Minimalistic 3D Chromatin Models: Sparse Interactions in Single Cells Drive the Chromatin Fold and Form Many-Body Units

Jie Liang[1,✉] **and Alan Perez-Rathke**[1]

[1]Center for Bioinformatics and Quantitative Biology & Richard and Loan Hill Department of Bioengineering, University of Illinois at Chicago, Chicago, IL, 60612, USA.

Computational modeling of 3D chromatin plays an important role in understanding the principles of genome organization. We discuss methods for modeling 3D chromatin structures, with focus on a minimalistic polymer model which inverts population Hi-C into high-resolution, high-coverage single-cell chromatin conformations. Utilizing only basic physical properties such as nuclear volume and no adjustable parameters, this model uncovers a few specific Hi-C interactions (15-35 for enhancer-rich loci in human cells) that can fold chromatin into individual conformations consistent with single-cell imaging, Dip-C, and FISH-measured genomic distance distributions. Aggregating an ensemble of conformations also reproduces population Hi-C interaction frequencies. Furthermore, this single-cell modeling approach allows quantification of structural heterogeneity and discovery of specific many-body units of chromatin interactions. This minimalistic 3D chromatin polymer model has revealed a number of insights: 1) chromatin scaling rules are a result of volume-confined polymers; 2) TADs form as a byproduct of 3D chromatin folding driven by specific interactions; 3) chromatin folding at many loci is driven by a small number of specific interactions; 4) cell subpopulations equipped with different chromatin structural scaffolds are developmental stage-dependent; and 5) characterization of the functional landscape and epigenetic marks of many-body units which are simultaneously spatially co-interacting within enhancer-rich, euchromatic regions. The implications of these findings in understanding the genome structure-function relationship are also discussed.

Minimalistic chromatin model | chromatin folding| polymer modeling| Chromatix| 3D chromatin| fractal Monte Carlo| Hi-C| specific interactions| proto-TADs| many-body units |Principal loops| Euchromatin| Super-enhancers

Correspondence: *jliang@uic.edu*

Introduction

Chromosome conformation capture (1–4) and imaging analysis (5–8) have generated a wealth of information on nuclear genome organization. Structural units, such as compartments, topologically associating domains (TADs), and loops, have been uncovered from block patterns in Hi-C frequency heatmaps (9–11). However, canonical Hi-C measures only population-averaged, pairwise interaction frequencies, which do not fully represent the underlying 3D conformational distributions in individual cells. This limits our understanding, as single-cell 3D chromatin structures do not automatically follow from Hi-C's population-averaged heatmap representation (12).

Veiled behind each Hi-C heatmap is a population of 3D chromatin conformations. While single-cell technologies can provide direct information on 3D spatial arrangements of genomic elements within a nucleus (5, 11), they are limited in both resolution and coverage. As an alternative, computational modeling of chromatin polymers provides a powerful means for uncovering 3D spatial configurations of genomic regions and plays important roles in deciphering physical principles of genome organization.

Optimization-Based 3D Chromatin Models. Among the many different approaches to modeling 3D chromatin from Hi-C, optimization methods aim to generate chromatin conformations maximally satisfying Hi-C derived distance-restraints (13–18) (reviewed in (19, 20)). However, they are *ad hoc* as conformations obtained do not follow a physically-governed, *a priori*-defined distribution, owing to the lack of a physical model underpinning these methods. In addition, there is no consideration to ensure adequate sampling such that a diverse structural ensemble is sufficiently represented. Furthermore, many tunable parameters are often employed to ensure a good fit to experimental data: as many distance thresholds as the number of restraints modeled may need to be adjusted. If inadequately regularized, they may lead to overfitting.

Understanding the spatial organization of chromatin requires a physical 3D chromatin model. With an accurate physical model, the sampled conformational ensemble will correctly reflect the conformational distribution in the cell population.

3D Heteropolymer Models with an Empirical Energy Function. We now briefly discuss physical models of heteropolymers with binding interactions among regions of different chromatin states (see excellent reviews of (21–23) for details). Here chromatin is modeled as heterogeneous blocks of 3D monomer chains. Monomers, representing $0.5 - 500$ kb chromatin, are grouped into different blocks by their epigenetically-defined chromatin states. An empirical energy function describes how these monomers interact according to their binding affinities (24–31). Molecular dynamics simulations then generate an ensemble of chromatin conformations. While realistic dynamics are not possible at this coarse-grained scale, the ensembles obtained can provide direct 3D structural information.

These powerful models allow detailed assessment of the effects of different mechanistic assumptions, which can reveal
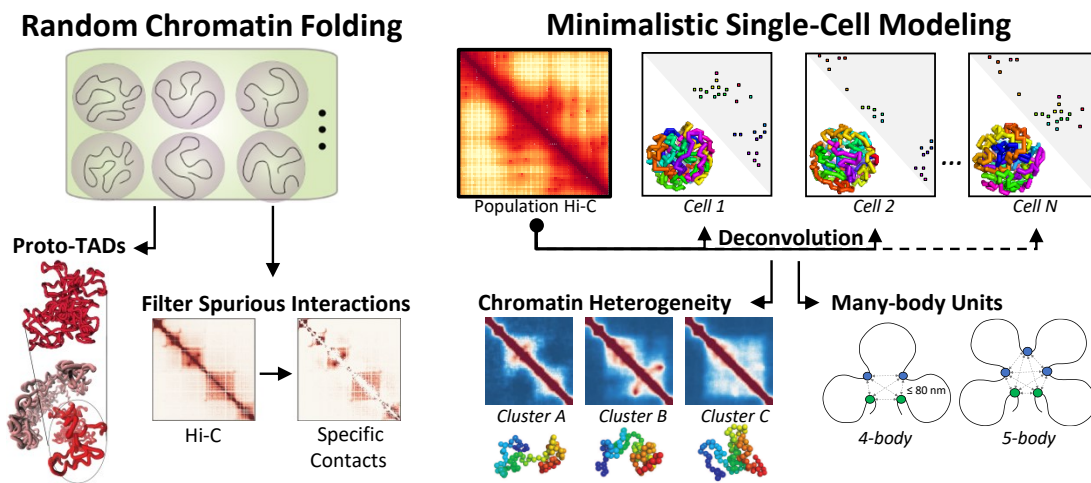
Fig. 1. *Graphic Abstract.*

