Supporting Information: Phase Separation and Correlated Motions in Motorized Genome

Zhongling Jiang, Yifeng Qi, Kartik Kamat, and Bin Zhang*

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA

E-mail: binz@mit.edu

Contents

The Diploid Human Genome Model Parameterized with Hi-C Data

The energy function of the genome model adopts the form

$$U_{\text{Genome}}(\boldsymbol{r}) = U(\boldsymbol{r}) + U_{\text{ideal}}(\boldsymbol{r}) + U_{\text{compt}}(\boldsymbol{r}) + U_{\text{Xi}}(\boldsymbol{r}), \tag{S1}$$

where $U(\boldsymbol{r})$ maintains the polymeric topology of chromosomes with bond and angular potentials, $u_{\text{bond}}(r_{i,i+1})$ and $u_{\text{angle}}(r_{i,i+1}, r_{i+1,i+2})$. It also accounts for excluded volumes effects among chromosomes and between chromosomes and the spherical confinement. Taken together, $U(\boldsymbol{r})$ is defined as

$$U(\boldsymbol{r}) = \sum_i \left[u_{\text{bond}}(r_{i,i+1}) + u_{\text{angle}}(\boldsymbol{r}_{i,i+1}, \boldsymbol{r}_{i+1,i+2}) + u_{\text{hc}}(r_i)\right] + \sum_{j>i} u_{\text{hc}}(r_{ij}), \tag{S2}$$

with

$$\begin{aligned}
&u_{\text{bond}}(r_{i,i+1}) = K_b[(r_{i,i+1} - R_0)^2 + (r_{i,i+1} - R_0)^3 + (r_{i,i+1} - R_0)^4], K_b = 20\epsilon, R_0 = 1\sigma \\
&u_{\text{angle}}(\boldsymbol{r}_{i,i+1}, \boldsymbol{r}_{i+1,i+2}) = K_a\left[1 - \cos(\theta - \pi)\right], K_a = 2\epsilon, \cos\theta = \frac{\boldsymbol{r}_{i,i+1} \cdot \boldsymbol{r}_{i+1,i+2}}{|\boldsymbol{r}_{i,i+1}| \cdot |\boldsymbol{r}_{i+1,i+2}|} \\
&u_{\text{hc}}(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \epsilon, r \leq 2^{1/6}\sigma \\ 0, r_i > 2^{1/6}\sigma \end{cases}
\end{aligned} \tag{S3}$$

where r_i is the radial distance of bead i from the origin and r_{ij} corresponds to the distance between beads i and j. $u_{\text{hc}}(r_i)$ forbids the escape of chromosomes from the nuclear envelope while $u_{\text{hc}}(r_{ij})$ prevents the cross-over of polymer chains.

The last three terms in Eq. S1 correspond to the ideal chromosome potential for beads from the same chromosome, $U_{\text{ideal}}(\boldsymbol{r})$, the compartment-specific nonbonded potentials for beads from the same or different chromosomes, $U_{\text{compt}}(\boldsymbol{r})$, and the X-chromosome inactivation potential $U_{\text{Xi}}(\boldsymbol{r})$. $U_{\text{Xi}}(\boldsymbol{r})$ was only applied to the inactive X chromosome to account for

its additional compaction compared to the active copy. These potentials are defined as :

$$
\begin{aligned}
U_{\text{ideal}}(\boldsymbol{r}) &= \sum_{I} \sum_{i,j \in I} \alpha_{\text{ideal}}(|i-j|) f(r_{ij}), \\
U_{\text{compt}}(\boldsymbol{r}) &= \sum_{I,J} \sum_{i \in I, j \in J} \alpha_{\text{compt}}(T_i^I, T_j^I) f(r_{ij}), \\
U_{X\text{i}}(\boldsymbol{r}) &= \sum_{i,j \in X\text{i}} w(r_{ij}) = \sum_{i,j \in X\text{i}} \alpha_{X\text{i}}(|i-j|) f(r_{ij}),
\end{aligned}
\tag{S4}
$$

where I and J loop over all 46 chromosomes and i and j go through all beads on one chromosome. $\alpha_{\text{ideal}}(|i-j|)$ varies with respect to the index difference between the bead-pair i and j. $\alpha_{\text{compt}}(T_i^I, T_j^I)$ is a function of the compartment types of beads i and j. Specifically, T_i^I and T_j^I can be type A, B or C. The contact probability between two loci distancing r_{ij} is depicted by the function $f(r_{ij})$, whose ensemble average is the same quantity evaluated in the Hi-C contact map. $f(r_{ij})$ is defined as:

$$
f(r_{ij}) = \begin{cases} \frac{1}{2}\left[1 + \tanh\left[\eta(r_c - r_{ij})\right]\right], & r_{ij} \le r_c \\ \frac{1}{2}\left(r_c/r\right)^4, & r_{ij} > r_c \end{cases}
\tag{S5}
$$

where $r_c = 1.5$ and $\eta = 2.5$.

As shown in Ref. S1, the above Hi-C data related terms were parameterized with an efficient maximum entropy optimization algorithm. Specifically, $\alpha_{\text{ideal}}(|i-j|)$, $\alpha_{\text{compt}}(T_i^I, T_j^I)$, and $\alpha_{X\text{i}}(|i-j|)$ were tuned to satisfy the following constraints:

$$
\begin{aligned}
\left\langle \sum_{I \neq X} \sum_{i,j \in I} f(r_{ij}) \delta_{|i-j|,s} \right\rangle_{U_{\text{Genome}}(\boldsymbol{r})} &= \sum_{I \neq X} \sum_{i,j \in I} f_{ij}^{\text{exp}} \delta_{|i-j|,s}, \quad \text{for } s = 1, \cdots, n-1 \\
\left\langle \sum_{I,J} \sum_{i \in I, j \in J} f(r_{ij}) \delta_{T_i^I, T_1} \delta_{T_j^J, T_2} \right\rangle_{U_{\text{Genome}}(\boldsymbol{r})} &= \sum_{I,J} \sum_{i \in I, j \in J} f_{ij}^{\text{exp}} \delta_{T_i^I, T_1} \delta_{T_j^J, T_2}, \quad \text{for } T_1, T_2 \in \{A, B, C\} \\
\left\langle \sum_{i,j \in X} f(r_{ij}) \delta_{|i-j|,s} \right\rangle_{U_{\text{Genome}}(\boldsymbol{r})} &= \sum_{i,j \in X} f_{ij}^{\text{exp}} \delta_{|i-j|,s}, \quad \text{for } s = 1, \cdots, n_X - 1
\end{aligned}
\tag{S6}
$$

where $\langle\cdot\rangle$ stands for the ensemble average, which is taken over the Boltzmann distribution $e^{-\beta U_{\text{Genome}}(\boldsymbol{r})}$. f_{ij}^{exp} correspond to the contact probability f_{ij}^{exp} between bead i and j measured in Hi-C experiments for GM12878 cells.[S2] $\delta_{T_i^I,T_1}$ is the Kronecker delta function adopting the form:

$$\delta_{T_i^I,T_1} = \begin{cases} 1, \text{if } T_i^I = T_1 \\ 0, \text{otherwise} \end{cases} \tag{S7}$$

An iterative algorithm was performed to derive the values for $\alpha_{\text{ideal}}(|i-j|)$, $\alpha_{\text{compt}}(T_i^I, T_j^I)$, and $\alpha_{X\text{i}}(|i-j|)$ that enforce the constraints defined in Eq. S6.

