Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/2/eaay4055/DC1

Supplementary Materials for

Physical and data structure of 3D genome

Kai Huang*, Yue Li, Anne R. Shim, Ranya K. A. Virk, Vasundhara Agrawal, Adam Eshein, Rikkert J. Nap, Luay M. Almassalha, Vadim Backman*, Igal Szleifer*

*Corresponding author. Email: khuang15@northwestern.edu (K.H.); v-backman@northwestern.edu (V.B.); igalsz@northwestern.edu (I.S.)

Published 10 January 2020, *Sci. Adv.* **6**, eaay4055 (2020)
DOI: 10.1126/sciadv.aay4055

This PDF file includes:

Section S1. Contact scaling of random fractal chain

It has been argued that chromatin is a crumpled globule (Ref. 25), latter termed fractal (self-similar) globule (FG), so that the long biopolymer could function without entangling and knotting. Computer simulation (Ref. 26) showed that FG has a contact probability (P_c) scaling factor s=-1, namely $P_c(L) \propto L^{-1}$, where L is contour distance between two loci on the polymer. Based on scale-invariance properties one can prove that for a random fractal chain, the P_c scaling factor s is inverse to the fractal dimension D of the chain

$$s = -\frac{E}{D} \tag{S1}$$

where E is the dimension of space where the fractal chain is embedded. Concerning models of chromatin structure, we have E=3 for three-dimensional space. As a special case, FG has a fractal dimension of D=3 and therefore s=-1 according to our theory, consistent with simulation result reported in literature (Ref. 26).

To test our scaling theory for more generic cases, we here carried out numerical calculations of the contact probability for random walks (RWs) and Lévy-flights (LFs) in both 2D and 3D space. The random walks have fixed fractal dimension of D=2. For LFs we used fractal dimensions 1.2, 1.5 and 1.8, respectively. For all the cases, we generated trajectories of 1 million steps. The RWs have a constant step size, and the LFs have power law distributions of their step size with the exponents determined by their fractal dimensions. We chose the fixed step size of RWs and the minimal step size of LFs to be 1 and employed a contact cutoff to be 3 for the scaling analyses. The asymptotic behaviors of all the cases tested are shown in the colored solid lines in fig. S1. The dash lines with their exponents predicted by Eq. S1 align well with the numerical results.

Since s and D determine the contact frequency and structural heterogeneity of the fractal chain, respectively, Eq. S1 means that it is impossible for a polymer to simultaneously satisfy (1) high contact frequency, (2) structural heterogeneity and (3) self-similarity. To understand this intuitively: a highly heterogeneous (small D) fractal chain such as the LF fills space with self-similar clusters of low inter-cluster contact frequencies, whereas a fractal chain of high contact frequencies, such as the FG, tends to occupy the space in a homogeneous way (large D) despite its compartmentalization along the polymer contour. Note that there is a subtle yet important difference between compartmentalization and clustering: clusters are spatially isolated compartments. In other words, compartmentalization is a weaker property than clustering, and without isolation it cannot render density

heterogeneity in space. The fact that interphase chromatin houses both frequent long-range contacts and heterogeneous local DNA densities thus leads us to conclude that its folding structure is not self-similar. It is also worth noting that since $D \leq E$, we have s=-1 as the upper bound for the scaling of fractal polymers. However, at small genomic range, it has been observed (Ref. 8) that chromatin has an abnormally slow decay of contact probability with s>-1, which cannot be explained by fractal chain models.

