Biophysical Journal
Article

Biophysical Society

Nonequilibrium Biophysical Processes Influence the Large-Scale Architecture of the Cell Nucleus

Ankit Agrawal,[1,2] Nirmalendu Ganai,[3] Surajit Sengupta,[4] and Gautam I. Menon[1,2,5,*]
[1]Computational Biology Group, The Institute of Mathematical Sciences, Taramani, Chennai, India; [2]Homi Bhabha National Institute, Mumbai, India; [3]Department of Physics, Nabadwip Vidyasagar College, Nabadwip, Nadia, India; [4]TIFR Centre for Interdisciplinary Sciences, Serilingampally Mandal, Ranga Reddy District, Hyderabad, India; and [5]Department of Physics, Ashoka University, Rajiv Gandhi Education City, National Capital Region P.O.Rai, Sonepat, India

ABSTRACT Model approaches to nuclear architecture have traditionally ignored the biophysical consequences of ATP-fueled active processes acting on chromatin. However, transcription-coupled activity is a source of stochastic forces that are substantially larger than the Brownian forces present at physiological temperatures. Here, we describe an approach to large-scale nuclear architecture in metazoans that incorporates cell-type-specific active processes. The model predicts the statistics of positional distributions, shapes, and overlaps of each chromosome. Simulations of the model reproduce common organizing principles underlying large-scale nuclear architecture across human cell nuclei in interphase. These include the differential positioning of euchromatin and heterochromatin, the territorial organization of chromosomes (including both gene-density-based and size-based chromosome radial positioning schemes), the nonrandom locations of chromosome territories, and the shape statistics of individual chromosomes. We propose that the biophysical consequences of the distribution of transcriptional activity across chromosomes should be central to any chromosome positioning code.

SIGNIFICANCE Biophysical models of nuclear architecture should be capable of predicting distribution functions for the gene density and center of mass of individual chromosomes obtained in chromosome painting experiments. However, models thus far have failed in this respect, also yielding little intuition for the fundamental biophysical principles responsible for large-scale chromatin organization. We describe a first-principles approach to large-scale nuclear architecture in metazoans that incorporates energy-consuming "active" processes acting on chromatin. Our computer simulations reproduce, within a combined framework, a number of common organizing principles underlying large-scale nuclear architecture across human cell nuclei in interphase that have so far resisted explanation.

INTRODUCTION

Chromosomes are not distributed at random within the interphase nucleus, an observation that is central to our current understanding of large-scale nuclear architecture in the interphase nuclei of metazoans (1–3). Gene-rich, more open, early-replicating euchromatin regions are typically distributed more centrally than gene-poor, relatively more compact, late-replicating heterochromatin (2). Chromosomes are organized territorially, with each being segmented into relatively more (A) and less (B) active compartments that are then further subdivided into topologically associated domains (4–6). In humans, gene-rich chromosome 19, containing a large number of housekeeping genes, is distributed more centrally across several cell types than the similarly sized but gene-poor chromosome 18 (7,8). This observation generalizes to a gene-density-based radial positioning schema for all chromosomes (9).

Gene-rich regions within chromosomes tend to orient toward the nuclear center, with expressed alleles often found further from the nuclear envelope than ones that are not expressed (9,10). In some human cell types, chromosomes appear to be positioned by size, with the centers of mass of smaller chromosomes disposed more centrally than those of larger ones (11–13). In female cells, the two X chromosomes are differentially positioned, with the more compact, inactive X chromosome found somewhat closer to the nuclear envelope than the active one (14,15). Actively transcribed chromosomes tend to have rougher, more elliptical

Submitted September 9, 2019, and accepted for publication November 11, 2019.

*Correspondence: menon@imsc.res.in

Editor: Anatoly Kolomeisky.
https://doi.org/10.1016/j.bpj.2019.11.017

territories than less active ones (15–19). The probability with which two loci along individual chromosomes are found in proximity to each other in ligation assays follows a power law $P(s) \sim (1/s^{\alpha})$ with $\alpha \simeq 1$ over an ~1–8 Mb range, consistent with a fractal globule picture of chromosome structure (4,20). Currently, experiments suggest that such organization is cell type dependent and that α ($1 \leq \alpha \leq 1.5$) also varies across chromosomes over a comparable range (21,22).

Most model approaches to nuclear architecture assume a priori that chromosomes are structured polymers in thermal equilibrium (23–28). Some models ignore thermal fluctuations altogether in favor of incorporating loop structure as derived from the Hi-C data while also requiring compatibility with physical restrictions on the overlaps of chromosomes (28–30). Others account for the domain structure of individual chromosomes (31–40). As summarized above, large-scale nuclear architecture exhibits generic features that are largely common across cell types. These should severely constrain potential models (41).

However, set against this stringent requirement, virtually all prior models for such architecture are incomplete: 1) these models fail to predict gene-density-based or size-based positioning schemes; 2) no simulations reproduce the chromosome-specific distribution functions for gene density or chromosome center of mass that FISH-based experiments provide; 3) the differential positioning of the active and inactive X chromosomes cannot be obtained using any model proposed so far; and 4) the spatial separation of heterochromatin and euchromatin, seen in interphase cell nuclei across multiple cell types, has not been reproduced in model calculations in which this information is not incorporated a priori. Understanding these discrepancies remains an outstanding problem.

All molecular machinery associated with chromatin remodeling, transcription, and DNA repair is energy consuming, relying on the hydrolysis of NTP molecules (42). Recently, we pointed out that this leads to the localized, irreversible consumption of energy at the molecular scale (43). This energy is transduced through chemomechanical "active" processes into mechanical work (44–46). Such processes can be modeled via recently developed biophysical theories of "active matter" (47–50). We argued that a description in terms of inhomogeneous, stochastic forces acting on chromatin, equivalent to an effective temperature reflecting local levels of activity, provided the right biophysical setting (51,52). If we describe each chromosome as a polymer composed of consecutive monomers, each representing a suitably averaged section of chromatin, different monomers can then be expected to experience different effective temperatures correlating to local active processes (43,53,54).

Here, extending these ideas, we propose a biophysical approach to predicting both cell-type-specific and cell-type-independent features of large-scale nuclear architecture, using data from RNA-seq experiments as a proxy for activity and a Hi-C-derived description of chromosome contacts in each cell type. The model attempts to provide a unified understanding of a number of common features of large-scale nuclear architecture observed across diverse cell types.

METHODS

Interactions of model chromosomes

Our model chromosomes (diploid, XX) occupy the interior of a spherical shell of radius R_0. The interaction between neighboring monomers is of the finitely extensible nonlinear elastic form. These monomers further interact with (non-neighboring) monomers via a Gaussian interaction, the Gaussian core potential used to model polymer brushes (55). Further details of these interactions and their benchmarking is discussed in the Supporting Materials and Methods, which also includes a section describing the biophysical intuition that informs our model.

Simulation methodology

We adapt the widely used LAMMPS code, implementing Brownian dynamics (56) for polymers of different lengths, with a local monomer-dependent effective temperature, and confined to the interior of a spherical domain; for more details, see Supporting Materials and Methods. For each monomer, LAMPPS applies a Langevin thermostat via an overdamped equation of motion, with a different "effective" temperature T_i associated with each monomer, reflecting its local level of activity. In thermal equilibrium, we have $T_i = T_{eq}$ for all monomers.

Units and normalization

We work in dedimensionalized units, discussed in the Supporting Materials and Methods.

Deriving effective temperatures

In the gene density model, the gene content of each 1 Mb region is obtained from the GENCODE database (57). Monomers containing a number of genes that fall below a preset cutoff are termed as "inactive" or "passive" and are characterized by an effective temperature $T = T_{ph} = 310$ K. Monomers with a larger number of associated genes are termed as "active" and assigned an effective temperature $T_a > T_{ph}$. For the gene expression model, we infer activities from transcriptome data using FPKM values from processed RNA-seq output (58). For the combined model, we use the same temperature assignments as for the gene expression model but, in addition, take the top 5% of monomers by gene density as inferred from GENCODE, promoting them to a temperature of $T = 12$.

Models for the looping of individual chromosomes

We use Hi-C data on GM12878, NHEK, IMR90, HUVEC, and HMEC cells, obtained from data made publicly available by the authors of (59), to represent the effects of strong long-range contacts within a chromosome. We represent the effects of such long-range strong contacts in terms of permanent connections between the contacting monomers, referring to them as "loops" for this reason. We ignore loops smaller than the 1 Mb scale because these are folded into our description of a single monomer. Across these cell types, we have 236 (GM12878), 50 (NHEK), 116 (IMR90), 51 (HUVEC), and 13 (HMEC) loops that are larger than the 1 Mb scale and that our model accounts for. These loops are represented by permanent

finitely extensible nonlinear elastic bonds, with an effective interaction strength that is the same as those of the springs that link consecutively connected monomers. For the GM12878 cell type, we incorporate 27 superloops only for the inactive chromosome (Xi), taking these from data in (59).

Summary of analysis

We calculate $S_i(R) = 4\pi R^2 P_i(R)$, where $P_i(R)dR$ is proportional to the probability of finding a monomer of chromosome i at a radial vector R from the origin. For a uniform distribution, $S_i(R) = 4\pi R^2$. We compute $S_i(R)$ for every model chromosome indexed by i. We measure activity in successive radial shells by performing a configurational average over the effective temperature of every monomer in that shell. From these, we extract a quantity similar to $S(R)$ but normalize by $4\pi R^2$ so that the quantity plotted in the cutaway sphere representation simply represents the activity at radial distance R. The quantity $S(R)$ measures the DNA density associated with a specific chromosome across a radial shell at distance R from the nuclear center, averaged over a large number of nuclei. The quantity $S_{CM}(R)$ measures a similar distribution, but of the chromosome center of mass. We calculate the distribution of centers of mass of each individual chromosome similarly. To visually examine configurations, we color-coded monomers belonging to individual chromosomes.

Calculation of 3D shape of chromosome territories

For each chromosome in our simulation, we draw a three-dimensional (3D) grid across the nucleus with a grid spacing of 0.2–0.6. We represent monomers as spheres about which the density decays as a Gaussian with defined width. Separating these monomers are cylindrical regions. The density about the axis of each cylinder is assumed to also fall off as a Gaussian with specified width. The density at any given grid point associated with a single chromosome can then be computed by adding up the contributions from all spherical and cylindrical regions as defined above. Once such a density field is obtained, we can find the surfaces on which it attains a fixed value, the "implicit surface." We adjust the scales governing the decay of the density distribution associated with the monomers and the cylindrical regions separating them, as well as the associated constant specifying the implicit surface to optimize geometrical quantities associated with chromosome territories vis-à-vis experiments. Once fixed, these parameters remain the same for all chromosomes.

