## SUMPLEMENTARY INFORMATION

### Overview of the heterogeneous loop model

*The expression for the mean square gyration radius of subchain* : Chromatin fiber was first regarded as a Gaussian polymer network with harmonic looping-restraints in heterogeneous loop model (HLM) (1). Given the energy potential of a chain composed of $N$ monomers,

$$U_{\mathcal{K}}(\mathbf{r}) = \sum_{i=0}^{N-1}\sum_{j=i+1}^{N-1}\frac{k_{ij}}{2}(\vec{r}_i-\vec{r}_j)^2 = \frac{3}{2}\mathbf{r}^T\mathbf{K}\mathbf{r}, \tag{S1}$$

where $\mathbf{r}=(\vec{r}_1,\vec{r}_2,\cdots,\vec{r}_{N-1})^T$ and $\mathbf{K}$ is the Kirchhoff matrix

$$\mathbf{K} = \begin{pmatrix} \sum_{j=0,j\neq 1}^{N-1} k_{1j} & -k_{12} & \cdots & -k_{1,N-1} \\ -k_{21} & \sum_{j=0,j\neq 2}^{N-1} k_{2j} & \cdots & -k_{2,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ -k_{N-1,1} & -k_{N-1,2} & \cdots & \sum_{j=0,j\neq N-1}^{N-1} k_{N-1,j} \end{pmatrix}, \tag{S2}$$

the probability density of the pairwise distance between the $i$-th and $j$-th monomers in 3D is (2)

$$P(r_{ij};\gamma_{ij}) = 4\gamma_{ij}^{3/2}/\sqrt{\pi}r_{ij}^2 e^{-\gamma_{ij}r_{ij}^2}, \tag{S3}$$

in which

$$\gamma_{ij} = \begin{cases} \frac{1}{2(\sigma_{ii}+\sigma_{jj}-2\sigma_{ij})}, & i>0 \\ \frac{1}{2\sigma_{jj}}, & i=0 \end{cases}, \tag{S4}$$

and $\sigma_{ij}(=(\Sigma)_{ij})$ is the elements of inverse matrix $\Sigma=\mathbf{K}^{-1}$. From Equation S3, the mean pairwise distance is (2)

$$\langle r_{ij}\rangle = \int_0^\infty r_{ij}P(r_{ij})dr_{ij} = \frac{2}{\sqrt{\pi}\gamma_{ij}^{1/2}}, \tag{S5}$$

and the mean square distance is

$$\langle r_{ij}^2\rangle = \int_0^\infty r_{ij}^2P(r_{ij})dr_{ij} = \frac{3}{2\gamma_{ij}}. \tag{S6}$$

The mean square gyration radius of a subchain between the $i$-th and $j$-th monomers ($i<j$), is

$$\begin{aligned}\langle r_{g,ij}^2\rangle &= 3\langle x_{g,ij}^2\rangle \\ &= 3\Big[\Big\langle\frac{1}{j-i+1}\sum_{p=i}^{j}x_p^2\Big\rangle - \Big\langle\Big(\frac{1}{j-i+1}\sum_{p=i}^{j}x_p\Big)^2\Big\rangle\Big] \\ &= 3\Big[\frac{1}{j-i+1}\sum_{p=i}^{j}\langle x_p^2\rangle - \frac{1}{(j-i+1)^2}\sum_{p=i}^{j}\sum_{q=i}^{j}\langle x_px_q\rangle\Big],\end{aligned} \tag{S7}$$

where $\langle x_px_q\rangle=\sigma_{pq}$ if $pq\neq 0$; otherwise, $\langle x_px_q\rangle=0$.

To validate the analytic result in Equation S7, we performed molecular dynamics (MD) simulations using three toy models ($N=20$) characterized with different intra-chain loops (i.e., different $\mathcal{K}$-matrix with elements $k_{ij}$): (i) a chain with a single loop (Figure S1A); (ii) a chain with two nested loops (Figure S1B); (iii) a chain composed of two blocks of monomers without any inter-block attraction (Figure S1C). The numeric estimates of $\langle r_{g,ij}^2\rangle$ (upper diagonal parts of Figure S1D-F), based on conformational ensembles from MD simulations, are close to the analytic results (lower diagonal parts), with a relative error of 0.011, 0.013 and 0.019 in three cases, respectively.

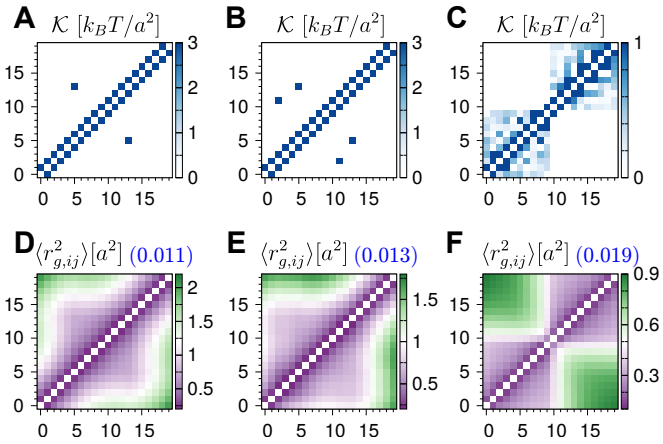


**Figure S1.** $\mathcal{K}$-matrix for three models of a gaussian polymer chain consisting of 20 monomers, each of which is characterized with (A) a single loop, (B) two nested loops, and (C) two blocks of monomers without inter-block loops. (D-F) The corresponding heatmap of the mean square gyration radius of a subchain between the $i$ and $j$-th monomers, $\langle r_{g,ij}^2\rangle$ for three different model. The upper and lower diagonal parts are from MD simulations and from Equation S7, respectively. The relative errors between the results from MD and Equation S7 are given inside the parenthesis on top.

*Heterogeneous loop model that integrates the constraints from Hi-C and FISH* : We have exploited the *one-to-one* analytic mapping between pairwise interaction strength

$k_{ij}$ and contact probability $p_{ij}$, as indicated by Eqs. S2-S4, to infer the 3D conformations of chromatin from Hi-C data (namely, a contact probability matrix $\mathcal{P}$ of elements $p_{ij}$). In short, the values of $2N$ selected elements in $\mathcal{K}$-matrix, $\{k_{ij}\}$, are iteratively optimized, which minimizes an objective function $\mathcal{F}_{\text{Hi-C}}(\mathcal{K})$

$$\mathcal{F}_{\text{Hi-C}}(\mathcal{K}) = \sum_{(i,j)}^{2N} \omega_{ij} \left( \frac{\langle r_{ij}^2(\{k_{\alpha\beta}\})\rangle}{\langle r_{ij}^{2*}\rangle} - 1 \right)^2, \tag{S8}$$

where $\langle r_{ij}^{2*}\rangle$ denotes the *reference* value of mean pairwise distance obtained from Hi-C, and $\langle r_{ij}^2(\{k_{\alpha\beta}\})\rangle$ is evaluated based on Equation S6. The weight factor $\omega_{ij}$, which is used to normalize the statistical bias from chromatin loops of different sizes, is defined as

$$\omega_{ij} = \omega(|i-j|) = \omega(s) = \frac{n^{-1}(s)}{\sum_s n^{-1}(s)}, \tag{S9}$$

where $n(s) = \sum_{(i,j)} \delta(|i-j|-s)$ is the number of loops of size $s$.

