Supplementary Information for

Super-resolution Imaging Reveals Spatiotemporal Propagation of Human Replication Foci Mediated by CTCF-organized Chromatin Structures

Qian Peter Su[a,b,1,2,3], Ziqing Winston Zhao[c,d,e,2,3], Luming Meng[f], Miao Ding[a,b], Weiwei Zhang[a,b], Yongzheng Li[a,b], Mengzhu Liu[a,b], Rongqin Li[a,b], Yi-Qin Gao[f], Xiaoliang Sunney Xie[a,b,g,3], and Yujie Sun[a,b,3]

[a] State Key Laboratory of Membrane Biology, Peking University, Beijing 100871, China.
[b] Biomedical Pioneering Innovation Center, School of Life Sciences, Peking University, Beijing 100871, China.
[c] Department of Chemistry, National University of Singapore, Singapore 117543, Singapore
[d] Centre for BioImaging Sciences, National University of Singapore, Singapore 117557, Singapore.
[e] Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore 138672, Singapore.
[f] Institute of Theoretical and Computational Chemistry, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China.
[g] Beijing Advanced Innovation Center for Genomics, Peking University, Beijing 100871, China.
[1] Present address: Institute for Biomedical Materials & Devices, Faculty of Science, University of Technology Sydney, NSW 2007, Australia.
[2] Q.P.S. and Z.W.Z. contributed equally to this work.
[3] To whom correspondence may be addressed. Email: qian.su@uts.edu.au; zhaozw@nus.edu.sg; sunneyxie@pku.edu.cn; sun_yujie@pku.edu.cn.

This PDF file includes:

Figs. S1 to S9
Supplementary text: Simulation of the spatio-temporal spread of RFi