important insights into principles of genome organization. As an example, studies based on the MiChroM method suggest the likely origin of chromosome territories, reproduce phase separation, and provide further evidence for the preferential localization of active genes (26, 27, 32). It was shown that epigenetic information can be used to predict the structural ensembles of multiple human cell lines, and short segments of chromatin make state transitions between closed conformations and open dumbbell conformations (33). The aggregation of denser and predominantly inactive chromatin was found to drive the positioning of active chromatin toward the surface of individual chromosomal territories (33). Recent efforts in scaling up MiChroM showed that inter-chromosomal interactions can now be studied through detailed molecular dynamics simulations (34). A study of the full diploid genome of all 46 human chromosomes at 1 Mb resolution was recently carried out to determine factors important for radial positioning of chromosomes (30). A separate study using a polymer model with only two epigenetic states and fixed loop anchors showed that compartments in Hi-C maps are due to microphase separation of these two states, leading to highly heterogeneous chromosome dynamics (28).

The String and Binders Switch (SBS) model and related methods (24) have been applied to study how binder mediated chromatin interactions can lead to 3D chromatin structures. In (35), the 3D structures of the *Sox9* locus and the whole mouse chromosome 7 were constructed. In (31), it was found that an increase in binder concentration can lead chromatin to adopt a coil-to-globule phase-separated state, where the intrinsically degenerate globule state corresponds to the large variation observed in modeled 3D chromatin conformations. Another recent study applied the SBS model to extract 3D chromatin structures from GAM, SPRITE and Hi-C data (36), and compared these experimental methods with *in silico* models. These experimental techniques were found to have different efficiency in capturing long-range interactions, requiring a different minimum number of cells (ranging from 250 to 800) for replicate experiments to return statistically consistent measurements (36).

Binding Factors and Affinities. The behavior of heteropolymer chromatin is fully determined by the energy landscape of the model (24–28). Unlike molecular mechanics, where physical forces such as electrostatic interactions have been thoroughly studied, first-principle understanding of the physical factors and interactions of chromatin at mesoscale is not available. There are likely over 2,600 different proteins expressed in a cell (37), many yet to be identified, which complex with DNA and each other at often unknown rates and affinities. Instead, coarse-grained chromatin states and empirical binding affinities must be inferred from Hi-C and epigenetic data. There are likely many different chromatin state and binding affinity assignments that are all consistent with experimental data. The non-identifiable nature of such assignments, combined with other assumptions of the energy model, may hinder precise inference and limit the biological interpretability in phenomena detected through simulations. Furthermore, while an *a priori* constructed energy model can effectively explore consequences of various model assumptions, making biological discoveries not encoded in the model input is challenging.

Importance of Thorough Sampling. A prerequisite for all 3D chromatin polymer methods is that conformational ensembles must be thoroughly sampled. This is exceedingly challenging as chromosomes are severely confined in the nuclear volume (38) and exhibit extraordinary heterogeneity (5, 6, 33, 39). Generating biologically-accurate chromatin ensembles using molecular dynamics is non-trivial. Without thorough sampling, it is difficult to ascertain if bias is present due to inadequate sampling, a misspecified energy model, or both.

Minimalistic Self-Avoiding Polymer Model of 3D Chromatin

Another approach is to model chromatin as a 3D self-avoiding polymer but with minimal physical properties and no adjustable parameters. Once its emerging behavior is characterized and deficiencies identified, additional ingredi-

ents are then introduced to refine the model. Such a minimalistic approach of polymer modeling has had great success in earlier studies of protein folding (40, 41).

The initial premises are that chromatin must be 1) connected, 2) self-avoiding, and 3) confined in the cell nucleus. This approach becomes feasible with recent deep-sampling algorithms (38, 42–45). These algorithms generate nuclear-confined, self-avoiding chromatin polymers by sequentially placing connected monomer units until the target polymer length is reached (Fig 2a). Advanced sampling techniques such as fractal Monte Carlo enable generation of large and complex ensembles consisting of 10^{4-5} single-cell 3D chromatin structures.

We next discuss recent findings using this minimalistic approach. First, their generic behavior without invoking Hi-C-derived information: experimentally measured chromatin scalings arise naturally, and incorporating telomere and centromere tethering results in 3D chromatin ensembles that reproduce Hi-C measured interactions in budding yeast.

Nuclear Confinement Is Intrinsic to Chromatin Scaling Behavior. For two loci separated by a genomic distance s, their mean-squared distance $R^2(s)$ and contact probability $p(s)$ scale characteristically. Specifically, $R^2(s) \sim s^{2v}$, with the exponent $v = 0.33$ (46) and with $R^2(s)$ further plateauing at longer s. The contact probability scales as $p(s) \sim \frac{1}{s^\alpha}$, with the exponent $\alpha = 1.08$ (1). Fractal globule was thought responsible for these behaviors (1), although subsequent studies suggested otherwise (47).

Imposing only nuclear volume confinement, randomly-generated ensembles of self-avoiding 3D chromatin polymers exhibit both scaling relationships and plateau behavior (Fig 2b) (38). These results show: 1) random and self-avoiding polymers in confinement give rise to scaling relationships and 2) nuclear size is a major determinant of chromatin folding. Indeed, nuclear morphology, such as a large ratio of nuclear to cytoplasmic volume, is likely important for the broad cellular reprogramming capabilities in stem cells (48). Further, nuclear size is often associated with diseases such as cancer (48).

Nuclear Confinement With Landmarks Explains Yeast 3D Genome. In budding yeast, centromeres and telomeres are tethered to the spindle-pole body and nuclear envelope, respectively. Incorporating physical tethering of these nuclear landmarks into a basic polymer model was sufficient to give rise to the preferential localization of functional loci in the nucleus (49, 50). The ensemble of confined chromatin polymers with landmarks reproduced intra- and inter-chromosomal Hi-C interaction frequencies (Pearson $R > 0.81$ and $R > 0.91$, respectively) (43, 49, 50) (Fig 2c). In addition, centromere tethering was found to be responsible for inter-chromosomal interactions (43). Furthermore, fragile sites spatially cluster together (43). These findings demonstrate that ensembles of self-avoiding polymers in confinement, along with simple landmarks, can explain many Hi-C frequencies observed in yeast. There are significant implications for mammalian cells with analogous tetherings: attach-

ment of heterochromatin to lamin, and association of actively transcribed regions to nuclear speckles (51–54). It would be interesting to quantitatively assess the extent to which these tetherings contribute to genomic 3D spatial organization, their functional consequences, and whether fragile sites similarly congregate spatially.