In order to characterize the intra-domain chromatin chain organization, Boettiger *et al.* have measured the mean gyration radii of subchains, $\langle r_{g,ij}^*\rangle$, for different types of domains (3). Thus, a corresponding objective function for FISH, $\mathcal{F}_{\text{FISH}}(\mathcal{K})$, can be defined as

$$\mathcal{F}_{\text{FISH}}(\mathcal{K}) = \sum_{(i,j)} \omega_{ij} \left( \frac{\langle r_{g,ij}^2(\{k_{\alpha\beta}\})\rangle}{\langle r_{g,ij}^{2*}\rangle} - 1 \right)^2, \tag{S10}$$

where the *reference* value $\langle r_{g,ij}^{2*}\rangle$ is approximated as $\langle r_{g,ij}^*\rangle^2$, $\langle r_{g,ij}^2(\{k_{\alpha\beta}\})\rangle$ and $\omega_{ij}$ are determined from Equation S7 and S9, respectively.

To incorporate the structural information from both Hi-C and FISH, our final objective function to determine the optimal value of $\mathcal{K}$ is

$$\mathcal{F}(\mathcal{K}; f) = (1-f) \times \mathcal{F}_{\text{Hi-C}}(\mathcal{K}) + f \times \mathcal{F}_{\text{FISH}}(\mathcal{K}), \tag{S11}$$

where the parameter $f$ tunes the relative weight of restraints from two experimental approaches.

After obtaining $\mathcal{K}$ at $f$, and hence $U_{\mathcal{K}(f)}(\mathbf{r})$, we sampled 3D structures using MD simulations with the full energy potential $U_{\text{HLM}}(\mathbf{r})$ for HLM,

$$U_{\text{HLM}}(\mathbf{r}) = U_{\mathcal{K}(f)}(\mathbf{r}) + (1-f) \times U_{\text{nb}}(\mathbf{r}). \tag{S12}$$

Note the excluded volume interaction is now considered in the non-bonded interaction term $U_{\text{nb}}(\mathbf{r})$ defined for all $i$ and $j$ monomer pairs,

$$\begin{aligned} U_{\text{nb}}(\mathbf{r}) &= \sum_{ij} \chi_{t_i,t_j} u_{\text{LJ}}(r_{ij}) \\ &= \sum_{ij} \chi_{t_i,t_j} \times \epsilon \left[ \left( \frac{a}{r_{ij}} \right)^{12} - 2 \left( \frac{a}{r_{ij}} \right)^{6} \right] \Theta(5a/2 - r_{ij}), \end{aligned} \tag{S13}$$

where $u_{\text{LJ}}(r_{ij})$ is the Lennard-Jones potential truncated for $r_{ij} \geq 5a/2$ with $\epsilon = 0.45\ k_B T$, and $\chi_{t_i,t_j}$ is a loci-pair-type-dependent prefactor which is evaluated based on Hi-C data. More details can be found in Ref. (1).

*Molecular dynamics simulations* : To generate a 3D conformational ensemble of chromatin by using HLM, we performed the low-friction Langevin simulations by integrating the following equation of motion (1, 4),

$$m \frac{d^2\mathbf{r}}{dt^2} = -\zeta_{\text{MD}} \frac{d\mathbf{r}}{dt} - \vec{\nabla}_{\vec{r}_i} U_{\text{HLM}}(\mathbf{r}) + \tilde{\xi}(\mathbf{t}), \tag{S14}$$

The friction coefficient $\zeta_{\text{MD}}$ and the integration time step $\delta t$ were chosen to be $1.0m/\tau_{\text{MD}}$ and $0.01\tau_{\text{MD}}$, respectively, with the characteristic time scale $\tau_{\text{MD}} = (ma^2/\epsilon)^{1/2}$. The whole simulation was carried out in three steps. (i) First, a Gaussian chain of only backbone connectivity was equilibrated for 500 $\tau_{\text{MD}}$ with the energy term $U_{\mathcal{K}}(\mathbf{r})$ ($k_{i,i+1} = 3\ k_B T/a^2$; otherwise $k_{i,j} = 0$). Due to the lack of the nonbonded interaction term $U_{\text{nb}}(\mathbf{r})$, there is no excluded volume interaction at this stage. (ii) The chain was simulated under the full HLM potential $U_{\text{HLM}}(\mathbf{r})$ for 100 $\tau_{\text{MD}}$ but with extra care. Excessive overlaps between monomers generated from the foregoing stage, were eliminated by gradually increasing the short-range repulsion. This was achieved by using a capped LJ potential term $u_{\text{LJ}}(r_{ij}) = \min\{u_c, u_{\text{LJ}}(r_{ij})\}$ with gradually increasing $u_c$. (iii) The production run was generated for $5 \times 10^5 \tau_{\text{MD}}$, except the longest genomic region on chr2L with a span over 3 Mb ($N = 610$, Figure S6). For the latter, 10 independent replicas were generated staring from the first step, which in total produced a sampling of $5 \times 10^6 \tau_{\text{MD}}$. Chain configurations were collected every 50 $\tau_{\text{MD}}$ for analysis. We carried out all simulations by using ESPResSo 3.3.1 package (5).

**Other classification of chromatin state**

In the classification of five (6) (or nine (7)) chromatin states in *Drosophila* $\text{Kc}_{167}$ (or S2R+) cells, most regions in I-type domains are BLACK chromatin (or state 9), and R-type domains are dominated by BLUE chromatin (or state 6) (Figure S2B). Similarly, RED chromatin corresponds to A-type domains. FAIRE-seq finds that BLACK chromatin is less accessible than BLUE chromatin (Figure 6D in Ref. (6)), and the state 9 has a lower level of chromatin accessibility than the state 6 determined by DNase-seq (Figure 5C in Ref. (7)), lending additional supports to our conclusion about domain accessibility and compactness of R-domains (Figure 2F).

**Mathematical expressions used for characterizing domain structures**

(i) Contact probability as a function of subchain size $s$:

$$P(s)=\frac{1}{L-s}\sum_{i=0}^{L-s-1}p_{i,i+s} \tag{S15}$$

where $L$ is the size of entire domain, and $p_{ij}$ is the pairwise contact probability that can be obtained from Hi-C data.

(i) The radius of gyration for a subchain between the $i$-th and $j$-th monomers ($i<j$):

$$r_{g,ij}=\left(\frac{1}{2(j-i+1)^2}\sum_{p=i}^{j}\sum_{q=i}^{j}(\mathbf{r}_i-\mathbf{r}_j)^2\right)^{1/2}. \tag{S16}$$

The mean radius of gyration for subchains of $s$ monomers in a genomic domain of length $L$ is calculated as follows.

$$r_g(s)=\frac{1}{L-s}\sum_{i=0}^{L-s-1}r_{g,i\ i+s}. \tag{S17}$$

(ii) When all the spatial coordinates of monomers comprising chromatin domain are given, the asphericity of the domain (*Asp.*) is calculated using

$$Asp.=\sum_{i=1}^{3}(\lambda_i-\bar{\lambda})^2/6\bar{\lambda}^2 \tag{S18}$$

where $\lambda_i$ ($i=1,2,3$) are the three eigenvalues of the moment of inertia tensor, and $\bar{\lambda}$ is their mean (8, 9).

**Correlations between structural properties characterizing epigenetic domains**

The correlations between $\nu$, density, three structural properties often quantified in FISH experiments ($L/R_g^3$, surface roughness $S/S_0$, asphericity), and the chromatin accessibility are further examined (Figure S5A).