www.pnas.org/cgi/doi/10.1073/pnas.2001521117

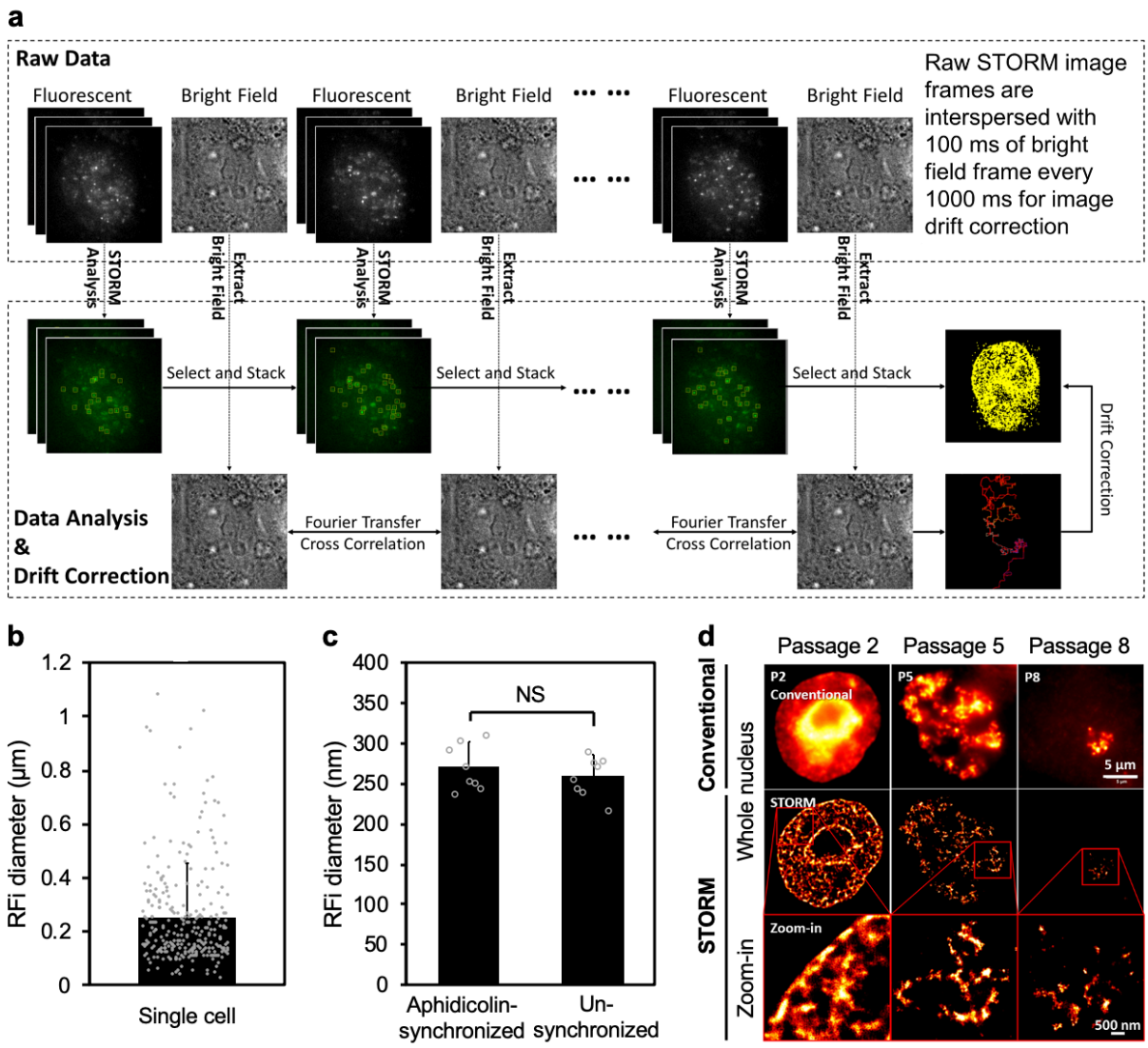

Figure S1 | STORM imaging of RFi and impact of cell cycle synchronization and metabolic labeling on cell growth and RFi morphology. a, Workflow of STORM imaging, drift correction and data analysis for reconstructing super-resolution images. **b,** Distribution of diameter of early S-phase RFi from a single Hela cell. **c,** Comparison of diameters of early S-phase RFi in aphidicolin-synchronized and un-synchronized cells indicates that the cell cycle synchronization does not affect RFi size and morphology. **d,** Metabolic labeling of RFi with EdU does not affect cell growth and division over multiple generations. By the 8th passage, only one chromosome in the nucleus is labeled. Error bars denote mean ± s.d. *p* values are determined by Student's *t* test; NS: not significant. n = 283 RFi for (**b**); and 8 cells for (**c**). Source data are provided as a Source Data file.

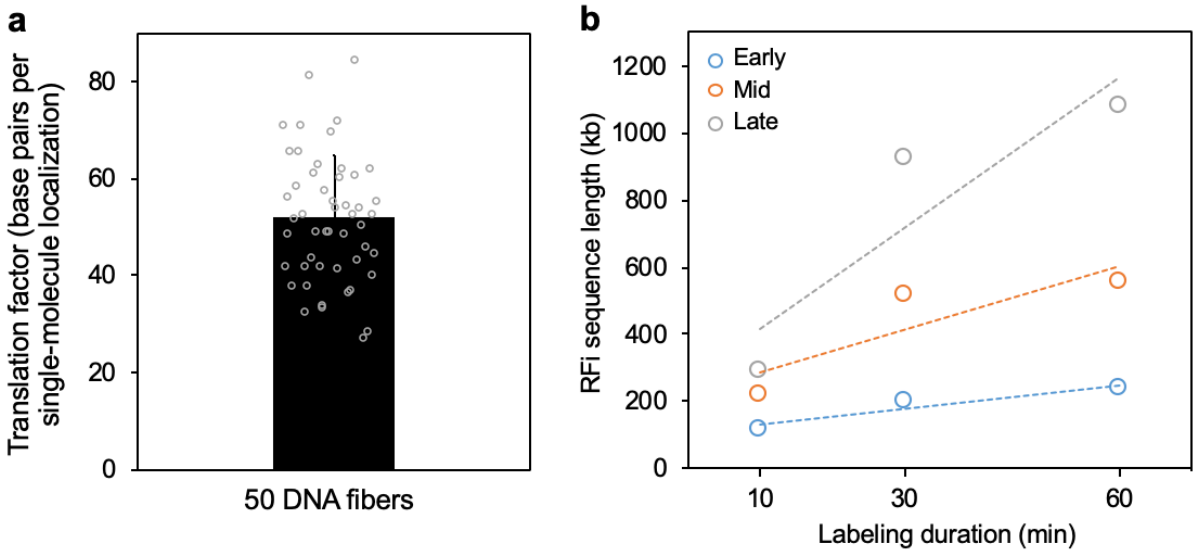

Figure S2 | Calibration of DNA combing experiments and measurement of replication speed across S-phase. a, Distribution of the value of translation factor between single-molecule localization and number of base pairs of replicated DNA, obtained from DNA combing experiment. **b,** Mean sequence lengths of RFi replicated during different labeling durations at early, mid and late S-phase (from **Fig. 2f**), together with their linear fits (dashed lines) which provide measure for replication speed. Error bars denote mean ± s.d. n = 50 separate DNA fibers for (**a**); and 1962, 1441, 1254 (10 min), 1443, 1214, 1125 (30 min), 1146, 865 and 709 (60 min) RFi from 5~8 cells at each stage for (**b**). Source data are provided as a Source Data file.