A Small Set of Specific Interactions Can Drive Chromatin Folding

The results from the yeast polymer model require no Hi-C inputs. This suggests that many measured Hi-C interactions are due to generic effects of confined polymers: many Hi-C interactions occur due to random collisions of chromatin fibers confined in the cell nucleus (38, 42, 55). However, a small number of interactions occur at significantly elevated frequencies than would be expected in a randomly-generated ensemble of confined chromatin polymers. These over-represented interactions are known as *specific interactions* (42). In Drosophila, they constitute only $5-7\%$ of measured Hi-C interactions. While their numbers are small, they appear to be critical in maintaining chromatin architecture, as simulations show that these interactions are sufficient to fold chromatin at both a locus and whole chromosome level. We discuss these below.

A Small Fraction of Specific Interactions Captures Overall Hi-C Pattern. Specific Hi-C interactions emerge after filtering of spurious interactions which occur due to generic effects of polymer connectivity, self-avoidance, and nuclear confinement. To filter these spurious interactions, one compares the frequency of a Hi-C contact against the corresponding frequency in a simulated ensemble of randomly-generated polymer contacts (Fig 2d). Specific Hi-C contacts will have high frequencies meeting statistical significance and hence unlikely to have resulted from random collisions (38, 42). In Drosophila, analysis of 10 loci showed that only $5-7\%$ of Hi-C interactions are specific ($2.0-2.3 \times 10^6$ out of $35-42 \times 10^6$ Hi-C interactions); however, they capture the overall Hi-C pattern (Fig 3b) (45). Furthermore, there are clear changes in specific interactions during embryo development, such as reductions in active-inactive interactions and increases in inactive-inactive interactions (45). These changes, however, are undetectable without filtering of spurious Hi-C interactions (Fig 3b). This study illustrates that specific interactions reveal important temporal changes in chromatin structure not apparent in the measured Hi-C data.

Specific Interactions Can Fold Chromatin in Drosophila. An important question about 3D chromatin is whether there is a set of critical interactions that drive chromosomal folding (Fig 3a) (42, 56). Provided they exist, identifying them will likely reveal crucial determinants of 3D chromatin organization and delineate the important interactions and corresponding chromatin structures which may be necessary for genomic function.