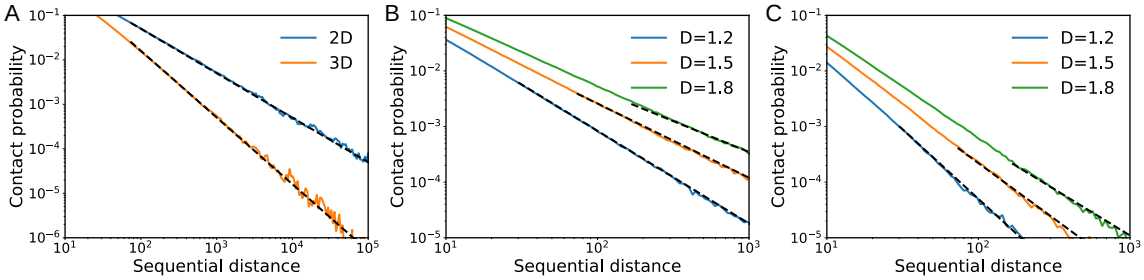

Fig. S1. Contact probabilities of RWs and LFs in 2D and 3D space. Numerical results are shown in solid lines and theoretical predictions in dash lines (**A**) RWs in 2D and 3D space. (**B**) LFs of varying fractal dimensions in 2D space. (**C**) LFs of varying fractal dimensions in 3D space.

Section S2. Granular and porous packing of SRRW

Here we highlight the heterogeneous packing of our model by comparing its typical cross-section with that of a confined RW with similar average density.

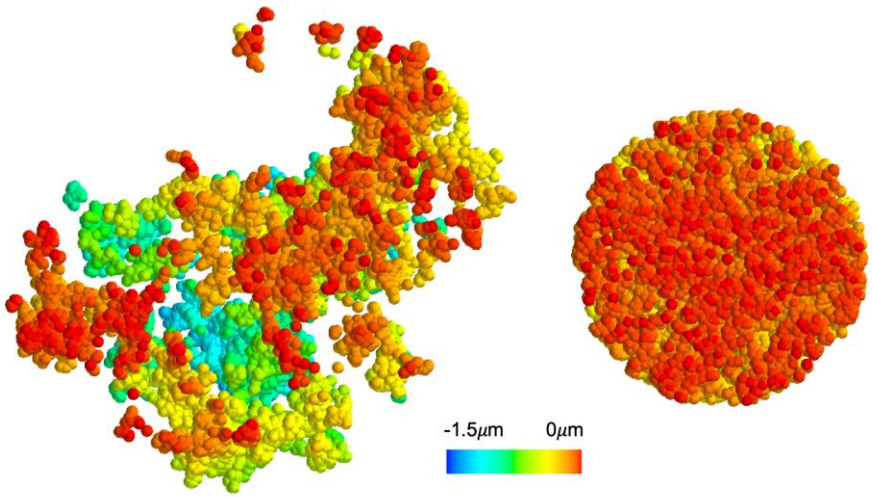

Fig. S2. Comparing packing cross sections of confined SRRW and RW. SRRW is shown on the left and RW on the right. The monomers are colored according to their distance to the cross-section.

Section S3. Tree domains and TADs

A typical 3D structure of a 3Mb segment of SRRW (1500 steps) is shown in fig. S3A. The local segment is clearly folded into a bundle of clusters at nano-scale. Allowing such conformation to relax with preserved topology offers a conformational set (1000 conformations) on which we can obtain a population-level contact map featuring TAD-like domains as shown in fig. S3B. Such 2D contact patterns are the statistical consequence of hierarchical 3D structures linked and isolated by stretched DNA. In this example, the 3Mb contact map divides itself into two packing domains isolated from each other by extended backbone segments. Inside these packing domains, similar genomic isolations recur at smaller scales, leading to a hierarchical organization of domains. The population-average TADs have been explained by the loop extrusion hypothesis (Ref. 9), in which large loops are formed by extrusion machinery (likely realized by CTCF and cohesin). However, at the single-cell level, loop extrusion does not produce clusters that are found to be ubiquitous under super-resolution microscopy even with cohesin knocked out (Ref. 10). In contrast, SRRW naturally assumes a clustered morphology at single-cell level.