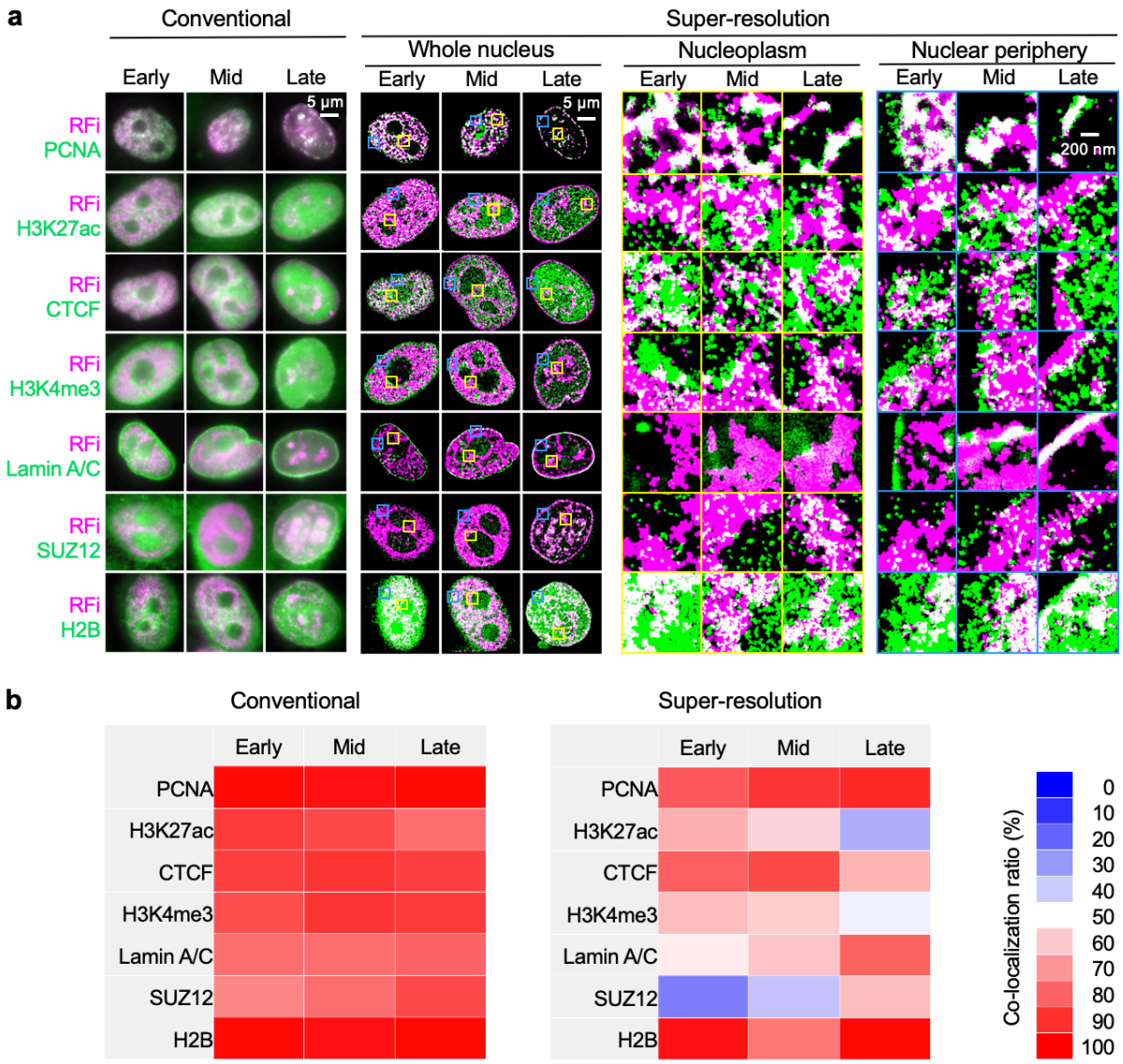

Figure S3 | Two-color conventional and STORM images of RFi and seven nuclear markers and quantification of their colocalization. a, Conventional and STORM images of RFi and seven nuclear markers at early, mid, and late S-phase. Insets show zoomed-in areas of nucleoplasm (yellow boxes) and nuclear periphery (blue boxes) in STORM images. All STORM images display enhanced contrast between the two colors for better visualization; white color indicates colocalization. **b,** Comparison of colocalization heat maps between RFi and each nuclear marker across S-phase using either conventional or STORM images from (**a**). Each co-localization ratio is the mean value of 5~8 cells.