Whereas the monomer density shows a strong (negative) correlation with the accessibility (Figure S5B), $\nu$ is only mildly correlated with it (Figure S5C). We note that this is caused by the presence of two different modes of Polycomb-repressed domain which depend on the genomic spacing between adjacent PRC1-bound loci (10). For Blue1 domain that has a large loop bridged by PRC1 (see Figure S6D), the exponents $\gamma$ and $\nu$ are smaller than those of the inactive Black2 domain, although Black2 domain has slightly higher density and lower accessibility. On the other hand, R-11 domain in Figure 3 corresponds to the second mode of Polycomb-repressed domain with densely spacing PRC1. The presence of two different modes of R-domain contributes to the weak Spearman correlation between $\nu$ and accessibility.

The significant correlation between asphericity and accessibility is noteworthy (Figure S5A and Figure S5D). As shown in Figure S5D, active domains have elongated aspherical shapes, and inactive domains are comparatively more spherical, albeit not perfect ($Asp.>0$). The considerable variation of domain shape results in small correlation between density and its estimate $L/R_g^3$ (Figure S5A).

**Conformational variations in domain structures**

We carried out PCA analysis on $N=10^4$ chromosome configurations for the three largest epigenetic domains. The clustering analysis was done on N chromosome configurations $(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N)$, each of which is characterized by $M$ features as $\mathbf{x}_k=(x_k^{(1)},x_k^{(2)},\ldots,x_k^{(M)})$. We use $M=100\times 99/2=4{,}950$ inter-loci pairwise distances as the feature characterizing each configuration. For the $M\times M$ covariance matrix

$$C_{ij}=\langle\delta x^{(i)}\delta x^{(j)}\rangle, \tag{S19}$$

where $\delta x^{(m)}=x^{(m)}-\overline{x^{(m)}}$ and $\overline{x^{(m)}}=\frac{1}{N}\sum_{k=1}^{N}x_k^{(m)}$, the eigenvectors were calculated, and ranked by a descending order of the corresponding eigenvalues. The PC1 and PC2 are the eigenvectors of two lowest eigenvalues. Figure S11A shows the projection of $N=10^4$ structures on the first two principal components. For the sake of visual clarity, we plotted only 2000 points.

As expected, the active A-23 domain has the largest, and I-14 domain has the smallest variations; yet none of the three domain types shows a clear pattern of clusters. Next, hierarchical clustering was performed based for each domain on the distance root mean squared distance (dRMSD), defined as

$$\mathrm{dRMSD}_{\alpha\beta}=\sqrt{\frac{2}{N(N-1)}\sum_{\{i,j\}}\left(r_{ij}^{\alpha}-r_{ij}^{\beta}\right)^2}, \tag{S20}$$

between any two structures $\alpha$ and $\beta$. Features in the dendrogram depicted in Figure S11B are summarized in Figure S11C. Compared with I-14 and R-11 domain, the structural ensemble of A-23 domain is decomposed into many small sized clusters. The histograms of density from the whole ensemble indicates the order of average packing density is A-23<R-11<I-14 (Figure S11D), however, some of R-11 structures have higher packing density than a subset of I-14 structures. Therefore, the contradictory result of R-11 and I-14 could have originated from the fact that Boettiger *et al.* detected only a subset of chromatin conformations in their measurement.

## REFERENCES


1. Liu, L., Kim, M. H., and Hyeon, C. (2019) Heterogeneous Loop Model to Infer 3D Chromosome Structures from Hi-C. *Biophys. J.,* **117**, 613–625.
2. Bohn, M., Heermann, D. W., and van Driel, R. (2007) Random loop model for long polymers. *Phys. Rev. E.,* **76**(5), 051805.

3. Boettiger, A. N., Bintu, B., Moffitt, J. R., Wang, S., Beliveau, B. J., Fudenberg, G., Imakaev, M., Mirny, L. A., Wu, C.-t., and Zhuang, X. (2016) Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature,* **529**, 418–422.
4. Veitshans, T., Klimov, D., and Thirumalai, D. (1997) Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Folding Des.,* **2**, 1–22.
5. Limbach, H. J., Arnold, A., Mann, B. A., and Holm, C. (2006) ESPResSo – An Extensible Simulation Package for Research on Soft Matter Systems. *Comput. Phys. Commun.,* **174**(9), 704–727.
6. Filion, G. J., van Bemmel, J. G., Braunschweig, U., Talhout, W., Kind, J., Ward, L. D., Brugman, W., de Castro, I. J., Kerkhoven, R. M., Bussemaker, H. J., and van Steensel, B. (2010) Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in *Drosophila* Cells. *Cell,* **143**(2), 212 – 224.
7. Kharchenko, P. V., Alekseyenko, A. A., Schwartz, Y. B., Minoda, A., Riddle, N. C., Ernst, J., Sabo, P. J., Larschan, E., Gorchakov, A. A., Gu, T., Linder-Basso, D., Plachetka, A., Shanower, G., Tolstorukov, M. Y., Luquette, L. J., Xi, R., Jung, Y. L., Park, R. W., Bishop, E. P., Canfield, T. K., Sandstrom, R., Thurman, R. E., MacAlpine, D. M., Stamatoyannopoulos, J. A., Kellis, M., Elgin, S. C. R., Kuroda, M. I., Pirrotta, V., Karpen, G. H., and Park, P. J. (2011) Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature,* **471**, 480–485.
8. Aronovitz, J.A. and Nelson, D.R. (1986) Universal features of polymer shapes. *J. Phys. (Paris),* **47**(9), 1445–1456.
9. Hyeon, C., Dima, R. I., and Thirumalai, D. (2006) Size, shape, and flexibility of RNA structures. *J. Chem. Phys.,* **125**(19), 194905.
10. Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J. T., Sadreyev, R. I., Dekker, J., and Kingston, R. E. (2017) Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. *Molecular Cell,* **65**(3), 432 – 446.
11. Szabo, Q., Jost, D., Chang, J.-M., Cattoni, D. I., Papadopoulos, G. L., Bonev, B., Sexton, T., Gurgo, J., Jacquier, C., Nollmann, M., Bantignies, F., and Cavalli, G. (2018) TADs are 3D structural units of higher-order chromosome organization in *Drosophila*. *Sci. Adv.,* **4**(2), eaar8082.
12. Eagen, K. P., Aiden, E. L., and Kornberg, R. D. (2017) Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proc. Natl. Acad. Sci. USA,* **114**(33), 8764–8769.
13. Li, L., Lyu, X., Hou, C., Takenaka, N., Nguyen, H., Ong, C.-T., Cubeñas-Potts, C., Hu, M., Lei, E., Bosco, G., Qin, Z., and Corces, V. (2015) Widespread Rearrangement of 3D Chromatin Organization Underlies Polycomb-Mediated Stress-Induced Silencing. *Molecular Cell,* **58**(2), 216 – 231.
14. Hou, C., Li, L., Qin, Z., and Corces, V. (2012) Gene Density, Transcription, and Insulators Contribute to the Partition of the *Drosophila* Genome into Physical Domains. *Mol. Cell,* **48**(3), 471–484.
15. Lesage, A., Dahirel, V., Victor, J.-M., and Barbi, M. (2019) Polymer coil–globule phase transition is a universal folding principle of Drosophila epigenetic domains. *Epigenetics & chromatin,* **12**(1), 1–15.