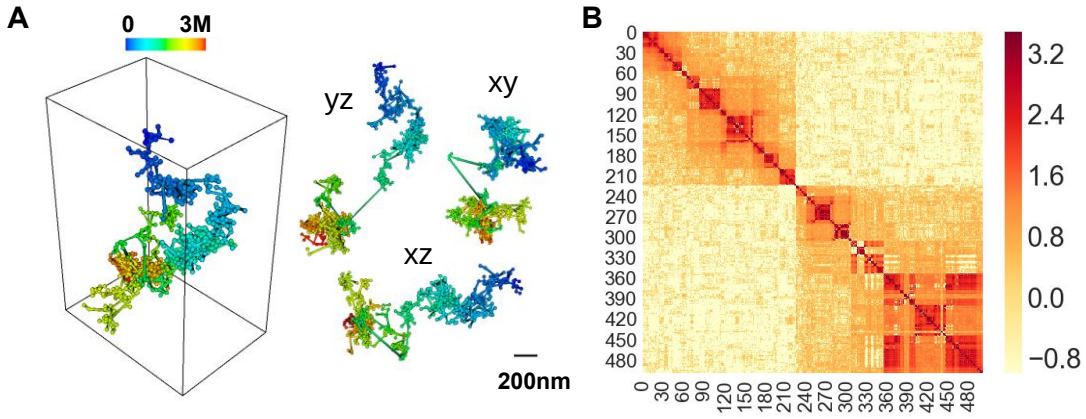

Fig. S3. Structure of local SRRW segment and ensemble-averaged contact map. (**A**) 3D structure of a 3Mb segment of SRRW and its xyz projections. (**B**) Contact map of the 3Mb SRRW segment over 1000 random conformations with preserved topology. Color bar in logarithm scale.

It is worth noting that tree domains are similar but not equivalent to Hi-C TADs: while Hi-C TADs are contact patterns that emerge over the sampling of millions of cells, trees domains predicted by our model are complexes of loops and clusters at the single-cell level. Due to the loss of information from 3D structures to 2D contact maps, qualitatively different 3D structural predictions can result in similar 2D contact patterns. Since Hi-C is sensitive to both topological contacts (contacts

underpinning topological constraints) and natural contacts (transient contacts with less functional significance), it is hard to identify specific folding modes based on only ensemble-averaged contact maps. For example, a 1Mb-level TAD pattern could be possibly resulted from either a large tree-like topological domain or from a small non-topological compartment. Therefore, it is important for modeling efforts to account for both Hi-C and imaging observations.

Recent experimental evidence shows that contact loops are dynamic (Ref. 38), meaning they can form and break, allowing rapid epigenetic reconfiguration. Given this dynamic picture, we posit that, like chemical reaction, small tree domains can group into large tree domains, which can in reverse ungroup into small ones. The contact map in fig. S2B is based on a frozen topology without any regrouping of tree domains and therefore could have overestimated the isolation across domains, which renders an over-hierarchical contact map at population-level. Nevertheless, as a proof of concept, it clearly demonstrated how single-cell level tree domains could lead to TAD patterns at population-level.

Section S4. Non-Gaussian statistics

Heterogeneity and non-Gaussian statistics are hallmarks of living systems. The intrinsic heterogeneity of SRRW is tightly associated with its non-Gaussian statistics. For example, the end-to-end distance (R) distribution of a sample DNA segment of 100kb predicted by SRRW is significantly deviated from normal, as shown in green in fig. S3, with high probability of small R due to self-returning and a heavy tail of stretched DNA with large R. Compared to the counterpart from a RW, the 100kb segment of SRRW is much shorter on average, signifying strong compaction. The existence of the long tail allows the compacted domains to be spatially separated.

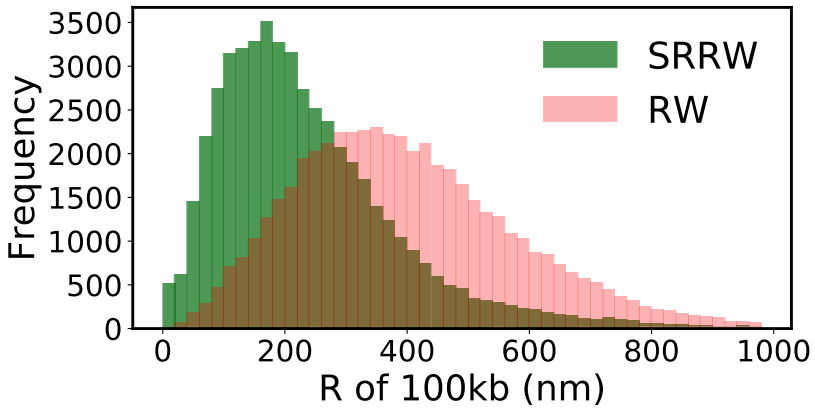

Fig. S4. End-to-end distance distributions of 100-kb SRRW segment (green) and of 100-kb RW segment (red).