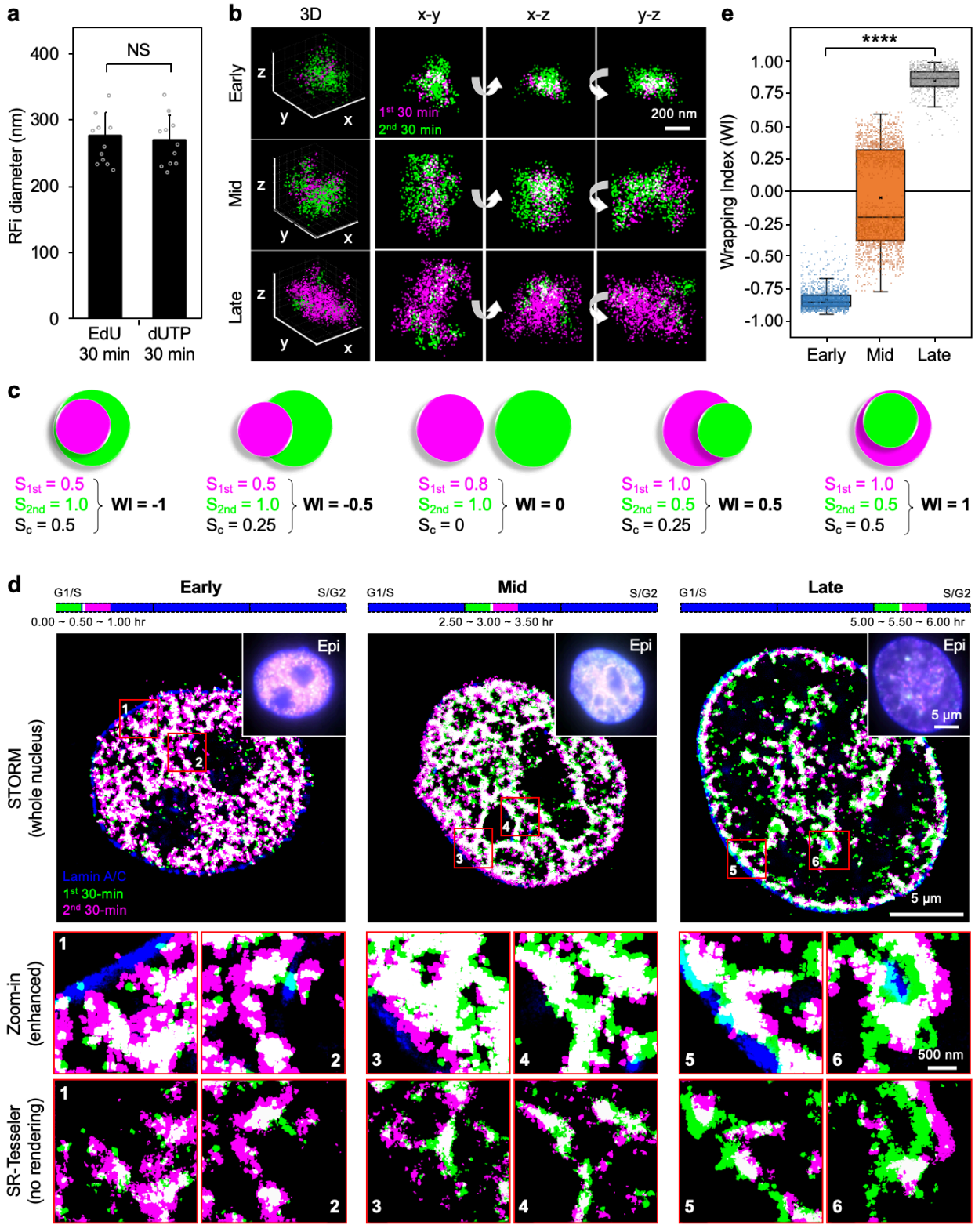

Figure S4 | Validation of the reversal in spatio-temporal propagation direction of RFi from early to late S-phase. a, Comparison of the diameter of RFi labeled using two nucleotide