**Table S1.** Genomic regions simulated in this work.

| Chr. | Start(bp) | End(bp) | $N$ | Pearson corr. [1] | Figure |
|---|---|---|---|---|---|
| chr2R | 19,600,000 | 20,100,000 | 100 | 0.98 | 2 |
| chr2R | 15,700,000 | 16,200,000 | 100 | 0.93 | 2 |
| chr3R | 2,450,000 | 2,950,000 | 100 | 0.95 | 2 |
| chr2L | 9,930,000 | 12,980,000 | 610 | 0.95 | S6 |
| chr3R | 12,300,000 | 13,250,000 | 190 | 0.95 | S7 |
| chr3R | 2,280,000 | 2,980,000 | 140 | 0.95 | S8 |
| chrX | 15,950,000 | 16,550,000 | 120 | 0.97 | S9 |
| chr2R | 8,700,000 | 9,200,000 | 100 | 0.97 | S10 |
| chr2R | 7,280,000 | 7,530,000 | 50 | 0.97 | S10 |
| chr2R | 1,490,000 | 1,740,000 | 50 | 0.98 | S10 |
| chr3R | 14,650,000 | 14,900,000 | 50 | 0.97 | S10 |

[1]The Pearson correlation between the contact probabilites from Hi-C and those from modeling by HLM, which measures the quality of the conformational ensembles.

**Table S2.** Epigenetic (sub)domains studied by super-resolution microscopy (3, 11), which are modeled in this work.

| Figure | Tag | Chr. | Start(bp) | End(bp) | $i_{\mathrm{S}}$ [2] | $i_{\mathrm{E}}$ |
|---|---|---|---|---|---|---|
| 2 | A-23 | chr2R | 19,726,615 | 20,092,780 | 25 | 98 |
| | I-14 | chr2R | 15,700,000 [3] | 16,128,463 | 0 | 85 |
| | R-11 | chr3R | 2,487,143 | 2,889,707 | 7 | 87 |
| S6 | Red1 | chr2L | 10,203,092 | 10,436,611 | 54 | 101 |
| | Black1 | chr2L | 10,540,472 | 10,717,954 | 122 | 157 |
| | Blue1 | chr2L | 11,317,986 | 11,468,388 | 277 | 307 |
| | Black2 | chr2L | 11,547,196 | 11,783,436 | 323 | 370 |
| | Blue2 | chr2L | 12,564,906 | 12,685,026 | 526 | 551 |
| | Red2 | chr2L | 12,690,712 | 12,739,196 | 552 | 561 |
| S7 | I-08 | chr3R | 12,312,771 | 12,438,748 | 2 | 27 |
| | A-03 | chr3R | 12,464,802 | 12,479,673 | 32 | 35 |
| | R-10 | chr3R | 12,481,406 | 12,810,708 | 36 | 102 |
| | R-10.3 | chr3R | 12,481,406 | 12,619,389 | 36 | 63 |
| | R-10.5 | chr3R | 12,683,230 | 12,810,708 | 76 | 102 |
| | A-17 | chr3R | 12,811,417 | 12,951,671 | 102 | 130 |
| | I-11 | chr3R | 12,977,555 | 13,217,431 | 135 | 183 |
| S8 | I-10 | chr3R | 2,287,031 | 2,473,421 | 1 | 38 |
| | A-04 | chr3R | 2,470,393 | 2,487,998 | 38 | 41 |
| | R-11 | chr3R | 2,487,143 | 2,889,707 | 41 | 121 |
| | R-11.1 | chr3R | 2,487,143 | 2,570,646 | 41 | 58 |
| | A-06 | chr3R | 2,890,667 | 2,916,994 | 122 | 127 |
| S9 | R-09 | chrX | 15,953,918 | 16,160,086 | 0 | 42 |
| | A-20 | chrX | 16,161,627 | 16,351,359 | 42 | 80 |
| | I-07 | chrX | 16,351,360 | 16,440,669 | 80 | 98 |
| S10 | R-03 | chr2R | 8,770,090 | 8,803,242 | 14 | 20 |
| | A-16 | chr2R | 9,019,018 | 9,125,165 | 63 | 85 |
| | A-02 | chr2R | 7,338,881 | 7,351,765 | 11 | 14 |
| | R-07 | chr2R | 7,350,135 | 7,469,804 | 14 | 37 |
| | R-05 | chr2R | 1,593,366 | 1,640,664 | 20 | 30 |
| | I-04 | chr3R | 14,754,844 | 14,804,163 | 20 | 30 |

[2]The corresponding range of monomer indices between $i_S$ and $i_E$ in the model.
[3]A short segment chr2R:15,602,615-15,700,000, which is missing in the Hi-C experiment (12), was omitted from modeling.

**Table S3.** Subchains in A-23, I-14, and R-11 domain, whose radii of gyration were measured in Ref. (3), and were compared with our modeling (see Figure 7).

| Domain | Tag | Start(bp) | End(bp) | $i_{\mathrm{S}}$ [4] | $i_{\mathrm{E}}$ |
|---|---|---|---|---|---|
| A-23 | A-23.1 | 19,726,615 | 19,787,585 | 25 | 37 |
| | A-23.2 | 19,809,874 | 19,888,410 | 41 | 57 |
| | A-23.3 | 19,906,491 | 19,976,552 | 61 | 75 |
| | A-23.4 | 20,000,481 | 20,092,780 | 80 | 98 |
| | A-23.5 | 19,726,615 | 19,888,410 | 25 | 57 |
| | A-23.6 | 19,809,874 | 20,092,780 | 41 | 98 |
| | A-23 | 19,726,615 | 20,092,780 | 25 | 98 |
| I-14 | I-14.2 | 15,702,616 | 15,802,615 | 0 | 20 |
| | I-14.6 | 15,702,616 | 15,852,615 | 0 | 30 |
| | I-14.7 | 15,802,616 | 15,952,615 | 20 | 50 |
| | I-14.8 | 15,852,616 | 16,128,463 | 30 | 85 |
| R-11 | R-11.1 | 2,487,143 | 2,570,646 | 7 | 24 |
| | R-11.2 | 2,568,667 | 2,653,159 | 23 | 40 |
| | R-11.3 | 2,641,835 | 2,721,984 | 38 | 54 |
| | R-11.4 | 2,721,063 | 2,827,254 | 54 | 75 |
| | R-11.5 | 2,825,712 | 2,889,707 | 75 | 87 |
| | R-11.6 | 2,487,143 | 2,653,159 | 7 | 40 |
| | R-11.7 | 2,568,667 | 2,827,254 | 23 | 75 |
| | R-11.8 | 2,487,143 | 2,889,707 | 7 | 87 |

[4]The corresponding range of monomer indices between $i_S$ and $i_E$ in the model.