Section S5. Effect of α on the statistics of tree domains

In the SRRW model, α governs the global folding architecture of the modeled chromatin and impacts the heterogeneity of the local DNA packing, which has been demonstrated and discussed in the main text. Here we show more numerical results on the effect of α. As shown in fig. S5, the distribution of the physical size of tree domains and the distribution of the genomic size of the tree nodes are both sensitive to α. In fig. S4A, the physical size of the tree domain is characterized by the radius of gyration (R_g) and normalized by the genomic size of the domain. Larger tree domains are more hierarchical in structure and contain more high-order contacts.

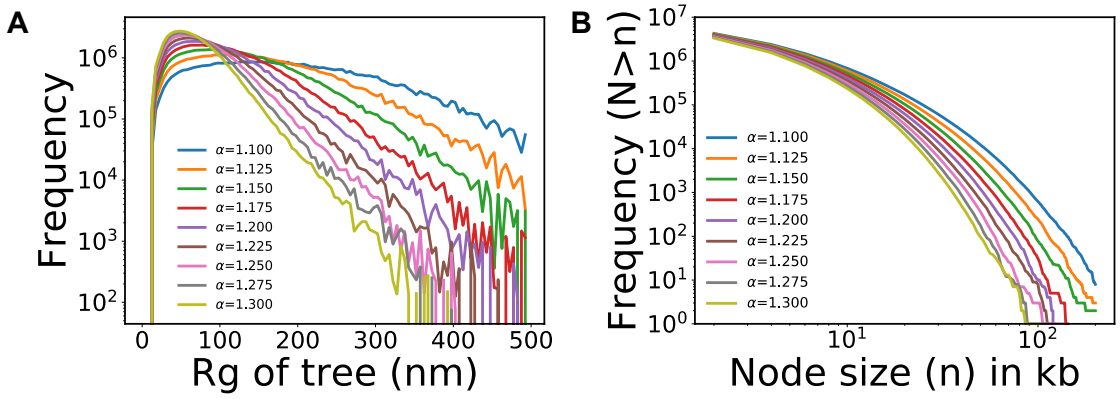

Fig. S5. Effect of α on the tree domains. (**A**) R_g distribution of tree domains normalized by their genomic sizes. (**B**) Genomic size distribution of the tree nodes.

As discussed in the main text, there is an anti-correlation between the effective mass scaling D and the contact scaling s. Figure S6 shows the data sets from which we extracted D and s.