analogues shows minimal difference in RFi morphology. Error bars denote mean ± s.d. **b,** Localizations constituting a representative RF from early, mid and late S-phase are displayed both in 3D and 2D profiles, demonstrating the reversal in the relative spatial spread between the RFi produced during the two labeling windows. **c,** Schematic definition of the Wrapping Index for RFi produced during the two labeling windows. **d,** Multi-color STORM images of single nuclei with newly replicated DNA in two consecutive 30 min windows at the beginning of early, mid and late S-phase labeled in opposite order from that in **Fig. 4a** to **e**. Insets show zoomed-in areas with enhanced contrast between the two colors (top), as well as direct plotting of coordinates of RFi localizations output by SR-Tesseler without pixel rendering (bottom) for comparison. The observation of the same propagation pattern as shown in **Fig. 4** eliminates the possibility that such pattern is the consequence of labeling and detection artefacts associated with specific dyes. **e,** Wrapping Index (WI) for RFi produced during the two labeling windows in (**d**) at early, mid, and late S-phase. Data are shown as box-and-whisker plots. Each dot denotes a single RF**.** *p* values are determined by two-tailed Student's *t* test; ****: $p < 0.0001$; NS: not significant. $n = 11$ cells for (**a**); and 1558, 2019 and 701 RFi from 8 cells at each stage for (**e**). Source data are provided as a Source Data file.

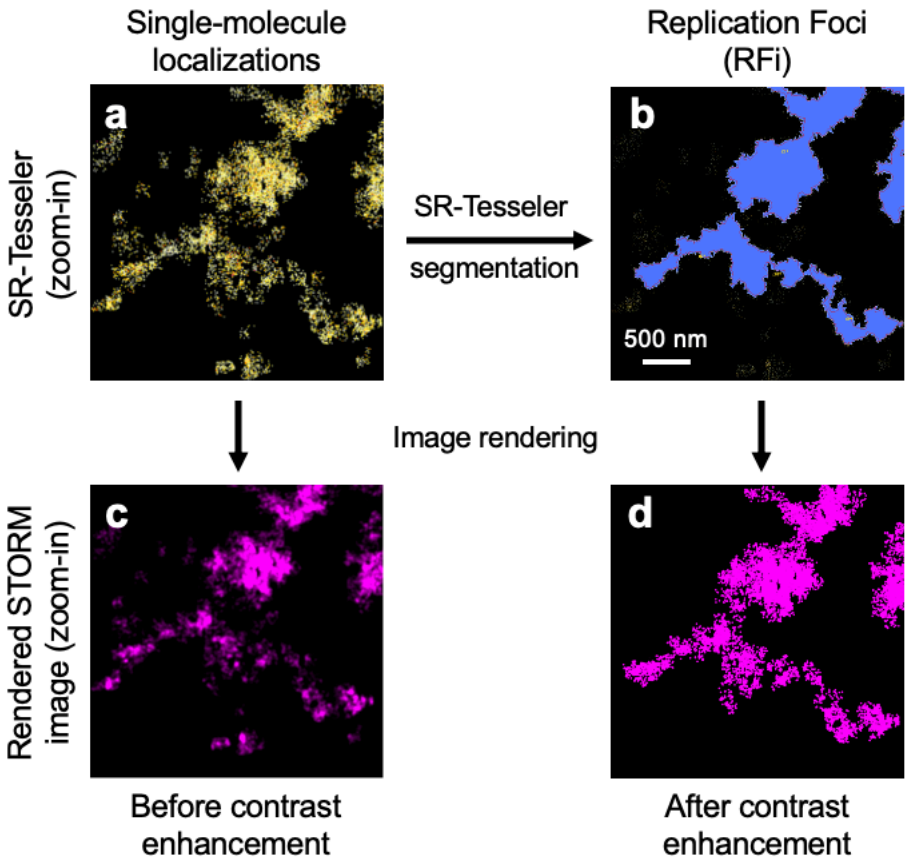

Fig. S5 | Comparison of images of RFi before and after contrast enhancement. The raw localizations (**a**) are first subjected to SR-Tesseler segmentation to identify the RFi (**b**); each is then rendered into STORM image with or without contrast enhancement (**c** and **d**). The remarkable similarity in overall shape, contour and RFi identification between the un-enhanced image from SR-Tesseler segmentation (**b**, same as shown in **Fig. 1b**) and the corresponding rendered image after contrast enhancement (**d**) demonstrates that our contrast enhancement does not yield image artefacts that will skew subsequent analysis and interpretation of our data.