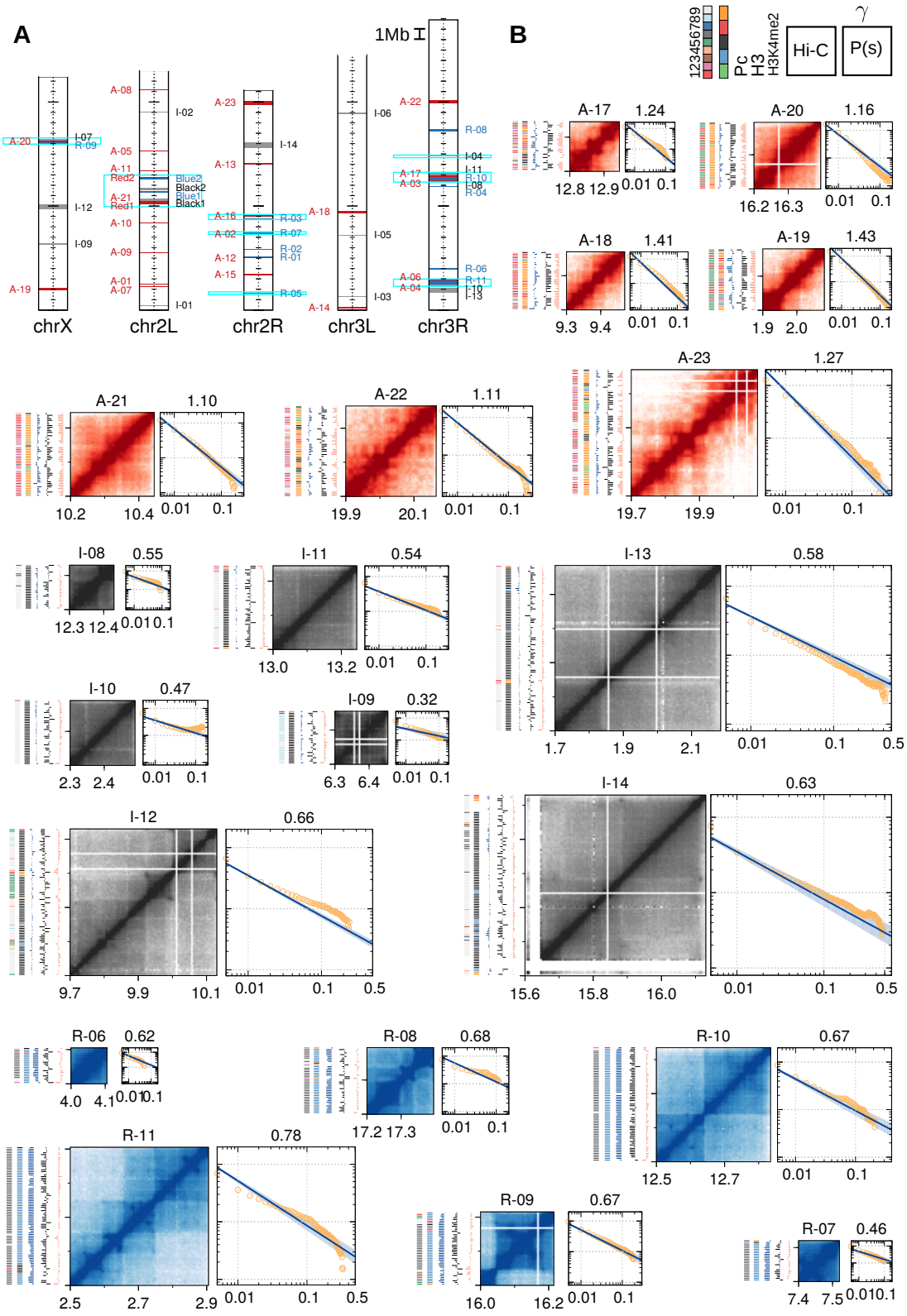


**Figure S2.** Analysis of the Hi-C data by Eagen *et al.*(12) for different types of epigenetic domains. (A) Presented is a diagram of *Drosophila* genome which labels the genomic positions of 54 epigenetic domains (25 active domains, 16 inactive domains, and 13 repressed domains). The epigenetic domains have been imaged by super-resolution FISH (3, 11), and are re-analyzed using Hi-C data in this work. In addition to the largest domain of each type (A-23, I-14, and R-11), we modeled 8 genomic regions each including several domains, highlighted by the wide cyan boxes, to study the intra-domain packaging and inter-domain mixing. (B) Heatmap of contact probabilities of large domains ($>100$ kb). For each domain, the mean contact probability ($P(s)$) is plotted on the right. The value of $\gamma$ obtained from the fit $P(s)\sim s^{-\gamma}$, is given on the top. The enrichment profiles of H3K4me2 (red), unmodified H3 (black), Pc (blue), five (6), and nine chromatin state classifications (7) are aligned with the heatmap on the left.

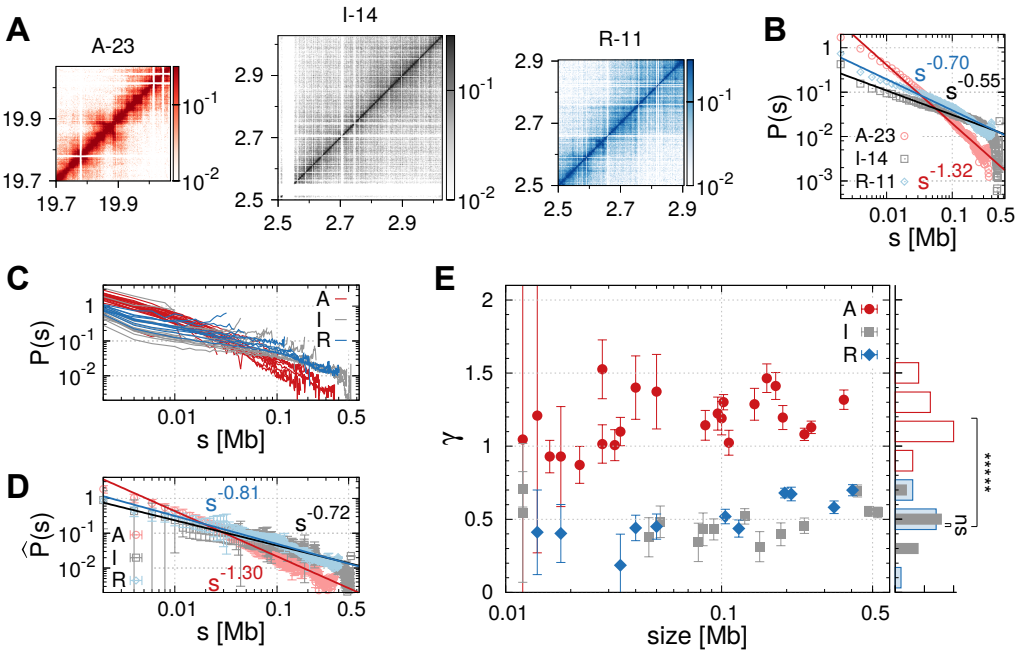


**Figure S3.** Analysis of the Hi-C library prepared with restriction enzyme DpnII, published by Li *et al.*(13), for different types of epigenetic domains. (A) Heatmap of contact probabilities, (B) Mean contact probability $P(s)$ of the largest domains in three epigenetic states. (C) $P(s)$ of all domains. (D) Mean contact probability further averaged over domains of the same epigenetic type $\widehat{P}(s)$. (E) $\gamma$ determined by fitting $P(s)$ to $s^{-\gamma}$ versus the genomic size of domain. The significance of similarity between the histograms of the values of $\gamma$, evaluated by one-side Mann Whitney U test, is shown on the graph with the notation: ns (not significant, $p>0.05$), * ($p<0.05$), ** ($p<0.01$), and ***** ($p<1\times10^{-5}$).

