Representative structures from each simulated ensemble fulfilling the maximum number of experimental restraints showed a differential spatial arrangement of *NANOG* and *STELLA* in the two cell types (Fig. 4e, f). hFibs exhibited a genomic 3D arrangement with a U-shape conformation, reflecting extended long-range interactions and the presence of *NANOG* and *STELLA* exposed at the surface of the U-turn (zoomed-in view, Fig. 4e; TAD_a, Fig. 3e, Extended Data Fig. 4c–e, and Supplementary Movie 1). In contrast, this prominent U-shape conformation is lost in hiPSCs, which have a reduced number of long-range contacts and have *STELLA* and *NANOG* located on the same contact domain, pointing towards the same spatial location (zoomed-in views, Fig. 4f and 3e; TAD_b, Extended Data Fig. 4c–e and Supplementary Movie 1). This was further supported by the reduced predicted distance between *NANOG* and *STELLA* in hiPSCs, as mentioned above (Fig. 4c, d).

Finally, we combined the iOS imaging data acquired in single cells with the chromosome conformations sampled by the restraint-based model. In particular, we incorporated spatial coordinates for H3 clusters and gene distances of individual alleles obtained by iOS to the models of the 2.3 Mb genomic region. A new ensemble of structures was derived by setting the physical distance between the *GAPDH–IFFO1* locus and the *NANOG* locus at the value measured by imaging for selected single nuclei, and the obtained structures were fitted to the H3 localization data considering a focal plane with a depth of 0.26 μm (see Supplementary Note). This allowed us to map the nucleosome localizations and to find possible paths in 3D for the whole genomic segment. The resulting models exhibited a high fitting sensitivity of the imaging and Hi-C data simultaneously, with more than 41% Hi-C restraints fulfilled, and with 52% to 76% overlap of H3 iOS signal for the best fitted individual structures in both hFibs and hiPSCs (Fig. 5a, b, and Extended Data Fig. 5; see representative structures

satisfying the measured *GAPDH–IFFO1 to NANOG* distances, Fig. 5c, d). Adding these new structures to the original (iOS-unbiased) ensemble increased the ability to reproduce experimental Hi-C restraints to more than 90% for both hFibs and hiPSCs (Fig. 5c, d) while also maintaining the correlation between predicted and experimental contact matrices (r_{HiCRep} = 0.87, r_{Sp} = 0.83 for hFibs and r_{HiCRep} = 0.72, r_{Sp} = 0.80 for hiPSCs). Hence, this result further validated our deconvolution approach and highlighted the importance of iOS data as a valuable input for achieving more realistic 3D chromatin models.

MiOS models genes at nucleosome resolution

We next aimed to integrate nucleosome positioning data from MNase-seq into our modeling strategy. For this, we first analyzed differential nucleosome positioning and enrichment using capture MNase-seq in hFibs and hiPSCs, using the 2.3 Mb region previously modeled and analyzed by capture Hi-C. Across the entire region, hFibs showed a general trend towards well-positioned nucleosomes; in contrast, hiPSCs had fuzzier or less stable nucleosome positioning (Extended Data Fig. 6a). In particular, *STELLA* and *NANOG* showed the most pronounced differences between hFibs and hiPSCs (Extended Data Fig. 6b–e); an in-depth analysis of nucleosome positioning with NucDyn [41] confirmed significant differences in evicted, shifted, or included nucleosomes for these genes in hFibs cells as compared to hiPSCs (Extended Data Fig. 6f–i).

Next, we focused on the MNase-seq results at the genomic areas labeled by OligoSTORM probes to determine the differential nucleosome distribution across the *GAPDH–IFFO1* and *NANOG* loci (Fig. 6). Since the MNase-seq results are representative of the average nucleosome occupancy across the cell population, a deconvolution of the signal was required to obtain a series of individual nucleosome distributions (i.e., configurations) which, combined, reproduced the cell-population MNase-seq data (see Methods) [24,42]. Clustering analysis showed that between 22 and 29 different nucleosome configurations were enough to recover the average experimental coverage for both cell types (Fig. 6b, d). According to these configurations, the total number of nucleosomes for the *GAPDH-IFFO1* genomic locus was ~72 in both hFibs and hiPSCs, and for the *NANOG* locus, ~57 in hiPSCs and ~53 in hFibs (Fig. 6). Overall, the increased fuzziness of hiPSCs resulted into a higher number of configurations (Extended Data Fig. 6).

Finally, we integrated nucleosome positions with the 3D conformations to dissect spatial gene organization at nucleosome resolution. Thousands of 3D structures of each nucleosome configuration were obtained through Monte Carlo simulations using a coarse-grained chromatin model [24,42] (see Supplementary Note). These ensembles of structures were fitted into the spatial localizations obtained from OligoSTORM imaging, considering again a focal plane with a depth of 0.26 μm, to extract the *GAPDH* and *NANOG* gene conformations in 3D at a nucleosome resolution, as displayed for one fitted structure of *GAPDH* in hFibs (zoomed-in view, Extended Data Fig. 7a). Since we designed the OligoSTORM probes to cover the genes at known genomic coordinates, only the base pairs of the nucleosome fiber corresponding to probe localizations were considered in the fitting procedure. We then binned the space in equally-sized rectangles and built densities from all individual OligoSTORM localizations that fell into a given bin (see Supplementary Note for further

details). The matching procedure was applied iteratively using all the structures, for any given OligoSTORM image, resulting in a ranking in terms of the fitting percentage. Of note, of the thousands of highly heterogeneous structures generated, we ended with filtered ensembles (see Supplementary Note), most of which had a very good fitting with the high-resolution OligoSTORM data (Extended Data Fig. 7b); this provided a good degree of confidence in the chosen nucleosome configurations. Best-fitted structures (top 1 displayed in Fig. 7) were then analyzed by examining the physical properties of the chromatin fiber and comparing the results between the two cell types. For *GAPDH*, up to 281 probes (of the 286 probes designed to cover the gene region) were fitted with an overlap percentage of 75.7%, and for *NANOG*, up to 137 probes (of the 185 probes designed) were fitted with an overlap percentage of 78.9% (see Extended Data Fig. 7b; Supplementary Table 2 shows the values for the top ten structures).

Modeling at this level of resolution obtaining nucleosome positioning at the single gene level provided us with an opportunity to estimate fine-scale structural features of gene organization, such as intragenic distances and nucleosome contacts, exposed surfaces, and accessibility to protein binding. Both between genes and between cell types, structures exhibited high degrees of variability in several parameters, such as the end–to–end distance or the contacts between nucleosomes (for *GAPDH*, Fig. 7a, b, d, e; for *NANOG*, Fig. 7g, h, j, k; see also Supplementary Table 2). In agreement with our previous results from STORM imaging, nucleosomes in the models were organized in clutches [21] (Fig. 7a, d, g, j). Interestingly, these clutches were formed and organized in TAD-like domains (Fig. 7b, e, h, k), which were highly variable between structures; this result is in line with previous observations at a larger scale range and with an assumed liquid-behavior of the chromatin fiber [18,43]. In the same way, the surface accessible to RNA polymerase II displayed a periodic behavior along the sequence, with higher accessibility at inter-nucleosomal spaces (Fig. 7c, f, i, l). Additionally, the 3D structural features that were estimated for the best-fitted structures revealed that the *NANOG* region in hFibs was more compact (in terms of the occupied volume) and less exposed to the solvent or to other biomolecular partners, such as RNA polymerase II, than that in hiPSCs, in which the gene is actively transcribed ($p = 0.00012$, Supplementary Table 2).

DISCUSSION

Our understanding of the 3D organization of the genome, and the role it plays in gene function, has tremendously increased in the past decade. However, our knowledge is still fragmented, as available imaging, biochemical, and genomic techniques leave gaps in coverage, resolution, and integration of data. Thus, there is an urgent need for approaches that bridge multiple data types and provide a comprehensive view of genome organization. MiOS aims to contribute to this collective scientific effort. iOS provides the possibility to resolve the conformation of specific genomic loci at the kilobase range and at a super-resolution scale (*i.e.* the scale of individual genes) and to simultaneously locate nucleosomes within a region and its surroundings. The development of iOS paves the way for studying the locus-specific recruitments of potentially any chromatin-associated protein. Recent development of fast DNA-PAINT and microfluidic setups [44] provides a wealth of opportunities to combine these technologies, which would allow us to investigate multiple

protein associations to chromatin and to correlate these with gene regulation. In the same way, the iOS approach can be multiplexed to trace larger number of loci, as recently reported for strategies based on OligoFISSEQ, ORCA, Hi-M, DNA-MERFISH or DNA seqFISH+ [13,14,17–20].

MiOS merges single-cell data from iOS with conformational and epigenetic information derived from deconvoluted Hi-C and MNase-seq data, generating one of the most comprehensive and highly resolved gene models available so far. We showed that our restraint-based modeling approach is versatile by using not only ensemble-derived Hi-C or capture-C maps but also distance maps from experimental single-cell imaging data. The modeled ensembles of structures reproduced population-based experimental average data without losing information about intercellular variability as determined by microscopy. Moreover, our model correctly captures the conformational rearrangement of the α-globin locus in erythroid cells as compared to ES cells, even using data from another species that has a different resolution and that spans a shorter genomic region. Our results suggest that our restraint-based model can be used widely to predict 3D conformation of any genomic region, starting from a variety of data sources. In line with state-of-the-art Hi-C-derived methods [26,28,35], which use imaging information from FISH or FISH-based techniques to validate their physics-based polymer models [26,35] or as additional integrated information [28], MiOS further integrates super-resolution OligoSTORM and DNA-PAINT imaging of histones with capture Hi-C, capture MNase-seq and data-driven and physical models of DNA implemented in molecular dynamics and Monte Carlo simulations. The unique multiscale perspective of the MiOS pipeline enables chromatin conformation to be investigated from the megabase scale of chromosome segments (micrometers) to the folding of specific genes with nucleosome resolution; this facilitates the simultaneous detection of large- and fine-scale folding rearrangements.

Applying MiOS to a genomic region with key housekeeping and pluripotency genes, such as *GAPDH* and *NANOG*, revealed structural rearrangements related to cell state identity and gene activity in human differentiated fibroblasts and pluripotent hiPSCs. Notably, during reprogramming, the investigated chromosome region underwent a switch to global decompaction and TAD re-positioning; locally, a reduced distance between the pluripotency genes *NANOG* and *STELLA* was observed in hiPSCs, reminiscent to what was reported for mouse ES cells and B-cell reprogramming [29,40,45]. Additionally, MiOS models revealed clutch-like nucleosome organization within individual genes, as well as local chromatin folding compatible with TADs and compartments at the single-cell level, as previously proposed [18,22].

The MiOS results highlight the importance of integrating multiple techniques as an instrument of cross-validation and data refinement, making more accurate and comprehensive models of genome organization possible. Although the coarse-grained model of the nucleosome fiber is able to reach base pair resolution, our current results are limited by nucleosome level techniques (MNase-seq, immuno-OligoSTORM). Furthermore, in the future, it will be interesting to integrate recently developed techniques (ensemble, single-cell or allele specific), such as Micro-C, Micro-Capture-C, Hi-CO or Dip-C data [3,46–49], into MiOS, by producing nucleosome contact–biased simulations. In principle, this could further

improve the prediction of the folded genes at nucleosome resolution, while reducing the simulation costs needed to find accurately folded structures of the desired loci. Additionally, this might ultimately confirm the biological truth, at the single-cell level, of the folded conformations of genes produced by MiOS, which taken together, can reproduce with high statistical power the average structural features measured experimentally.

By combining the multiscale capabilities of MiOS with the recently developed strategies for high-throughput imaging and multiplexing at high-resolution, it should be soon possible to model the chromatin fiber over entire chromosomes and chromatin compartments, and to carry out fine-scale functional epigenetic analyses at specific loci across these structures. For instance, this could make possible to experimentally test, at multiple size scales, how perturbing specific genomic elements affect gene conformation and function, and how this is related to specific changes in the epigenetic state of the chromatin fiber or in the accessibility to transcriptional machinery.

METHODS

Cell lines used and culture conditions

Human IMR90 fibroblasts (hFibs, ATCC CCL-186) were grown in Dulbecco's Modified Eagle Medium (DMEM) (Thermo Fisher Scientific [TFS], #41965039) with 10% fetal bovine serum (FBS) (TFS, #10270106), 1% Penicillin-Streptomycin (TFS, #15140122), 1% GlutaMAX (TFS, #35050038). Human IMR90-derived induced pluripotent stem cells (hiPS(IMR90)-4, WiCell, #WISCi004) were cultured in mTeSR1 medium (Stemcell #85850) in Matrigel (BD, #356231) coated dishes. Cells were grown in humidified incubator at 37°C, 5% CO_2. For imaging, cells were plated in borosilicate glass-bottom 8-well chambers (Lab-TekI Nunc #155411 or µSlide Ibidi #80827). At 1 h before fixation, fluorescent amino-yellow beads (1:3500 dilution, Spherotech, #AFP-0252-2) were added to the culture medium to be later used for drift correction and further adjustments of STORM images.

Oligopaint probe design and probe synthesis

Oligopaint probes were designed *in silico* by using Oligominer balanced setting (blockParse.py with balanced flag '-l 35 -L 41 -t 42 -T 47' or with coverage flag -l 26 -L 32 -t 37 -T 42) [50] to a repeat-masked hg19 human genome assembly against the following coordinates Chr12: 6,641,500–6,666,000, Chr12: 7,854,000–7,881,000, Chr12: 7,931,000–7,970,000, consisting of 286, 158 and 185 probes and corresponding to *GAPDH-IFFO1*, *STELLA* and *NANOG* loci, respectively (see Extended Data Fig. 1 and Supplementary Table 3). The genomic distance between the *GAPDH-IFFO1* and *STELLA* probes was 1.19 Mb, and between the *GAPDH-IFFO1* and *NANOG* probes, 1.28 Mb. Oligominer scripts are available through Github: (https://github.com/brianbeliveau/OligoMiner).

The library was synthesized as 12K oligopool (CustomArray) after a quality check assessment by qPCR (see Supplementary Table 4). Oligopaint probes were prepared according to the T7 amplification followed by reverse transcription method given in Beliveau *et al.* [51]. For PCR amplification steps we used Kapa Taq (Kapa Bioscience,

#BK1002). ssDNA probes were obtained by degrading RNA through alkaline hydrolysis: RT reactions were mixed with 0.5 M NaOH and 0.25 M EDTA solution (1:1 v/v) and incubated at 95°C for 10 min. See Supplementary Table 4 for T7 promoter sequences and 5′ AF405 labeled primers.