Simulation details

We carried out an independent Brownian dynamics simulation for every combination of parameters, including the interaction strength between B compartments and the magnitude of active forces. These simulations were initialized with an equilibrium structure obtained from a long-timescale trajectory performed in our previous study.[S1] We carried out 5×10^6-step-long relaxation simulations before switching to production runs that lasted for 2×10^7 steps. The relaxation simulations were performed with Langevin dynamics, during which the temperature was gradually annealed from $3T$ to T. The soft-core potential introduced in Ref. S1 was used for nonbonded interactions between chromosomes to facilitate polymer conformational rearrangement. To ensure simulation stability, we used smaller timesteps at larger active forces. The values are 0.01, 0.005, 0.003, 0.002 for zero, five, ten, and twenty-fold active forces. These trajectories were used to compute the density profiles and the displacement correlation functions with $\Delta t > 15$s.

Additional 4×10^6-step-long simulations were carried out with a time step of 0.001 to compute the displacement correlation functions $C(r, \Delta t)$t for $\Delta t \leq 15$s. Each simulation was initialized with the last configuration of the trajectory for the density profile calculations.

Analysis details

Effective temperature calculations

We computed the effective temperature of A/B compartments using the fluctuation-dissipation ration[S3–S6]

$$T_{\text{eff}} = \frac{C(t_0) - C(t)}{\chi(t)}. \tag{S8}$$

The correlation function is defined as:

$$C(t) = \langle O(t + t_0)O'(t_0)\rangle - \langle O(t_0)\rangle \langle O'(t_0)\rangle . \tag{S9}$$

We used $O(t) = \frac{1}{N}\sum_{j=1}^{N} \epsilon_j \exp[i\boldsymbol{\kappa} \cdot \boldsymbol{r}_j(t)]$ and $O'(t) = 2\sum_{j=1}^{N} \epsilon_j \cos[\boldsymbol{\kappa} \cdot \boldsymbol{r}_j(t)]$, and the summation goes over all particles of the same compartment type. The set of ϵ_j was sampled from ± 1 with equal probability. The wave vector $|\boldsymbol{\kappa}| = 6.28$ corresponds to the first peak the structure factor, $S(\boldsymbol{\kappa}) = \frac{1}{N}\sum_{j=1}^{N}\sum_{l=1}^{N} \langle \exp\{i\boldsymbol{\kappa} \cdot [\boldsymbol{r}_j(t_0) - \boldsymbol{r}_l(t_0)]\}\rangle$

We produced a total of 100 sets of random realizations for ϵ_j. For each set, 50 independent simulations were carried out starting from configurations uniformly sampled from the second half of the production trajectories for density profile calculations. These simulations lasted 5×10^4 steps with a time step of 0.001. The correlation function was determined from the ensemble average of these trajectories.

In addition, for each set of ϵ_j, we carried out 50 independent simulations with the presence of a perturbation force to each particle i in the form of $\boldsymbol{F}_i(t) = \frac{\partial h O'(t)}{\partial \boldsymbol{r}_i}$ with $h = 0.02$. From these simulations, we computed the response function as

$$\chi(t) = \frac{\langle O(t) - O(t_0)\rangle}{h} \tag{S10}$$

by averaging over all realization and all trajectories.

Radial density profiles

The radial density profile $\rho(r)$ is defined as

$$\rho(r) = \frac{\langle n_i(r) \rangle}{4\pi r^2 \Delta r \cdot N_i}, \tag{S11}$$

where $n_i(r)$ is the number of type i beads lying between the spherical shell r to $r + \Delta r$ and N_i is the total number of type i beads. $\langle \cdot \rangle$ stands for the ensemble average. We sampled 5000 configurations from the second half of the production trajectories at equal time intervals to compute the density profiles. These configurations were separated into five non-overlapping blocks to estimate the error bars as the standard deviation of the mean.

Chromosome clustering with Voronoi tessellation

To provide a quantitative characterization of chromosome clusters, we first adopted the Voronoi tessellation method[S7] to assign regions of non-overlapping space to each genomic segment, i.e., the coarse-grained one-MB-in-size particles. Two particles are connected as nearest neighbors if they share a common cell wall. The largest cluster was then identified as the largest connected component for the network whose nodes include particles of the same compartment type and edges only defined for nearest neighbors.[S8] See Figure S7 for an illustration of the procedure. From the identified clusters, we defined the average volume per bead as $\frac{\sum_i V_i}{N_c}$, where V_i is the volume for the region of space assigned for particle i and N_c is the total number of particles in the cluster. The radius of gyration was computed as $R_g = \sqrt{\frac{\Sigma_{i=1}^{N_c} r_i^2}{N_c}}$.[S9,S10] We repeated the analyses for 500 configurations uniformly selected from the trajectories used for the density profile calculations, and the average results were reported in the main text.

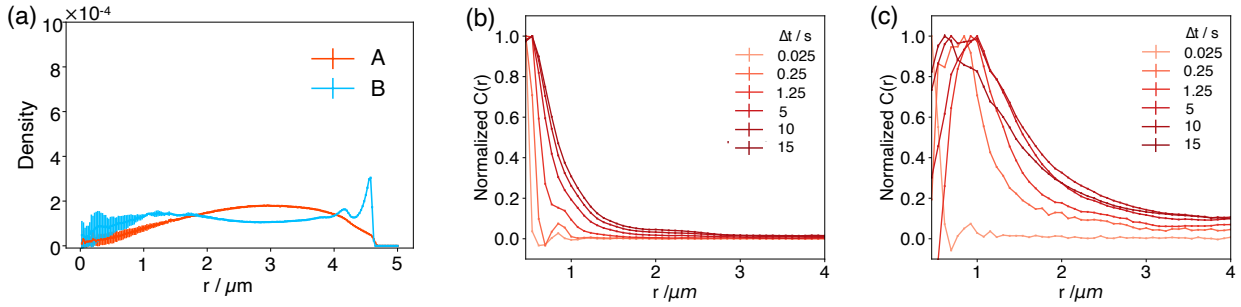

Figure S1: (a) Radial density profiles calculated with a soft-core potential support the role of active forces in positioning B compartments towards the nuclear periphery. Unlike the Lennard-Jones potential, the soft-core potential for non-bonded interactions allows polymer chain crossing with a finite energetic barrier.[S11] Besides a change in the non-bonded potential, other simulation setups are identical to the one presented in Figure 2b of the main text. (b, c) Displacement correlation functions for systems with $T_a = 0$ (b) and $T_a = 10T$ (c). The simulation setups are identical to the ones presented in Figures 4a and 4b of the main text.