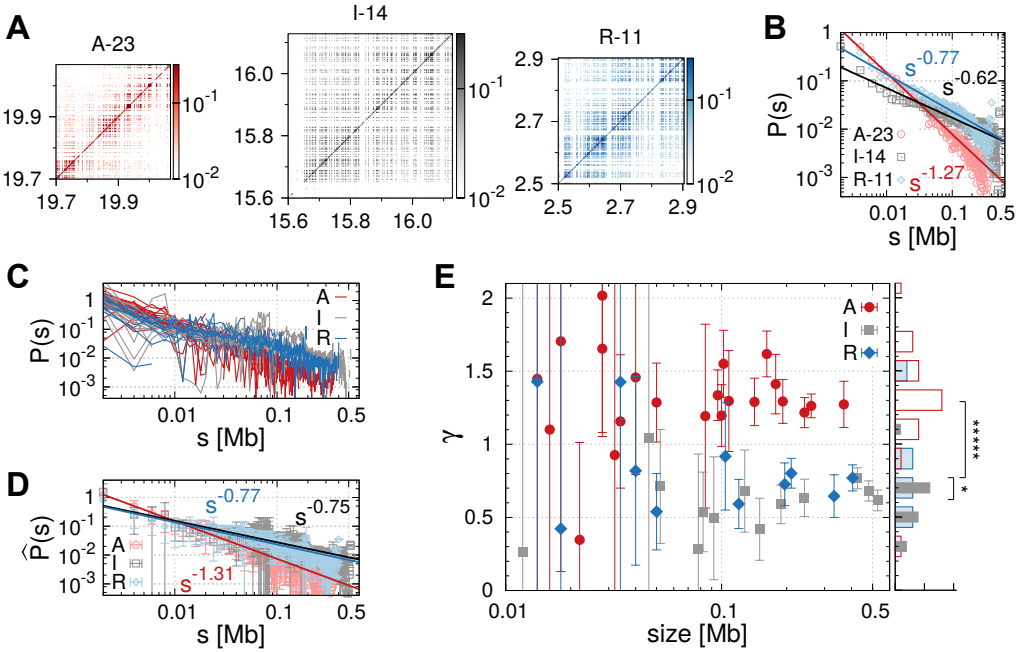


**Figure S4.** Analysis of the Hi-C library prepared with restriction enzyme HindIII, published by Hou *et al.*(14), for different types of epigenetic domains. See also the caption of Figure S3.

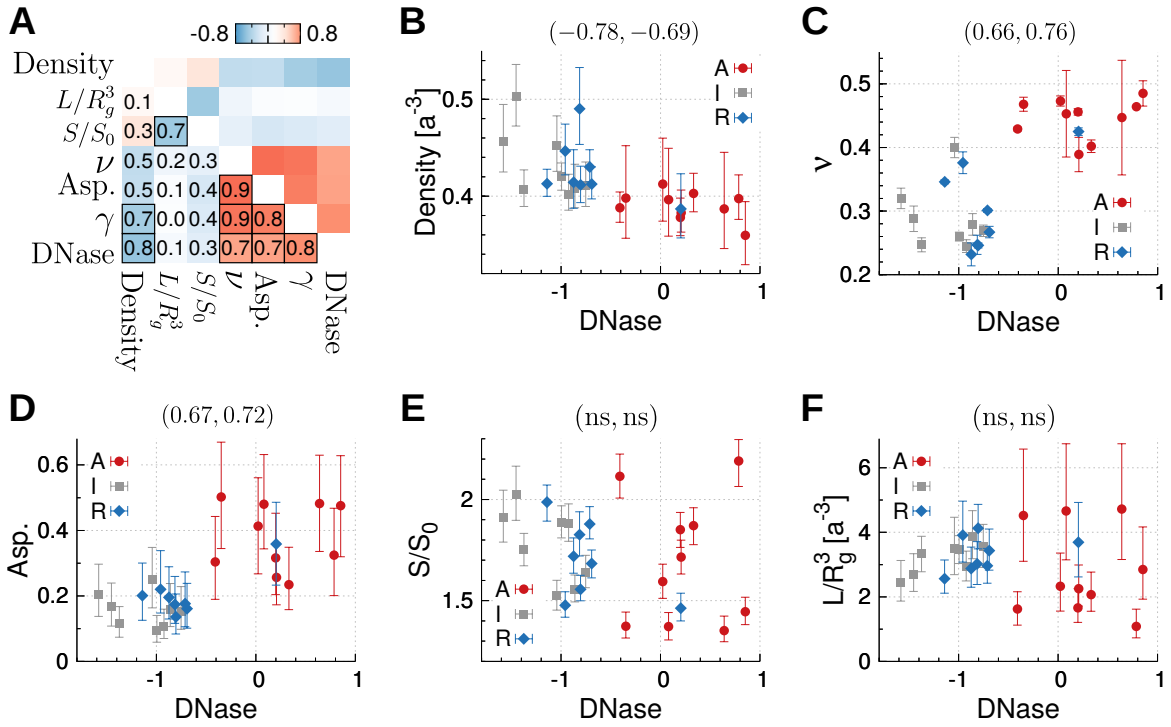


**Figure S5.** Correlations between different structural properties. (A) Spearman correlation between different structural properties. The numbers in the lower diagonal part denote the absolute value of correlations, and the black boxes mark the correlations of confidence ($p<0.001$). (B) Density, (C) $\nu$, (D) Asphericity, (E) Surface roughness $S/S_0$, (F) $L/R_g^3$ versus chromatin accessibility probed by DNase assay. The corresponding Spearman correlation and Pearson correlation are given inside the parenthesis on the top.