Geometric properties of chromosome territories

To calculate the two-dimensional (2D) properties of chromosome territories, we use the algorithm of (19). To compare our simulation data with data from 2D FISH, we project 3D chromosome territories into the xy plane. We use the ellipticity calculations of (60). In the 3D case, once we associate an implicit surface to a chromosome, that surface can further be triangulated using standard methods, such as the ISOSURFACE command in MATLAB (The MathWorks, Natick, MA). The total surface area of the chromosomes is obtained by adding the area of these triangles. To calculate the volume of the chromosome, we count the number of grid points whose grid density values are more than the given isovalue density c. The asphericity Δ and shape (or prolateness) Σ parameters of a particular chromosome are calculated from the semiaxis lengths a, b, and c of the smallest ellipsoid that encloses all the monomers (61,62). The Khachiyan algorithm is used to find the smallest ellipsoid that encloses all the data points (63).

Contact probability

The contact probability is computed using numerical calculations of the contact frequency of monomers of a given chromosome, averaged over a large number of configurational snapshots. When two monomers i and j of the same chromosome are separated in 3D space by 2.5 units in terms of our scaled unit distance, we assume that they are in contact.

RESULTS AND DISCUSSION

We model human chromosomes in diploid female (XX) cells within interphase, describing each as a polymer made up of monomers linked along a chain. These polymers are confined within a spherical shell that models the nuclear envelope. Each monomer represents a 1 Mb section of chromatin (13,64). Our model chromosomes are dynamic and explore different configurations based on the forces they experience. Such forces arise from the dense, nonequilibrium, and fluctuating environment of the cell nucleoplasm; the interactions of chromosomes; and chromosome-nuclear envelope interactions. A final input to our model is a representation of contacts between sections of chromosomes, as derived from Hi-C experiments. We count Hi-C-derived contacts averaged over 1 Mb regions, connecting the appropriate monomers if the strength of these contacts exceeds a threshold; we call such permanent connections loops. A number of simulation snapshots of both homologs of chromosomes 18 and 19 against a background of all other chromosomes represented in grayscale are shown in Fig. 1 *A*. From such snapshots, we compute a variety of statistical properties of chromosomes accessed in experiments (Fig. 1, *B–G*).

We describe three models that associate local levels of nonequilibrium transcriptional activity to an effective temperature. In the gene density model proposed in our earlier work, we chose the top 5% of monomers by gene density, assigning them an active temperature in excess of the physiological temperature T_{ph} (43). The gene density model yields fairly accurate representations of the measured distribution function of DNA density S(R) in GM12878 cells, leading to very different distributions for the chromosome pairs 18 and 19 vis-à-vis chromosomes 12 and 20, as seen experimentally (43). However, such a model is insensitive to cell-type-dependent features of nuclear architecture (41).

Accordingly, in the gene expression model, we focused on transcriptomes across a number of model systems, exploring varied ways of associating transcript levels to effective temperatures. Fig. 1 *H* shows RNA-seq-derived FPKM values summed over 1 Mb intervals, indexing transcript levels, across GM12878, HMEC, HUVEC, IMR90, and NHEK cell types. Their distribution follows a Gumbel form (Fig. S1). The values in Fig. 1 *H* are plotted in increasing order of expression on a logarithmic scale. We chose structured effective temperature assignments that reflect the overall shape of this curve. The gene expression models provide more accurate predictions for S(R) and S_{CM}(R) in comparison to experiments than the gene density model.

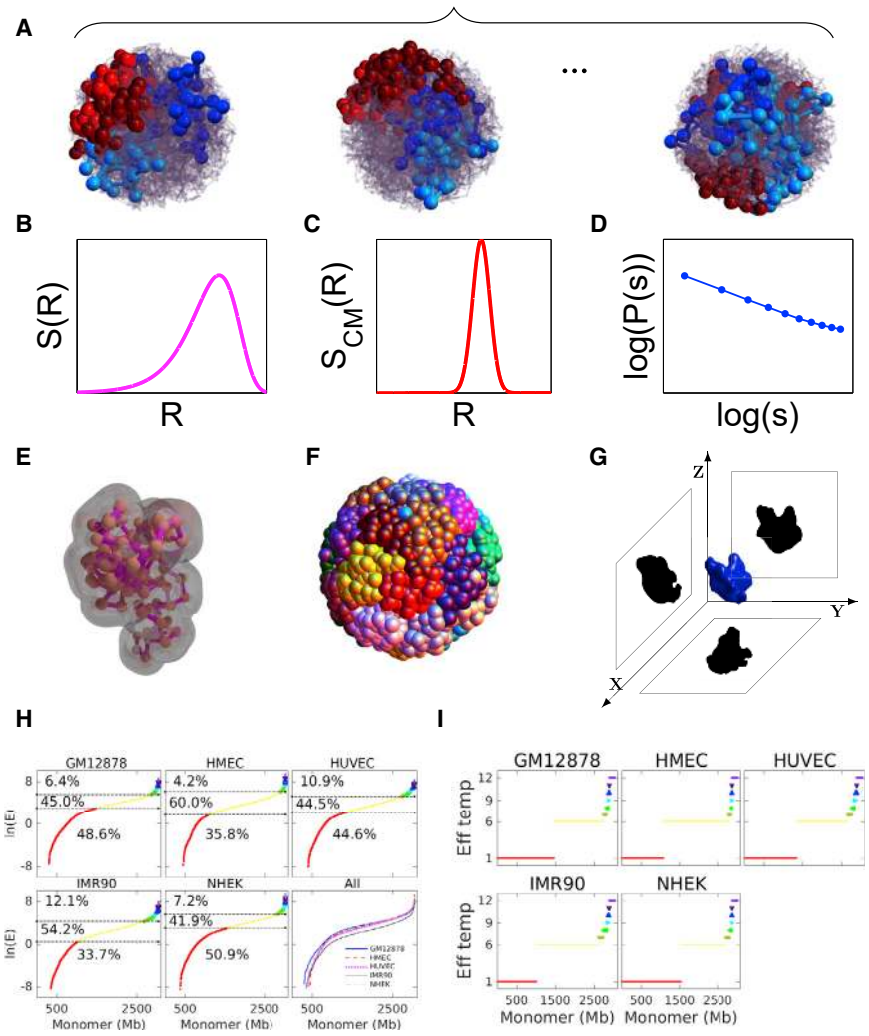

FIGURE 1 Model schematics and active temperature assignments. (*A*) Several simulated configurations of 23 pairs of chromosomes within a spherical nucleus, with pairs of chromosomes 18 and 19 highlighted in the background of other chromosomes, are shown in grayscale. Each bead represents a 1 Mb section on each chromosome. We average all calculated quantities, such as distribution functions, over a large number of such configurations in steady state. (*B*) A schematic of the DNA distribution S(R) of each chromosome, plotted against the radial coordinate R and averaged over many nuclei in our simulations, is given. (*C*) A schematic of the center-of-mass distribution of each chromosome, $S_{CM}(R)$, plotted against the radial coordinate R and computed from an average over many simulated nuclei, is given. (*D*) A schematic of the contact probability P(s) between beads of chromosomes, for two monomers separated by an internal (genomic) distance s along the polymer, is given. (*E*) The shapes of individual chromosome territories extracted from simulation configurations are shown. Such shapes are used to compute a number of geometrical properties of chromosome territories, e.g., their volume, surface area, asphericity, and other shape parameters. (*F*) A typical image of chromosome territories computed in our simulations is given, with each chromosome colored a different color, illustrating the emergence of territoriality. (*G*) A schematic illustrating a 2D projection of a 3D chromosome territory, projected along the *xy*, *yz*, and *xz* planes, is shown. The ellipticity and regularity parameters can be computed from such 2D projections and compared to 2D FISH data. (*H*) The logarithm of gene expression values for each 1 Mb monomer, plotted in order of increasing gene expression, is given. These are computed from transcriptome data. Data are shown for five cell types, as indicated in the title to each panel. The horizontal lines drawn motivate our assignment of effective temperatures as discussed in the text and correspond to our assignment of activity in proportion to gene expression. The last panel plots these data together, illustrating that the shape of the activity profile is largely similar, even though individual monomers in different cell types can be classified differently on the basis of their activity. (*I*) Assignment of effective temperature to each monomer for the combined model is hsown. The red monomers are simulated at $T = 1$, yellow at $T = 6$, yellow-green at $T = 7$, green at $T = 8$, cyan at $T = 9$, blue at $T = 10$, indigo at $T = 11$, and violet at $T = 12$ times the physiological temperature T_{ph}. To see this figure in color, go online.

Our final model attempts to account for the observation that transcript levels need not directly correlate to activity because RNA-seq methods account only for steady-state RNA production and because our description averages over the typical timescales associated with transcriptional "bursts" (65,66). We felt that a model that included features of both gene density and gene expression models should provide a more accurate representation of inhomogeneous cell-type-dependent activity (67). Accordingly, we decided to combine features of both the gene density and the gene expression model into what we call a combined model. In the combined model, we assign active temperatures based on activity as obtained from the gene expression model as well as by including the most active monomers indicated by the gene density model.

All the results presented in this work are for the combined model. Fig. 1 *I* shows temperature assignments within the combined model for five cell types. Such inhomogeneous (effective) temperature assignments, correlating both to gene density and transcription levels averaged over consecutive 1 Mb sections of each chromosome, lie at the core of our work.