Sample preparation; OligoSTORM and Immuno-OligoSTORM (iOS) labeling

hFibs or hiPSCs cells were prepared (e.g., fixed, permeabilized, denatured, hybridized, washed) according to Beliveau *et al.* [51]. After post-hybridization washes, cells were either stored for up to one week in 2× SSCT, 50% formamide at 4° C or used immediately for labeling. For hybridization, 50 pmoles of 5′ AF405 labeled primary probes, and 1 µl 100 µM 3′ AF647- or AF488-labeled secondary probes, were used (see Supplementary Table 4 for primary and secondary probes); samples were incubated at 42 °C for 16 to 24 h. After hybridization, samples were washed twice with 2× SSCT at 60 °C for 10 min each, and OligoSTORM samples were then washed once with 2× SSCT for 10 min at room temperature (RT) and once with 2× SSC, and then imaged immediately, while iOS samples were then washed once with 2× SSCT for 2 min at RT before proceeding with immunolabeling.

For immunolabeling, cells were i) blocked and permeabilized with 2% BSA (Fisher Scientific, #9048468), 0.2% Triton X-100 in PBS, 30 min at RT; ii) incubated with an anti-H3 primary antibody (Abcam, #ab1791) in blocking buffer at 1:50 dilution overnight, 4 °C; iii) washed three times (10 min each) in wash buffer (2% BSA, 0.04% Triton X-100 in PBS); iv) incubated with anti-rabbit docking strand–labeled secondary antibody for DNA-PAINT imaging (Ultivue-2, goat anti-rabbit D2) at a 1:100 dilution in antibody dilution buffer (Ultivue-2) for 2 h at RT; iv) washed three times (10 min each) with wash buffer; and v) used directly for imaging.

OligoSTORM and iOS imaging

OligoSTORM and iOS samples were imaged on a N-STORM 4.0 microscope (Nikon) with a CFI HP Apochromat TIRF 100× 1.49 oil objective and a iXon Ultra 897 camera (Andor), under highly inclined and laminated optical sheet (HILO) illumination mode. NIS elements (4.60 and 5.21) software was used for image acquisition. Conventional fluorescence images were taken at the beginning of each imaging cycle to register the position of loci and fiduciary beads. OligoSTORM images were acquired under continuous image acquisition mode with simultaneous illumination at 405 nm (with power gradually increasing) and 647 nm (at constant ~2 kW/cm^2 power density), with 16 ms exposure times, for 60,000 frames. Every 100 frames, one frame of 488 nm illumination (~ 0.05 kW/cm^2 power density) was taken to image fiduciary beads for drift correction. For iOS samples, a second acquisition cycle was performed to image H3 signal with a DNA-PAINT approach, whereby a 560 nm laser (~0.6 kW/cm^2 power density) was used to excite the Cy3-equivalent dye attached to the imager strand. H3 images were acquired under continuous image acquisition mode with 80 ms exposure time for 60,000 frames. Every 100 frames, one frame of 488 nm illumination (~0.012 kW/cm^2 power density) was taken to image fiduciary beads used for drift correction and for accurate overlap between the OligoSTORM and the H3 images.

The imaging buffer composition for OligoSTORM imaging was 100 mM cysteamine MEA (Sigma-Aldrich, #30070), 1% Glox solution (0.5 mg/ml glucose oxidase, 40 mg/ml catalase; Sigma-Aldrich, #G2133 and #C100), 5% glucose (Sigma-Aldrich, #G8270) in PBS. The imaging buffer composition for iOS (combined OligoSTORM and DNA-PAINT imaging) was 100 mM cysteamine MEA, 1% Glox solution, 5% glucose. and 0.75 nM Imager strand (I2–560, Ultivue) in Ultivue Imaging Buffer.

OligoSTORM and iOS images were analyzed and rendered in Insight3 v4.29.8 (a kind gift of Dr. Bo Huang) and ImageJ. Localizations were identified based on a threshold and fit to a simple Gaussian to determine the x and y positions[11,52]. Drift correction and overlap between OligoSTORM and the H3 DNA-PAINT SR images were performed using the localizations derived from the fiducial beads as reference in x and y using a linear affine transformation, with custom-made Matlab scripts[53]. Conventional and SR images were overlapped to verify the position of *GAPDH-IFFO1*, *STELLA* and *NANOG* loci. Spatial position and distances between loci (brightest pixel to brightest pixel) were measured in ImageJ 2.0.0. Loci areas were manually selected in ImageJ and used for cluster identification. Cluster analysis of histone H3 was carried out in Matlab with an algorithm developed in Ricci *et al.*[21], using the following analysis parameters (threshold 5; factor 5; minimum localizations per cluster 5 and iterative segmentation). Localization precision was estimated from acquired super-resolution images by estimating the standard deviation of the localizations derived from individual fluorophores over multiple frames[54].

Confocal imaging

To measure the nuclear volume of hFibs and hiPSCs, cells were fixed as described above and stained with DAPI. Full nuclear volumes were acquired on a Leica TCS SP8 confocal microscopy equipped with HC Plan-Apochromat CS2 63x/1.40 oil lens, using LAS X Software (v. 3.5.7.23225, Leica), at 400 Hz, Pinhole 1, and optimized z stack steps of 300 nm. Nuclear volumes were then estimated from confocal images using Imaris 9 software (Oxford Instruments).

Compartment analysis and epigenetic profiling of the chromosome segment (chr12: 6,141,500–8,460,000, hg19)

A/B compartments computed from Hi-C data in hFibs IMR90[9] were downloaded from the 4D Nucleome Portal (https://data.4dnucleome.org/files-processed/4DNFIHM89EGL/#details) and plotted for the chr12: 6,141,500–8,460,000, hg19 genome assembly after applying liftOver to convert from hg38 to hg19 (Extended Data Fig. 1a). ChIP-seq data in human IMR90 and hiPS-20b cells, obtained by the Human Reference Epigenome Mapping Project [30,31], were downloaded from GEO under accession numbers GSE17312 and GSE16256. The wig files obtained were imported into R to plot enrichment of each epigenetic mark in the target region (Extended Data Fig. 1a).

Gene expression analysis

RNA extraction from hFibs and hiPSCs was carried out with RNeasy Mini Kit (QIAGEN, #50974106). Reverse transcription was performed with iScript cDNA Synthesis Kit (BIO-RAD, #170–8891). qRT-PCR were run on a Lightcycler 480

(Roche) with Lightcycler 480 SYBR green I master (Roche, #4887352001) and the primers (*β-ACTIN:* ATAGCAACGTACATGGCTGG – CACCTTCTACAATGAGCTGC; *GAPDH:* AGCCACATCGCTCAGACAC – GCCCAATACGACCAAATCC; *IFFO1:* GACGTGCAGATGGAGACCTG – CGCAGTGAAAGCAGGAGACTT; *STELLA:* ACGCCGATGGACCCATCACAGTTT – TCTCGGAGGAGATTTGAGAGGCCC; *NANOG:* TGCTGAGATGCCTCACACGGA - TGACCGGGACCTTGTCTTCCTT).

Hi-C pre-capture library preparation

Each Hi-C library was generated from 20 million cells to reach a quantity of DNA high enough to perform the capture step, following the protocol from Belaghzal et al.[55]. Briefly, cells were crosslinked during 10 min at RT with 1% formaldehyde, the crosslinking reaction was then quenched by 0.125 M glycine during 5 min at RT. Chromatin was digested overnight with DpnII and DNA ends were repaired and biotinylated. *In situ* ligation was performed overnight at 16 °C. After 1h incubation at 37°C with RNase A (1 μg/ml) (TFS, #EN0531), the crosslink was reverted by incubation at 65°C overnight with proteinase K (10μg/ml final) and DNA was purified by phenol-chloroform extraction and ethanol precipitation. Biotin was then removed from unligated ends following the protocol from Lieberman-Aiden et. al. [1]. Finally, DNA was fragmented using a Covaris™ E220 to reach final fragment sizes of ~150–300 bp. Sheared DNA was end-repaired and adenylated, and biotinylated Hi-C products were pulled down with Dynabeads™ MyOne™ Streptavidin C1 (TFS, #65001), ligated to IDT adaptors with unique dual-matched indexes (Integrated DNA Technologies), and amplified with 8 PCR cycles using the KAPA HiFi HotStart ReadyMix (2×) (Roche Kapa Biosystems).

MNase digestion

After fixation with 1% formaldehyde for 10 min under agitation, 3 million cells were lysed 10 min on ice and resuspended in MNase buffer (10 mM Tris-HCl pH 7.4, 15 mM NaCl, 60 mM KCl). DNA was digested for 5 and 10 min at 37 °C using 0.1u of MNase (Sigma-Aldrich, #N3255)/500 000 cells in 100 μl (see [56] for details).The MNase fragmented genomic DNA was used to build the pre-capture libraries using the KAPA Library Preparation Kit with PCR Library Amplification/Illumina series (Roche Kapa Biosystems) with minor modifications together with IDT adaptors with unique dual-matched indexes (Integrated DNA Technologies) and amplified with 7 PCR cycles.

Custom capture of Hi-C and MNase-seq libraries

For DNA capture, the Roche's Customer Services was used to design a custom SeqCap EZ Choice Library (Roche NimbleGen) that targets the complete 2.318 Mbp genomic sequences of the *GAPDH-IFFO1*, *STELLA* or *NANOG* loci (chr12: 6,141,500–8,460,000, hg19). After masking repetitive DNA elements, the total targeted DNA represented 1,866,648 bp (80.5% of target bases covered). The PCR-enriched libraries were pooled with a combined mass of 2360 ng for Hi-C and 5000 ng for MNase-seq libraries, and hybridized to the custom SeqCap EZ Choice baits (Roche Nimblegen) at 47 °C for 72 h. After capture with beads, DNA was amplified by 14 cycles of post-capture PCR.

Paired-end next generation sequencing (NGS), mapping and filtering

Eight target-enriched libraries (two biological replicates for each of the two cell lines) were prepared in different capture pools. Each library pool was sequenced on an Illumina HiSeq2500 using the v4 flowcell chemistry or a HiSeq4000 instrument, in a fraction of a sequencing lane following the manufacturer's protocol, with a paired end run of 2×76+8+8 bp. Image analysis, base calling and quality scoring of each run was processed using the manufacturer's software Real Time Analysis (RTA 1.18.66.3) for HiSeq2500 and RTA 2.7.7 for HiSeq4000: FASTQ sequence files were then generated.

Quantification and statistical analyses

Quantification and analyses were performed in ImageJ (2.0.0), Matlab (2015, 2016b), and Imaris 9 as previously specified for each case. Statistical analysis was performed in Graphpad Prism (v9.0). For every dataset, normality tests (using D'Agostino-Pearson and Shapiro-Wilk) were run to assess normal distribution. Datasets: i) with Gaussian distribution of values, parametric tests were applied, two-tailed tests were run, and multiple comparison corrections were applied for datasets with > 2 groups and multiple comparisons; ii) with non-Gaussian distribution, non-parametric tests were applied. The type of statistical test is specified in each case. Not significant (ns), $p > 0.05$; *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$; **** $p \leq 0.0001$.

Nucleosome calling

MNase-seq paired-end reads were mapped to the human genome (hg19, Feb. 2009 GRCh37) using BowTie aligner [57], allowing 1 or 2 mismatches and maximum insert size of 500 bp. Output BAM files were imported into R[58], and quality control was performed with htSeqTools package to remove PCR artifacts [59]. Filtered reads were processed with nucleR package[41,60] as follows (the script with all steps to run nucleR can be found at (https://github.com/nucleosome-dynamics/nucleosome_dynamics/blob/master/bin/nucleR.R): mapped fragments were trimmed to 50bp maintaining the original center and transformed to reads per million. Noise was filtered through Fast Fourier Transform, keeping 1% of the principal components. For peak calling, peak width = 147 bp, peak detection threshold = 10, maximum overlap = 80 bp, and dyad length = 50 bp. Well-positioned nucleosomes (e.g., coverage from the cell population localized around a clear and well-defined peak) were low, while scores of fuzzy nucleosomes (e.g., blurry coverage with high variability in the cell population) were high.

Hi-C data processing and normalization

Capture Hi-C data was processed using TADbit [61] (https://github.com/3DGenomes/tadbit) for quality control, mapping, and filtering. First, quality control was performed with the FastQC protocol implementation in TADbit. Reads were then mapped to the reference human genome (hg19, Feb. 2009 GRCh37) with a fragment-based strategy. Afterwards, non-informative contacts (self-circle, dangling-end, error, duplicated, and random breaks) identified by TADbit were filtered out, leaving 44–87 million valid interactions per experiment. Off-target contacts (e.g., neither end of the read mapped to one of the capture regions) were also discarded (see Supplementary Table 5). Finally, contact matrices were

created from valid reads at 5-kb resolution with the corresponding TADbit module, and low frequency bins were removed.

Contact matrices for the captured region (chr12: 6140000–8460000) were balanced with the Iterative Correction (ICE) method [62] using the implementation provided in HiTC R package [63] with parameters max_iter = 1000, eps = 1e-5, sparse.filter = 0.01. For Fig. 3c, d, normalized contact matrices were transformed into Binary Upper TrianguLar MatRix (BUTLR) file format, using BUTLRtools (https://github.com/yuelab/BUTLRTools) suited for 3D Genome Browser (http://3dgenome.org) to visualize contact maps together with genome annotations [64].

TAD boundaries were detected computing insulation score in the balanced contact matrices as described [39]. Briefly, the degree of separation by a given locus between the two adjacent regions was computed as the relative frequency of contacts over that bin (j) at distance s:

$$R_j(s) = log2\left(\sum_{k=j-s/2}^{j+s/2} \frac{C_{k,\ k+s}}{M}\right)$$

where

$$M = mean_s\left(\sum_{k=j-s/2}^{j+s/2} C_{k,\ k+s}\right)$$

Negative values of $R_j(s)$ indicate insulation at bin j. Then, TAD boundaries were defined at local minima of smoothed $R_j(s)$, with a cubic smoothing spline in R (R Core Team 2016).

Software

Insight3 v4.29.8 software used for SR image processing has been kindly provided by Dr Bo Huang (UCSF) [73]. ImageJ 2.0.0 software used for SR image analysis can be found at: https://imagej.net/downloads. Graphpad Prism v 9.0 software used for statistical analysis can be found at https://www.graphpad.com/scientific-software/prism/. MatLab (2015 and 2016b) software used for image drift correction, overlap and data analysis can be found at: https://www.mathworks.com/products/matlab.html. OligoMiner scripts used for library design are available at: https://github.com/brianbeliveau/OligoMiner. Bitplane Imaris 9 software (Oxford Instruments) used to measure nuclear volumes can be found at: https://imaris.oxinst.com. Amber 18 software [66] was used for distance restraint-based chromatin simulations. Simulation details are provided as supplementary material.