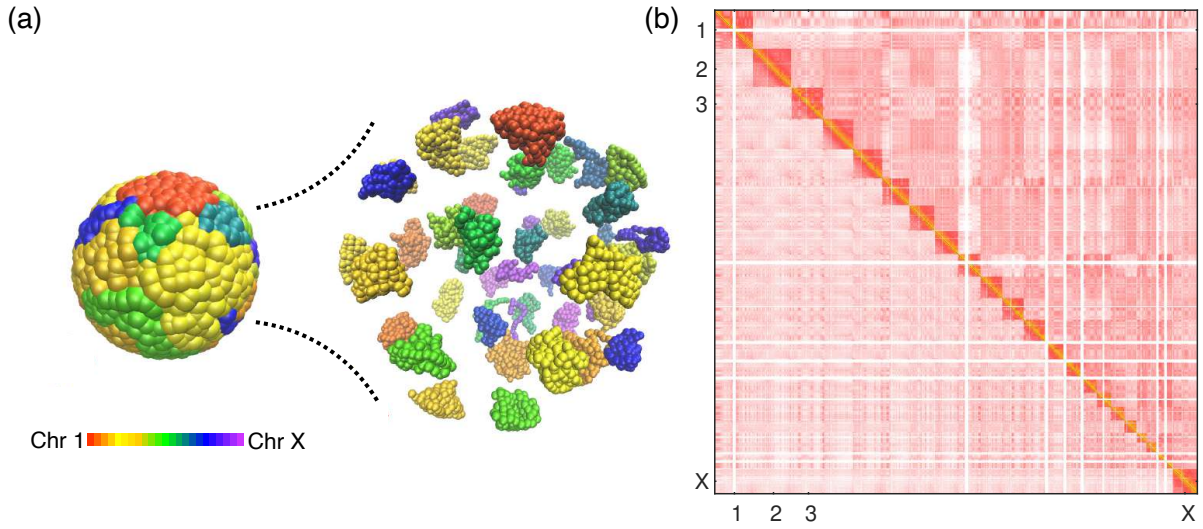

Figure S2: The human genome model presented in the main text succeeds in reproducing the formation of chromosome territories and the phase separation of A/B compartments. (a) A representative configuration of the genome with individual chromosomes assigned with unique colors. Homologs are shown in the same colors. A blow up of the configuration is shown on the right to better highlight chromosome territories. (b) Comparison between simulated (top right) and experimental[S2] (bottom left) contact maps for the genome.

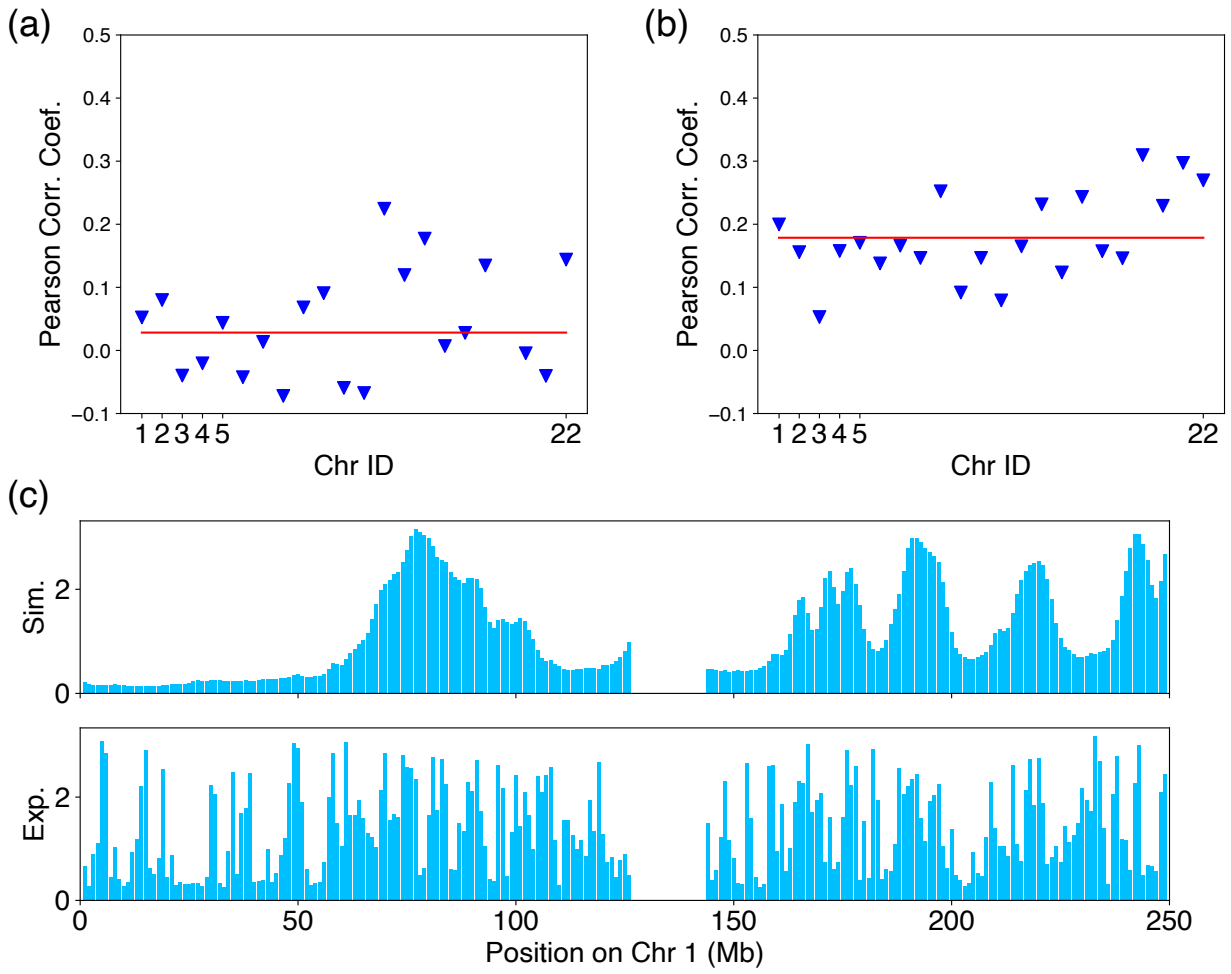

Figure S3: Active forces improve the the correlation between simulated and experimental contact frequency between chromosomes and the nuclear envelope. For the simulated configurations, a coarse-grained bead is determined as in contact with the nuclear envelope if its radial distance from the spherical confinement is less than 1.5σ. We averaged the contact probabilities over two alleles to obtain the profiles for haploid chromosomes. The observed over expected (OE) score was then determined by normalizing the contact probabilities with the genome-wide average. Since no DamID data are available of the GM12878 cell line, we compared simulation results with data for HFF cells generated by the van Steensel lab[S12,S13] (https://data.4dnucleome.org, accession number 4DNESXZ4FW4T). (a, b) Pearson correlation coefficients between simulated and experimental OE score for chromosome-lamina contacts. The in silico data were obtained from simulations without (a) and with (b) active forces of $T_a = 10T$. (c) Simulated (top) and experimental (bottom) OE score as a function of the genomic position for chromosome one.