Inhomogeneous activity underlies large-scale nuclear architecture

Fig. 2 *A* shows simulation-derived chromosome territories for the GM12878 cell type, with each chromosome colored a different color. Simulations recover such territorial organization robustly, illustrating how territoriality is an emergent consequence of our model. Fig. 2 *B* shows a cutaway profile

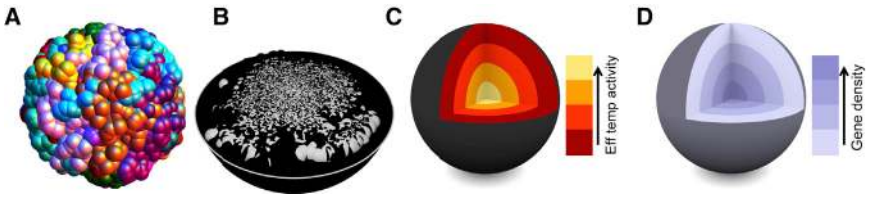

FIGURE 2 Model predictions for large-scale features of nuclear architecture. (*A*) Chromosome territories computed in our simulations are shown, with each chromosome colored a different color. Note the tendency of each chromosome to overlap relatively little, visually representing territoriality. (*B*) A cutaway sphere representation of the average spatial distribution of euchromatin (or active *white*) and heterochromatin (or inactive *black*) monomers as computed for the GM12878 cell type is given. Here, the active monomers are defined as those having an effective temperature in excess of the physiological one. Heterochromatin is found more peripherally compared to euchromatin, which is located toward the nuclear interior. (*C*) A cutaway sphere representation of average effective temperatures within the simulated nucleus, as computed for the GM12878 cell type, is shown. This illustrates the larger effective temperatures, indicating enhanced activity, obtained toward the center of the nucleus, in comparison to a lower effective temperature in the nuclear periphery. (*D*) A cutaway sphere representation of the average gene density within the simulated nucleus, computed for the GM12878 cell type, is shown. This illustrates the excess in gene density seen toward the center of the nucleus in comparison to the gene density in the nuclear periphery. This separation of gene-dense and gene-poor 1 Mb segments of chromatin correlates to the distinction in the spatial positioning of euchromatin and heterochromatin. To see this figure in color, go online.

showing the averaged spatial distribution of active (white, $T/T_{ph} > 1$) and inactive (black, $T/T_{ph} = 1$) monomers in the GM12878 cell type, extracted from snapshots of a typical configuration. Low-gene-density monomers, shown in black and representing heterochromatin, are enriched toward the boundaries of the nucleus, whereas high-gene-density euchromatin regions, shown in white, preferentially occupy the bulk. In Fig. 2 *C*, we show a cutaway profile of the time-averaged effective temperature within our simulated nucleus, an indicator of local activity in each spherical shell centered around the origin. In Fig. 2 *D*, we show the time-averaged gene density across spherical shells in a similar visualization. Gene densities, as well as activity, increase toward the nuclear center.

Our ab initio biophysical description of chromosomes and their structuring reproduces the different spatial distributions of euchromatin and heterochromatin, a feature seen across multiple cell types. A central consequence of our model is that gene expression should correlate to a larger strength of mechanical fluctuations, i.e., activity, and that the distribution of both these quantities should be attenuated toward the boundaries of the nucleus. This is an emergent property, arising from the combination of differential activity and confinement, that could not have been inferred from how the model was constructed.

The model predicts positional distributions of individual chromosomes

Chromosome-specific distribution functions S(R) are obtained experimentally using confocal slices of FISH images from an ensemble of fixed nuclei. Our computed S(R) for chromosomes 18 and 19 in the five cell types we study are shown in Fig. 3 *A*. All data are averaged over the two autosomal homologs because their positioning was found to be equivalent. For the GM12878 cell type, we compare our results with experimental results extracted from (68). S(R) for chromosomes 18 and 19 exhibits well-separated peaks, a feature that holds across cell types. For comparison, the R^2 rise of S(R) toward the nuclear envelope, expected for uniformly distributed chromosomes, is shown in magenta, specified as "Random" in the legend of the panel.

Fig. S2 shows our calculated S(R) for all chromosomes in the five cell types. Fig. S3 shows S(R) for the GM12878 cell type in which we compare the predictions of the gene expression model and the combined model. In Fig. S4, we show how S(R) for the GM12878 cell type varies when we include or exclude looping and when we include or exclude activity. The predictions of the different models differ substantially for both the gene-rich chromosomes as well as the smallest chromosomes. In general, both looping and differential activity are needed to best represent available experimental data.

Fig. 3 *B* shows S(R) for chromosomes 12 and 20. For the GM12878 cell type, we compare our results with results from (68). Note that simulation data for the different cell types all yield similar plots for S(R), with the exception of the GM12878 cell type. Here, although the simulation and experimental data peak at different locations, the overall shape of the curve is rendered accurately, including the relative shift in peak positions.

Fig. 3 *C* shows the distribution of centers of mass of specified chromosomes, $S_{CM}(R)$, for chromosomes 18 and 19. For GM12878 cells, we compare our results with experimental data from (69). The center-of-mass distributions are captured well, especially for chromosome 19. The broader distribution of $S_{CM}(R)$ for chromosome 18 is also in agreement with the left tail of the experimental data, although the experiments show a weaker and more outward-shifted peak than the simulation prediction. Differences in positioning of chromosomes across cell types are more apparent in $S_{CM}(R)$ compared to S(R).

Fig. S5 shows $S_{CM}(R)$ for all cell types across all chromosomes. We compare the predictions for $S_{CM}(R)$ in the gene expression and combined models in Fig. S6. Results for $S_{CM}(R)$ with different combinations of activity and loops are shown in Fig. S7. Overall, apart from the gene-rich chromosomes, the predictions of the gene expression and

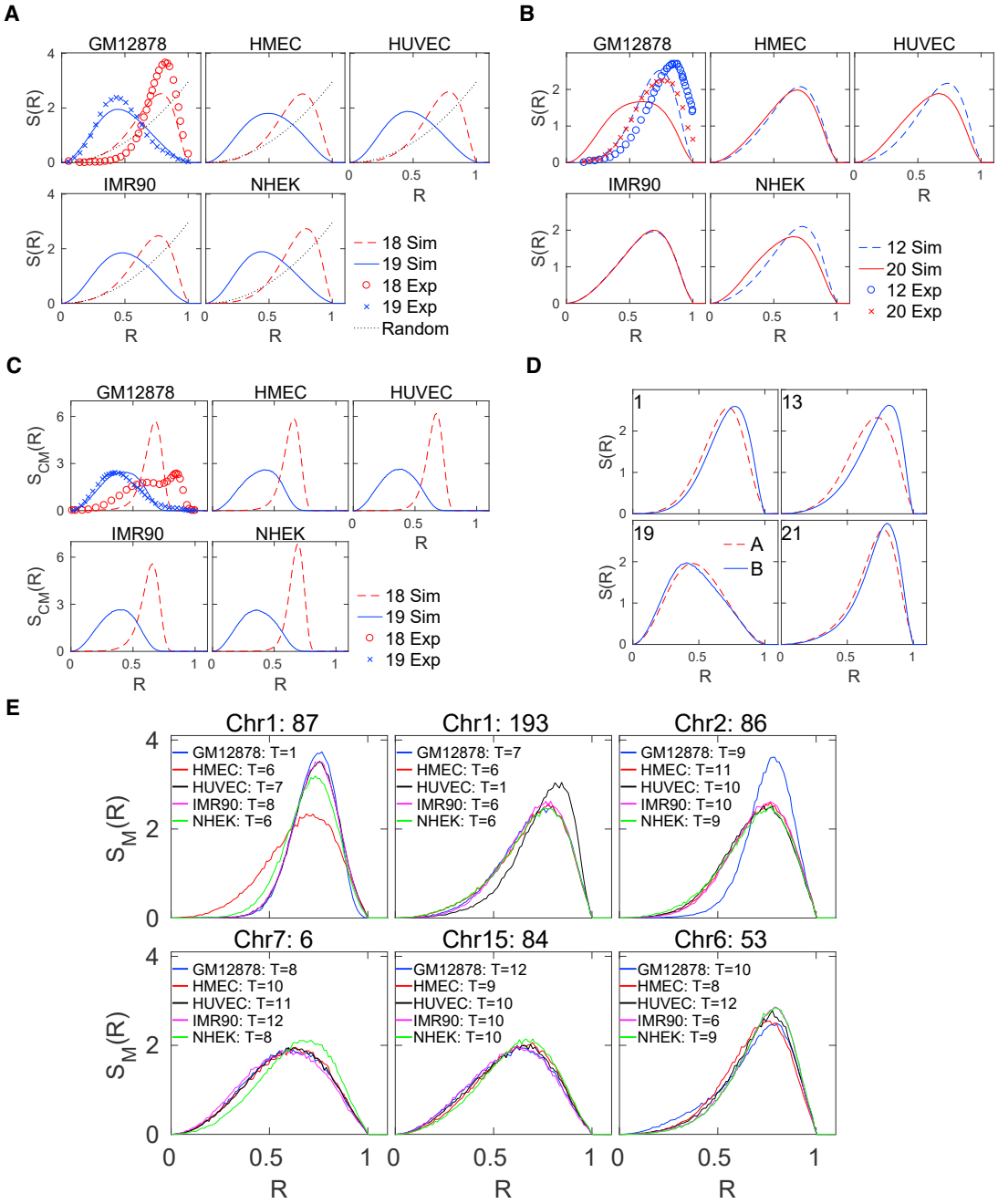

FIGURE 3 Predicted radial distribution functions S(R) compared to experimental data. (*A*) Distribution of monomer density S(R), reflecting the local density of DNA, is shown for chromosomes 18 and 19 (*red* (*dashed*) and *blue* (*smooth*) *lines*, respectively) across five cell types as indicated in the titles of each panel. Experimental data obtained from (68) for the GM12878 cell type are plotted together with the simulation predictions (*red ovals*: Chr 18, *blue crosses*: Chr 19). If chromosomes are distributed randomly across the nucleus, $S(R) \sim R^2$ is expected, as shown with magenta (*dotted*) lines. (*B*) Distribution of the density of monomers, reflecting the local density of DNA, is shown for Chr 12 and 20 (*blue* (*dashed*) and *red* (*smooth*) *lines*, respectively) across five cell types as indicated in the titles of each panel. Experimental data obtained from (68) for the GM12878 cell type are plotted (*red ovals*: Chr 18, *blue crosses*: Chr 19), together with the simulation prediction. (*C*) Distribution of chromosome centers of mass for Chr 18 and 19 (*red* (*dashed*) and *blue* (*smooth*) *lines*, respectively) is shown for five cell types, as indicated in the titles of each panel. Experimental data obtained from (69) for the GM12878 cell type are plotted (*red ovals*: Chr 18, *blue crosses*: Chr 19) together with the simulation prediction. (*D*) Density distribution S(R) of overall numbers of active (*red*, *dashed*) and inactive (*blue*, *smooth*) monomers is shown for the GM12878 cell type. These are plotted for four chromosomes: the largest, Chr1; the smallest, Chr 21; gene-poor Chr 13; and gene-rich Chr 19. The distribution of active monomers is more interior with respect to inactive monomers. Here, inactive monomers refer to those monomers assigned a temperature of $T = 1$; all other monomers are active. (*E*) Density distribution $S_M(R)$ of specific monomers as indicated, on chromosomes 1, 2, 7, 15, and 6, is plotted for five cell types studied here. These monomer-specific distributions can differ depending on cell type, suggesting that loci associated to these monomers can be positioned differently depending on their levels of activity but also on the levels of inhomogeneous activity of the chromosome they belong to. To see this figure in color, go online.