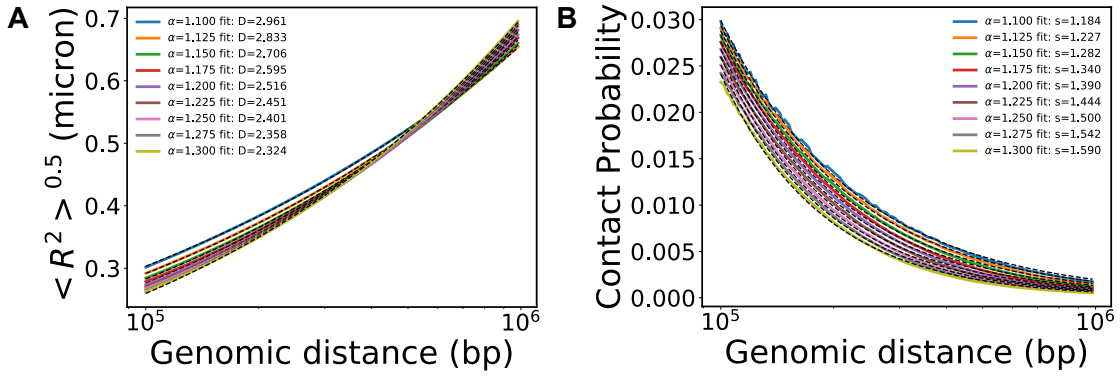

Fig. S6. *D* and *s* from the modeled chromatin at varying α. (**A**) End-to-end distance scalings and their power-law fittings from 100kb to 1000kb. (**B**) Contact probability scalings and fittings for the same genomic range.

Section S6. Chromatin scanning transmission electron microscopy

1. EM sample preparation

Cells were rinsed in Hanks's balanced salt solution without calcium and magnesium then fixed with 2.5% EM grade glutaraldehyde (EMS) in 5mM $CaCl_2$, 0.1M sodium cacodylate buffer, pH 7.4 for 5 min at room temperature. Then fresh fixative was used to continue fixing the cells on ice for an hour. From this step, the cells were kept cold on ice or on a cold stage with temperature ranges from 4°C to 10°C. The ChromEM staining and embedding follows published protocol (Ref. 3). 100 nm A549 cell sections and 50 nm BJ cell sections were made by ultramicrotomy (Leica, UC7) and mounted on copper mesh grids with Formvar/carbon film (EMS).

2. EM imaging and tomography reconstruction

Solution of 10 nm colloidal gold fiducial markers was deposited onto both slide of the sample prior to imaging. A 200 kV STEM (HD2300, HITACHI) was employed to collect the dual-tilt series in high angle annular dark field (HAADF) mode. During each tilt series, the sample was tilted from -60° to 60° with a step size of 2°. Between the two-tilt series, the sample was taken out of the microscope and rotated around 90° manually. A low dose imaging condition was conducted to minimize beam damage. The tilt series were aligned in IMOD (Ref. 57) using fiducial markers. After alignment, each tilt series were reconstructed separately using penalized maximum likelihood algorithm in Tomopy (Ref. 58). IMOD was used again to combine the two reconstructed tomograms to minimize artifacts from missing wedge. The BJ cell sections were imaged by a TEM (HT7700, HITACHI) operated at 80 kV. The brightfield micrographs were recorded with pixel resolution of 5 nm for the nucleus region for 20 cells.

3. Chromatin volume concentration (CVC) calculation and nanocluster size analysis

The automated chromatin segmentation was conducted using CLAHE (100 nm) – Li in FIJI as previously described. We used the same definition of the chromatin volume concentration (CVC) as Ou's work but with a moving window summation with a three-dimensional window size of 99 nm and a stride of 3 nm. From the CVC map, we observed nanoclusters with similar CVC values, and measured the full width half maximum (FWHM) of the line profile of CVC as the cluster diameter. 29 clusters were selected, and the histogram of the cluster size was plotted.

4. Mass scaling analysis with TEM images

The 2D-mass scaling was calculated for each tomogram slice. Starting with a random position around the center of chromatin a circle was drawn with radius r and the total amount of chromatin (M) encapsulated by the circle can be calculated. The

mass-scaling is the relationship between the radius r and the total amount of chromatin (M) encapsulated. The mass scaling M(r) follows a power law and for fractal structure in specific, the power is smaller than 2 for the 2D slice and smaller than 3 for the 3D case.