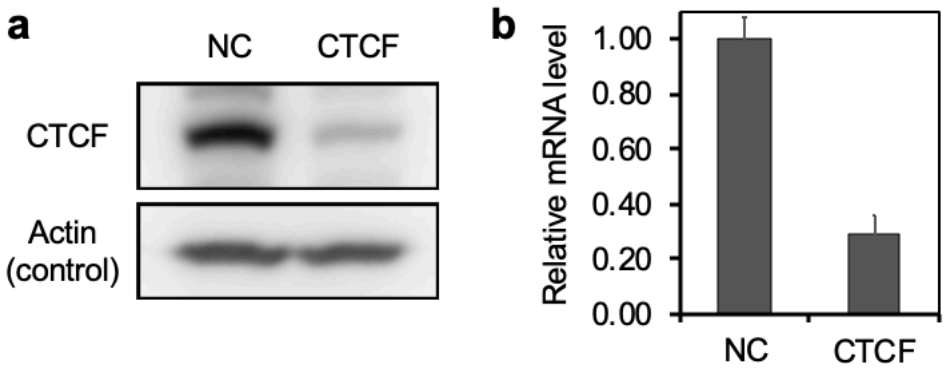

Figure S6 | Quantitative validation of CTCF knockdown. Western blot (**a**) and qPCR (**b**) quantification of CTCF expression at the protein and mRNA levels in cells treated with either NC siRNA or CTCF siRNA are shown. Both measurements were performed in triplicates, and error bars denote mean ± s.d.

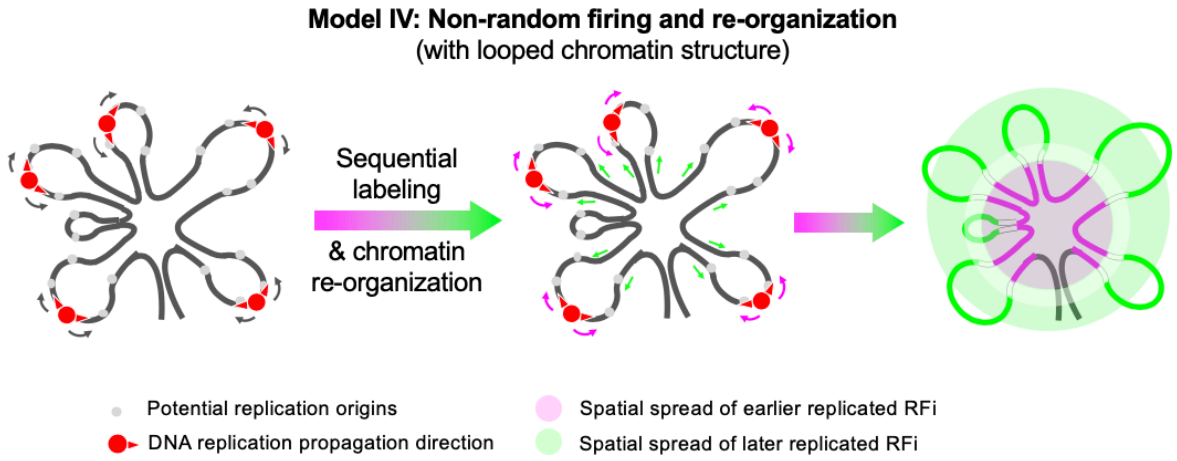

Figure S7 | Non-random firing and chromatin re-organization model for replication activation and propagation. DNA replication could activate at the periphery of RDs, and chromatin structural re-organization then drives the DNA that are replicated earlier towards the interior of the RD and those replicated later towards the exterior, thereby generating a propagation pattern similar to that of the "CoREP" model.

Simulation of the spatio-temporal spread of RFi

To gain further insights into the spatio-temporal propagation pattern of DNA replication observed (as shown in **Fig. 4**), we simulated the spatio-temporal spread of replicated chromatin in 3D using different models of replication activation. Briefly, we first used coarse-grained models for nucleosome and linker DNA to sample short chromatin chains, and then used them to construct long chromatin chains with internal loops to simulate chromatin structures organized by CTCF and cohesin. Such long chromatin chain was then split into two segments to represent DNA replicated during the two consecutive labeling time windows, and their radii of gyration were calculated for both the "CoREP" and random firing models.

1. Coarse-grained models of short chromatin chains

In our simulation, we first decomposed chromatin into nucleosomes and linker DNA chains, which are respectively described by Four-Sphere (FS) and Worm-Like-Chain (WLC) models (shown in **Fig. S8**), and then used these coarse-grained models as building blocks to construct short chromatin chains. In the FS model, one nucleosome is coarse-grained as the coalescence of four identical spheres with a diameter of 5.5 nm; centers of four spheres are aligned at four vertices of a square with a diagonal length of 5.5 nm. With these settings, the FS model roughly depicts a cylinder-shaped object with a diameter of 11 nm and a thickness of 5.5 nm. Terminal segments of linker DNA chains are placed tangent to the side surface and parallel to the top surface of the cylinder, with the two contact points positioned at 1/4 and 3/4 along the cylinder height, respectively. The angle formed by the two contact points and the central axis is 120°, roughly consistent with the wrapping of 1.67 turns of DNA in a nucleosome.