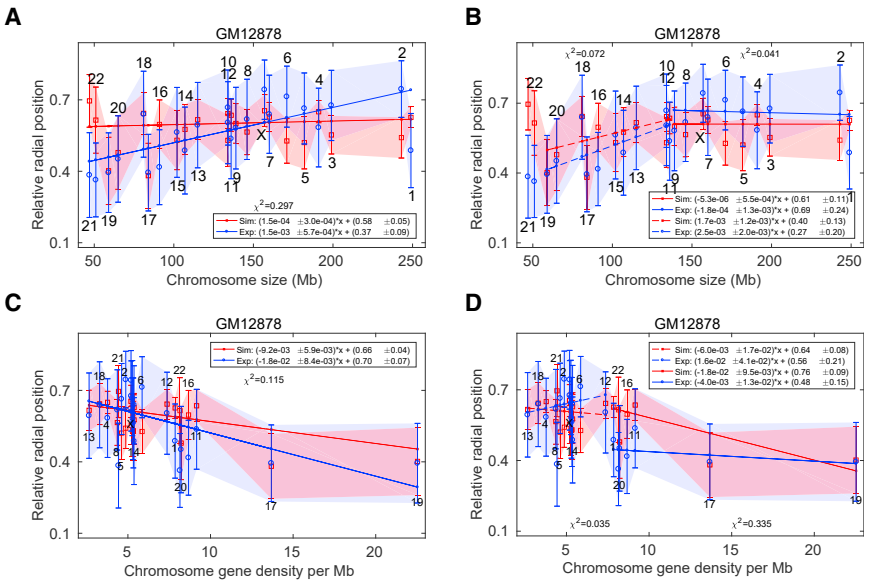

FIGURE 4 Predicted chromosome center-of-mass locations compared to experimental data. Predictions from simulations for the mean center-of-mass location for each chromosome, for the GM12878 cell type, as a function of chromosome size in the upper row (*A* and *B*), and as a function of chromosome gene density per Mb in the lower row (*C* and *D*) are shown. These predictions are compared to experimental data on the average radial position of the center of mass of each chromosome as obtained from (69). Simulation and experimental points are shown using red squares and blue circles respectively, together with error bars indicating one standard deviation from the mean with filled colors. The relative radial position 0 and 1 represent the center and periphery of the nucleus. Chromosome numbers are indicated above or below each error bar. The simulation and experimental points fitted to a straight line including all chromosomes are shown in panels (*A*) and (*C*); fits excluding the two smallest chromosomes 21 and 22 are shown in panel (*B*). The smaller and larger size chromosomes in panel (*B*) are fitted with two separate fits. The slope and intercept value with error for experimental and simulation fitted lines are provided in each panel. The χ^2 error associated to the fitted lines is provided in each panel. To see this figure in color, go online.

combined models are comparable. The largest variability across cell types is seen for chromosomes 1, 4, 7, 11, 12, 16, 21, and 22. $S_{CM}(R)$ for gene-poor chromosomes are sharply peaked, whereas gene-rich chromosomes have broader distributions across all cell types.

Fig. 3 *D* shows the partial distribution functions S(R) for inactive and active monomers in the GM12878 cell type for chromosomes 1, 13, 19, and 21. The distribution for active monomers is shifted toward the nuclear center, whereas for the inactive monomers, it is seen to be shifted toward the nuclear periphery. These results relate to the experimental observation that active alleles are positioned more toward the interior of the nucleus, an effect strong enough to be apparent in our simulations (10,70). In Fig. 3 *E*, we show monomer-specific distribution functions $S_M(R)$ for six monomers across chromosomes 1, 2, 6, 7, and 15 across all cell types; these monomers contain multiple genetic loci and typically show differential activity across the cell types shown. We note that such monomer distributions are not identical but depend on both their active temperature as well as the overall activity and loop content of the chromosomes that contain them.

These results indicate that cell-type-dependent signatures of activity can be especially prominent at the level of individual monomers and thus loci. They are overall less prominent in chromosome-specific DNA density distributions and the distributions of their centers of mass but display subtle differences nevertheless. These differences originate both in differences in activity profiles across different cell types as well as variations in their contact structure (loop content in our model), suggesting that these should be essential components of any biophysical description of large-scale nuclear architecture.

Model predictions for size- and gene-density-dependent chromosome positioning

Fig. 4 shows our computation of the mean center of mass of each chromosome within the combined model. The data shown here are for the GM12878 cell type; data for other cell types are provided in the Supporting Materials and Methods. Our results are shown for each chromosome as a function of chromosome size in the upper row (Fig. 4, *A* and *B*) and as a function of chromosome gene density in the lower row (Fig. 4, *C* and *D*). These predictions are compared to experimental data on the average radial position of the center of mass of each chromosome, as obtained from (69). Our fits of the data to straight lines are shown in blue for the experimental data and in red for the simulations. The filled colors, in blue and red, represent the standard deviation for experimental and simulation data, respectively. The fitting parameters, slope, and intercept are mentioned in each panel. The relative radial positions 0 and 1 represent the center and periphery of the nucleus. Chromosome numbers are provided above or below each error bar.

The simulations reproduce most of the experimental systematics. The positions in the experimental and simulation data coincide for some chromosomes or lie well within error bars. The positions of chromosomes 14, 16, 18, 19, 20, and X are very close to the experimental data, reproducing the unusual nonmonotonicity in their positions. The slight overall shift between the positions between experiment and simulation arises from the fact that our simulations are performed for spherical nuclei, whereas the experiments are performed on more flattened, ellipsoidal nuclei as well as averaged over an ensemble of such shapes.

The fits to the data are weakly consistent with an approximate size dependence of chromosome positions relative to the nuclear center, although the correlation between such positioning and chromosome gene density is stronger (see below). The activity associated with each individual chromosome also plays a role in determining its position. The mean center-of-mass locations for chromosomes in different cell types are similar but not identical. Chromosomes 18 and 19, although similarly sized, have very different positions relative to the nuclear center, as also seen in the data of Fig. 3, *A* and *C*. Note that chromosomes 21 and 22 in Fig. 4 *A* are positioned more toward the exterior of the nucleus in the simulations than in the experimental data.

We experimented with excluding a small set of chromosomes to see if the fits might be improved, examining whether fitting the smaller and larger chromosomes independently might improve the overall agreement. We present two sets of fits. In the first, shown in Fig. 4, *A* and *C*, we fit a straight line to the center-of-mass positions of all chromosomes. The fits shown in Fig. 4 *B* exclude the two smallest chromosomes, 21 and 22; we fit the centers of mass of the smaller and larger chromosomes, excluding these, to two independent straight lines as a function of chromosome size. We show a similar plot in Fig. 4 *D*, in which we fit the center-of-mass positions for low- and high-gene-density chromosomes separately. The slope and intercept value, with associated error bars, for experimental and simulation fitted lines are provided in each panel. For completeness, the χ^2 error for the fit lines is also provided in each panel. These fits are further quantified in Table S1. When the positions of all chromosomes are fitted to their size (Fig. 4 *A*), the fit is definitely inferior. The best *p*-value for fits are obtained when chromosomes 21 and 22 are dropped from the fit (Fig. 4 *B*) and the smaller and larger chromosomes are fitted separately. For fits of the positions of all chromosomes to the gene density, a single straight line fits well, as shown in Fig. 4 *C*, and would appear to best model the data overall.

Fig. S8 show similar plots for simulation calculations of chromosome centers of mass plotted for HMEC, HUVEC, IMR90, and NHEK cell types. When chromosome centers of mass are plotted against individual gene densities, the slope of the straight line is negative in all cell types. Thus, depending on the region that is fitted, one can have fits to both size dependence and gene density dependence of chromosome centers of mass relative to the nuclear center; our methods predict that the gene density dependence is the stronger one. The fact that the smallest chromosomes, 21 and 22, lie outside of the fit to chromosome size may reflect aspects of their activity that our method does not resolve, as well as variations in loop assignments.

Fig. S9 shows the mean center-of-mass position as computed for the GM12878 cell type across a variety of simulation conditions, including for the gene density model as well as for the combined model with various choices for the incorporation of loops and activity. Fig. S9 *A* shows results for the gene expression model. In Fig. S9 *B*, we show results for the case in which we allow differential activity but ignore looping. In Fig. S9 *C*, we show results for the case in which differential activity is absent but looping, as prescribed by the Hi-C data, is retained. All monomers then experience the same effective temperature, which we take to be the thermodynamic temperature. Finally, in Fig. S9 *D*, we show results for the case in which looping is absent as well, so that this case corresponds to the case of chromosomes without loops at thermal equilibrium.

From these, we conclude that in the absence of both activity and looping, chromosome positioning is only weakly structured. Our simulations indicate that chromosome positioning is very weakly size dependent or even independent of size in all conditions in which activity is switched off. Allowing for loops induces some changes in positioning, but these results do not match with experiment. Allowing for activity but ignoring loops leads to a differential positioning of chromosomes, but, if anything, the size dependence of chromosome positions is opposite to that seen in the data. Only models that incorporate both activity and looping are successful in both reproducing the approximate size dependence of radial positioning while also accounting for those specific cases that fall outside this general trend, such as chromosome 19.

We also checked to see which model gave the best fit to relative center-of-mass positions data. We did a relative center-of-mass position fit for all chromosomes against gene density and size dependence and computed the χ^2 error and *p*-value. These values are mentioned in Table S1 for different models. It is clear that the combined model gives the best fit to such data as are available for the relative center-of-mass positions.

Thus, our model predicts the center-of-mass positions of almost all chromosomes, with the exception of the smallest chromosomes 21 and 22, with reasonable accuracy. Our simulation predictions are well within error bars on the measurements in almost all cases. Finally, the fact that a number of broad features of the experiments are reproduced in the model suggests that the large-scale structure and positioning of individual chromosomes are principally determined by inhomogeneous activity across chromosomes, the presence of loops, and confinement, as we have suggested.