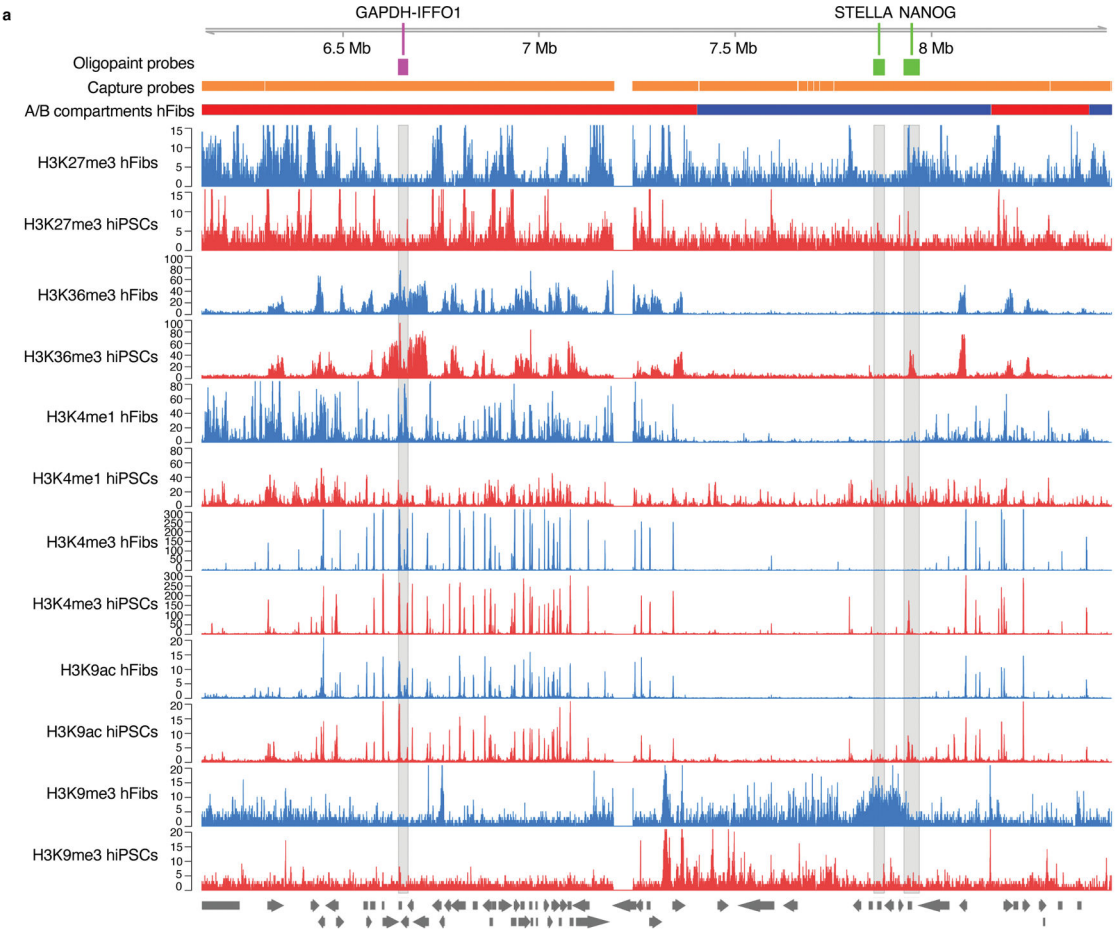
DATA AVAILABILITY

Raw data for the capture MNase-seq and Hi-C sequencing experiments generated in this study were deposited at the European Nucleotide Archive under accession number PRJEB42293. Imaging and modeling datasets generated in this work are available upon request. We provide raw data related to plots and statistical source data in the source data section.

CODE AVAILABILITY

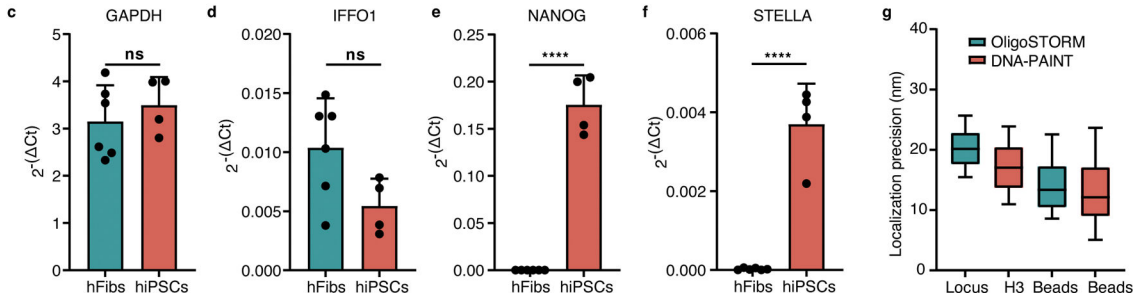
Stand-alone versions of the softwares used for chromatin coarse-grained simulations and for the fitting algorithms developed herein are available in the following repositories: Chromatin Dynamics (http://mmb.irbbarcelona.org/gitlab/juanpablo/chrom_dyn) and Chromatin Fitting (http://mmb.irbbarcelona.org/gitlab/juanpablo/fit_chrom).

Extended Data

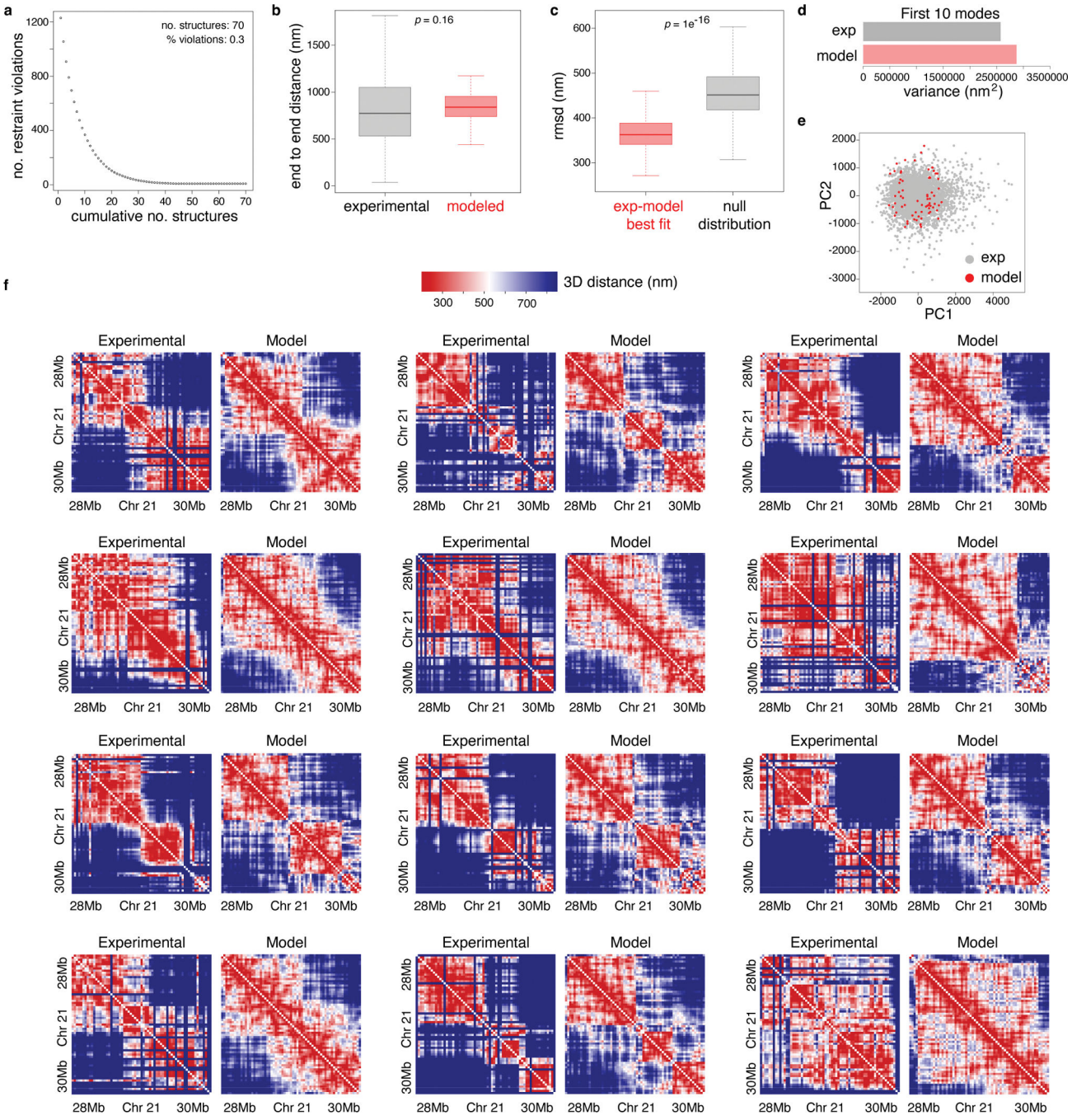


b

Region / Name	Start	End	Size (kb)	Number	Probes density (probes/kb)
GAPDH - IFFO1	6641500	6666000	24.5	286	11.67
STELLA	7854000	7881000	27	158	5.85
NANOG	7931000	7970000	39	185	4.74



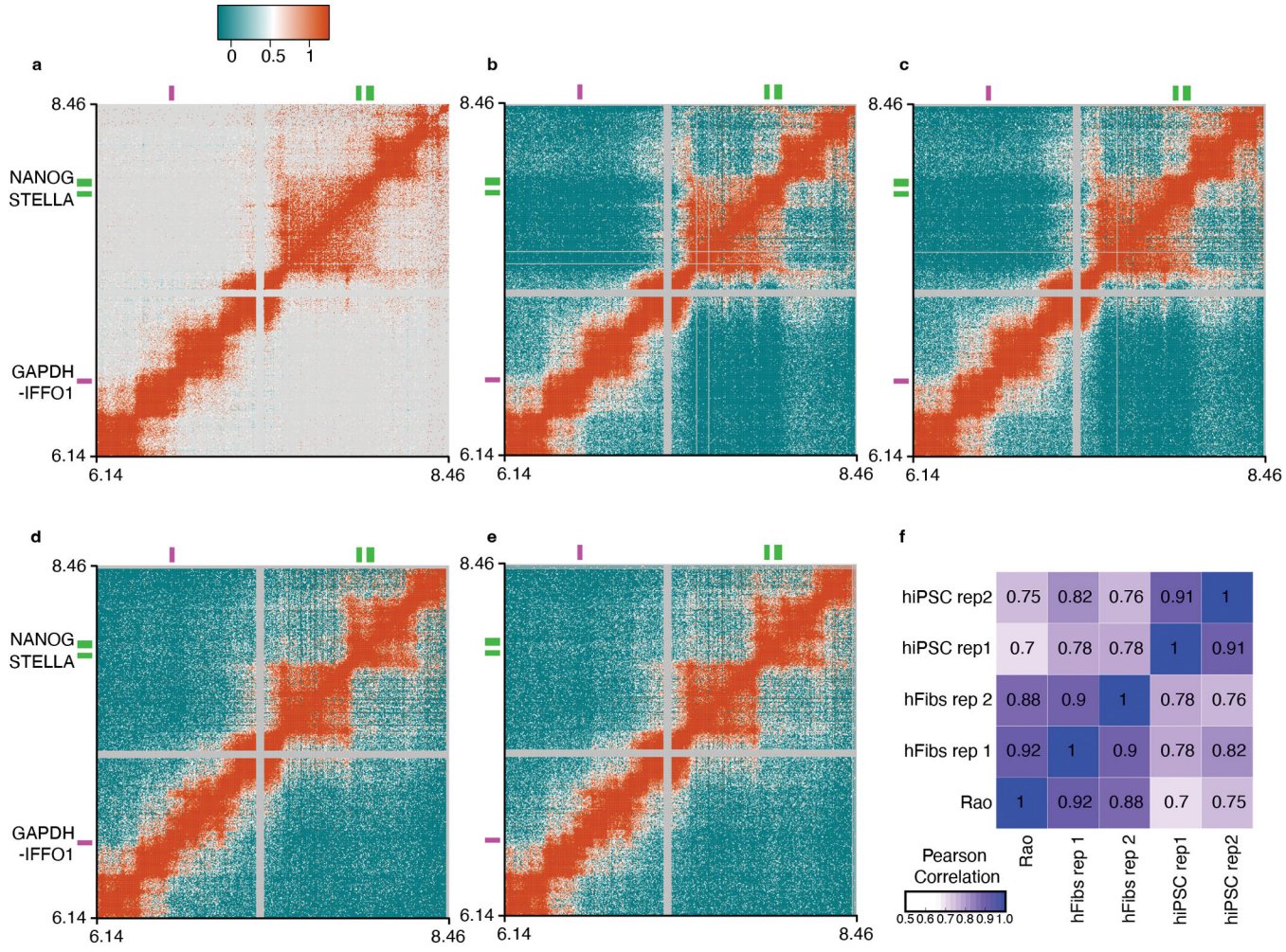
Extended Data Fig. 1. Schematic overview of human chromosome 12 region analyzed by MiOS. **a** Schematic representation of chr12:6,140,000:8,460,000 region, showing the position of genes (grey arrows), Oligopaint probes (in green for *NANOG*, *STELLA* and magenta for *GAPDH-IFFO1*), and capture probes (orange). The A/B compartment track shows active A (red) and repressed B (blue) compartments for hFibs (from Hi-C; taken from [9]. Epigenetic marks for hFibs IMR90 (blue) and hiPSCs 20-b (red) are displayed (ChIP-seq tracks taken from [30,31]. The positions of the regions analyzed for the target genes (from left to right: *GAPDH-IFFO1*, *STELLA*, and *NANOG*) are highlighted in grey. **b** Genomic coordinates of Oligopaint probes. **c-f** qRT-PCR analysis in hFibs and hiPSCs for expression of (**c**) *GAPDH*, (**d**) *IFFO1*, (**e**) *NANOG*, and (**f**) *STELLA*. Mean and standard deviation (SD) of 2^-dCt values to *B-ACTIN* are shown; n = 6, and n = 4, independent replicates for hFibs and hiPSCs, respectively; two-tailed unpaired *t*-test; p = 0.4692 (c), p = 0.0672 (d), p = 1.4e-13 (e), p = 1.02e-7 (f). **g** Quantification of localization precision of super-resolution images. Boxplots (median with interquartile range) and whisker plots (10–90 percentile) are shown for oligoSTORM (locus, n = 8995), DNA-PAINT (H3, n = 23023), and oligoSTORM and DNA-PAINT beads (n = 135 and n = 158 localization tracks, respectively).