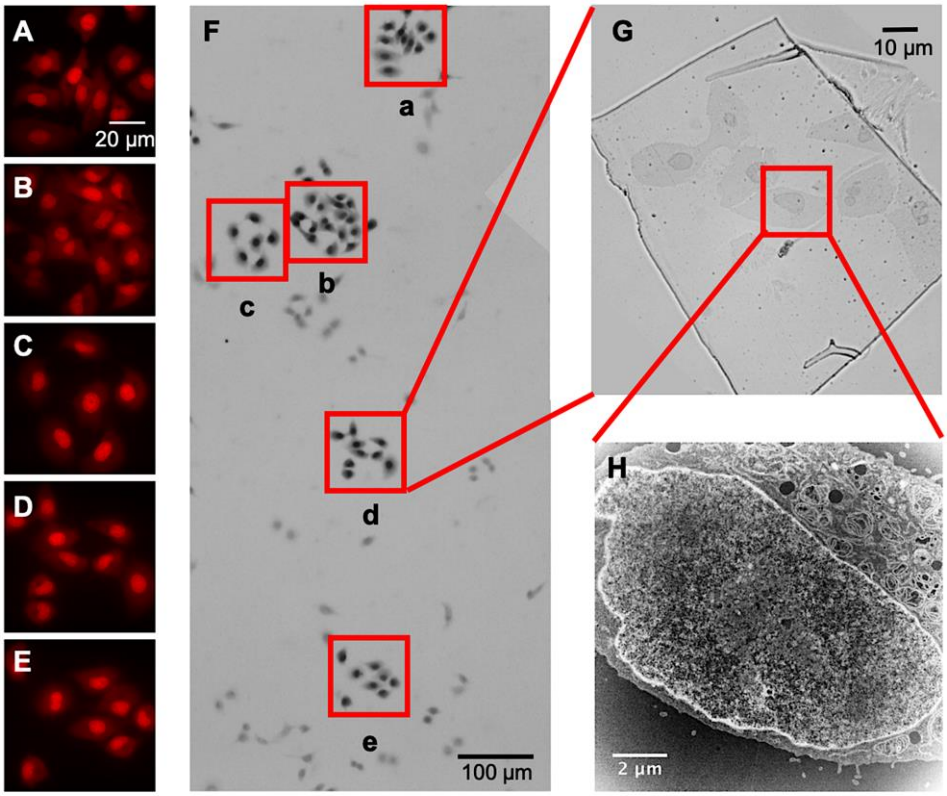

Fig. S7. ChromEM sample preparation and STEM imaging on A549 cells. (**A**) to (**E**), fluorescence image of A549 cells with DRAQ5 label during photo-bleaching. (**F**) Optical bright field image of A549 cells after resin embedding. Spots after photo-bleaching (a to e) showed darker contrast compare to neighboring cells without photo-bleaching. (**G**) Optical bright field image of one thin section of A549 cells prepared by ChromEM protocol. (**H**) STEM HAADF imaging of 100 nm ultra-thin resin section of one A549 cell.