Shapes and statistical features of individual chromosome territories compare well to experiments

Fig. 5, *A* and *B* show territories for the two autosomes corresponding to chromosomes 12, 20, 18, and 19. In Fig. 5 *C*, we show comparisons between 2D FISH data for chromosome regularity and ellipticity on WI38 cells, for which data are available (19), to predictions from our simulations for the GM12878 and IMR90 cell types. Both IMR90 and

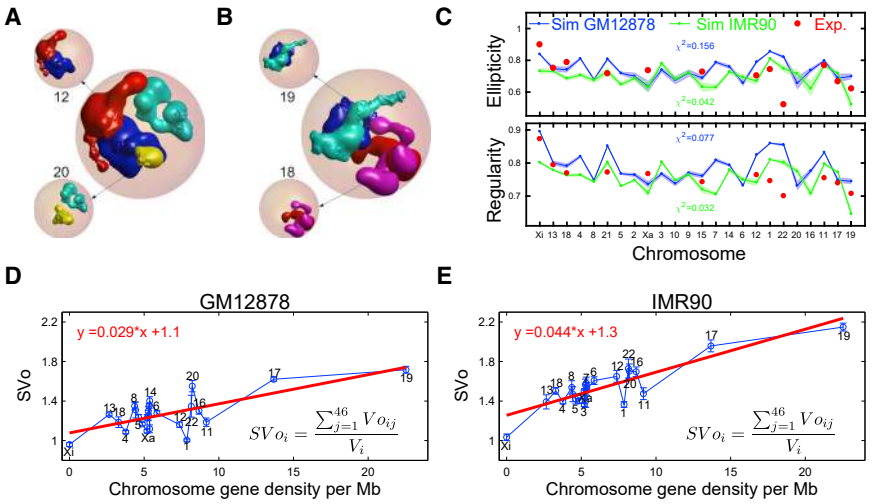

FIGURE 5 Structural properties of individual simulated chromosomes in our model. (*A*) A snapshot of simulated configurations of both homologs of chromosomes 12 and 20 is given. Each chromosome is colored differently so that they can be separately visualized. (*B*) A snapshot of simulated configurations of both homologs of chromosomes 18 and 19 is shown. Each chromosome is colored differently so that they can be separately visualized. (*C*) Ellipticity and regularity for each chromosome as predicted by the model and obtained from simulations representing the GM12878 (*blue*) and IMR90 (*green*) cell types are shown. These are compared to experimental data (*red ovals*) from 2D FISH experiments (19) for a cell type closely related to the IMR90 cell type. Ellipticity values of 1 represent a perfect elliptical chromosome, and regularity values of 1 refer to a perfectly regular chromosome, without roughness. The *x* axis is plotted in order of increasing gene density. The χ^2-value and its *p*-value are mentioned for the GM12878 and IMR90 cell types, respectively, in blue (*dark*) and green (*light*) colors. (*D* and *E*) Summed volume overlap of chromosomes in GM12878 and IMR90 cell types is shown with the *x* axis, plotted in order of increasing gene density per chromosome. There is a weak increase with gene density in both cell types, shown as the solid line, representing the best linear fit to the data. The IMR90 cell shares more volume overlaps with other chromosomes compared to the GM12878 cell type. The (self-) volume overlap for the same chromosome is taken to be 0. To see this figure in color, go online.

WI38 are lung fibroblast cell lines. Chromosomes are indexed along the *x* axis in order of their gene density.

The simulations and experimental data track each other, with the simulations finding the same dip and subsequent rise of both ellipticity and regularity around chromosome 22. Both ellipticity and regularity peak for chromosome 11, a feature of both the simulations and the experiments. The ellipticity and regularity also appear to decrease weakly with increasing gene density, although individual chromosomes may deviate from this general trend (Fig. 5 *C*). Our computation of χ^2-values indicates a closer connection between simulation and experimental data for the IMR90 cell type ($\chi^2 = 0.042$) than for the GM12878 cell type ($\chi^2 = 0.156$) as regards ellipticity. This correlation is also strong for the regularity data, in which $\chi^2 = 0.032$ (IMR90) against a $\chi^2 = 0.077$ (GM12878). Because WI38 and IMR90 are both fibroblast cell lines, this gives us some confidence that our methods might be able to predict cell-type-based variations of CT shape.

Fig. 5, *D* and *E* show the summed volume overlap, sometimes referred to as the intermingling and used to understand chromosome-chromosome interactions in *trans*, of different chromosomes in our model. The ordering of chromosomes according to their gene density per chromosome as shown on the *x* axis is the same as the ordering used for the 2D projected data in Fig. 5 *C*. The largest overlap is for the most gene-rich chromosome. There are perceptible differences in the overlaps of chromosomes in the GM12878 and the IMR90 cell types.

In summary, the simulations reproduce broad features of individual chromosome territories. More active chromosomes deviate more from a spherical shape and have rougher territories (17). The summed volume overlap increases approximately linearly with chromosome gene density, with the Xi being an exception to this trend. Activity and looping have countervailing trends because activity expands chromosome territories, whereas looping contracts them.

Simulations reproduce the differential positioning of the active and inactive X chromosomes

Experiments investigating the positioning of the active and inactive homologs of the X chromosomes within interphase have consistently found that they are differentially positioned. The inactive X chromosome (the heterochromatic Barr body) is located most often toward the periphery of the nucleus (15). This contrasts to the more central disposition of the active X chromosome Xa, which is larger and more extensively transcribed than the more compact Xi. We thus specifically investigated the positioning and other structure properties of the Xa and Xi chromosomes because we expected that they would provide an example of when our methods, which emphasize the role and importance of activity, would yield predictions that other models could not.

Fig. 6 *A* shows a simulation snapshot of active and inactive X chromosome territories. Fig. 6 *B* shows our predictions for how these chromosomes are differentially positioned across all the cell types we study through S(R). For the GM12878 cell type, we calculate S(R) for the Xi in two ways from simulations: in the first, shown with a red dashed line, we ignore the presence of "superloops." These are recently studied large-scale loops that provide additional compaction in inactive X chromosome. In the

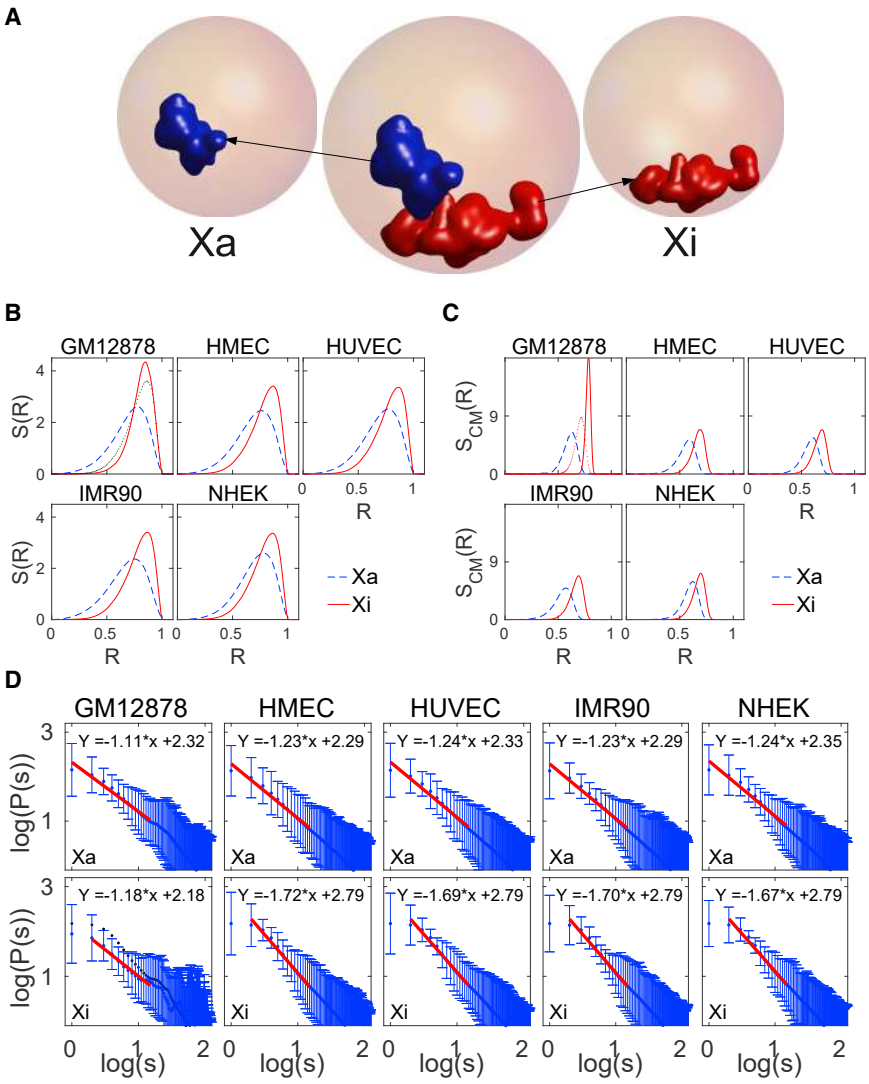

FIGURE 6 Predicted distributions and structural features of the inactive and active X chromosomes. (*A*) A snapshot of a typical configuration of the active Xa and inactive Xi chromosomes obtained from simulations is given. (*B*) Monomer density distribution S(R) vs. R for the Xi and Xa chromosomes is shown, as obtained from simulations across five cell types, named in the header to each panel. The inactive X chromosome, Xi, is shown in red (*solid* or *dotted line*), and the active X chromosome, Xa, is shown with a blue dashed line. Loops on the Xi in the GM12878 cell type can include (*red solid line*) or exclude (*red dotted line*) "superloops" as seen in recent experiments (59,76). (*C*) Distribution of the location of the center of mass of the Xi and Xa chromosomes is shown as obtained from simulations across five cell types, named in the header to each panel. The inactive X chromosome, Xi, is shown in red, and the active X chromosome, Xa, is shown in blue (*dashed line*). Superloops on the Xi in the GM12878 cell type can include (*red solid line*) or exclude (*red dotted line*). (*D*) Contact probability P(s) vs. s is shown for the active (*top row*) and inactive (*bottom row*) X chromosomes, computed for five cell types within our simulations. The Xa chromosome exhibits a reasonable power-law decay of P(s) with an exponent α between 1.1 and 1.25. The Xi chromosome shows a reduced region of power-law scaling, with an exponent across this reduced range that is between 1.5 and 1.7. Red lines show the power-law fit in both cases, with the fit parameters indicated within each panel. Error bars indicate one standard deviation about the mean. In the absence of superloops on the Xi (GM12878), $\alpha \simeq 1.5$ (*fitted line* not shown), whereas the fit in the presence of superloops reduces α to $\alpha \simeq 1.18$. Note the base of the log is 10. To see this figure in color, go online.

second, shown with a red solid line, our simulations account for such superloops. The inactive X chromosome (Xi) has an S(R) that is sharply peaked close to the nuclear periphery.