For the linker DNA, we adopted the WLC model by coarse-graining DNA chains into a sequence of DNA beads with a diameter of 2 nm, in which each bead roughly corresponds to 6 bp of DNA. The DNA beads, sharing similar diameter with the double-helical structure of DNA, ensures similar steric effect in the DNA chain. To restrict the bending of DNA chains, potential energy is added in the form of $E = \sum_i g\,(\pi - \theta_i)^2/2$, in which θ_i is the i^{th} bending angle between three neighboring beads along the chain, and g denotes the harmonic spring constant that generates the flexibility of the chain. In our system, the value of g is set as 60.68 kJ/(mol · rad^2) to give DNA a persistence length of 50 nm at 300 K. The value of the bending angle is set as π at the end of the DNA chain to mimic the restriction imposed by linker histone. For bending angles located at other positions, their values are chosen according to Boltzmann distribution. In our model, we assume each linker DNA chain initially contains 8 DNA beads (or 42 bp) throughout the chromatin, and increase in linker DNA length comes from exposure of unwrapped DNA upon the removal of its neighboring nucleosome. Thus, the length of linker DNA can be expressed as 8+33n with the probability $P(n) = (1-\alpha)\alpha^n$, where n is the number of nucleosomes that are removed in a row and α is the removal percentage of nucleosomes in the system.

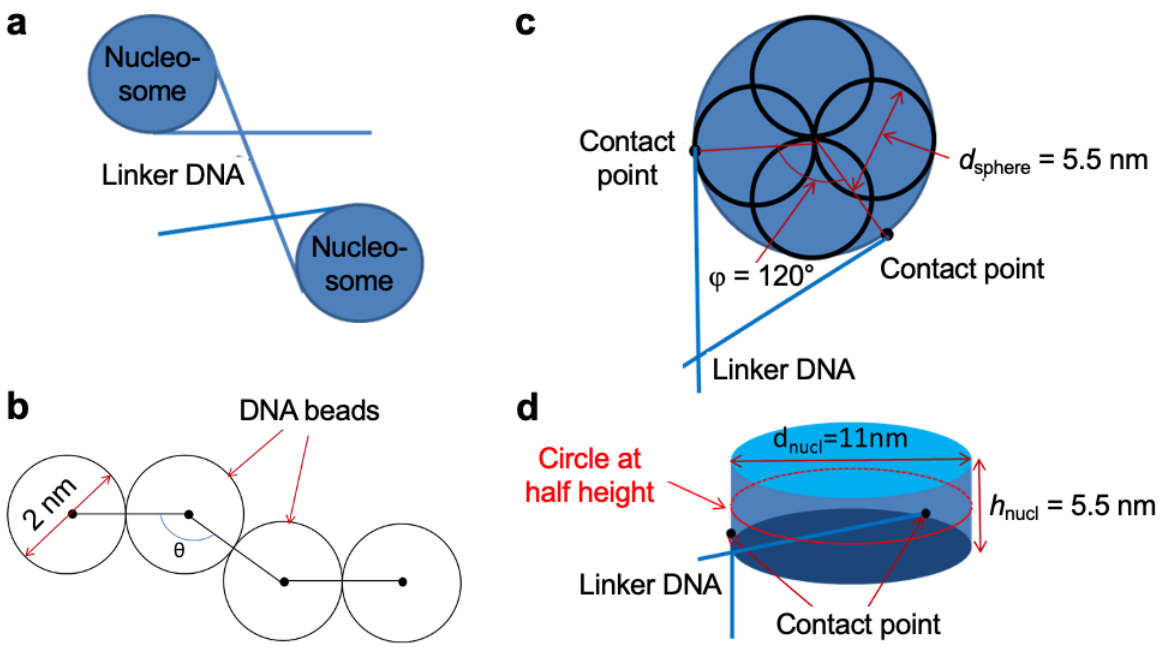

Figure S8 | Coarse-grained model of chromatin. a, Schematic illustration of the decomposition of chromatin into nucleosomes and linker DNA that connects them. **b,** Worm-Like-Chain model for linker DNA. **c,** Top view of the Four-Sphere model, in which the four spheres are partially overlapping in space. The blue circle represents the projection of the cylinder-shaped profile, and linker DNA is tangent to the side surface of the cylinder at contact point. **d,** Side view of the cylinder-shaped profile for the Four-Sphere model, where contact points of the linker DNA are placed at different heights of the cylinder.

Based on the coarse-grained models, the sampling of chromatin chains was conducted by iteratively adding nucleosomes and linker DNA through a self-avoiding random walk, which was performed in three steps: (1) one nucleosome with two neighboring linker DNA chains is randomly placed in space as one terminus of the chromatin chain; (2) another nucleosome together with its linker DNA is added to the existing chromatin chain at the end of the previous linker DNA chain with a random orientation, while avoiding overlapping between structural elements; (3) the second step is repeated until a desired length of chromatin chain is obtained. In our simulation, 10 sets of short chromatin chains are simulated with nucleosome removal percentage α varying from 0.05 to 0.5 at an interval of 0.05. In each set, 1000 chromatin chains of 30 kb long and 1000 chromatin chains of 1 kb long are generated as the basic components for longer chromatin chains.