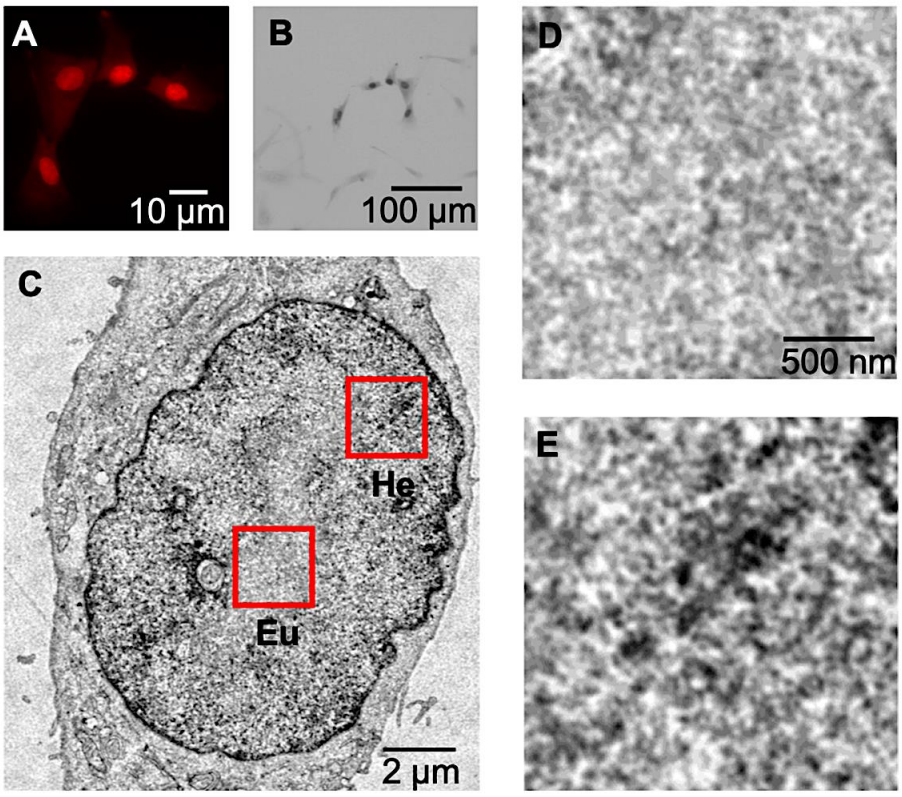

Fig. S8. ChromEM sample preparation and TEM imaging on BJ cells. (**A**) Fluorescence image of BJ cells labeled with DRAQ5 during photo-bleaching. (**B**) Optical bright field image after resin embedding for corresponding cells in (A). Compared to neighboring cells, the cells with photo-bleaching showed significantly darker contrast. (**C**) TEM image of a BJ cell with ChromEM preparation. Euchromatin (Eu) and heterochromatin (He) were identified based on the DNA density. (**D**) Magnified image of the euchromatin (red square, Eu) in (C). (**E**) Magnified image of the heterochromatin (red square, He) in (C). The euchromatin showed lighter contrast compared to the heterochromatin.

Section S7. Heat shock experiments across multiple cell lines

Using our heat shock protocol described in our Materials and Method section, we performed PWS experiments on two different cancer cell lines (colon cancer HCT116 cells and lung adenocarcinoma A549 cells) as well as differentiated BJ fibroblast cells. Our results demonstrate that the effect of heat shock on fractal dimension of chromatin is cell line independent.