Accounting for superloops leads to a narrower S(R) distribution. Although the active X chromosome has a peak at a comparable location, its distribution has a long tail toward the nuclear center. Fig. 6 *C* shows the calculated distribution of the center of mass $S_{CM}(R)$ for these multiple cell types, verifying this essential distinction. Here again, for the GM12878 cell type, the red dashed line is the case without superloops, whereas the red solid line is for the simulations that include them. This distinction between the distribution of Xa (*blue dashed line*) and Xi (*red solid line*) in Fig. 6, *B* and *C* suggests that differential positioning should be more readily seen in $S_{CM}(R)$ than in S(R).

We can compute the contact probabilities P(*s*) by applying a cutoff to the monomer-monomer distance distributions obtained in our simulation, averaging across a large number of simulation configurations. Fig. 6 *D* shows our computation of the contact probability P(*s*) for both Xa and Xi, across the five cell types we study. The cell type is listed at the top of each panel. The active X chromosome shows more prominent power-law scaling of the contact probability than the inactive X chromosome, for which any fit to a power law can only be obtained over a far shorter genomic scale. Exponents for the power-law scaling of P(s) range from 1.11 to 1.24, with the smallest values obtained for the GM12878 cell type. For the Xi chromosome, accounting for superloops leads to a comparable scaling. However, in other cell types, such superloop information is unavailable. Accounting for loops as obtained through conventional Hi-C leads to a power-law exponent varying from 1.52 to 1.72 in those cell types. The variation in the scaling of P(*s*) between Xa and Xi should be accessible experimentally, but the presence of superloops in other cell types as well might lead to a smaller divergence between these cases. Finally, the volume/surface area ratio between Xa and Xi chromosomes from our calculation is consistent with results

from (16). The ratio of the roundness factor between Xi and Xa is consistent with observations in (16,19).

The fact that our model consistently predicts that Xa should occupy a more interior location compared to Xi, reproducing results from a large number of experiments over the past several decades, is a point we stress.

Large-scale structural features of individual chromosomes are obtained

Fig. 7 *A* exhibits our results for the contact probability P(s) for chromosome 1, across the five different cell types we study here. The data for small s show a power law P(s) $\sim 1/s^{\alpha}$ behavior over approximately a decade. The exponent is smallest for the GM12878 cell type, for which our fits yield $\alpha \approx 1.06$. This value is very close to that obtained experimentally across the same region of genomic separation (4,21). Values of α for all other cell types are consistently larger, with the exception of the IMR90 cell type.

A behavior P(s) $\sim 1/s^{\alpha}$ with $\alpha \simeq 1$ is predicted by the fractal globule model. Our simulations display such scaling even though our model lacks virtually all the requisite ingredients for the fractal globule model to be applicable. Our model requires that activity is differentially distributed along the chromosome, that we account for looping as drawn from the Hi-C data, and that we account for crowding by other chromosomes, all features that previous work, including the fractal globule model, elides. Figs. S10 and S11 show plots of P(s) for all chromosomes computed for the GM12878 and IMR90 cell types. This P(s), for each chromosome, is best described in terms of a range of exponents between $0.97 \leq \alpha \leq 1.40$ across the 1–10 Mb range. The fact that best-fit values for α span such a scale is now increasingly recognized in analyses of experimental Hi-C data.

Our model specification can be relaxed in several ways so that we can examine and quantify independent contributions to this behavior. For the specific case of chromosome 1, we have also investigated the predictions of the gene expression model, as shown in Fig. S12, as well as the effects of varying both activity and looping in the combined model. Both the gene expression and combined models exhibit values of α that lie close to the experimental data, for which $\alpha \simeq 1$. In Fig. S12, we also show results for the combined model with varying combinations of activity and looping. In the absence of both activity and loops, the exponent is close to the $\alpha = 1.5$ expected for simple polymers. Adding loops or activity reduces this exponent. However, only the combined model, which includes both activity and looping, obtains α-values closest to those in experiments.

In Fig. 7 *B*, we show the spread of the asphericity parameter Δ and the shape parameter Σ across chromosomes in the GM12878 cell type. (We exclude Xi because its behavior appears to depend sensitively on whether superloops are included.) The simulations yield a linear relationship between Δ and Σ. Larger chromosomes have smaller values of Δ and Σ. Thus, an outcome of our model is that larger chromosomes are more spherical. The regularity and ellipticity indices calculated for the 2-d projections are in reasonable agreement with experimental trends (Fig. 5 *C*). We find that the data appear to fall into two classes: one a more compact set corresponding to all chromosomes, with the exception of 1 and 21 contained within an elliptical domain, as shown in Fig. 7 *B*. Values of Δ and Σ for these special three chromosomes appear to be somewhat displaced from the locations for the other chromosomes, falling approximately onto the periphery of a larger ellipse. We show similar plots for other cell types in Fig. S13. In the absence of activity, both whether loops are present or absent, the Δ- and Σ-values for these chromosomes falls within the inner elliptical region across the five cell types we consider (Fig. S14).

Fig. 7 *C* shows a heat map of monomer distances of chromosomes, indexed in increasing order of gene density for the GM12878 cell type. A similar plot is shown for the IMR90 cell type in Fig. 7 *D*. One feature of the data is that the more active chromosomes show smaller values of interchromosomal distance, likely reflecting the fact that more active regions are enriched toward the nuclear center. In Fig. 7, *E* and *F*, we show the enlarged distance maps for chromosome 1. Applying a cutoff to such data, we can derive the likelihood of contacts arising from intrachromosomal interactions, yielding P(s). Solid lines outside the figure body indicate those permanent attachments between different monomers that the Hi-C data provide. Note that regions connected by such loops exhibit a larger overlap. Fig. 7, *G* and *H* show the contact maps inferred after applying a cutoff to the corresponding distance map. The borders of the axes show, in black and green, the active temperatures associated to specific monomers belonging to those chromosomes. The black color refers to the most active monomers, with an effective temperature of 12 in units of the physiological temperatures, whereas the green color shows monomers with an effective temperature in the range 6–11. Monomers with a lower effective temperature are not shown. Regions with the same high effective temperature appear to contact each other more, but these are further modulated by the presence of internal loops. Note the presence of a dark banded region toward the center of chromosome 1, associated with a large inactive region on this chromosome. This is a prominent feature of the experimental data, also seen in other cell types (59).

To summarize, our model yields structural information for chromosome structures and shapes that are broadly in agreement with available data. Our simulated distance maps lack the fine detail of distance maps computed in Hi-C experiments, which provide data for contacts at the smaller scales of 10–100 kB, but nevertheless are relevant to experiments that probe large-scale structuring. Our computed P(s) contains about a decade or so of power-law

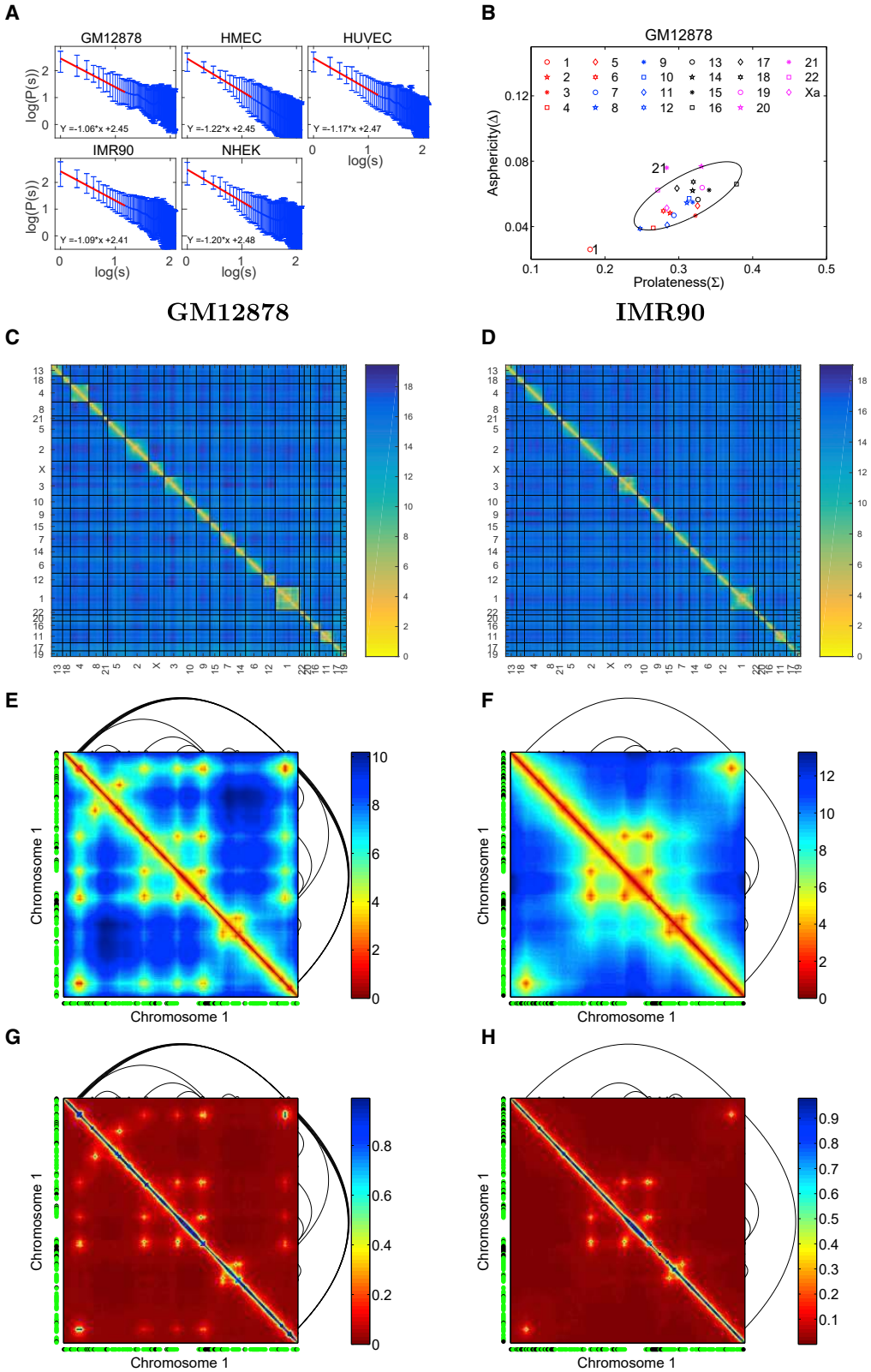

FIGURE 7 Predicted chromosome contact probabilities, shape parameters, distance maps, and contact maps. (*A*) Contact probability P(s) as a function of genomic distance for Chr 1 is shown computed across a range of 1–15 MB and plotted for five different cell types. Our data are plotted with blue dots displayed with error bars. Error bars indicate one standard deviation about the mean. Depending on the region that is fitted, a power-law scaling is obtained with an exponent between roughly 1.17 and 1.22; these fits are shown with red colors. Note the base of the log is 10. (*B*) Calculated average values of the

(legend continued on next page)

decays, with exponents that are comparable to those seen in experiments. Our model-based predictions for trends in the asphericity and prolateness of chromosomes with chromosome size and gene density are testable.