Extended Data Fig. 2. Restraint-based model reproduces cell-to-cell structural variability.

a Evolution of the number of input distance restraint violations from the experimental median distance matrix (Fig. 2a, left panel) when adding subsequent modeled structures to the ensemble obtained with the restraint-based approach. **b** End-to-end distance distributions for the experimental (gray) and modeled (red) ensembles. The boxes highlight the first, second and third quartiles, while the whiskers extend 1.5 times the interquartile range away from the box edges. Outliers are omitted. The plots come from 3,496 experimental and 70 modeled conformations. $p = 0.16$, two-sided Mann-Whitney test. **c** Root mean

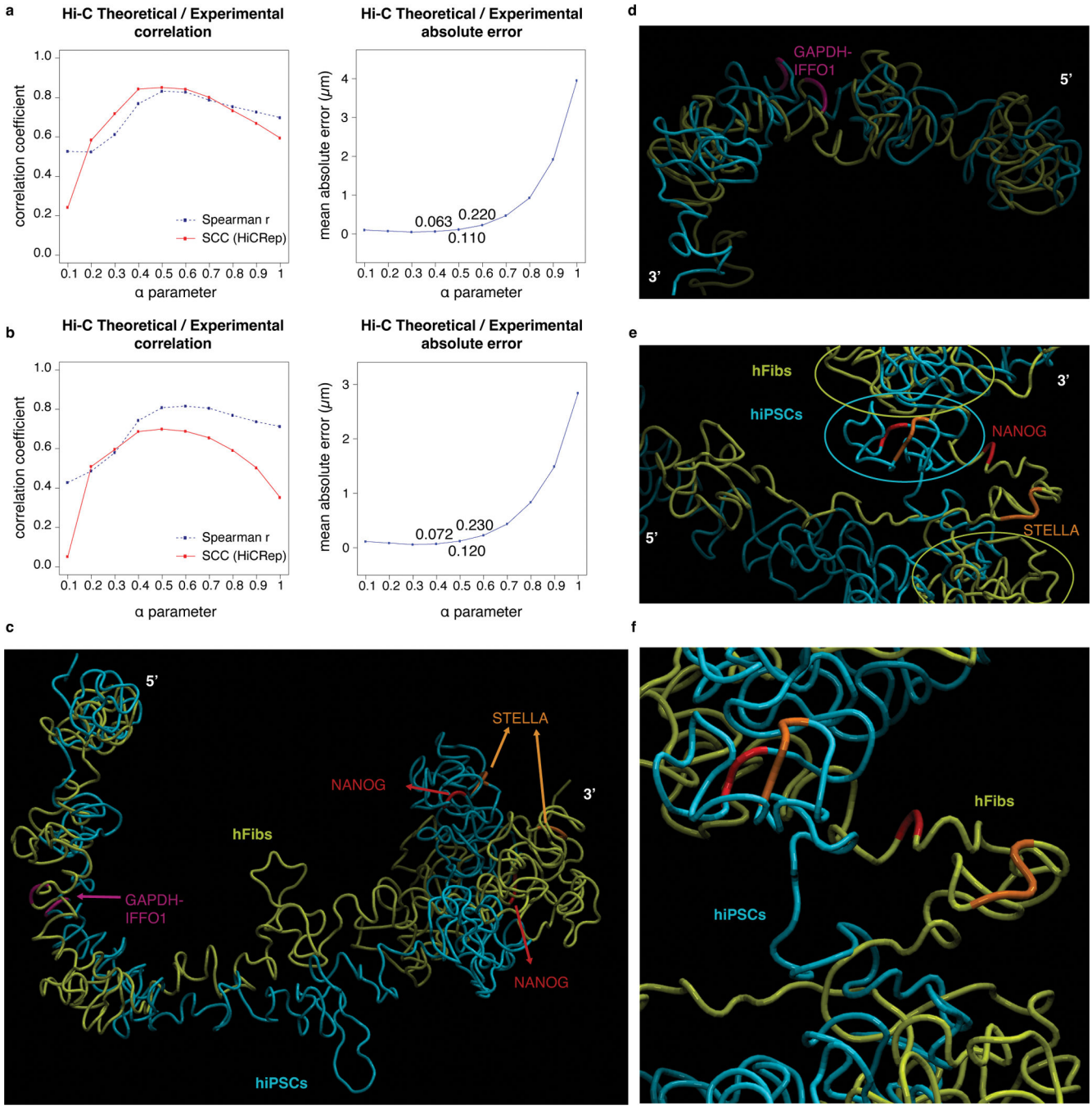
square deviation (rmsd) of bead/probe positions for best fitted modeled structures against each experimental structure (red, $N_{exp\text{-}model}$ = 3,496), and null distribution of rmsd values between all experimental structures, after fitted / aligned (gray, N_{null} = 6,109,260). $p < 1e^{-16}$, two-sided Mann-Whitney test. The boxes highlight the first, second and third quartiles, while the whiskers extend 1.5 times the interquartile range away from the box edges. Outliers are omitted. **d** Variance from the first 10 principal components from PCA. **e** Projection of the displacement vectors onto the first 2 principal components from PCA. **f** 3D distance matrices for single structures extracted from experimental microscopy data [18] (left) and from the ensemble obtained with the restraint-based model (right). The color scale ranges from 200 nm to 850 nm.



Extended Data Fig. 3. Contact matrices and correlation analyses between capture Hi-C replicates in hFibs, hiPSCs and published datasets.

a-c Contact matrices for the region chr12:6,140,000:8,460,000 in hFibs, displayed at 5-kb resolution, for (**a**) Hi-C data from Rao et al. (2014), (**b**) capture Hi-C for replica 1, and (**c**) capture Hi-C for replica 2. Plotted values are log10 of iteratively corrected interaction counts scaled to sum 1 million. The position of genes *GAPDH-IFFO1* (magenta) and of *STELLA* and *NANOG* (green) are marked on X and Y axes. **d, e** Replicates 1 and 2, respectively, of

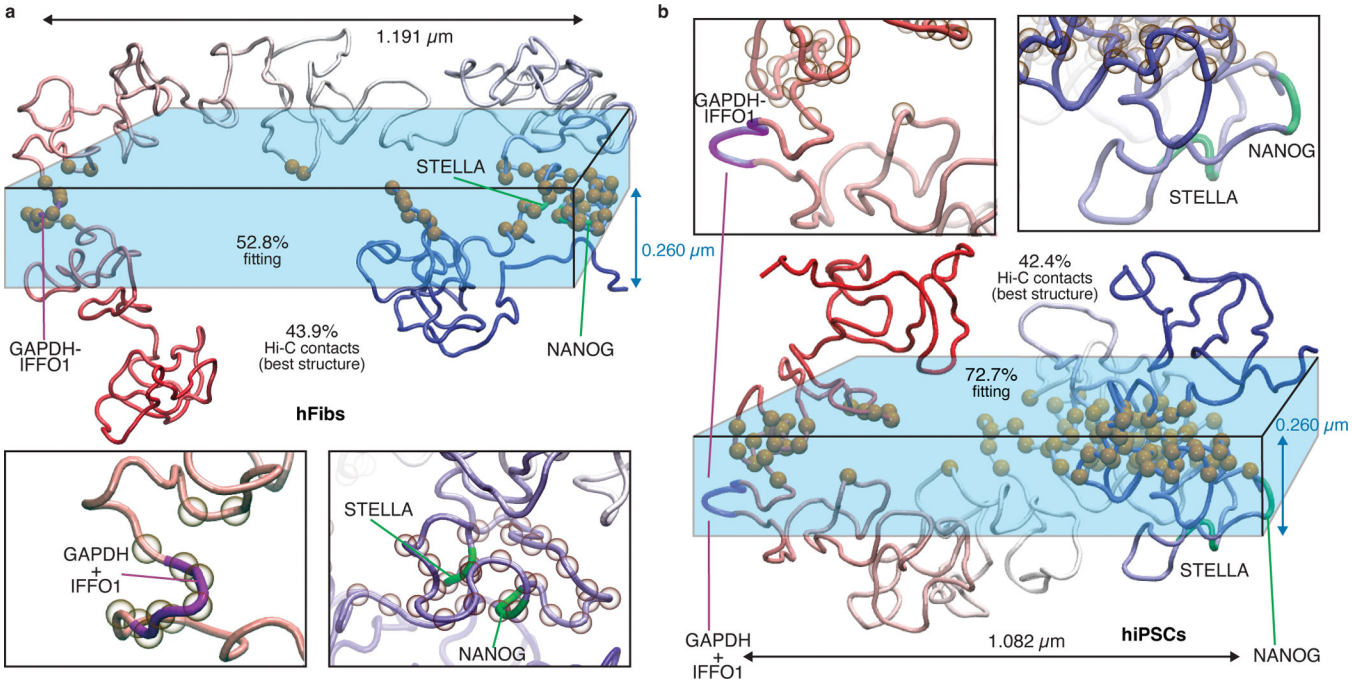
the capture Hi-C from hiPSCs. Plotted values are the same as in (a-c). **f** Pearson correlation coefficient of the contact matrices between every pair of experiments: Rao et al. 2014, in-house hFibs (replicates 1 and 2), and hiPSCs (replicates 1 and 2).



Extended Data Fig. 4. Parameter selection and structure overlap for restraint-based models.

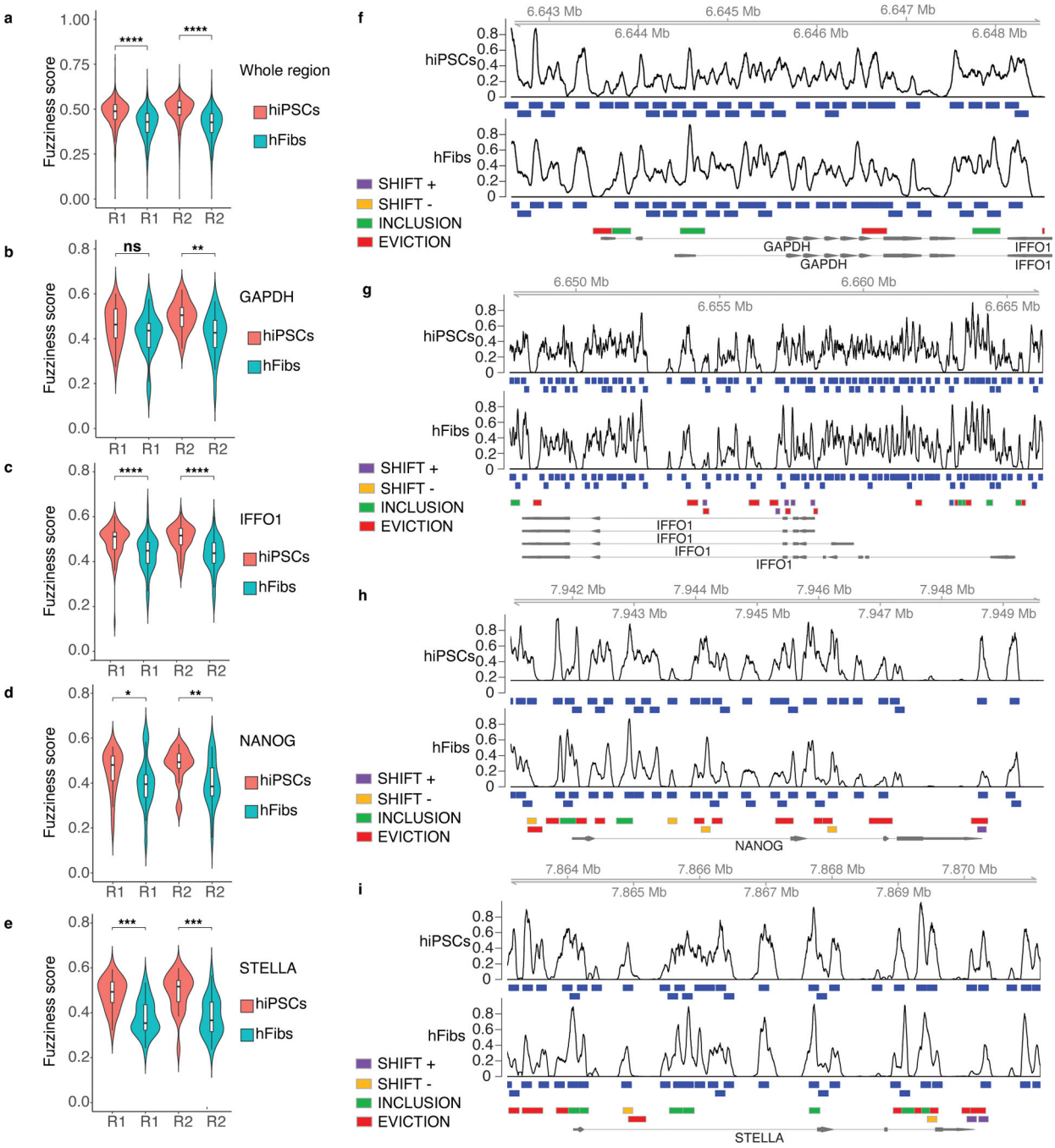
a, b Tuning of α parameter used in the distance restraint-based model for hFibs (a) or hiPSCs (b). For each α, correlation between experimental Hi-C interaction matrix and the modeled contact matrix (left) and mean absolute error between Hi-C derived average

distances and predicted ensemble mean distances (right). Both Spearman and stratum-adjusted (HiCRep) [68] correlation coefficients are shown. **c** Representation of the 2.3 Mb region of human chr12 segment from hiPSCs (cyan) and hFibs (yellow) cells. The *GAPDH-IFFO1*, *NANOG*, and *STELLA* loci are colored in pink, red, and orange, respectively. **d** Close-up of the *GAPDH-IFFO1* region. **e, f** Close-ups of the *STELLA/NANOG* region highlighting the relative location of the two genes with respect to TAD formation in hiPSCs (cyan circle) and hFibs (yellow circles) cells.



Extended Data Fig. 5. Fittings using iOS localizations, capture Hi-C contacts, and the restraint-based model of chromatin.