Taking only the specific interactions for a locus ($5-7\%$ of all interactions) as distance restraints to be imposed on the self-
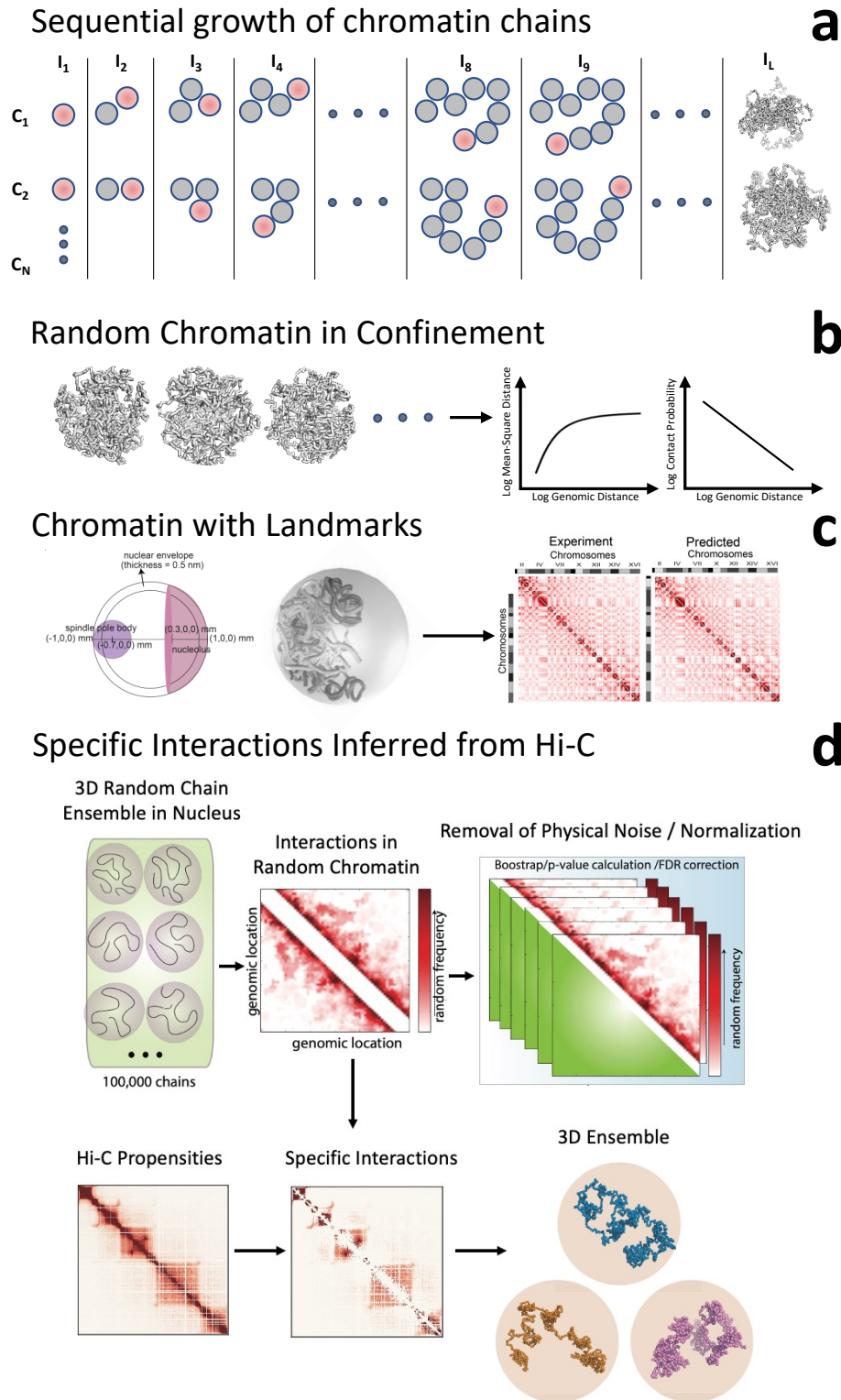
Fig. 2. *Minimalistic 3D self-avoiding polymer model.* **(a)** Illustration of sequential polymer growth process. An ensemble of N polymer chains, $C_1 \ldots C_N$, is grown by iterative addition of self-avoiding monomer units until target length L is reached. **(b)** A confined ensemble of random polymers gives rise to the characteristic scaling relationships of mean-square distance and contact probability with genomic distance. **(c)** An ensemble of confined random polymers with nuclear landmarks (telomere and centromere attachments) imposed reproduces the heatmap of Hi-C measured frequencies, although no Hi-C information was used (from (38)); **(d)** The ensemble of confined random polymers can be used to remove spurious interactions in Hi-C measurements and to identify specific interactions, which have significantly elevated interaction frequencies. Upper subpanel illustrates that simulated random polymers, when aggregated, exhibit a heatmap of random interactions. The lower subpanel shows that spurious interactions can be removed from the experimental Hi-C maps through bootstrapping and FDR correction of the random heatmap, resulting in the identification of a small set of specific interactions. These specific interactions can then be used to fold chromatin into 3D structures (from (42)).
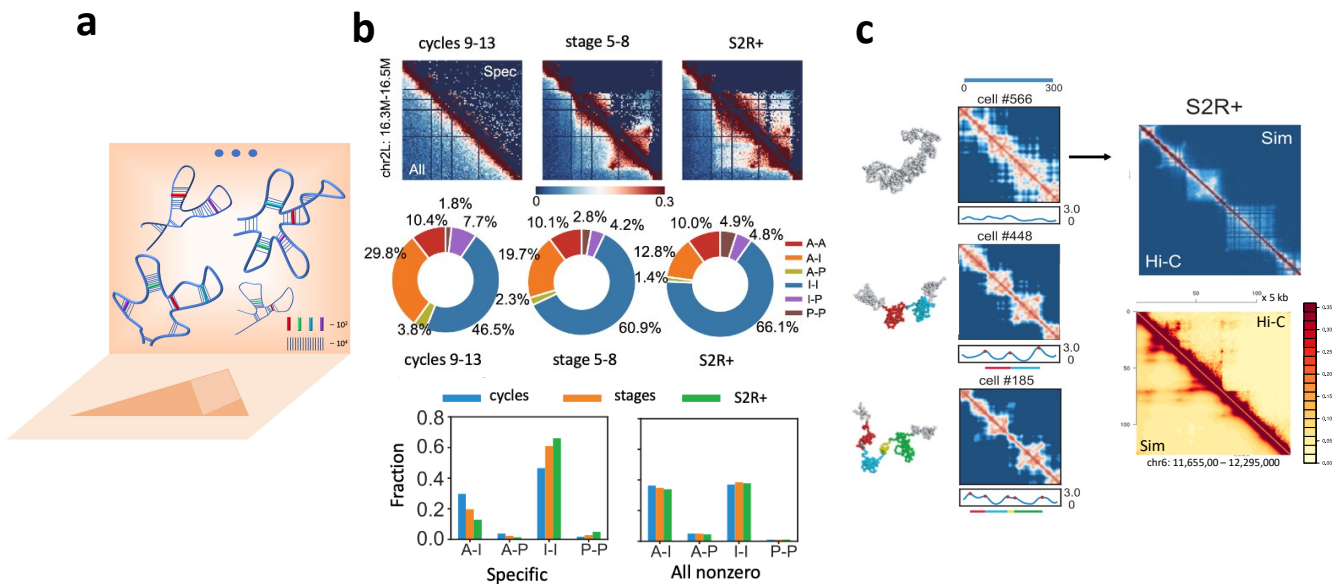
Fig. 3. *Chromatin can be folded using specific interactions* (**a**) Overview. A small set of interactions (thick colored sticks, $10 - 100$ for a typical locus of $1 - 2$ Mb) can fold chromatin in individual cells, where they appear in different combinations. Their aggregation gives rise to the 2D projection of the Hi-C heatmap. The Hi-C measured chromatin interactions ($\sim 10^4$) include both specific (colored sticks) and likely-biologically relevant, as well as bystander interactions (blue vertical lines); (**b**) Specific interactions in chr2L: $16.3 - 16.5$ Mb of Drosophila (upper triangles) capture the overall Hi-C interaction patterns (lower triangles) of cells at pre-MBT stages (cycles $9 - 13$), post-MBT (stages $5 - 8$), and S2R+. Pie charts and bar charts show % of different types of specific interactions (A: active, I: inactive, P: polycomb-repressed) in the three cell types. As the embryo develops, A-I interactions (orange in pie chart) decreases, while I-I interactions (blue in pie chart) increase. In contrast, no such changes are detected when all Hi-C interactions are considered (bar chart, right) (from (45)); (**c**) Examples of single-cell chromatin conformations (chr2L: $11.0 - 12.0$ Mb, 2 kb resolution, Drosophila S2R+ cells) and their spatial-distance heatmaps. When aggregated, the simulated population Hi-C heatmap (upper triangle, top right column) resembles the measured Hi-C heatmap (lower triangle, Pearson correlation $R = 0.95$). (bottom right column) Experimental (top triangle) (2) and simulated (lower triangle) Hi-C heatmaps of human GM12878 cells (Chr6: $11.65 - 12.29$ Mb, 5 kb) reconstructed by aggregation of single-cell conformations (lower triangle) folded using a small set of ≤ 35 specific interactions ($R = 0.98$) (from (44, 45)).

avoiding 3D model, locus conformations generated using a deep-sampling algorithm reproduce Hi-C heatmaps with high accuracy (10 loci of length 0.2–2 Mb at different developmental stages at 2 kb resolution, Pearson correlation $R = 0.91$–0.98, Fig 3b) (45). In addition, details of loops and known long-range interactions are recovered (45). Furthermore, whole chromosome X of Drosophila can be folded at 5 kb resolution solely using specific interactions (45). Similar success was reported for the human α-globin locus, where independent ChIA-PET measurements were predicted from chromatin folded by 5C-derived specific interactions.