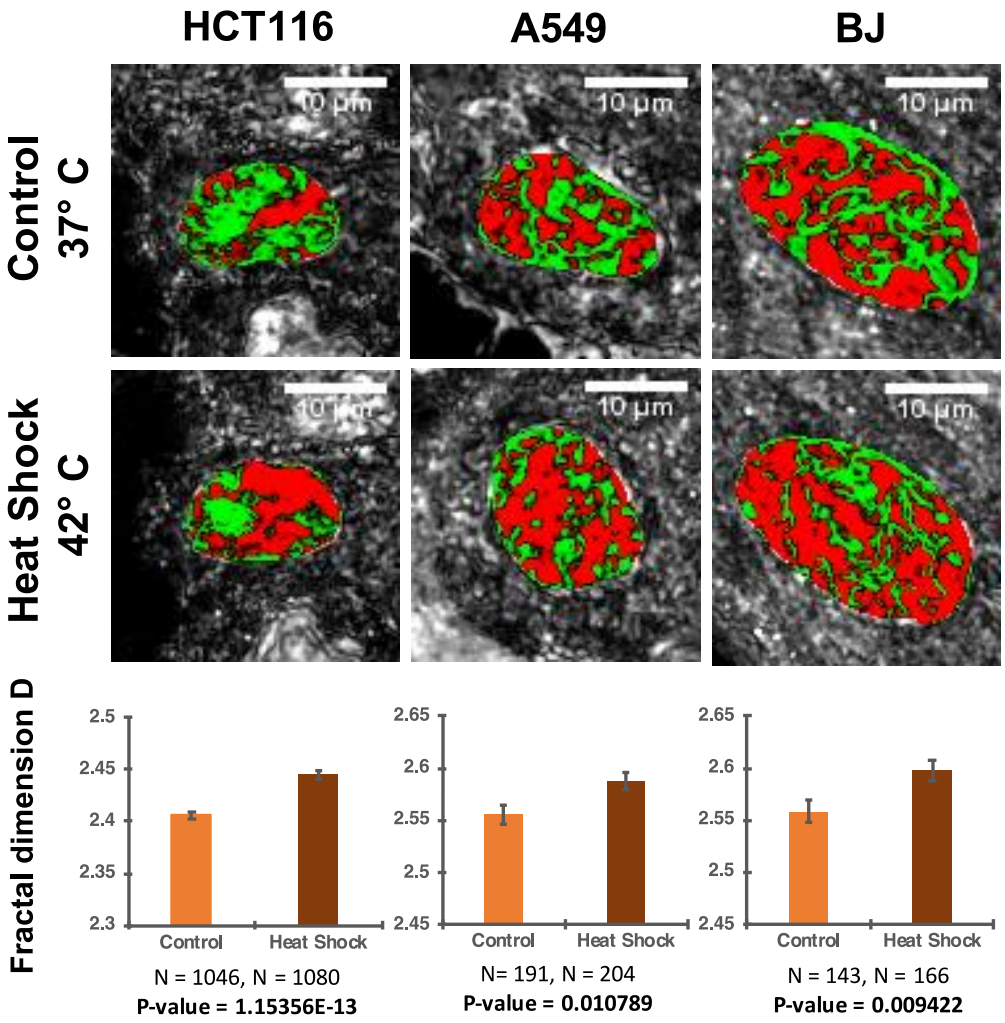

Fig. S9. Heat shock increases fractal dimension *D* across multiple cell lines. The first two row show the control group and heat shock group of HCT116, A549 and BJ cell lines, respectively. The analyzed chromatin fractal dimension *D* values for all the three cell lines are shown in the third row.

Section S8. Histone modification affects fractal dimension *D*

We performed PWS experiments on two different ovarian cancer cell lines with valproic acid (VPA) treatment. Since VPA increases acetylation level throughout the nucleus, we would expect more open/decondensed chromatin segments. With respect to the SRRW model, this would correspond to a larger alpha value and a larger portion of chromatin backbone. Our PWS experiments demonstrate a decrease in fractal dimension upon VPA treatment, which is in line with our prediction based on the model.

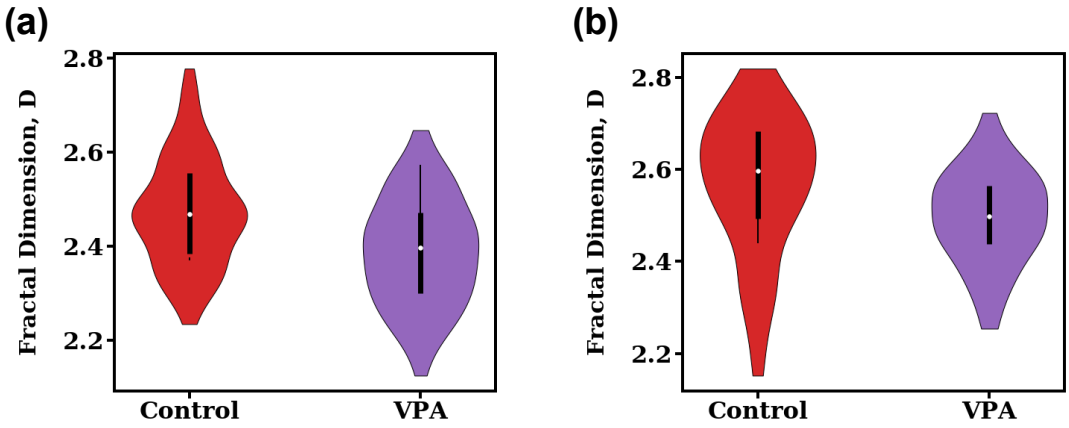

Fig. S10. Histone deacetylase inhibitor valproic acid decreases fractal dimension *D*. Quantification of change to fractal dimension in A2780 (**A**) and M248 (**B**) cells treated with 100 μM valproic acid, which decreases within 30 minutes of treatment in cells treated with VPA ($p=4.7x10^{-13}$ for A2780, $5.7x10^{-7}$ for M248). Significance was determined using Student's t-test with unpaired, unequal variance on the average nuclear *D* between the conditions.