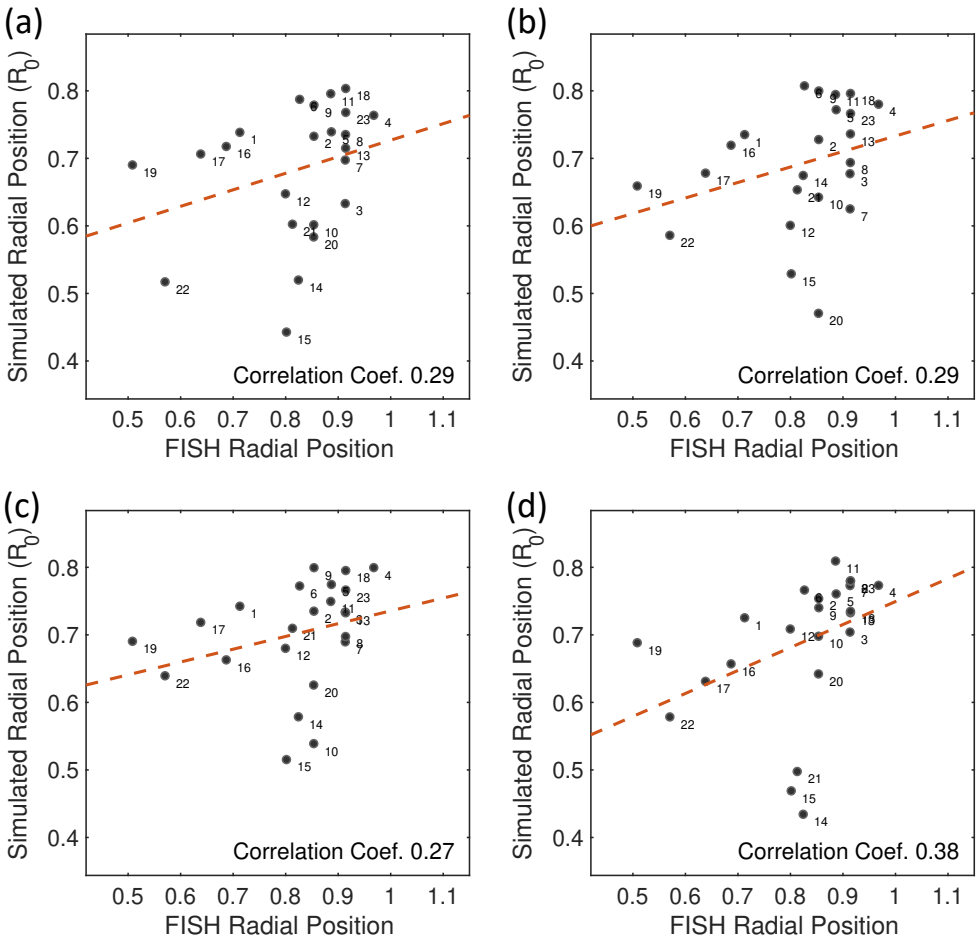

Figure S4: Comparison between experimental[S14] and simulated chromosome radial positions. The four panels were determined from simulations performed with $T_a = 0$ (a), $5T$ (b), $10T$ (c) and $20T$ (d). Chromosome radial positions were estimated using the center of masses of the coarse-grained polymers. The Pearson correlation coefficients between the two data sets are shown in legends. From a to d, it is evident that the agreement between simulation and experiment improves for most chromosomes as T_a increases. Indeed, after removing the three outliers, chr 14, 15 and 21, the correlation coefficients between data shown in part d increases to 0.7. From these plots, it is also evident that more active chromosomes, including 16 and 17, move more towards the nuclear interior at higher active forces, while the inactive ones, including 3, 20, 22, position themselves towards the periphery.

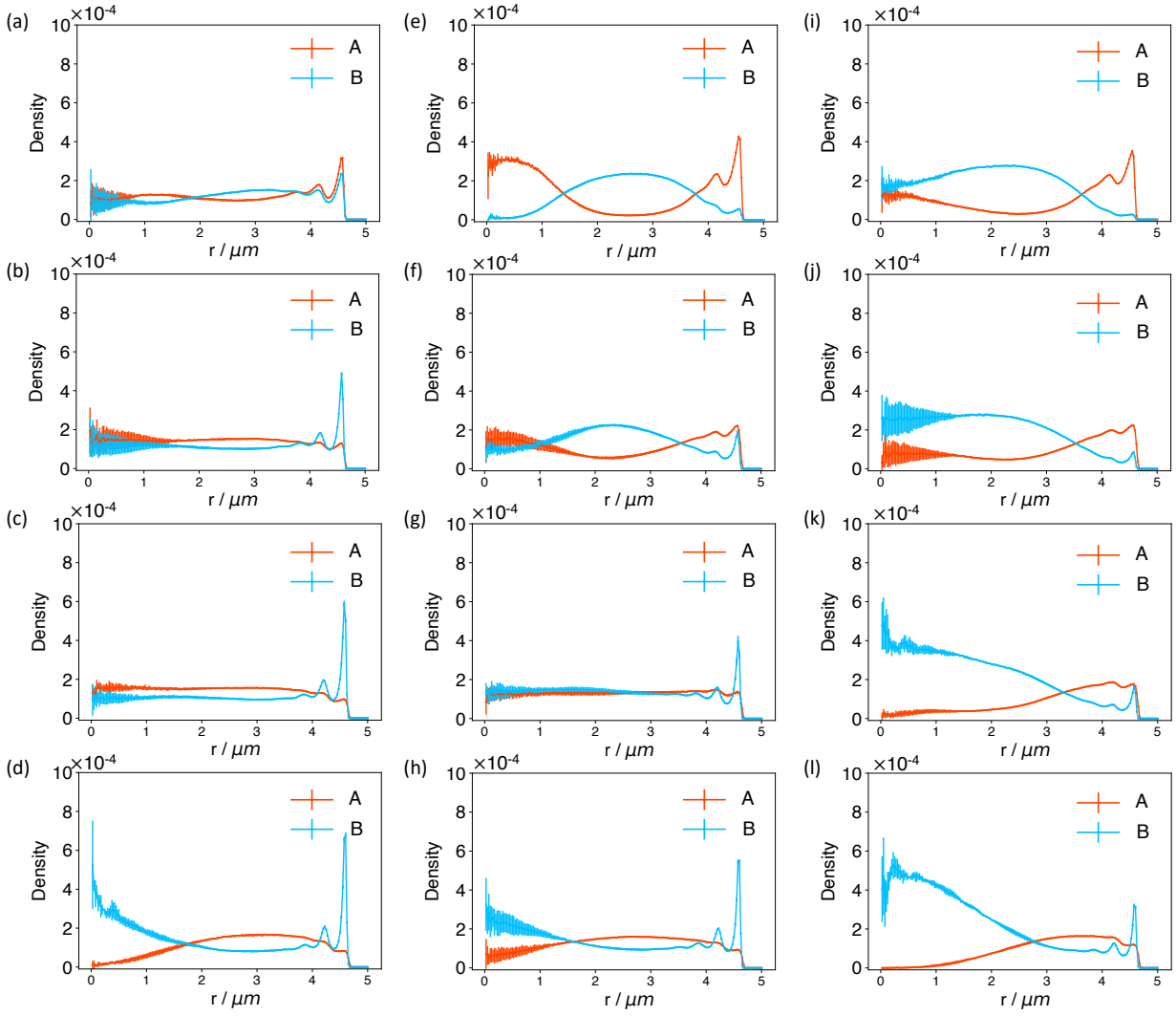

Figure S5: Radial density profiles of A/B compartments determined from simulations performed at different parameter sets. From left to right, the interaction strength among B compartments was scaled by a factor of 1, 2, and 3. From top to bottom, the strength of the active force was varied from $T_a/T = 0, 5, 10$, and 20.