While 3D chromatin ensembles can be constructed by several methods, the success of the minimalistic self-avoiding polymer model is compelling, as only elementary physical considerations of fiber density, nuclear volume, and ligation distance are the model input. There are no adjustable parameters, no chromatin state assignments, and no *a priori* assumptions on loop anchors. Further, the small number of specific Hi-C interactions emerge naturally after spurious Hi-C interactions are removed using an ensemble of random conformations under the same volume confinement. These successes suggest that a small number of interactions, not necessarily indistinct chromatin states with unclear biological interpretation *per se*, may be sufficient to generate functional 3D chromatin ensembles.

An Even Smaller Set of Specific Interactions Can Fold Chromatin Loci in Human Cells. The existence of driver interactions was further explored in a study examining mul-

tiple TAD-bounded regions with ≥ 2 super-enhancers in human GM12878 cells (44). Remarkably, only 14–35 specific interactions are sufficient to fold 39 loci (480 kb $- 1.94$ Mb) into Hi-C consistent ensembles at 5 kb resolution ($R = 0.970 \pm 0.003$) (Fig 3c, bottom right). These are enriched with functional associations and active marks (44). They represent $0.024 - 1.3\%$ (median 0.67%) of $2,414 - 62,785$ measured Hi-C interactions, and $0.7 - 11\%$ (median 5.7%) of the $301 - 2,112$ specific interactions (44).

Interestingly, while removal of cohesin complexes and associated loops can abolish TADs (57), looping interactions alone (as defined by HiCCUPS (2)) are insufficient to drive chromatin folding for 10 Drosophila loci analyzed in (45). We expect similar results in the human loci studied in (44), as only 0.1% of the interactions in this smaller set are on domain boundaries or are loops identified by HiCCUPS. These results suggest that looping interactions occurring at TAD boundaries alone do not drive chromatin folding. Rather, a small set of specific interactions can fold chromatin into structural ensembles that naturally exhibit TADs.

While the absolute minimum set of driver interactions is unknown, it is intriguing whether this finding is general for other loci. Further analysis of these key interactions will be worthwhile to define the genomic elements involved and elucidate their functional roles.

TADs Form as a Byproduct of 3D Chromatin Folding Driven by Specific Interactions

Topologically associating domains (TADs) (9, 10) are important structural units in our current understanding of 3D genome organization. However, their origin and role in genome function is of considerable debate (58–61). Minimalistic polymer models have shed some light on these.

Proto-TADs Exist in Random, Volume-Confined Chromatin. When random chromatin polymers are confined within the nuclear volume, 3D domain structures with heightened intra-domain interactions arise naturally (Fig 4a)(38). These domains, called *proto-TADs*, are uniformly distributed along the genome (Fig 4c) (45), and appear frequently: on average, they cover ∼21% of the random chromatin polymer (38). The entropic origin of TAD formation was also pointed out in (62). These results show that volume confinement induces a chromatin folding landscape with a propensity for TAD formation. This may greatly simplify the task of TAD formation during evolution: strategically placed protein factors may alter the folding landscape sufficiently so certain pre-existing proto-TADs are probabilistically favored and become fixed into TADs (Fig 4a).

TAD Formation Is Driven by a Small Number of Specific Interactions. Studying TAD formation during embryo development has provided important insights. For a 1 Mb region of Drosophila, TADs are found in Hi-C maps of later cells (stages $5 - 8$ and S2R+) but not in earlier pre-MBT cells (Fig 4b) (45). This pattern is also observed in modeled 3D conformations, which reproduce Hi-C maps at all stages (Fig 4b). Specifically, TAD-like structures are found in a large portion (60%) of individual S2R+ cells (Fig 4b), consistent with imaging studies (6, 63). However, despite there being no TADs in the Hi-C maps (Fig 4b, left), a large fraction (54%) of predicted single-cell chromatin conformations of early embryos contain TAD-like structures (Fig 4b). This paradox of TAD-like structures appearing in modeled individual cells but absent in population Hi-C can be understood from structural analysis of modeled single-cell chromatin conformations. Random polymers already possess TAD-like structures (Fig 4a) that are distributed uniformly (Fig 4c, top). At early embryo stages, simulations show TAD-like structures appear in individual cells, but with highly variable boundary positions and sizes (Fig 4b,left); reminiscent of random chromatin polymers and hence no TADs are detectable when aggregated. At later embryo stages, Hi-C domain boundaries become sharply peaked at distinct positions; correspondingly, aggregation of simulated single-cell conformations reproduces well-defined TADs with sharp boundaries (Fig 4c). These peaks are likely due to increased insulator binding: at TAD boundaries predicted from 3D conformations, binding of insulator complexes BEAF-32 and Chromator are highly enriched (Fig 4c) (45). Furthermore, boundaries in pre-MBT embryos associated with genes expressed zygotically before MBT (64) are in excellent agreement with boundary probabilities in modeled single-cell conformations (45).

These results suggest that TADs largely arise because of 3D chromatin folding driven by a small number of specific spatial interactions. The polymer folding landscape induced by the nuclear volume is already prone to form proto-TADs. With gradual introduction of strategically placed protein mediators of spatial interaction during development, TADs become favored, fixed, and appear as 2D patterns in population Hi-C. TADs are a byproduct of 3D interactions induced during folding, rather than a cause of 3D genome organization. This is consistent with recent findings that there is no simple relationship between TAD structures and gene expression (58–60).

Minimalistic Single-Cell Chromatin Models Quantify Chromatin Heterogeneity and Uncover Many-Body Units

Hi-C reports only population averages, which hinders the detection of functional cellular subpopulations that may be important for developmental progression (45). Similarly challenging is detection of *many-body units*, three or more genomic regions simultaneously co-interacting within an individual cell, likely important for super-enhancer condensation (65, 66). Experimental single-cell techniques can address these issues but are limited in genomic coverage and sequencing depth (67). Alternatively, minimalistic chromatin models can now invert high-quality, population Hi-C into single-cell conformations whose aggregation is consistent with Hi-C measurements. We next briefly explain how this is accomplished (Fig 4a).