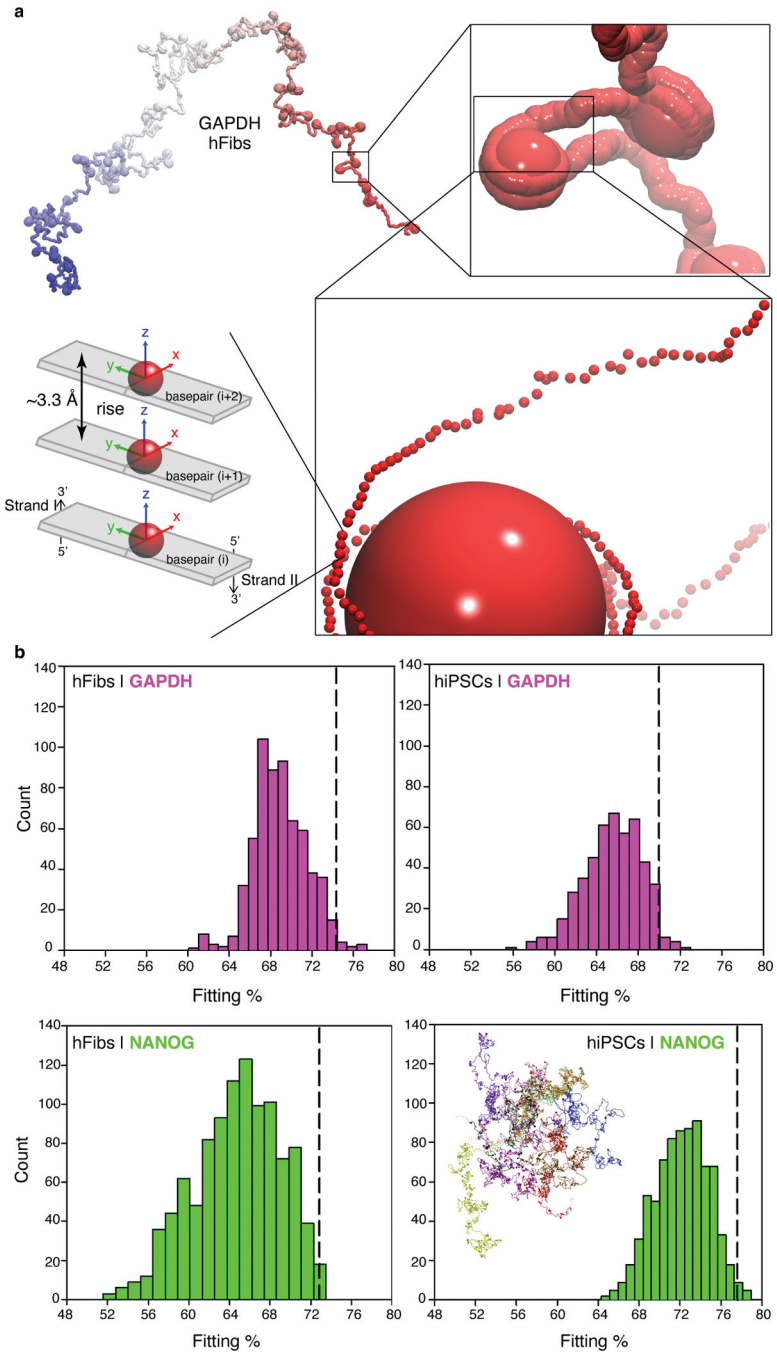
Two specific cells with a given distance between both gene regions (*GAPDH-IFFO1* and *NANOG*) are shown. **a** Structure from the simulated ensemble (distance *GAPDH-IFFO1* to *NANOG* fixed at 1.191 μm) that best fits the iOS localizations within the confocal plane (considering a depth of 0.260 μm) of one hFib cell. Note that this single structure of the chr12 segment connecting the genes of interest fit to 52.8% of the iOS localizations (5-kb beads fitted are shown as orange spheres) and fulfills 43.9% of the Hi-C contacts simultaneously. Zoom-in of the genes showing the beads fitted to iOS localizations. **b** Same as (a) for a hiPS cell, where the *GAPDH-IFFO1* to *NANOG* distance was fixed to 1.082 μm, fulfilling 72.7% of iOS localizations and 42.4% of the Hi-C contacts.



Extended Data Fig. 6. Nucleosome positioning in hiPSC and hFibs cells determined from capture MNase-seq.

a-e Comparison of fuzziness score obtained with nucleR (0:well-positioned - 1:fuzzy nucleosome) between hiPSCs and hFibs for nucleosomes detected in the complete captured region (chr12:6140000-8460000). (**a**) and at the individual genes (**b-e**) Replica 1 (R1) and 2 (R2) are shown. Box plots include a marker for the median of the data and a box indicating the interquartile range. Whiskers show minimum and maximum values. Wilcoxon rank sum test; (a) $p < 2.22e-16$ (188 vs 172 nucleosomes over two independent experiments, in

hiPSCs and hFibs, respectively), (b) R1: p = 0.17, R2: p = 0.0012 (35 vs 34 nucleosomes over two independent experiments, in hiPSCs and hFibs, respectively) (c) R1: p = 4.8e-8, R2: p = 4.2e-10 (92 vs 83 nucleosomes over two independent experiments, in hiPSCs and hFibs, respectively), (d) R1: p = 0.04, R2: p = 0.0069 (32 vs 29 nucleosomes over two independent experiments, in hiPSCs and hFibs, respectively), (e) R1: p = 0.00022, R2: p = 0.0014 (29 vs 26 nucleosomes over two independent experiments, in hiPSCs and hFibs, respectively). **f-i** Nucleosome positioning around the following genes: (**f**) *GAPDH*, (**g**) *IFFO1*, (**h**) *NANOG*, and (**i**) *STELLA*. Black lines represent normalized (0–1) nucleosome coverage. Blue boxes are the nucleosome positions detected by nucleR [60]. Changes in nucleosome organization from hiPSCs to hFibs, detected with NucDyn [41] are represented as color-coded boxes for inclusion (green), eviction (red), positive shifts (purple), and negative shifts (yellow).



Extended Data Fig. 7. Amplified views of the bottom-up first-principle coarse-grained model of the nucleosome fiber and the distribution of fitting values when the sampled conformations are confronted to iOS localizations.

a A representative folded GAPDH gene is amplified until consecutives centroids are shown. Each individual DNA centroid is located at the base pair reference frame (BPRF) following Cambridge and Tsukuba conventions [72,75]. These centroids represent the monomer length defined in our implementation, whose arbitrariness was based on a detailed knowledge on the structure and dynamics of B-DNA at the atomistic level [42,70]. In the first amplification, each DNA centroid is roughly represented considering its spherical exclusion volume (radius

of van der Waals) centered at the BPRF. In the last two amplifications, the DNA-excluded volume is no longer depicted, and only the center of the BPRF is shown. Note that on average, being B-DNA, the distance between two consecutives base pairs is ~3.3 Å, although the experimentally falsifiable resolution of our predictions ranges from nucleosome clutches to near single nucleosome particles. **b** Distribution of the fitting values obtained from the filtered ensembles of GAPDH and NANOG folded conformations in hFibs and hiPSCs when confronted to iOS localizations. The top 10 structures with the highest fitting numbers, for which physical descriptors were computed and reported in Fig. 7 and Supplementary Table 2, are found at the right of the vertical dashed lines.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors acknowledge the support from the Barcelona Institute of Science and Technology (BIST) Ignite Grants [Seeding Stage 2017 and Second Phase 2018, to MVN and PDD]. European Union's Horizon 2020 Research and Innovation Programme [CellViewer No 686637 to ML and MPC; ERC SimDNA No 676556 to MO and under the Marie Skłodowska-Curie grant agreement No. 754510 to JPA]; Ministerio de Ciencia e Innovación, grant [008506 - PID2020-114080GB-I00 to MPC], and an AGAUR grant from Secretaria d'Universitats i Recerca del Departament d'Empresa iConeixement de la Generalitat de Catalunya [2017 SGR 1110 to MO and 006712 BFU2017-86760-P (AEI/FEDER, UE) to MPC]; Centro de Excelencia Severo Ochoa [CEX2020-001049-S, MCIN/AEI/10.13039/501100011033 to CRG and CNAG authors and awarded to IRB Barcelona 2020–25]; CERCA Programme/Generalitat de Catalunya [to CRG and CNAG authors]; People Program (Marie Curie Actions) FP7/2007–2013 under REA grant [608959 to MVN]; Juan de la Cierva-Incorporación 2017 [to MVN]; INTREPiD Postdoctoral Programme co-funded by the European Commission [under grant agreement 754422, to XG]; Grant for the recruitment of early-stage research staff FI-2020 [Operational Program of Catalonia 2014–2020 CCI 2014ES05SFOP007 of the European Social Fund to LM] and La Caixa" Foundation fellowship [LCF/BQ/DR20/11790016 to LM]; Spanish Ministry of Science [for the EMBL partnership to CRG and CNAG authors and RTI2018-096704-B-100 to MO]; Instituto de Salud Carlos Tercero [to CNAG authors; PT17/0009/0007 to MO]; The Biomolecular and Bioinformatics Resources Platform (ISCIIIPT 13/000/0030 co-funded by the Fondo Europeo de Desarrollo Regional FEDER) [grants Elixir-Excelerate: 676559; BioExcel2:823830 and MuG: 676566 to MO]; NIH R01GM123289 and R01HD091797 [to JAA and C-tW]; Bruker Inc [to C-tW]; PEDECIBA (Programa de Desarrollo de las Ciencias Básicas) and SNI-ANII (Sistema Nacional de Investigadores, Agencia Nacional de Investigación e Innovación, Uruguay) [to PDD]; ICREA (Institucio Catalana de Recerca i Estudis Avançats) [to MO and MPC]. The authors acknowledge the advanced light microscopy unit (ALMU) unit from CRG for their excellent technical support.

REFERENCES

1. Lieberman-Aiden E et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289–293, doi:10.1126/science.1181369 (2009). [PubMed: 19815776]
2. Fullwood MJ et al. An oestrogen-receptor-alpha-bound human chromatin interactome. Nature 462, 58–64, doi:10.1038/nature08497 (2009). [PubMed: 19890323]
3. Hsieh TH et al. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. Cell 162, 108–119, doi:10.1016/j.cell.2015.05.048 (2015). [PubMed: 26119342]
4. Rowley MJ & Corces VG Organizational principles of 3D genome architecture. Nature reviews. Genetics 19, 789–800, doi:10.1038/s41576-018-0060-8 (2018).
5. Dixon JR et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380, doi:10.1038/nature11082 (2012). [PubMed: 22495300]
6. 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, doi:10.1038/nbt.2057 (2011). [PubMed: 22198700]

7. Sexton T et al. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell 148, 458–472, doi:10.1016/j.cell.2012.01.010 (2012). [PubMed: 22265598]
8. Nora EP et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485, 381–385, doi:10.1038/nature11049 (2012). [PubMed: 22495304]
9. Rao SS et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680, doi:10.1016/j.cell.2014.11.021 (2014). [PubMed: 25497547]
10. Lakadamyali M & Cosma MP Visualizing the genome in high resolution challenges our textbook understanding. Nat Methods 17, 371–379, doi:10.1038/s41592-020-0758-3 (2020). [PubMed: 32123395]
11. Rust MJ, Bates M & Zhuang X Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat Methods 3, 793–795, doi:10.1038/nmeth929 (2006). [PubMed: 16896339]
12. Beliveau BJ et al. Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. Nat Commun 6, 7147, doi:10.1038/ncomms8147 (2015). [PubMed: 25962338]
13. Cardozo Gizzi AM et al. Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. Molecular cell 74, 212–222 e215, doi:10.1016/j.molcel.2019.01.011 (2019). [PubMed: 30795893]
14. Takei Y et al. Integrated spatial genomics reveals global architecture of single nuclei. Nature 590, 344–350, doi:10.1038/s41586-020-03126-2 (2021). [PubMed: 33505024]
15. Boettiger AN et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature 529, 418–422, doi:10.1038/nature16496 (2016). [PubMed: 26760202]
16. Nir G et al. Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. PLoS Genet 14, e1007872, doi:10.1371/journal.pgen.1007872 (2018). [PubMed: 30586358]
17. Mateo LJ et al. Visualizing DNA folding and RNA in embryos at single-cell resolution. Nature 568, 49–54, doi:10.1038/s41586-019-1035-4 (2019). [PubMed: 30886393]
18. Bintu B et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. 1783, doi:10.1126/science.aau1783 (2018).
19. Su JH, Zheng P, Kinrot SS, Bintu B & Zhuang X Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. Cell 182, 1641–1659 e1626, doi:10.1016/j.cell.2020.07.032 (2020). [PubMed: 32822575]
20. Nguyen HQ et al. 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. Nat Methods 17, 822–832, doi:10.1038/s41592-020-0890-0 (2020). [PubMed: 32719531]
21. Ricci MA, Manzo C, Garcia-Parajo MF, Lakadamyali M & Cosma MP Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. Cell 160, 1145–1158, doi:10.1016/j.cell.2015.01.054 (2015). [PubMed: 25768910]
22. Szabo Q et al. Regulation of single-cell genome organization into TADs and chromatin nanodomains. Nat Genet 52, 1151–1157, doi:10.1038/s41588-020-00716-8 (2020). [PubMed: 33077913]
23. Dans PD, Walther J, Gomez H & Orozco M Multiscale simulation of DNA. Curr Opin Struct Biol 37, 29–45, doi:10.1016/j.sbi.2015.11.011 (2016). [PubMed: 26708341]
24. Buitrago D et al. Impact of DNA methylation on 3D genome structure. Nature Communications 12, 3243, doi:10.1038/s41467-021-23142-8 (2021).
25. Di Stefano M, Paulsen J, Jost D & Marti-Renom MA 4D nucleome modeling. Current opinion in genetics & development 67, 25–32, doi:10.1016/j.gde.2020.10.004 (2020). [PubMed: 33253996]
26. Conte M et al. Polymer physics indicates chromatin folding variability across single-cells results from state degeneracy in phase separation. Nat Commun 11, 3289, doi:10.1038/s41467-020-17141-4 (2020). [PubMed: 32620890]
27. Fiorillo L et al. Comparison of the Hi-C, GAM and SPRITE methods using polymer models of chromatin. Nat Methods 18, 482–490, doi:10.1038/s41592-021-01135-1 (2021). [PubMed: 33963348]