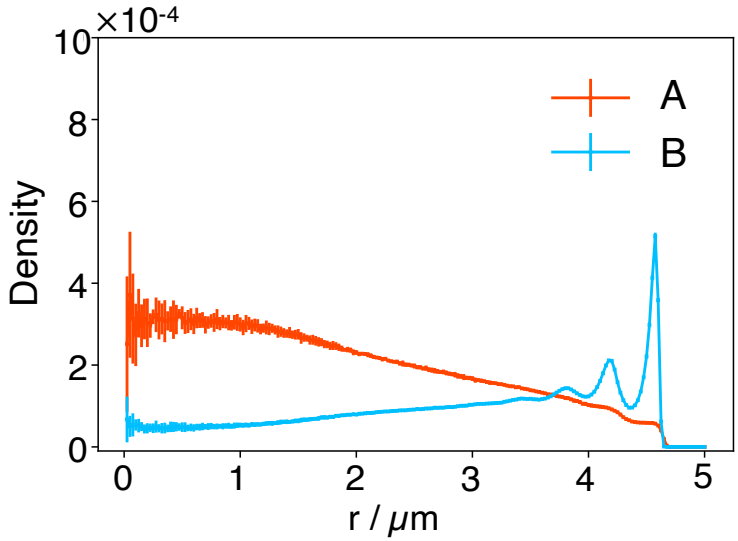

Figure S6: Radial density profiles determined from simulations in which the active forces were limited to the top half of most active A compartments. We quantified the activity of each bead using the fraction of sub-compartment A_1, which was known to exhibit high transcriptional activity.[S2,S15] Specifically, we used the sub-compartment assignments from Ref. S2 at the 50kb resolution to determine the fraction of A_1 in each one MB-sized bead. The top 1212 beads with the highest A_1 fraction were selected as the most active beads to apply active forces on.

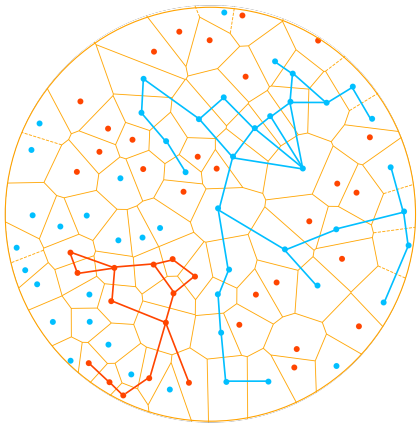

Figure S7: Two-dimensional illustration of the Voronoi tesselation method for identifying the largest cluster. The method tiles the space into non-overlapping cells (orange) occupied by all chromatin particles, with A compartments shown in red and B compartments in blue. Nearest neighbors are defined as particles whose Voronoi cells share boundaries. With chromatin particles as nodes, and edges introduced for nearest neighbors, a network can be defined to find the largest connected component. These components, with examples shown for A/B compartments, were used to quantify the degree of clustering. See text *Section: Chromosome clustering with Voronoi tessellation* for more details

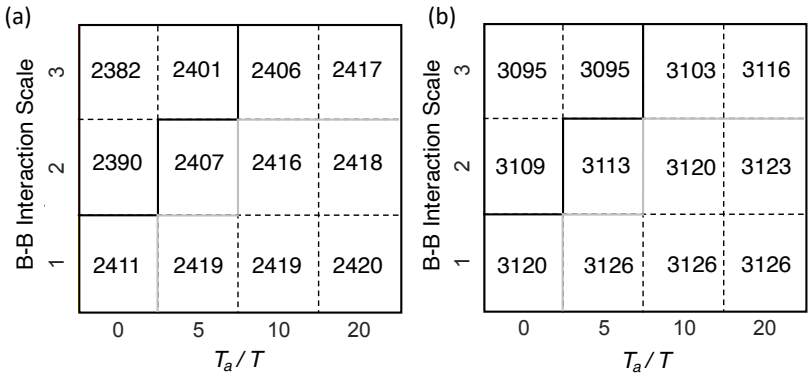

Figure S8: The average size of the largest connected components for A (a) and B (b) compartments at different parameter sets.

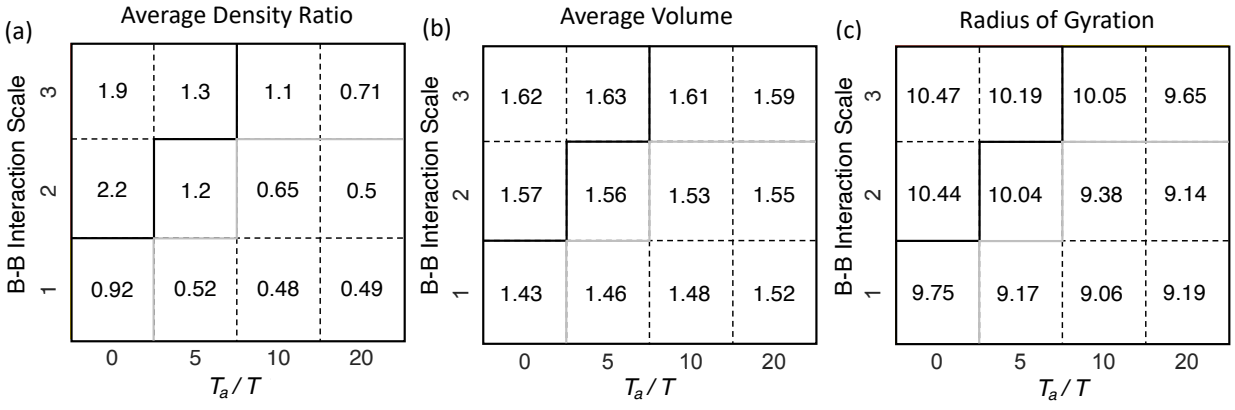

Figure S9: The same plots as in Figure 3 of the main text but for A compartments.