Minimalistic modeling of single-cell 3D chromatin must first predict which specific interactions identified from population Hi-C are present within an individual cell. Once specific contacts in individual cells are identified, 3D conformations are generated following the self-avoiding polymer growth approach (Fig 2a), with a distance restraint placed between loci of assigned specific contacts (Fig 5a). Central to this method is characterizing how certain interactions may cooperatively induce or exclude formation of other interactions and then account for these geometric dependencies when predicting co-occurrence of specific contacts within individual nuclei (44). This is accomplished through extensive polymer simulations using a Bayesian generative model (44). An alternative approach is described in (45). We next discuss the accuracy of these single-cell models as well as insights gained from them.

Single-Cell 3D Chromatin Is Accurately Modeled from Population Hi-C. Individual chromatin conformations generated are in excellent agreement with experimental single-cell measurements (44, 45). This concordance is seen when comparing with both single-cell imaging and Dip-C measurements (Fig 5b-5c). While reconciling FISH and Hi-C measurements is challenging (69), aggregation of individual 3D chromatin conformations can also reproduce the distance distributions of genomic regions by 3D-FISH (45) (Fig 4d). Overall, minimalistic 3D polymer modeling can quantitatively invert statistical patterns in Hi-C heatmaps into highly-
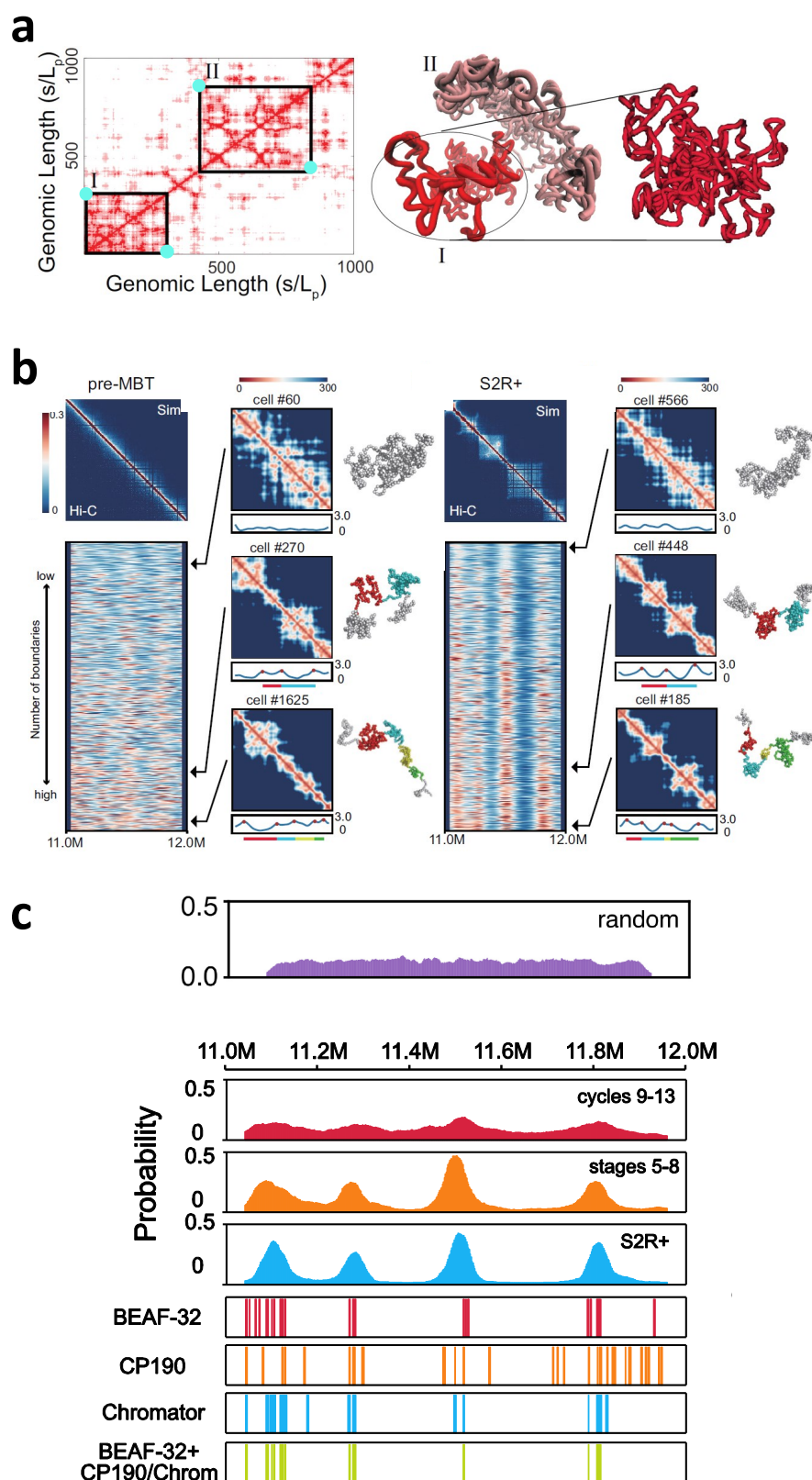
Fig. 4. *TAD formation as a byproduct of 3D chromatin folding driven by specific interactions.* **(a)** Nuclear volume confinement induces a folding landscape where random polymers form domain-like 3D structures (right, domain I and II) which correspond to TAD-like patterns (left) in the 2D distance map (from (38)). Strategically placed protein factors (cyan dots) at the TAD boundaries for loop-formation may sufficiently tilt the chromatin folding landscape such that pre-existing proto-TADs are favored and become fixed; **(b)** Aggregated single-cell conformations of pre-MBT cells do not form population TADs, while those in S2R+ cells do. However, single-cell conformations of pre-MBT cells do form TAD-like structures, as also seen in single-cell spatial-distance heat maps. Boundary strength profiles of 5,000 conformations show most conformations in pre-MBT cells have ≥ 1 domain(s), but their boundaries have highly variable positions. In contrast, S2R+ cells have fixed boundaries; **(c)** Distributions of domain boundary probabilities along genomic position. Domains do exist in random polymers and are uniformly distributed with no preferred positions. Boundary probabilities increase at preferred positions as embryo cells develop. These boundary positions predicted from 3D single-cell conformations are preferentially localized at insulator binding sites such as BEAF-32, Chromator, and CP190.