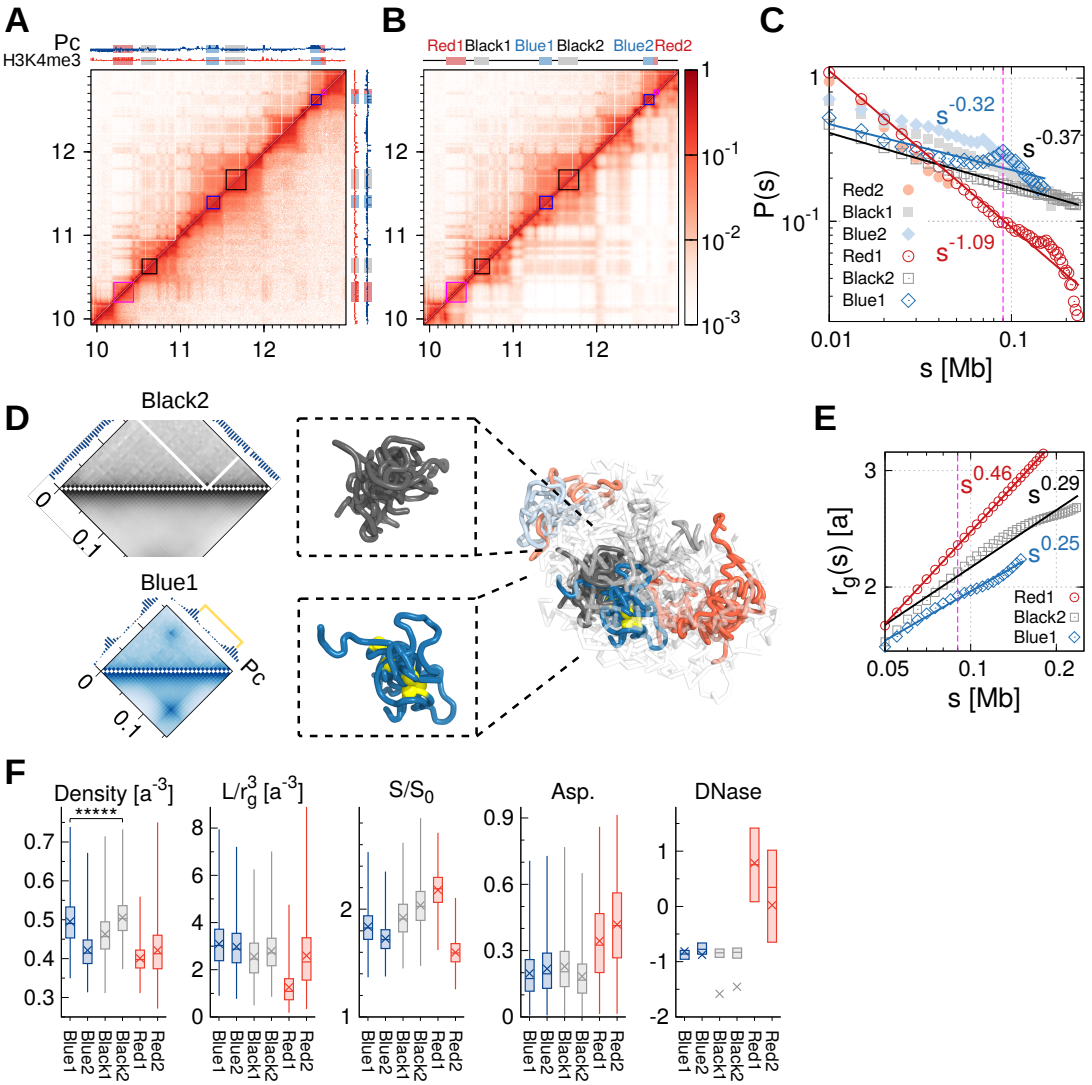


**Figure S6.** A 3.05 Mb region at chr2L:9,930,000–12,980,000 modeled by HLM. (A) Heatmap of contact probabilities of $Kc_{167}$ cells (upper diagonal part) and S2R+ cells (lower diagonal part). The enrichment profiles of H3K4me3 (red tack) and Pc (blue track) are shown at the top. (B) Heatmap of contact probabilities from Hi-C of $Kc_{167}$ cells (upper diagonal part) and from HLM (lower diagonal part). (C) Mean contact probability $P(s)$ within each domain. (D) An ensemble of chromatin structures, where the inactive domain Black2 and the Polycomb-repressed domain Blue1 are zoomed in, as well as heatmaps of their contact probability and Pc enrichment profiles. In Blue1 domain, the two loci of the strongest Pc binding are rendered as yellow spheres. (E) Mean radius of gyration $r_g(s)$ calculated for each domain type. (F) Different epigenetic domains are compared in terms of their various structural properties and chromatin accessibility probed by DNase.

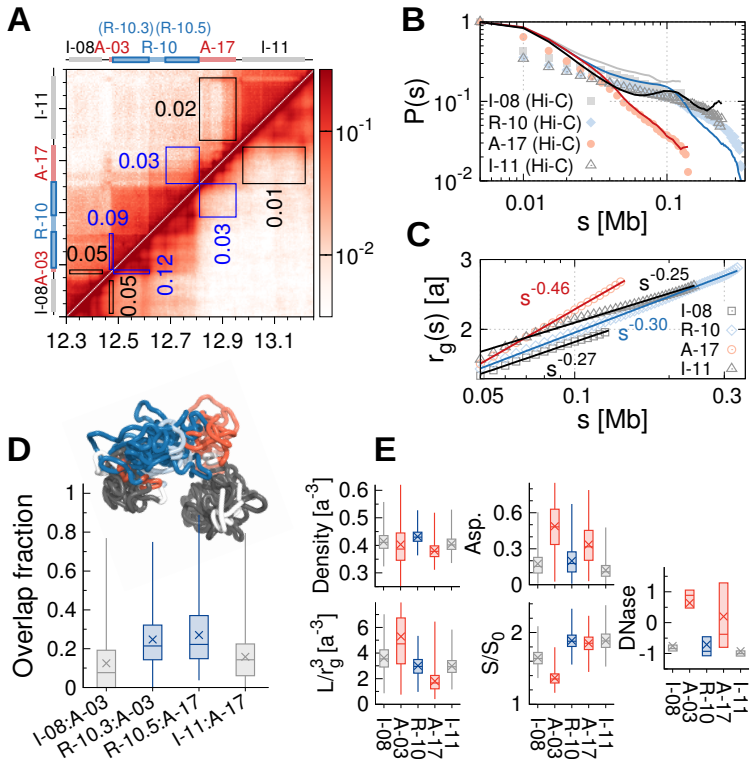


**Figure S7.** A 0.95 Mb region at chr3R:12,300,000–13,250,000 modeled by HLM. (A) Heatmap of contact probabilities from Hi-C (upper diagonal part) and HLM (lower diagonal part). The inter-domain contacts are enclosed by rectangular boxes, beside which the labels denote the corresponding mean value of inter-domain contact probabilities $\langle p \rangle_{\text{Hi-C}}$ and $\langle p \rangle_{\text{HLM}}$. (B) Intra-domain mean contact probability $P(s)$. (C) Intra-domain mean gyration radius $r_g(s)$. (D) Overlap fraction of inactive and repressed domain with adjacent active domains. (E) Structural properties and accessibility of different epigenetic domains.