CONCLUSIONS

Model descriptions of chromosomes must bridge multiple scales, ranging from microscopic length scales of a few angstroms to scales of microns, on the order of the nuclear size. For now, brute-force atomistic simulations of the 23 pairs of chromosomes in human nuclei contained within the densely crowded, fluid, and confined environment of the nucleoplasm are impossible. They are likely to remain so, at least for the foreseeable future. Understanding how microscopic descriptions connect to macroscopic ones thus requires intuition for the processes that act to couple these scales so that model building, which is as much about what to leave out as it is about what to leave in, can proceed.

The model described in this work stresses a specific biophysical effect, ignored in previous work, of relevance to the modeling of chromosomes in living cells. We began by emphasizing the relevance of nonequilibrium effects arising from local transcriptional activity for descriptions of nuclear architecture (46,71). We proposed that the intensity of active processes should increase with increased transcription levels. We mapped a reasonable measure of local transcriptional activity, inferred from combining population-level measures of local RNA output with estimates of the local gene expression, into an effective temperature seen by each monomeric unit in our polymer model of chromosomes.

We then performed simulations of these confined polymers with properties chosen to reflect generic biophysical aspects of chromosomes. The monomers in our simulation represented 1 Mb sections of chromosomes, although we could have defined our model at the smaller scales of 0.1 or even 0.01 Mb. However, the averaging inherent in summing transcriptional output over a 1 Mb scale renders the model relatively less sensitive to errors and noise in this input. Further, the 1 Mb scale is believed to be an appropriate building block for chromosome territories. A more detailed and explicit model for nonequilibrium activity and its consequences for an active temperature description would be useful, but the form such a model ought to take is presently unclear and best left to more extensive investigations. Irrespective of potential quantitative improvements on the model front, the broad trends we describe here should be largely robust.

We made a number of other modeling choices. First, we ignored the important role of lamin proteins in anchoring specific lamin-associated domains to the nuclear lamina, as well as the interactions of specific gene loci with nuclear pore complexes (72). Although these are important omissions, they can at least be qualitatively justified by the biophysical intuition that the activity-based physical segregation of chromosomes is a bulk or "volume" effect that should dominate, at the simplest level of description, over "surface" effects arising from interactions with the nuclear envelope. Thus, modeling the effects of interactions of lamin-associated domains with the nuclear lamina by introducing weak monomer-specific interactions with the inner surface of the confining sphere in our simulations might be expected to modify the results we present here for specific chromosomes, but hopefully in a controllable manner. Second, we ignored nucleoli, formed around nucleolus organizer regions containing multiple copies of rRNA genes, with such regions located on the short arms of the acrocentric chromosomes 13, 14, 15, 21, and 22 (73). Third, we simulated the nucleus as a spherical shell containing our model chromosomes, although nuclear shapes exhibit considerable variability and much of the experimental data comes from experiments on the relatively flattened nuclei of fibroblasts (12). Our model could be generalized to account for the effects of variable nuclear shapes. Fourth, we ignored potential interactions in *trans* between chromosomes. Such interactions could potentially arise from the looping out of loci on different chromosomes to interact at transcription factories (74). We could account for this by making designated monomers on different chromosomes "sticky" with respect to each other, thus coupling regions of different chromosomes that are known to physically localize together when co-transcribed. Fifth, in using RNA-seq data as a proxy for activity, we ignored noncoding transcription because RNA-seq largely provides steady-state gene expression. Inferring activity from methods such as GRO-seq, which also extracts nascent and rapidly degraded transcripts, may help to provide a more accurate view of transcription-coupled activity. Last, the role of nuclear actin

prolateness parameter (Σ) and the asphericity parameter (Δ) for the GM12878 cell type are shown. Larger (smaller) chromosomes have smaller (larger) values of Σ and Δ, implying that larger chromosomes are more close to spherical, whereas smaller chromosomes prefer a more prolate, rod-like shape. The data suggest that values of Σ and Δ for Chr 1 and 21 take more extremal values than for the other chromosomes, as shown by the ellipse drawn together with the data. (*C*) A heat map of mean distances between monomers, the distance map, is given, in which chromosomes are ordered by their gene density, shown for the GM12878 cell type. (*D*) A heat map of mean distances between monomers, the distance map, is given, in which chromosomes are ordered by their gene density, shown for the IMR90 cell type. (*E* and *F*) A heat map of the distance matrix for chromosome 1, expanded out from (*C*) and (*D*), is shown. The locations of the permanent loops inferred from the Hi-C data are plotted in black. Individual monomers at $T = 6$ and $7 \leq T \leq 12$ are shown in green and black, adjacent to the *x* and *y* axes, respectively. (*G* and *H*) A contact map inferred from the distance matrix for chromosome 1, (*C*) and (*D*), is shown. The locations of the permanent loops inferred from the Hi-C data are plotted in black. Individual monomers at $T = 6$ and $7 \leq T \leq 12$ are shown in green and black, adjacent to the *x* and *y* axes, respectively. To see this figure in color, go online.

and associated motors remains unclear, although they could potentially contribute additional source of nonequilibrium noise (75). Indeed, all the possible improvements on our model that we list above could be incorporated, but only at the expense of more model detail and a number of further assumptions. These modifications of our model would have obscured the core argument of this work, that the primary driver of many features of nuclear architecture is nonequilibrium activity that is inhomogeneous across chromosomes, so we choose to leave these questions to future work.

If such biophysical approaches have any truth to them, they indicate that a small set of initial model assumptions, argued for on general grounds, must yield consistent explanations and descriptions for all data, not just those the model abstracts in its construction. The advantage of simple models is that they enable us to concentrate on underlying principles that are often obscured by the complexity of real data, including intrinsic heterogeneities across cell populations, varied experimental and analysis procedures, and the lack of sufficient statistics in some cases. Prior models for nuclear architecture in mammalian cells fail to reproduce many general attributes of nuclear architecture known from experiments.

These properties—certainly their important trends—are emergent in our calculations because they were not directly encoded in our model specification. This suggests that our methodologies provide hitherto unavailable biophysical insights into the determinants of large-scale nuclear architecture in metazoans.

SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2019.11.017.

AUTHOR CONTRIBUTIONS

Conceptualization, G.I.M.; methodology, G.I.M., S.S., and A.A.; formal analysis and validation, A.A., N.G., and G.I.M.; investigation, A.A. and N.G.; writing—original Draft, G.I.M. and A.A.; writing—review and editing, G.I.M., A.A., S.S., and N.G.; visualization, A.A. and G.I.M.; software, A.A. and N.G.; supervision, G.I.M.; project administration, G.I.M.

ACKNOWLEDGMENTS

We thank Ana Pombo, G. V. Shivashankar, Jean-Francois Joanny, Jacques Prost, Joe Howard, Sriram Ramaswamy, Mehran Kardar, B. J. Rao, and Frank Jülicher for useful discussions. We are especially grateful to Kundan Sengupta, Sandhya Koushika, Eldon Emberly, and P. Ranjith for detailed input and discussions regarding our manuscript.

The work of G.I.M. was supported in its early stages through a DAE-SRC Outstanding Researcher Fellowship. We acknowledge the use of the Annapurna high-performance computing cluster at the Institue of Mathematical Sciences.

REFERENCES

1. Meaburn, K. J., and T. Misteli. 2007. Cell biology: chromosome territories. *Nature.* 445:379–781.
2. Cremer, T., and M. Cremer. 2010. Chromosome territories. *Cold Spring Harb. Perspect. Biol.* 2:a003889.
3. Bickmore, W. A., and B. van Steensel. 2013. Genome architecture: domain organization of interphase chromosomes. *Cell.* 152:1270–1284.
4. Lieberman-Aiden, E., N. L. van Berkum, …, J. Dekker. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 326:289–293.
5. Dixon, J. R., S. Selvaraj, …, B. Ren. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature.* 485:376–380.
6. Fraser, J., I. Williamson, …, J. Dostie. 2015. An overview of genome organization and how we got there: from FISH to Hi-C. *Microbiol. Mol. Biol. Rev.* 79:347–372.
7. Croft, J. A., J. M. Bridger, …, W. A. Bickmore. 1999. Differences in the localization and morphology of chromosomes in the human nucleus. *J. Cell Biol.* 145:1119–1131.
8. Boyle, S., S. Gilchrist, …, W. A. Bickmore. 2001. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum. Mol. Genet.* 10:211–219.
9. Takizawa, T., K. J. Meaburn, and T. Misteli. 2008. The meaning of gene positioning. *Cell.* 135:9–13.
10. Therizols, P., R. S. Illingworth, …, W. A. Bickmore. 2014. Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science.* 346:1238–1242.
11. Sun, H. B., J. Shen, and H. Yokota. 2000. Size-dependent positioning of human chromosomes in interphase nuclei. *Biophys. J.* 79:184–190.
12. Bolzer, A., G. Kreth, …, T. Cremer. 2005. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* 3:e157.
13. Kölbl, A. C., D. Weigl, …, S. Dietzel. 2012. The radial nuclear positioning of genes correlates with features of megabase-sized chromatin domains. *Chromosome Res.* 20:735–752.
14. Dyer, K. A., T. K. Canfield, and S. M. Gartler. 1989. Molecular cytological differentiation of active from inactive X domains in interphase: implications for X chromosome inactivation. *Cytogenet. Cell Genet.* 50:116–120.
15. Jégu, T., E. Aeby, and J. T. Lee. 2017. The X chromosome in space. *Nat. Rev. Genet.* 18:377–389.
16. Eils, R., S. Dietzel, …, T. Cremer. 1996. Three-dimensional reconstruction of painted human interphase chromosomes: active and inactive X chromosome territories have similar volumes but differ in shape and surface structure. *J. Cell Biol.* 135:1427–1440.
17. Berezney, R., K. S. Malyavantham, …, R. Acharya. 2005. Spatio-temporal dynamics of genomic organization and function in the mammalian cell nucleus. *Adv. Enzyme Regul.* 45:17–26.
18. Khalil, A., J. L. Grant, …, A. Arnéodo. 2007. Chromosome territories have a highly nonspherical morphology and nonrandom positioning. *Chromosome Res.* 15:899–916.
19. Sehgal, N., A. J. Fritz, …, R. Berezney. 2014. Gene density and chromosome territory shape. *Chromosoma.* 123:499–513.
20. Mirny, L. A. 2011. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.* 19:37–51.
21. Sanborn, A. L., S. S. Rao, …, E. L. Aiden. 2015. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. USA.* 112:E6456–E6465.
22. Kang, H., Y. G. Yoon, …, C. Hyeon. 2015. Confinement-induced glassy dynamics in a model for chromosome organization. *Phys. Rev. Lett.* 115:198102.