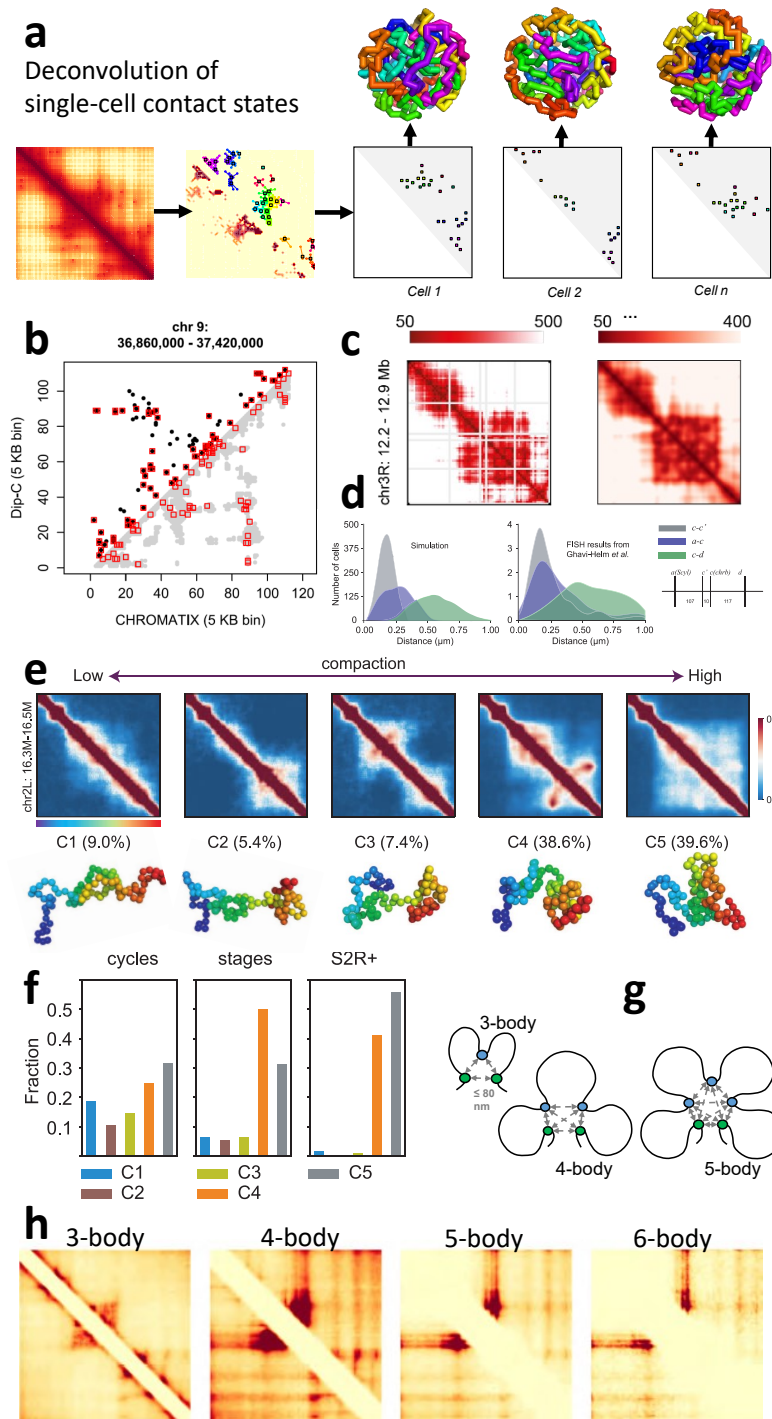
Fig. 5. *Single-cell chromatin conformations folded by specific driver interactions are accurate and reproduce Dip-C, imaging, and FISH studies.* They can quantify chromatin heterogeneity and uncover interacting many-body units. **(a)** Illustration of inverting Hi-C measurements into an ensemble of single-cell chromatin conformations. Specific interactions are first identified from population Hi-C (see also Fig 2). Single-cell contact states (right) are then obtained through Bayesian deconvolution. Each contact state is then used to generate single-cell chromatin conformations, whose aggregation reproduces the poulation Hi-C heatmap. **(b)** Comparison with Dip-C single-cell data (GSE117874) (68). Pairwise contacts from a representative Dip-C single-cell locus (upper triangle, black dots) and the corresponding modeled single-cell locus conformation (lower triangle, gray dots). Contacts present in both models are outlined in red (from (44)). **(c)** Correspondence between distance maps of a modeled single-cell conformation (chr3R: $12.20\check{}12.90$ Mb of Drosophila S2R+ at 2 kb resolution) with a conformation constructed from imaging studies in Mateo et al (63) ($R = 0.75$) (from (45)). **(d)** The distributions of spatial distances between chrb and Scyl genes and two other control regions in Drosophila derived from simulated single-cell chromatin conformations (Left) are highly consistent with DNA FISH measurements (Right) (59)(from (45))). **(e)** 3D single-cell chromatin conformations in a polycomb-repressed region of Drosophila can be grouped into 5 clusters, whose reperative conformations are also shown. **(f)** The distributions of the 5 clusters in cell types of pre-MBT(cycles $9\check{}13$), post-MBT (stages $5\check{}8$) and S2R+. Collectively, these conformations reproduce the measured Hi-C heatmaps. **(g)** Diagrams of 3-, 4-, and 5-body chromatin interactions. All have pairwise Euclidean distance of ≤ 80 nm. The principal loop is the longest loop (in bp) among chromatin regions forming a many-body interaction. Green dots represent anchors of principal loops. **(h)** Principal loop heatmaps of k-body interaction units for the TAD (arrowhead) region containing the SH3KBP1 locus (chr X: $19,560,000\check{}20,170,000$). Principal loop interaction frequencies are captured from deconvolved single-cells after aggregation (from (44, 45)).

informative 3D single-cell chromatin conformations. The modeled conformations are not restricted in genome coverage, thus bridging the gap between population Hi-C and single-cell studies (6, 7, 63, 68, 70). This opens new avenues for modeling how genomic interactions in individual cells are related to cellular functions such as *cis* regulation of gene expression and replication.