2. Construction of long chromatin chains with internal loops

To generate target chromatin chains with internal loops, the short chromatin chains are first combined to form loop-like chromatins which are further connected to construct long chromatin chains (**Fig. S9**). In the first step, six 30-kb chromatin chains are randomly picked out from 1000 samples and connected to form one 180-kb chromatin chain by linker DNA with the length of 41 DNA beads (the same length of DNA with one nucleosome removed). The 180-kb chromatin resembles the loop structure with an end-to-end distance of less than 50 nm. A total of 200 chromatin chains of 180 kb are generated in the first step. In the second step, five randomly selected 180-kb chromatins are further connected together by four 1-kb chromatin chains to construct a 900-kb chromatin chain through a self-avoiding random walk process. 500 chromatin chains of 900 kb were generated for each chromatin set.

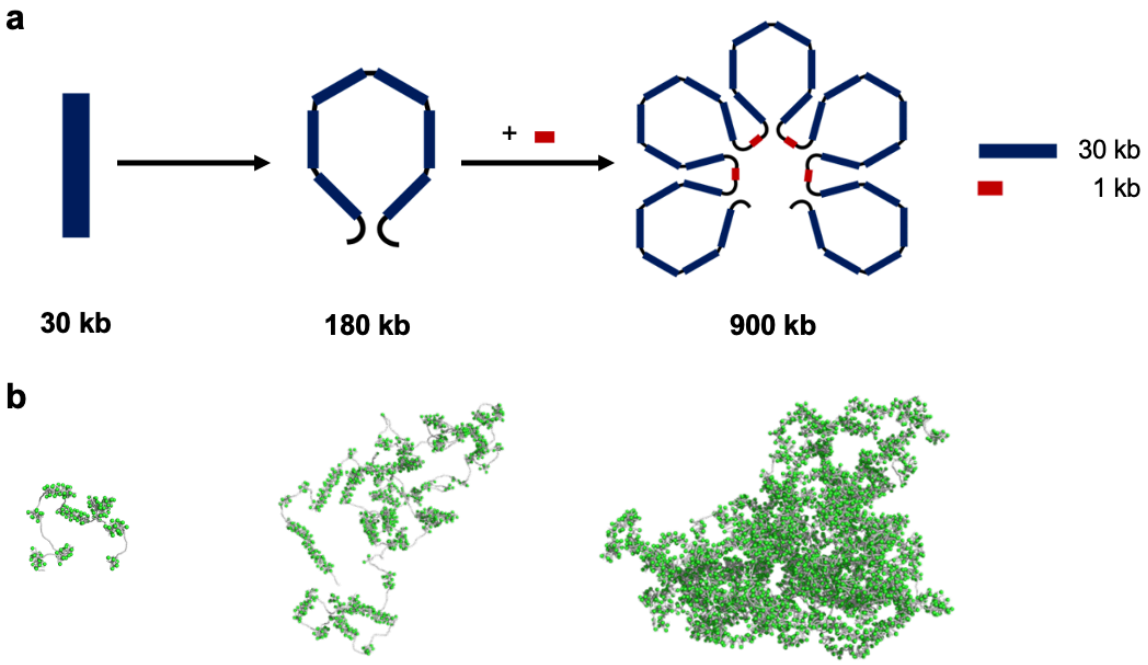

Figure S9 | Construction of long chromatin chains with internal loops from short chromatin chains. a, Schematic illustration of the construction process. **b,** Representative conformation of chromatin chains with lengths 30 kb, 180 kb, and 900 kb, respectively (not drawn to scale). Nucleosomes are shown as green spheres and DNA beads as gray threads.

3. Simulating the spatial spread of consecutively replicated chromatin

Starting from different sets of long chromatin chains with internal loops, we then simulated the spatial spread of chromatin produced during the two consecutive labeling time windows. In our simulation, we used both the "CoREP" model for replication activation as well as random firing

model. For each model, we calculated the radius of gyration of chromatin DNA replicated during the two labeling time windows (**Figs. 6e**). We found that for the "CoREP" model, the radius of gyration of later replicated chromatin is significantly larger than that of earlier replicated chromatin, in good agreement with the pattern observed in **Fig. 4**. On the other hand, for random activation model, the radii of gyration of chromatin in both time windows are similar.