28. Abbas A et al. Integrating Hi-C and FISH data for modeling of the 3D organization of chromosomes. Nat Commun 10, 2049, doi:10.1038/s41467-019-10005-6 (2019). [PubMed: 31053705]
29. Levasseur DN, Wang J, Dorschner MO, Stamatoyannopoulos JA & Orkin SH Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells. Genes & development 22, 575–580, doi:10.1101/gad.1606308 (2008). [PubMed: 18283123]
30. Hawkins RD et al. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell 6, 479–491, doi:10.1016/j.stem.2010.03.018 (2010). [PubMed: 20452322]
31. Bernstein BE et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol 28, 1045–1048, doi:10.1038/nbt1010-1045 (2010). [PubMed: 20944595]
32. Yu J et al. Induced pluripotent stem cell lines derived from human somatic cells. Science 318, 1917–1920, doi:10.1126/science.1151526 (2007). [PubMed: 18029452]
33. Jungmann R et al. Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. Nat Methods 11, 313–318, doi:10.1038/nmeth.2835 (2014). [PubMed: 24487583]
34. Schnitzbauer J, Strauss MT, Schlichthaerle T, Schueder F & Jungmann R Super-resolution microscopy with DNA-PAINT. Nat Protoc 12, 1198–1228, doi:10.1038/nprot.2017.024 (2017). [PubMed: 28518172]
35. Chiariello AM et al. A Dynamic Folded Hairpin Conformation Is Associated with alpha-Globin Activation in Erythroid Cells. Cell Rep 30, 2125–2135 e2125, doi:10.1016/j.celrep.2020.01.044 (2020). [PubMed: 32075757]
36. Oudelaar AM et al. Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. Nat Genet 50, 1744–1751, doi:10.1038/s41588-018-0253-2 (2018). [PubMed: 30374068]
37. Oudelaar AM, Beagrie RA, Kassouf MT & Higgs DR The mouse alpha-globin cluster: a paradigm for studying genome regulation and organization. Current opinion in genetics & development 67, 18–24, doi:10.1016/j.gde.2020.10.003 (2021). [PubMed: 33221670]
38. Brown JM et al. A tissue-specific self-interacting chromatin domain forms independently of enhancer-promoter interactions. Nat Commun 9, 3849, doi:10.1038/s41467-018-06248-4 (2018). [PubMed: 30242161]
39. Mizuguchi T et al. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. Nature 516, 432–435, doi:10.1038/nature13833 (2014). [PubMed: 25307058]
40. Blinka S, Reimer MH Jr., Pulakanti K & Rao S Super-Enhancers at the Nanog Locus Differentially Regulate Neighboring Pluripotency-Associated Genes. Cell Rep 17, 19–28, doi:10.1016/j.celrep.2016.09.002 (2016). [PubMed: 27681417]
41. Buitrago D et al. Nucleosome Dynamics: a new tool for the dynamic analysis of nucleosome positioning. Nucleic Acids Res 47, 9511–9523, doi:10.1093/nar/gkz759 (2019). [PubMed: 31504766]
42. Walther J et al. A multi-modal coarse grained model of DNA flexibility mappable to the atomistic level. Nucleic Acids Res 48, e29, doi:10.1093/nar/gkaa015 (2020). [PubMed: 31956910]
43. Maeshima K, Ide S, Hibino K & Sasai M Liquid-like behavior of chromatin. Current opinion in genetics & development 37, 36–45, doi:10.1016/j.gde.2015.11.006 (2016). [PubMed: 26826680]
44. Schueder F et al. An order of magnitude faster DNA-PAINT imaging by optimized sequence design and buffer conditions. Nat Methods 16, 1101–1104, doi:10.1038/s41592-019-0584-7 (2019). [PubMed: 31591576]
45. Stadhouders R et al. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. Nat Genet 50, 238–249, doi:10.1038/s41588-017-0030-7 (2018). [PubMed: 29335546]
46. Krietenstein N et al. Ultrastructural Details of Mammalian Chromosome Architecture. Molecular cell 78, 554–565 e557, doi:10.1016/j.molcel.2020.03.003 (2020). [PubMed: 32213324]
47. Hua P et al. Defining genome architecture at base-pair resolution. Nature 595, 125–129, doi:10.1038/s41586-021-03639-4 (2021). [PubMed: 34108683]

48. Ohno M, Ando T, Priest DG & Taniguchi Y Hi-CO: 3D genome structure analysis with nucleosome resolution. Nat Protoc 16, 3439–3469, doi:10.1038/s41596-021-00543-z (2021). [PubMed: 34050337]
49. Tan L, Xing D, Chang CH, Li H & Xie XS Three-dimensional genome structures of single diploid human cells. Science 361, 924–928, doi:10.1126/science.aat5641 (2018). [PubMed: 30166492]
50. Beliveau BJ et al. OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. Proceedings of the National Academy of Sciences of the United States of America 115, E2183–E2192, doi:10.1073/pnas.1714530115 (2018). [PubMed: 29463736]
51. Beliveau BJ et al. In Situ Super-Resolution Imaging of Genomic DNA with OligoSTORM and OligoDNA-PAINT. Methods Mol Biol 1663, 231–252, doi:10.1007/978-1-4939-7265-4_19 (2017). [PubMed: 28924672]
52. Bates M, Huang B, Dempsey GT & Zhuang X Multicolor super-resolution imaging with photo-switchable fluorescent probes. Science 317, 1749–1753, doi:10.1126/science.1146598 (2007). [PubMed: 17702910]
53. Otterstrom J, Garcia AC, Vicario C, Cosma MP & Lakadamyali M Nanoscale decompaction of nucleosomal DNA revealed through multi-color super-resolution microscopy. bioRxiv, 470823, doi:10.1101/470823 (2018).
54. Gomez-Garcia PA, Garbacik ET, Otterstrom JJ, Garcia-Parajo MF & Lakadamyali M Excitation-multiplexed multicolor superresolution imaging with fm-STORM and fm-DNA-PAINT. Proceedings of the National Academy of Sciences of the United States of America 115, 12991–12996, doi:10.1073/pnas.1804725115 (2018). [PubMed: 30509979]
55. Belaghzal H, Dekker J & Gibcus JH Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. Methods 123, 56–65, doi:10.1016/j.ymeth.2017.04.004 (2017). [PubMed: 28435001]
56. Cui XJ, Li H & Liu GQ Combinatorial patterns of histone modifications in Saccharomyces cerevisiae. Yeast 28, 683–691, doi:10.1002/yea.1896 (2011). [PubMed: 21815215]
57. Langmead B, Trapnell C, Pop M & Salzberg SL Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10, R25, doi:10.1186/gb-2009-10-3-r25 (2009). [PubMed: 19261174]
58. R: A Language and Environment for Statistical Computing (2014).
59. Planet E, Attolini CS, Reina O, Flores O & Rossell D htSeqTools: high-throughput sequencing quality control, processing and visualization in R. Bioinformatics 28, 589–590, doi:10.1093/bioinformatics/btr700 (2012). [PubMed: 22199381]
60. Flores O & Orozco M nucleR: a package for non-parametric nucleosome positioning. Bioinformatics 27, 2149–2150, doi:10.1093/bioinformatics/btr345 (2011). [PubMed: 21653521]
61. Serra F et al. Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. PLoS Comput Biol 13, e1005665, doi:10.1371/journal.pcbi.1005665 (2017). [PubMed: 28723903]
62. Imakaev M et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods 9, 999–1003, doi:10.1038/nmeth.2148 (2012). [PubMed: 22941365]
63. Servant N et al. HiTC: exploration of high-throughput 'C' experiments. Bioinformatics 28, 2843–2844, doi:10.1093/bioinformatics/bts521 (2012). [PubMed: 22923296]
64. Wang Y et al. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol 19, 151, doi:10.1186/s13059-018-1519-9 (2018). [PubMed: 30286773]
65. Adhikari B, Trieu T & Cheng J Chromosome3D: reconstructing three-dimensional chromosomal structures from Hi-C interaction frequency data using distance geometry simulated annealing. BMC Genomics 17, 886, doi:10.1186/s12864-016-3210-4 (2016). [PubMed: 27821047]
66. Case DA et al. AMBER 2018. (2018).
67. Tjong H et al. Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. Proceedings of the National Academy of Sciences of the United States of America 113, E1663–1672, doi:10.1073/pnas.1512577113 (2016). [PubMed: 26951677]

68. Yang T et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. Genome Res 27, 1939–1949, doi:10.1101/gr.220640.117 (2017). [PubMed: 28855260]

69. Meaburn KJ, Misteli T & Soutoglou E Spatial genome organization in the formation of chromosomal translocations. Semin Cancer Biol 17, 80–90, doi:10.1016/j.semcancer.2006.10.008 (2007). [PubMed: 17137790]

70. Dans PD et al. The static and dynamic structural heterogeneities of B-DNA: extending Calladine-Dickerson rules. Nucleic Acids Res 47, 11090–11102, doi:10.1093/nar/gkz905 (2019). [PubMed: 31624840]

71. Ivani I et al. Parmbsc1: a refined force field for DNA simulations. Nat Methods 13, 55–58, doi:10.1038/nmeth.3658 (2016). [PubMed: 26569599]

72. da Rosa G et al. Sequence-dependent structural properties of B-DNA: what have we learned in 40 years? Biophysical Reviews, doi:10.1007/s12551-021-00893-8 (2021).

73. Huang B, Wang W, Bates M & Zhuang X Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. Science 319, 810–813, doi:10.1126/science.1153529 (2008). [PubMed: 18174397]

74. Lun AT & Smyth GK diffHic: a Bioconductor package to detect differential genomic interactions in Hi-C data. BMC Bioinformatics 16, 258, doi:10.1186/s12859-015-0683-0 (2015). [PubMed: 26283514]

75. Olson WK et al. A standard reference frame for the description of nucleic acid base-pair geometry. J Mol Biol 313, 229–237, doi:10.1006/jmbi.2001.4987 (2001). [PubMed: 11601858]

76. Wieczor M, Hospital A, Bayarri G, Czub J & Orozco M Molywood: streamlining the design and rendering of molecular movies. Bioinformatics 36, 4660–4661, doi:10.1093/bioinformatics/btaa584 (2020). [PubMed: 32573714]

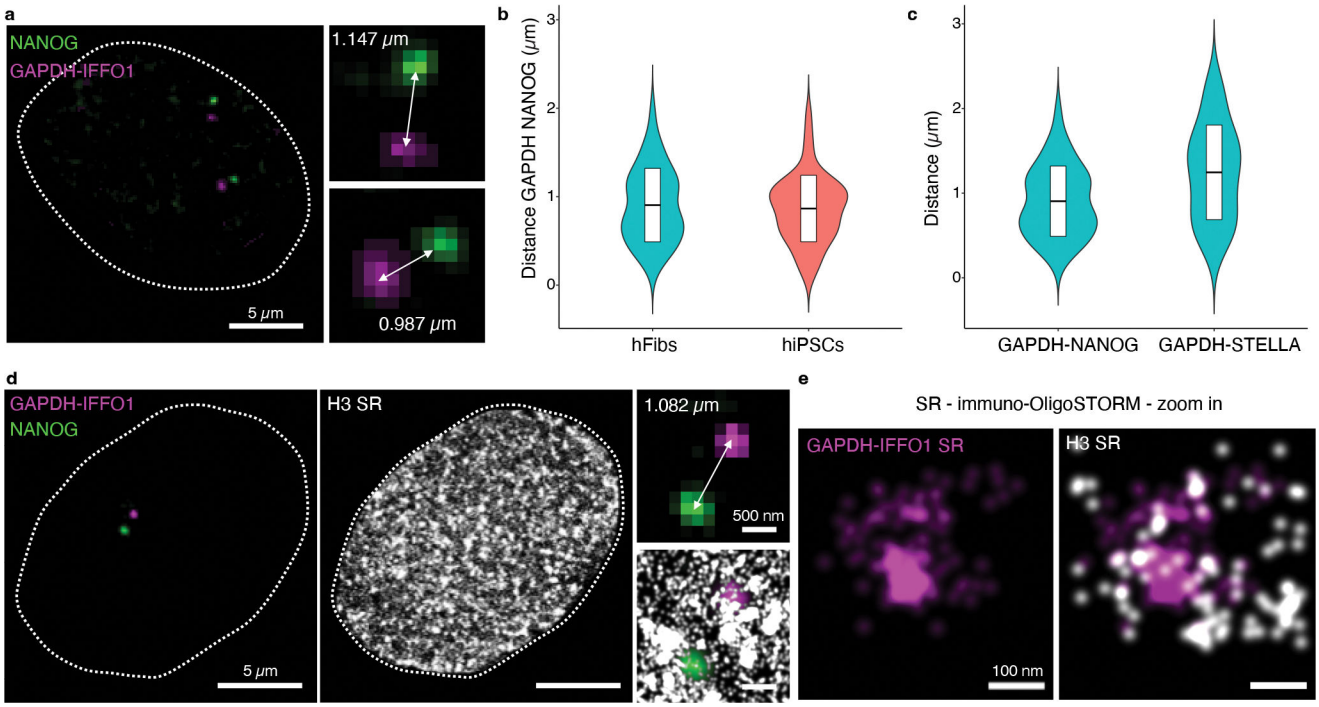


Fig. 1. iOS allows simultaneous labeling of genes and proteins in super-resolution.

a Representative example of dual color Oligopaint labeling of *GAPDH-IFFO1* (magenta) and *NANOG* (green) loci detected with AF488 and AF647-labeled Oligopaint probes, respectively. Zoomed-in views (right panels) show intergene distances. **b** *GAPDH-IFFO1* to *NANOG* distance distribution measured in dual color Oligopaint images in hFibs and hiPSCs ($n = 60$ or $n = 38$ loci distances measured in 6 hFibs or 3 hiPSCs independent experiments, respectively; mean +/− standard deviation [SD], in the central box). **c**. Distances from *GAPDH-IFFO1* to *NANOG* or to *STELLA*, as measured in hFibs ($n = 60$ and $n = 26$ loci distances, respectively; mean +/− SD, in the central box). **d** Representative example of immuno-OligoSTORM (iOS) imaging. Localization of super-resolution imaging are shown as Gaussians of fixed width for visualization purposes. *NANOG* (green) and *GAPDH-IFFO1* (magenta) loci were detected with AF488- or AF647-labeled Oligopaint probes, respectively, and H3 (white) was labeled with DNA-PAINT (560 signal). **e** Zoomed-in views. Left: STORM super-resolution imaging for the *GAPDH-IFFO1* locus (magenta) and histone H3 histone (white).