23. Cook, P. R., and D. Marenduzzo. 2009. Entropic organization of interphase chromosomes. *J. Cell Biol.* 186:825–834.

24. Tark-Dame, M., R. van Driel, and D. W. Heermann. 2011. Chromatin folding–from biology to polymer models and back. *J. Cell Sci.* 124:839–845.

25. Marti-Renom, M. A., and L. A. Mirny. 2011. Bridging the resolution gap in structural modeling of 3D genome organization. *PLoS Comput. Biol.* 7:e1002125.

26. Heermann, D. W., H. Jerabek, …, Y. Li. 2012. A model for the 3D chromatin architecture of pro and eukaryotes. *Methods.* 58:307–314.

27. Vasquez, P. A., and K. Bloom. 2014. Polymer models of interphase chromosomes. *Nucleus.* 5:376–390.

28. Imakaev, M. V., G. Fudenberg, and L. A. Mirny. 2015. Modeling chromosomes: beyond pretty pictures. *FEBS Lett.* 589:3031–3036.

29. Amitai, A., and D. Holcman. 2017. Polymer physics of nuclear organization and function. *Phys. Rep.* 678:1–83.

30. Tjong, H., W. Li, …, F. Alber. 2016. Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. *Proc. Natl. Acad. Sci. USA.* 113:E1663–E1672.

31. Odenheimer, J., G. Kreth, and D. W. Heermann. 2005. Dynamic simulation of active/inactive chromatin domains. *J. Biol. Phys.* 31:351–363.

32. Jost, D., P. Carrivain, …, C. Vaillant. 2014. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res.* 42:9553–9561.

33. Jost, D., C. Vaillant, and P. Meister. 2017. Coupling 1D modifications and 3D nuclear organization: data, models and function. *Curr. Opin. Cell Biol.* 44:20–27.

34. Chiariello, A. M., S. Bianco, …, M. Nicodemi. 2015. Polymer models of the organization of chromosomes in the nucleus of cells. *Mod. Phys. Lett. B.* 29:1530003.

35. Haddad, N., D. Jost, and C. Vaillant. 2017. Perspectives: using polymer modeling to understand the formation and function of nuclear compartments. *Chromosome Res.* 25:35–50.

36. Ghosh, S. K., and D. Jost. 2018. How epigenome drives chromatin folding and dynamics, insights from efficient coarse-grained models of chromosomes. *PLoS Comput. Biol.* 14:e1006159.

37. Zhang, B., and P. G. Wolynes. 2017. Genomic energy landscapes. *Biophys. J.* 112:427–433.

38. Tiana, G., A. Amitai, …, L. Giorgetti. 2016. Structural fluctuations of the chromatin fiber within topologically associating domains. *Biophys. J.* 110:1234–1245.

39. Di Pierro, M., B. Zhang, …, J. N. Onuchic. 2016. Transferable model for chromosome architecture. *Proc. Natl. Acad. Sci. USA.* 113:12168–12173.

40. Di Pierro, M., R. R. Cheng, …, J. N. Onuchic. 2017. De novo prediction of human chromosome structures: epigenetic marking patterns encode genome architecture. *Proc. Natl. Acad. Sci. USA.* 114:12126–12131.

41. Bickmore, W. A. 2013. The spatial organization of the human genome. *Annu. Rev. Genomics Hum. Genet.* 14:67–84.

42. Flaus, A., and T. Owen-Hughes. 2011. Mechanisms for ATP-dependent chromatin remodelling: the means to the end. *FEBS J.* 278:3579–3595.

43. Ganai, N., S. Sengupta, and G. I. Menon. 2014. Chromosome positioning from activity-based segregation. *Nucleic Acids Res.* 42:4145–4159.

44. Weber, S. C., A. J. Spakowitz, and J. A. Theriot. 2012. Nonthermal ATP-dependent fluctuations contribute to the in vivo motion of chromosomal loci. *Proc. Natl. Acad. Sci. USA.* 109:7338–7343.

45. Zidovska, A., D. A. Weitz, and T. J. Mitchison. 2013. Micron-scale coherence in interphase chromatin dynamics. *Proc. Natl. Acad. Sci. USA.* 110:15555–15560.

46. Chu, F. Y., S. C. Haley, and A. Zidovska. 2017. On the origin of shape fluctuations of the cell nucleus. *Proc. Natl. Acad. Sci. USA.* 114:10338–10343.

47. Menon, G. I. 2010. Active matter. *In* Rheology of complex Fluids. J. Murali Krishnan, A. Deshpande, and P. B. S. Kumar, eds. Springer, pp. 193–218.

48. Prost, J., F. Jülicher, and J.-F. Joanny. 2015. Active gel physics. *Nat. Phys.* 11:111–117.

49. Marchetti, M. C., J. Joanny, …, R. A. Simha. 2013. Hydrodynamics of soft active matter. *Rev. Mod. Phys.* 85:1143–1189.

50. Needleman, D., and Z. Dogic. 2017. Active matter at the interface between materials science and cell biology. *Nat. Rev. Mater.* 2:17048.

51. Fodor, É., M. Guo, …, F. van Wijland. 2015. Activity-driven fluctuations in living cells. *EPL.* 110:48005.

52. Hameed, F. M., M. Rao, and G. V. Shivashankar. 2012. Dynamics of passive and active particles in the cell nucleus. *PLoS One.* 7:e45843.

53. Agrawal, A., N. Ganai, …, G. I. Menon. 2017. Chromatin as active matter. *J. Stat. Mech.* 2017:014001.

54. Wang, S., and P. G. Wolynes. 2011. Communication: effective temperature and glassy dynamics of active matter. *J. Chem. Phys.* 135:051101.

55. Stillinger, F. H. 1976. Phase transitions in the Gaussian core system. *J. Chem. Phys.* 65:3968–3974.

56. Plimpton, S., P. Crozier, and A. Thompson. 2007. LAMMPS-large-scale atomic/molecular massively parallel simulator. *Sandia National Laboratories.* 18. https://lammps.sandia.gov/.

57. Harrow, J., A. Frankish, …, T. J. Hubbard. 2012. GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res.* 22:1760–1774.

58. ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 489:57–74.

59. Rao, S. S., M. H. Huntley, …, E. L. Aiden. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell.* 159:1665–1680.

60. Žunić, D., and J. Žunić. 2013. Shape ellipticity based on the first Hu moment invariant. *Inf. Process. Lett.* 113:807–810.

61. Millett, K. C., P. Plunkett, …, A. Stasiak. 2009. Effect of knotting on polymer shapes and their enveloping ellipsoids. *J. Chem. Phys.* 130:165104.

62. Rawdon, E. J., J. C. Kern, …, K. C. Millett. 2008. Effect of knotting on the shape of polymers. *Macromolecules.* 41:8281–8287.

63. Todd, M. J., and E. A. Yıldırım. 2007. On Khachiyan's algorithm for the computation of minimum-volume enclosing ellipsoids. *Discrete Appl. Math.* 155:1731–1744.

64. Jackson, D. A., and A. Pombo. 1998. Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J. Cell Biol.* 140:1285–1295.

65. Fraser, P., and W. Bickmore. 2007. Nuclear organization of the genome and the potential for gene regulation. *Nature.* 447:413–417.

66. Chubb, J. R., T. Trcek, …, R. H. Singer. 2006. Transcriptional pulsing of a developmental gene. *Curr. Biol.* 16:1018–1025.

67. Murmann, A. E., J. Gao, …, J. D. Rowley. 2005. Local gene density predicts the spatial position of genetic loci in the interphase nucleus. *Exp. Cell Res.* 311:14–26.

68. Kreth, G., J. Finsterle, …, C. Cremer. 2004. Radial arrangement of chromosome territories in human cell nuclei: a computer model approach based on gene density indicates a probabilistic global positioning code. *Biophys. J.* 86:2803–2812.

69. Kalhor, R., H. Tjong, …, L. Chen. 2011. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* 30:90–98.

70. Fedorova, E., and D. Zink. 2009. Nuclear genome organization: common themes and individual patterns. *Curr. Opin. Genet. Dev.* 19:166–171.

71. Almassalha, L. M., G. M. Bauer, …, V. Backman. 2017. Macrogenomic engineering via modulation of the scaling of chromatin packing density. *Nat. Biomed. Eng.* 1:902–913.

72. Mattout, A., D. S. Cabianca, and S. M. Gasser. 2015. Chromatin states and nuclear organization in development–a view from the nuclear lamina. *Genome Biol.* 16:174.

73. Németh, A., and G. Längst. 2011. Genome organization in and around the nucleolus. *Trends Genet.* 27:149–156.

74. Maharana, S., K. V. Iyer, …, G. V. Shivashankar. 2016. Chromosome intermingling-the physical basis of chromosome organization in differentiated cells. *Nucleic Acids Res.* 44:5148–5160.

75. de Lanerolle, P. 2012. Nuclear actin and myosins at a glance. *J. Cell Sci.* 125:4945–4949.

76. Darrow, E. M., M. H. Huntley, …, E. L. Aiden. 2016. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc. Natl. Acad. Sci. USA.* 113:E4504–E4512.