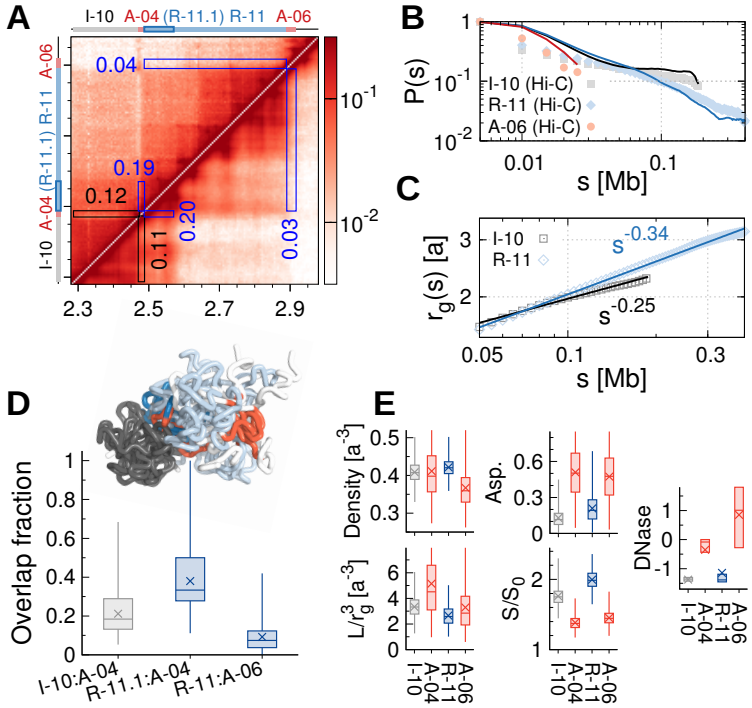


**Figure S8.** A 0.7 Mb region at chr3R:2,280,000–2,980,000 modeled by HLM. See the caption of Figure S7.

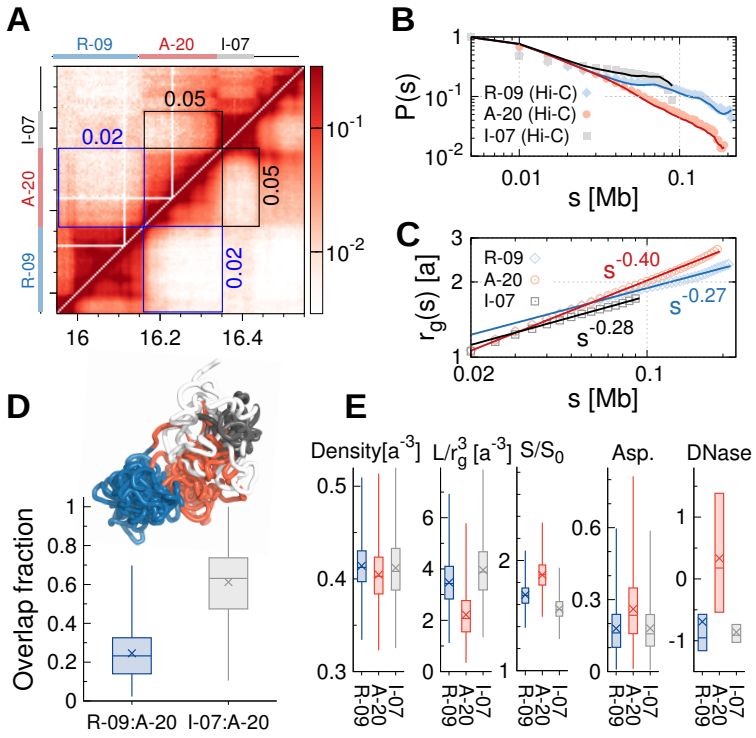


**Figure S9.** A 0.6 Mb region at chrX:15,950,000–16,550,000 modeled by HLM. See the caption of Figure S7.

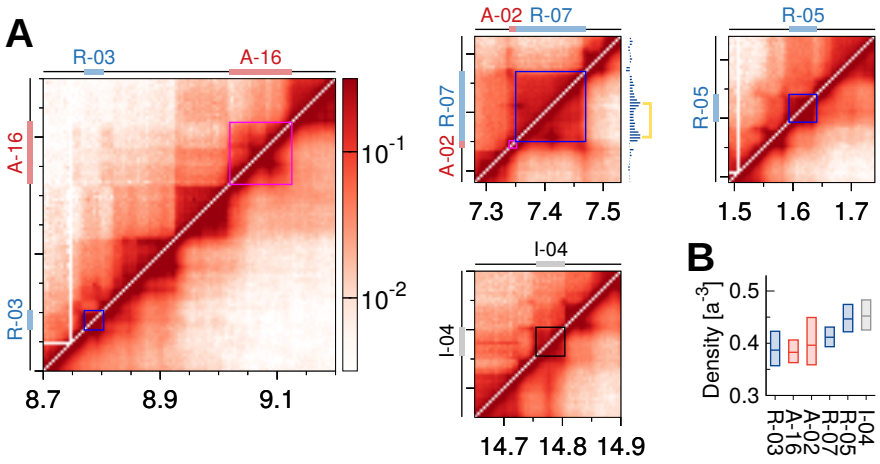


**Figure S10.** (A) Heatmaps of contact probabilities from HiC (upper diagonal part) and HLM (lower diagonal part) in four genomic regions, which include a few small-sized epigenetic domains. See Table S1 and S2 for their detailed genomic positions. Pc enrichment profile of the region including A-02 and R-07 domains is shown on the right of the contact map. (B) Density of each domain based on reconstructed structures.

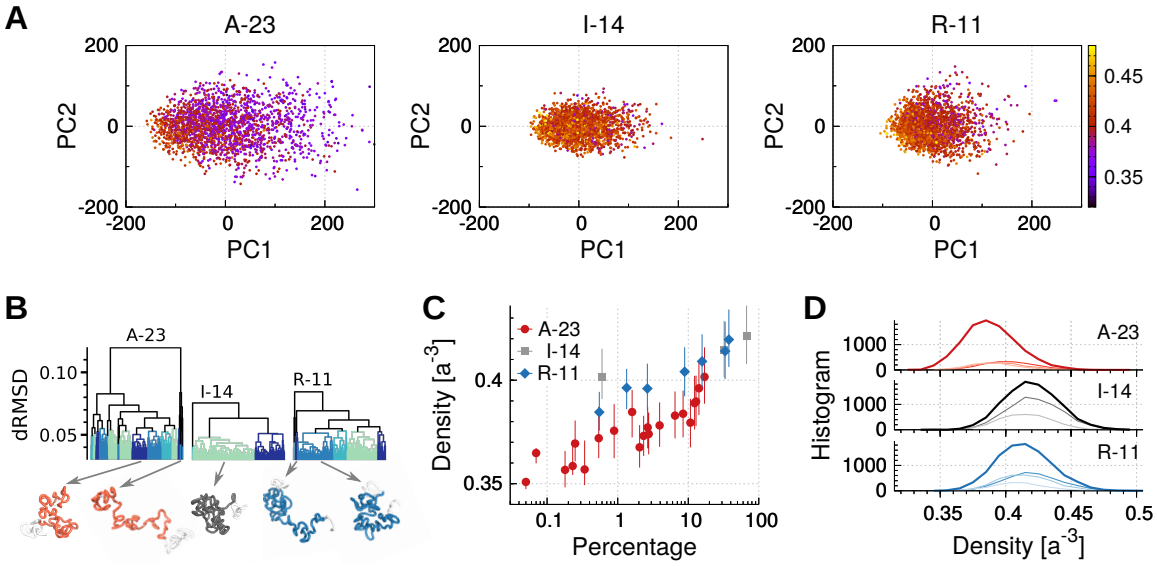


**Figure S11.** Comparing the packing density of structures in different clusters. (A) The projection of structures on the PC1 and PC2 by using PCA on structural ensembles of three domain types. The color-code represents the density ($L/V$) of each structure. (B) Dendrogram of structures by using hierarchical clustering, with a representative structure in each cluster being illustrated. (C) The density of structures in each cluster versus the percentage of the cluster in the whole ensemble. (D) Histogram of density of the whole ensemble (the thick lines) and that of the three largest clusters (the thin lines).

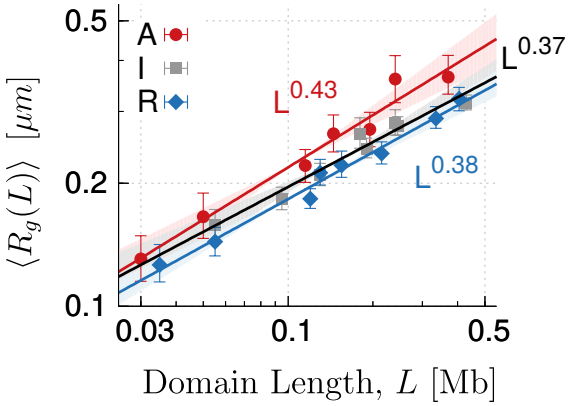


**Figure S12.** The gyration radius $R_g(L)$ of whole domain when the epigenetic type-dependent Kuhn lengths are takin into account (15). $R_g(L)$ are recalculated by mapping the HLM length unit $a$ to $k_{\rm nm}(5000/k_{\rm bp})^c$ nm, where $k_{\rm nm}$ and $k_{\rm bp}$ are the Kuhn segment lengths in units of nm and bp, respectively. The domain-type dependent values of $k_{\rm nm}$ and $k_{\rm bp}$ are taken from Table 3 in Ref. (15).