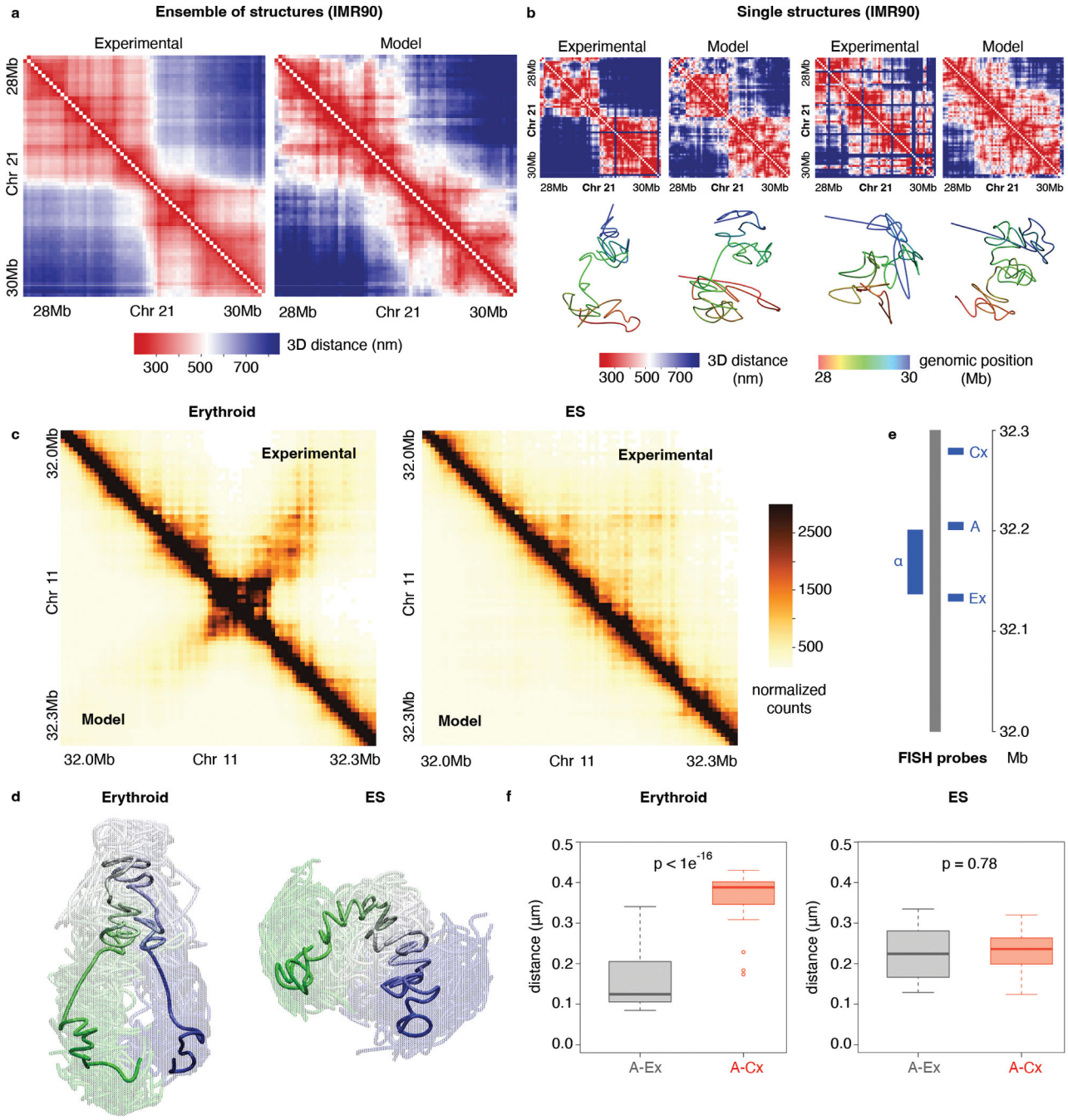


Fig. 2. Restraint-based model reproduces ensemble average data and cell–to–cell structural variability.

a Median 3D distance matrix for: (left) n = 3,496 single structures of IMR90 lung fibroblasts from multiplexed FISH-based 3D diffraction-limited imaging data[18] of a 2-Mb region (chr21:28Mb-30Mb) at 30-kb resolution, and (right) the ensemble of n = 70 structures modeled by the distance restraint–based method. Correlation coefficients between both matrices are $r_{Spearman}$ = 0.98 and r_{HiCRep} = 0.96. Color scale ranges from 200 to 850 nm. **b,** 3D distance matrices (top) and corresponding 3D conformations (bottom) of two single

structures extracted from the experimental (left) or modeled (right) structural ensembles. Color scale as in (**a**) and structure colors indicate genomic positions (red indicates 28 Mb and blue 30 Mb). **c** Capture-C matrices at 4-kb resolution from [36] (above diagonal), and contact matrices from structures obtained with the restraint-based method (below diagonal), for a 300-kb region around the alpha genes (chr11, mm9) for erythroid (left, $n = 60$) and embryonic stem (ES) cells (right, $n = 68$). Spearman correlation coefficient: 0.96 for erythroid cells, 0.95 for ES cells; stratum-adjusted correlation coefficient from HiCRep: 0.93 for erythroid cells, 0.78 for ES cells. **d** 3D tube representation of modeled representative structure for erythroid and ES cells. The full ensemble is superimposed as translucent tubes. Genomic sequence indicated using a green to blue color code. **e** Localization of FISH probes on the genomic sequence (alpha, self-interacting domain) as used in[38]. **f** Distance distribution in modeled ensembles between the centroids of FISH probes (A-Ex vs. A-Cx) (as used in [38]) for erythroid and ES cells. Boxplots highlight the median and first and third quartiles; whiskers extend to minimum and maximum values when no outliers are present. Outliers beyond 1Q-1.5 interquartile range depicted as circles. *p*-values are indicated for two-sided Wilcoxon rank-sum test between the sampled distributions. The ensemble size (number of data points in boxplots) is 60 for erythroid, and 68 for ES cells. Mean experimental interquartile ranges: erythroid cells, [0.13, 0.24] for A-Ex and [0.19, 0.40] for A-Cx; ES cells, [0.16, 0.30] for A-Ex and [0.17, 0.31] for A-Cx.

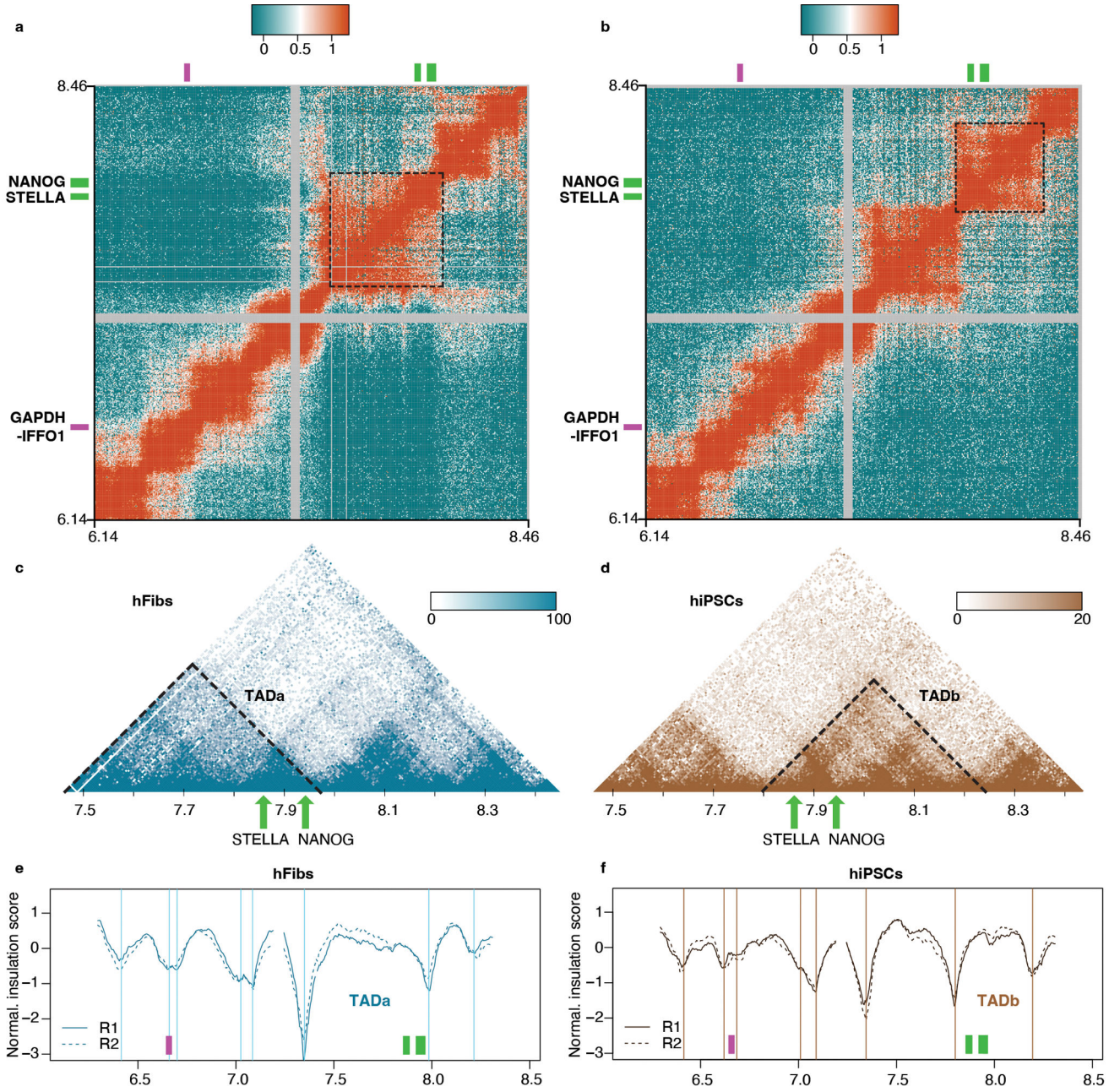


Fig. 3. Hi-C reveals changes in TAD organization between hFibs and hiPSCs.

a, b Capture Hi-C contact matrices for the region chr12:6,140,000:8,460,000 displayed at 5-kb resolution, for (**a**) human fibroblasts (hFibs, replicate 1); and (**b**) human induced pluripotent stem cells (hiPSCs, replicate 1). Log10 plots of iteratively corrected interaction counts scaled to sum 1 million are displayed. The position of the genes *GAPDH-IFFO1* (magenta), *STELLA* (green) and *NANOG* (green) are marked on the axes. **c,d** Zoomed-in view of the chr12 region containing the *STELLA* and *NANOG* genes in hFibs (**c**) and hiPSCs (**d**). **e** Normalized insulation score showing TADs and boundaries between TADs

(vertical lines) for hFibs (blue) and hiPSCs (brown). Note that two replicates (R1, solid lines, and R2, dotted lines) are shown.

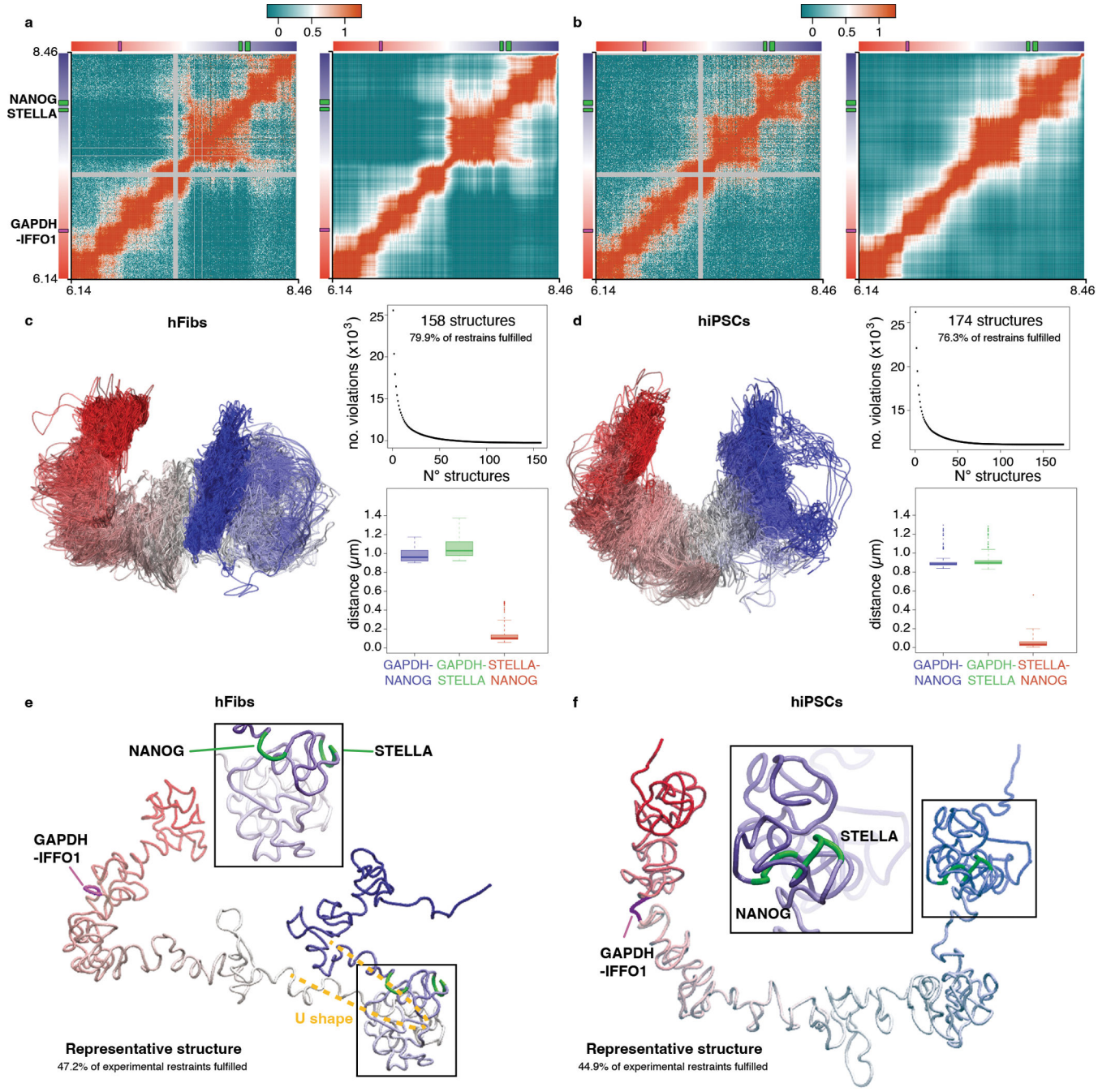


Fig. 4. Hi-C derived models of a 2.3 Mb region of human chr12 segment reveals conformational changes between hFibs and hiPSCs.

a,b Log10 contact matrices (iteratively corrected interaction counts scaled to sum 1 million) obtained from capture Hi-C (left) and from the modeled ensemble of structures (right) for the region chr12:6,140,000:8,460,000 displayed at a 5-kb resolution, for hFibs (**a**) and hiPSCs (**b**). The modeled maps show the mean value of 158 structures for hFibs, and 174 structures for hiPSCs. The correlation coefficients between experimental and theoretical matrices are $r_{Spearman}$ = 0.83, r_{HiCRep} = 0.86 for hFibs and $r_{Spearman}$ = 0.80, r_{HiCRep} = 0.70 for hiPSCs. The positions of the genes *GAPDH-IFFO1*, *STELLA* and *NANOG*

are highlighted in both axes. **c,d** Ensemble of modeled structures fulfilling the maximum number of experimental restraints coming from Hi-C, along with the absolute number of restraint violations in the ensemble and the 3D distances between genes for hFibs (**c**) and hiPSCs (**d**). The fiber is colored according to the DNA sequence with a gradient from 5′-red-white-blue-3′. The percentages of restraint fulfilment refer to the complete ensemble of structures. The boxplots highlight the median, first, and third quartile, while the whiskers extend to minimum and maximum values if no outliers are present (n = 158 and 174 structures of the ensembles from (c) and (d), respectively). Outliers (beyond 3Q+1.5 interquartile range) are depicted as dots. **e,f,** Single representative structures taken from the simulated ensemble, fulfilling on its own the largest number of Hi-C contacts, along with an amplified view of the *STELLA–NANOG* subregion, for hFibs (**e**) or hiPSCs (**f**).