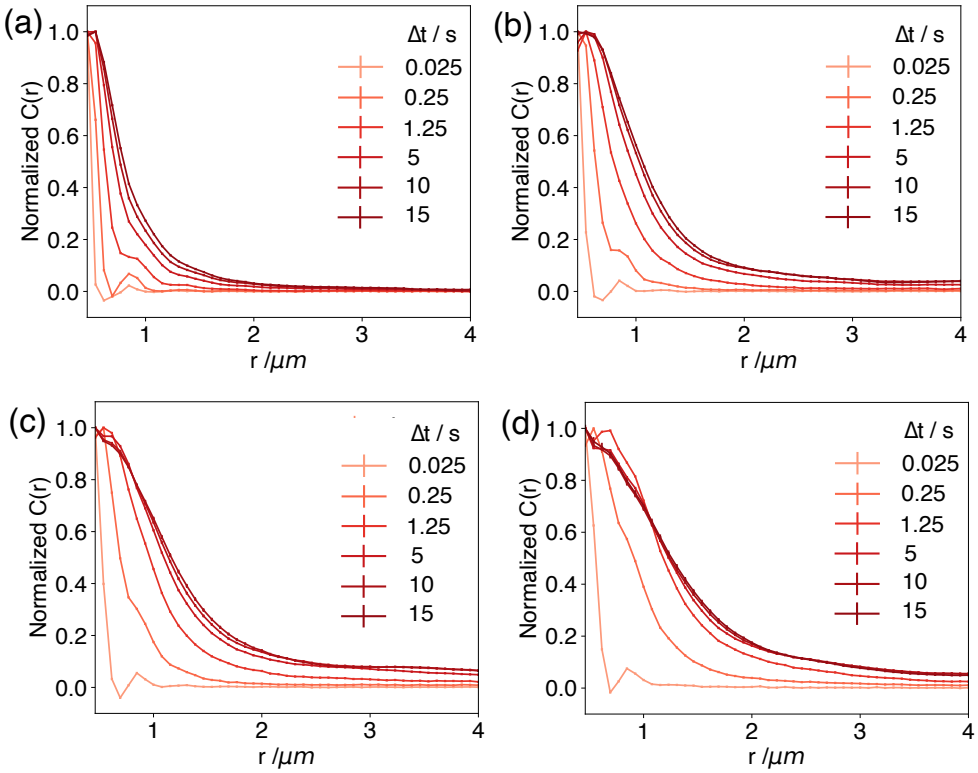

Figure S10: Active forces enhance chromosome spatial correlation. The normalized displacement correlation shown in a-d were obtained from simulations performed with non-correlated active forces with $T_a = 0$, 5, 10 and $20T$ respectively.

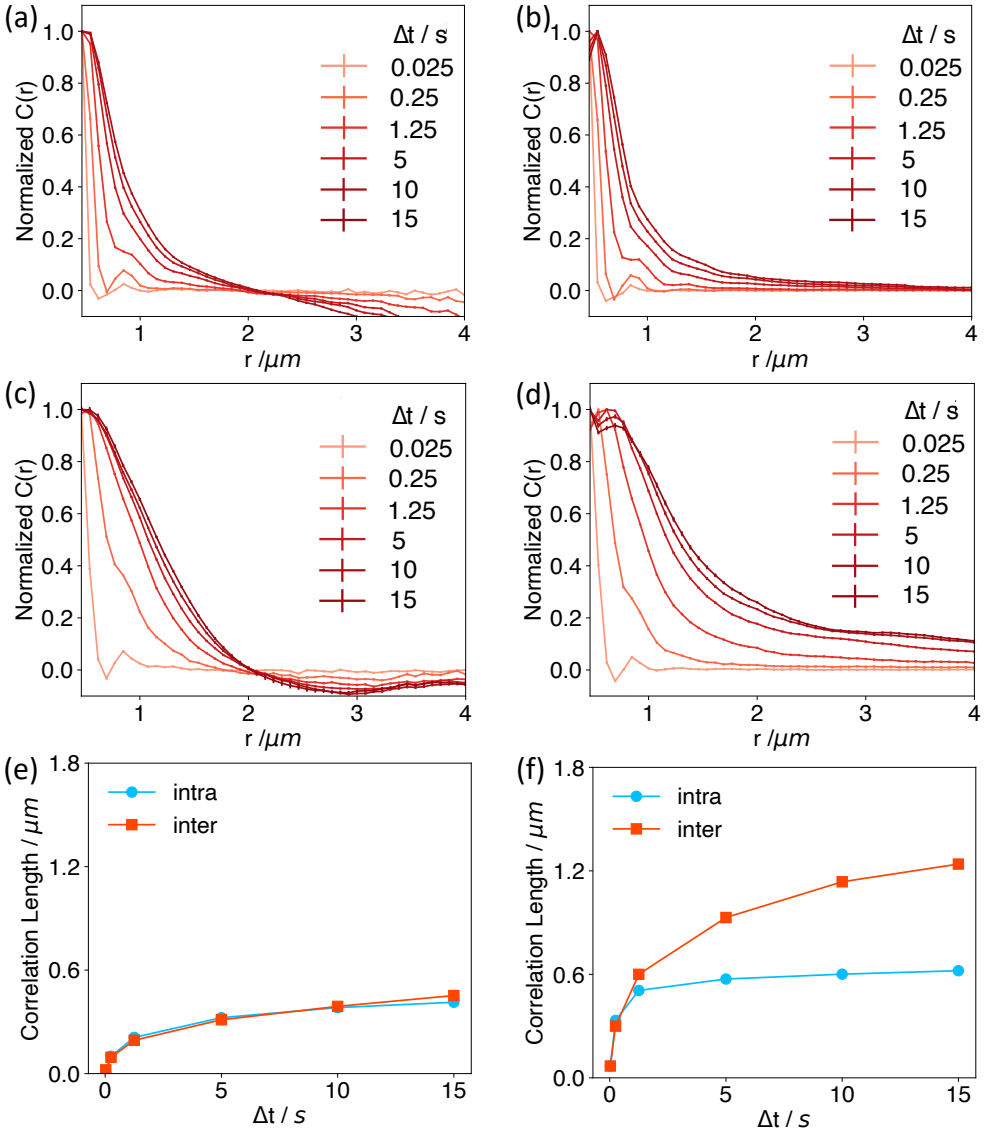

Figure S11: Contributions from intra and inter-chromosome genomic pairs to chromosome spatial correlation. (a,b) Normalized displacement correlation plots for the passive system without active forces calculated using genomic pairs from the same (a) and different (b) chromosomes. (c,d) The same plots as parts a and b but for simulations performed with non-correlated active forces $T_a = 10T$. (e,f) Correlation lengths obtained from exponential fitting to curves from passive (e) and active (f) simulations.