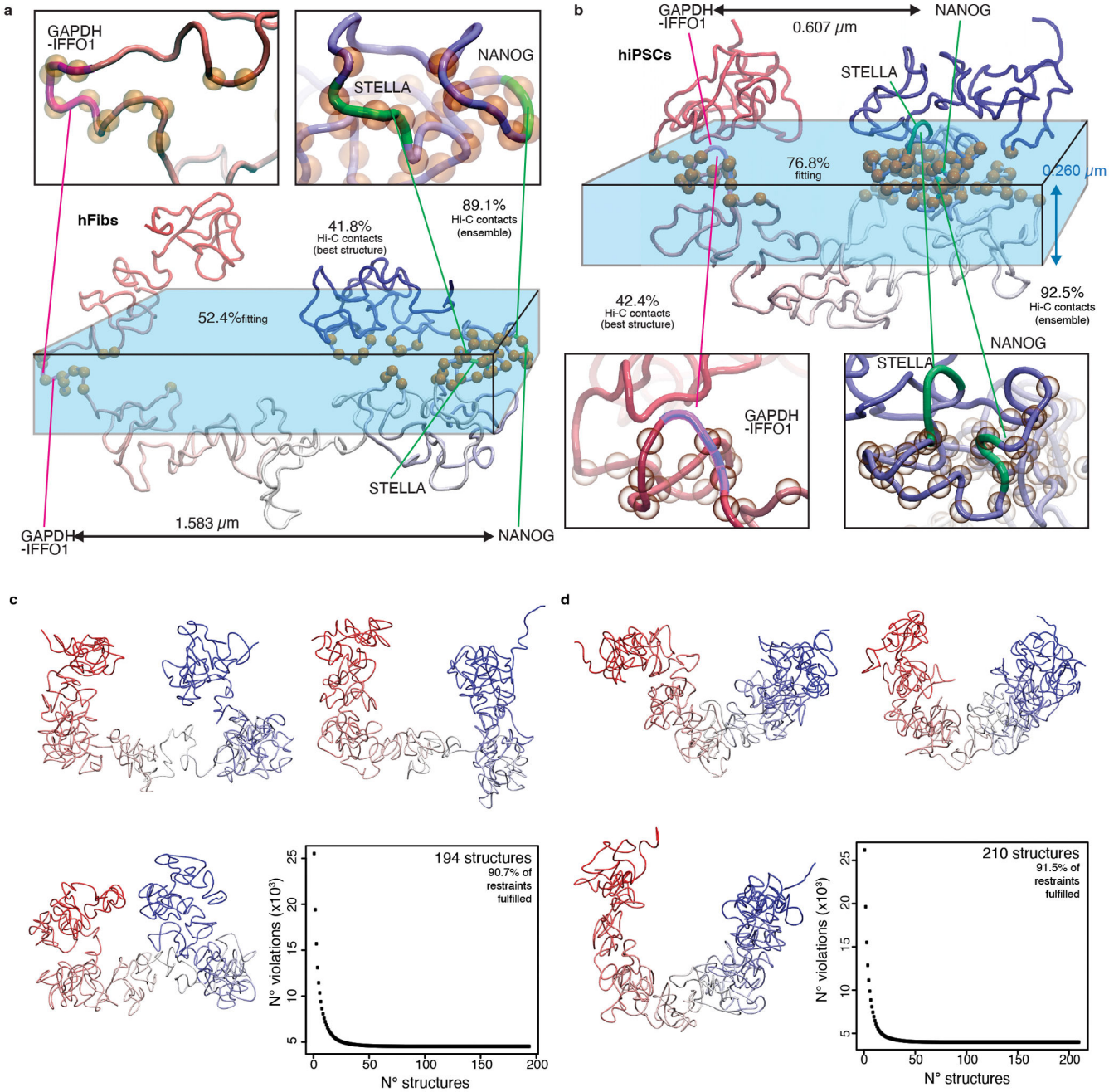


Fig. 5. Combined iOS and Hi-C modeling in hFibs and hiPSCs.

Fittings using immuno-OligoSTORM (iOS) localizations, capture Hi-C contacts and the restraint-based model of chromatin showing two specific cells with a given distance between the gene regions of *GAPDH-IFFO1* and *NANOG.* **a** Structure from the simulated ensemble, using a fixed distance for *GAPDH–NANOG* of 1.583 μm, that best fits the iOS localizations within the confocal plane (using a depth of 0.260 μm) for one hFib cell. This single structure of the chr12 segment connecting the genes of interest fits to 52.4% of the iOS localizations (5-kb fitted beads are shown as orange spheres) and fulfills 41.8% of the

Hi-C contacts simultaneously. Zoomed-in view of the genes showing the beads fitted to iOS localizations. **b,** As for (**a**) but using a fixed distance for *GAPDH–NANOG* of 0.607 μm in hiPSCs. This individual structure fits to 76.8% of iOS localizations and fulfills 42.4% of the Hi-C contacts. **c,d,** Representative structures taken from the conformationally highly-diverse ensembles generated by fixing intergene (*GAPDH–NANOG*) distances on the restraint-based model in hFibs cells (**c**) or hiPSCs (**d**). Cumulative violations of distance restraints when adding structures to the modeled ensemble are shown in the bottom right panel.

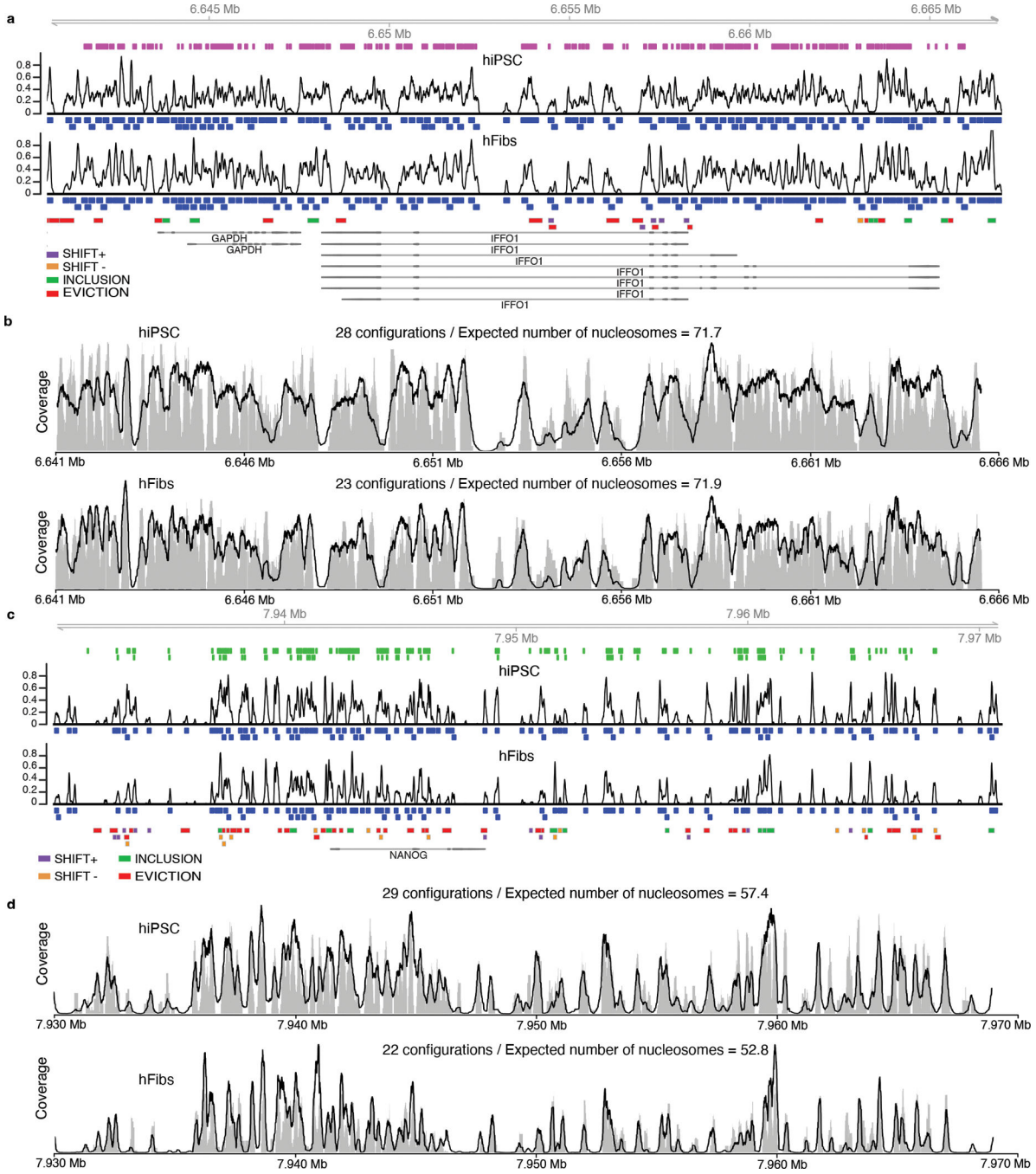


Fig. 6. Single cell–like deconvolution of capture MNase-seq–derived nucleosome in the *GAPDH-IFFO1* and *NANOG* regions labeled with OligoSTORM.

a–d Nucleosome positioning in hFibs and hiPSCs determined from capture MNase-seq for the regions used in the coarse-grained nucleosome-level model (**a, c**) Black lines represent the MNase-seq experimental signal normalized (0–1) to coverage. Upper track: nucleosome positions are represented as (a) magenta or (c) green boxes. Bottom track: boxes representing nucleosome positions and dynamics are color-coded as: inclusions (green), evictions (red), positive shifts (purple) and negative shifts (yellow). (**b,d**) Comparison of

nucleosome coverage from experimental data (black line) and prediction from nucleosome deconvolution (grey area) between hiPSCs and hFibs. Regions modeled in chr12 around (**a,b**) *GAPDH-IFFO1* (6,641,500–6,666,000) and (**c,d**) *NANOG* (7,931,000–7,970,000) regions labeled with OligoSTORM.

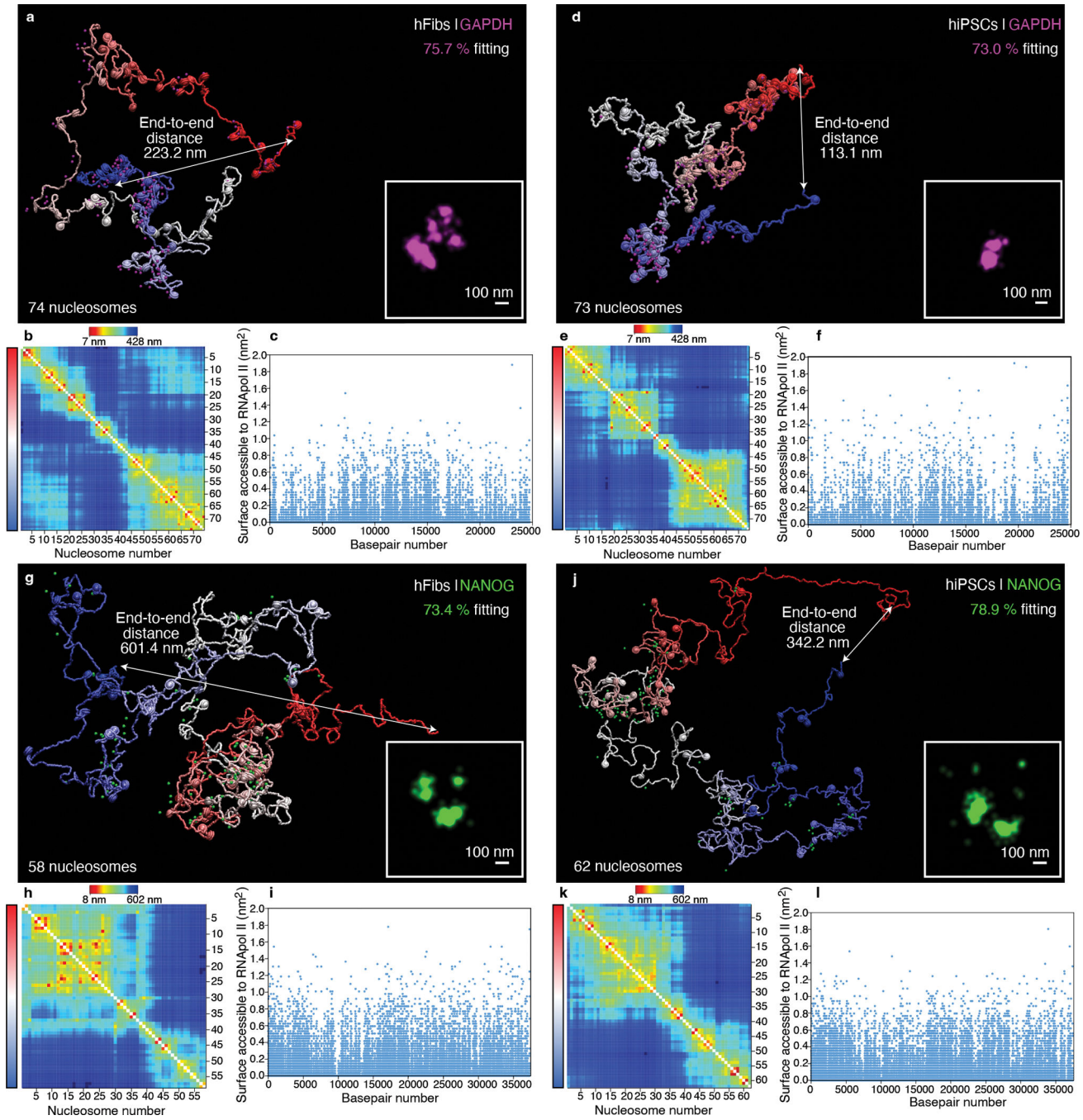


Fig. 7. Integrated MiOS models of *GAPDH-IFFO1* and *NANOG* genes in hFibs and hiPSCs.

Coarse-grained chromatin structures fitted to high-resolution OligoSTORM localizations. Structures with the highest percentage of fitting are shown (top 1) for the two genes (*GAPDH-IFFO1* in magenta and *NANOG* in green) in both cell types (hFibs, left quadrants; hiPSCs, right quadrants). **a** Best structure fitted to the OligoSTORM region containing *GAPDH-IFFO1* in hFibs. The chromatin segment is colored from red (5′-end), to white, to blue (3′-end). **b** Nucleosome-nucleosome contact map in log scale showing in the y-axis the same color range representing the 5′-3′ direction of the fiber. Note that the shortest and

longest interaction distances are shown under the gradient legend using a linear scaling. **c** Surface accessible to RNA polymerase II (RNApol II) per basepair for the *GAPDH-IFFO1* region. See Supplementary Table 2 for quantitative data. **d, e,** and **f,** same as (**a**), (**b**) and (**c**) for *GAPDH-IFFO1* region in hiPSCs; **g, h** and **i,** same as (**a**), (**b**), and (**c**) for *NANOG* in hFibs. **j, k,** and **l,** same as (**a**), (**b**), and (**c**) for *NANOG* in hiPSCs.