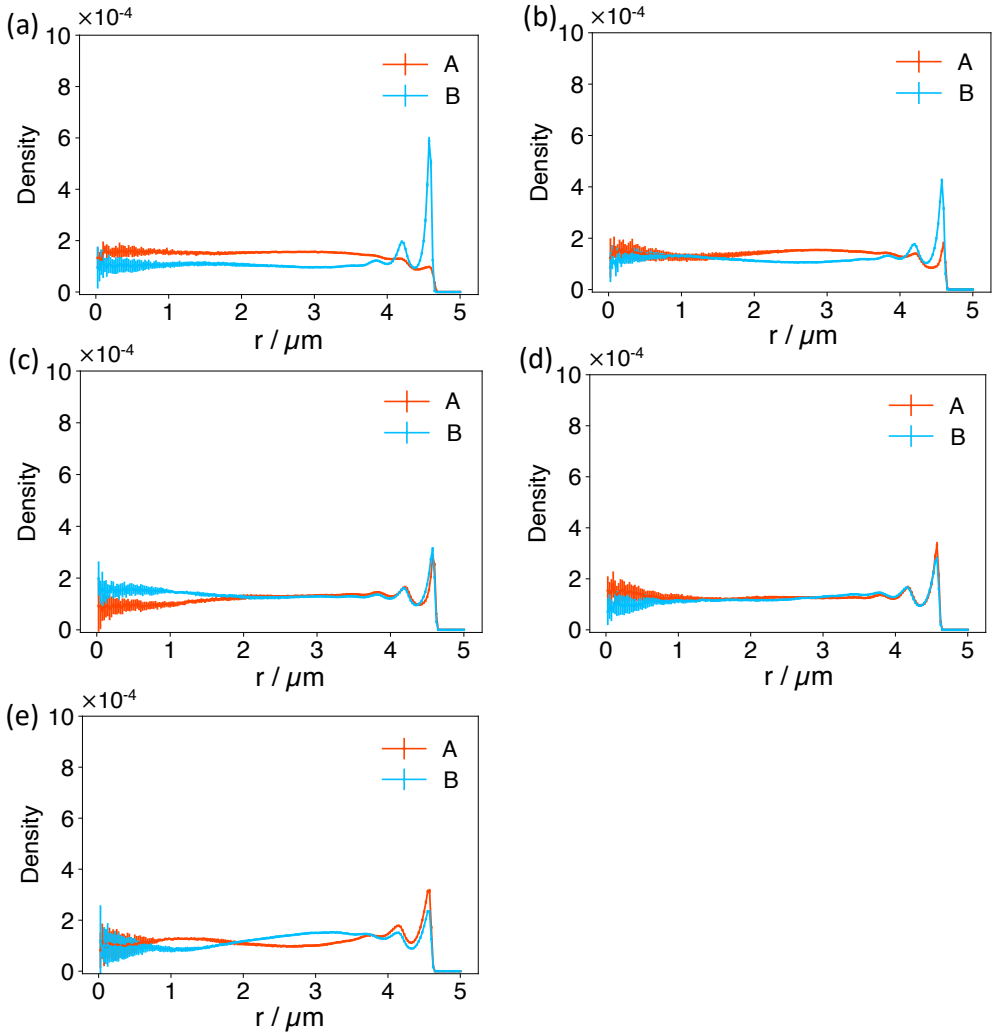

Figure S12: Radial density profiles calculated from simulations with active forces at $T_a = 10T$ and the correlation time $\tau = 0s$ (a), $0.125s$ (b), to $1.25s$ (c), and to $12.5s$ (d). Results in Figure 2a of the main text obtained from simulations without active forces are shown in part e for reference. It's evident that at larger values of the correlation time τ, the impact of active forces in positioning B compartments towards the nuclear periphery decreases.

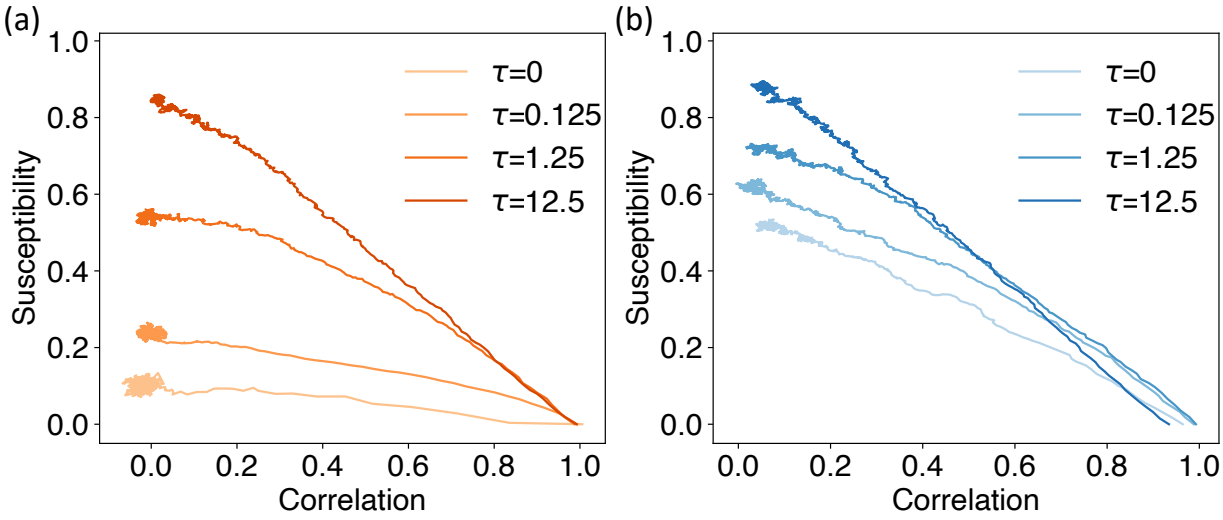

Figure S13: Susceptibility-correlation curves for A (a) and B (b) compartments obtained from simulations with correlated active forces at $T_a = 10T$. Values for the correlation time τ are provided in the legends.

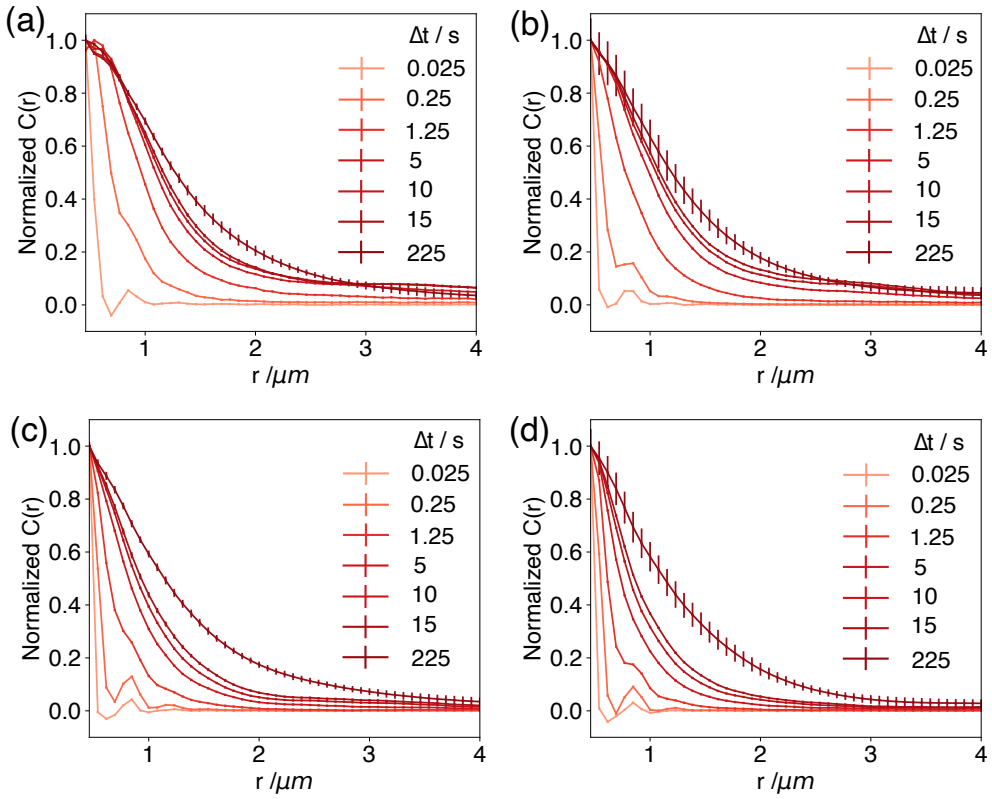

Figure S14: Normalized displacement correlation plots determined from simulations with active forces of $T_a = 10T$ and correlation time $\tau = 0$, $0.125s$, $1.25s$ and $12.5s$.

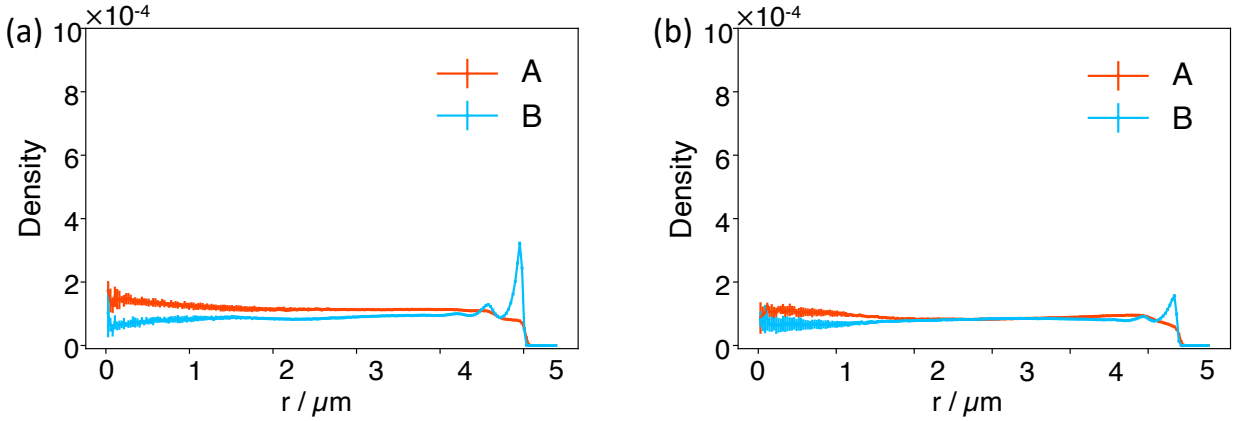

Figure S15: Radial density profiles calculated from simulations with spherical confinements of radius 5.4 μm (a) and 5.8 μm (b). Besides the change in the confinement radius, other simulation setups are identical to the one presented in Figure 2b of the main text. Compared to the profiles shown in Figure 2b, these results support that impact of active forces in positioning B compartments towards the nuclear periphery diminishes at larger volumes.

References

(S1) Qi, Y.; Reyes, A.; Johnstone, S. E.; Aryee, M. J.; Bernstein, B. E.; Zhang, B. Data-Driven Polymer Model for Mechanistic Exploration of Diploid Genome Organization. *Biophysical Journal* **2020**, *119*, 1905–1916.

(S2) Rao, S. S.; Huntley, M. H.; Durand, N. C.; Stamenova, E. K.; Bochkov, I. D.; Robinson, J. T.; Sanborn, A. L.; Machol, I.; Omer, A. D.; Lander, E. S.; Aiden, E. L. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **2014**, *159*, 1665–1680.

(S3) Barrat, J. L.; Berthier, L. Fluctuation-dissipation relation in a sheared fluid. *Phys. Rev. E* **2001**, *63*, 0125031–0125034.

(S4) Berthier, L.; Barrat, J. L. Nonequilibrium dynamics and fluctuation-dissipation relation in a sheared fluid. *Journal of Chemical Physics* **2002**, *116*, 6228–6242.

(S5) Berthier, L.; Barrat, J. L. Shearing a Glassy Material: Numerical Tests of Nonequilibrium Mode-Coupling Approaches and Experimental Proposals. *Phys. Rev. Lett.* **2002**, *89*, 1–4.

(S6) Loi, D.; Mossa, S.; Cugliandolo, L. F. Non-conservative forces and effective temperatures in active polymers. *Soft Matter* **2011**, *7*, 10193–10209.

(S7) Rycroft, C. H. VORO++: A three-dimensional Voronoi cell library in C++. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **2009**, *19*, 041111.

(S8) Hagberg hagberg, A. A.; Los, l.; Schult, D. A.; Swart swart, P. J. Exploring Network Structure, Dynamics, and Function using NetworkX. **2008**,

(S9) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domański, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, O. MDAnaly-

sis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proceedings of the 15th Python in Science Conference* **2016**, 98–105.

(S10) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **2011**, *32*, 2319–2327.

(S11) Zhang, B.; Wolynes, P. G. Topology, structures, and energy landscapes of human chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*, 6062–6067.

(S12) Dekker, J.; Belmont, A. S.; Guttman, M.; Leshyk, V. O.; Lis, J. T.; Lomvardas, S.; Mirny, L. A.; O'Shea, C. C.; Park, P. J.; Ren, B.; Ritland Politz, J. C.; Shendure, J.; Zhong, S. The 4D nucleome project. *Nature 2017 549:7671* **2017**, *549*, 219–226.

(S13) Reiff, S. B.; Schroeder, A. J.; Kırlı, K.; Cosolo, A.; Bakker, C.; Lee, S.; Veit, A. D.; Balashov, A. K.; Vitzthum, C.; Ronchetti, W.; Pitman, K. M.; Johnson, J.; Ehmsen, S. R.; Kerpedjiev, P.; Abdennur, N.; Imakaev, M.; Öztürk, S. U.; Çamoğlu, U.; Mirny, L. A.; Gehlenborg, N.; Alver, B. H.; Park, P. J. The 4D Nucleome Data Portal as a resource for searching and visualizing curated nucleomics data. *Nature Communications 2022 13:1* **2022**, *13*, 1–11.

(S14) Boyle, S.; Gilchrist, S.; Bridger, J. M.; Mahy, N. L.; Ellis, J. A.; Bickmore, W. A. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Human molecular genetics* **2001**, *10*, 211–219.

(S15) Chen, Y.; Zhang, Y.; Wang, Y.; Zhang, L.; Brinkman, E. K.; Adam, S. A.; Goldman, R.; Van Steensel, B.; Ma, J.; Belmont, A. S. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol.* **2018**, *217*, 4025–4048.