Single-Cell Chromatin Models Quantify Chromatin Heterogeneity. A cell population may have functional subpopulations with distinct chromatin conformations. However, it is not possible to directly inspect chromatin heterogeneity and identify subpopulations from Hi-C heatmaps. This requires an ensemble model of single-cell conformations with a properly defined distribution and thorough sampling.

Such single-cell 3D ensembles can now be generated (33, 42, 45). In Drosophila, chromatin heterogeneity of a polycomb-repressed region was quantified using 5.0×10^4 single-cell conformations (45), which are found to form five clusters (Fig 5e). Their occupancies change dramatically as the embryo develops: evenly populated at an early stage, but two clusters dominate at the later S2R+ stage, accounting for $>96\%$ of the conformations (Fig 5f). Conformational subpopulations have also been quantified in single-cell models of the α-globin locus (42): K562 cells are homogeneous with a single cluster accounting for 97% of the cells, whereas GM12878 cells have many clusters and are far more heterogeneous.

The ability to characterize chromatin heterogeneity allows quantification of cellular subpopulations according to their intrinsic 3D chromatin structures. This will facilitate understanding of the relationship between genome structure and function. Important questions to be examined include: whether there are structural scaffolds facilitating *cis*-regulatory control of gene expression, are these scaffolds related to the distinct conformations representing cellular subpopulations, and can such a connection be substantiated with single-cell transcriptomics.

Many-Body Units and Genome Function. Many-body ($\geq$ 3) or multivalent spatial interactions likely play important roles in condensing super-enhancer (SE) regions into a transcriptional apparatus (65, 66). However, Hi-C (1) records only population-averaged, pairwise genomic interactions and therefore obscures which many-body interactions are present within individual cells. Single-cell Hi-C (70), Dip-C (68), Tri-C (71), MC-4C (72), GAM (73), and SPRITE (54) have great promise in uncovering multivalent chromatin interactions, but are currently limited in sequencing depth, genomic coverage, or inability to resolve direct versus indirect spatial interactions. Furthermore, it is challenging to evaluate if these spatial relationships are significant or simply explained by elementary physical effects of polymer connectivity, volume exclusion, and nuclear confinement.

The recent CHROMATIX method allows identification of specific many-body ($\geq$ 3) interactions from Hi-C (44). Extending deep-sampling methods of (38, 42), it folds chromatin through fractal Monte Carlo sampling (44, 45, 74) and utilizes a Bayesian deconvolution approach to identify specific many-body units which are: i) fully spatially interacting, where all participating loci are within a Euclidean distance threshold (Fig 4g), and ii) not arising from aforementioned elementary polymer effects.

The functional landscape of many-body interactions of 39 transcriptionally-active TADs in human GM12878 cells (2) was constructed using CHROMATIX. Many-body units were found to occur frequently in these euchromatic loci. Compared to randomly formed many-bodies, specific many-bodies were enriched in promoters, enhancers, and super-enhancers. In addition, anchor loci of principal loops - the longest spanning loops within many-body units (Fig 4g), exhibit banding patterns when projected as a 2D-heatmap and are enriched in super-enhancers. These results show principal loops likely bridge enhancers and promoters to enable spatial coupling of functional regions. As reported in a recent Drosophila study (75), there is now emerging evidence that analogous multi-way interactions among enhancers and promoters are pre-formed in early embryo development and then become activated or repressed during developmental progression.

Principal loop anchors of specific many-bodies can be directly predicted using 1D biomarkers (44). DNase accessibility was found to be the most predictive biomarker. POLR2A occupancy and nuclear fraction RNA abundance are also important predictors, indicating these specific many-bodies may help facilitate transcription. This is consistent with a subsequent study (76) proposing RNA accumulation as a mechanism of microphase separation, resulting in co-segregation of transcriptionally-active chromatin. CTCF and cohesin subunit RAD21 were modestly predictive of specific principal loops, indicating that while loop extrusion (47, 77) may occur within the examined TAD regions, there are likely other important mechanisms at play in the formation of many-body units enriched in functional elements.

Outlook

Minimalistic self-avoiding 3D chromatin modeling with few tunable parameters has revealed insights into 3D genome organization: 1) chromatin scaling rules are a result of volume-confined polymers, 2) chromatin folding at many loci is driven by a small number of specific interactions, 3) TADs emerge from ensembles of single-cell chromatin folded according to these small number of specific interactions, 4) heterogeneous structural scaffolds help define intrinsic cellular subpopulations whose relative representations are likely important to developmental progression, and 5) the extent and functional roles of many-body spatial interactions in enhancer-rich regions, which are enabled by principal loops. These findings point to several interesting future directions. A first task is to identify specific driver interactions in a wide range of loci. This can be done broadly for different cell types to define the genomic elements involved and their functional roles. Furthermore, it will be highly informative to assess which specific interactions are conserved and which are

unique across different cell types, so the structure-function relationship of chromatin can be characterized at the tissue level.

In addition, there may be better representations of 3D chromatin structure. While the relationship of function and form is of extraordinary importance to understanding 3D genome organization, the prevailing representations of 3D genome structure in terms of TADs and subTADs (78, 79) are inadequate, as they are indirect 2D projections of 3D chromatin and more direct representations may be feasible. These representations could be based on the set of interactions that drive chromatin folding, the principal structures within clusters of 3D conformations, or by characterization of the folding mechanisms among pioneering interactions and other functional interactions enabled by them. Such new representations can stimulate investigation into the most relevant molecular players, their interactions, and the biological processes involved.

A practical utility of such a representation is to allow targeted perturbation of 3D genome organization. For genomic regions that are poorly characterized, current perturbations may involve reversal/deletion of large (≥ 100 kb) genomic intervals such that many elements are disrupted, and their accumulative effects detected. If a small set of critical driver interactions can be identified for a specific locus, it would facilitate experimental perturbations of much smaller intervals to assess the extent of 3D chromatin alteration, and thereby precisely identify regions important for nuclear organization, enabling further investigation of their molecular functions. Such a perturbation strategy may be generally applicable to any arbitrary locus. The small number of specific interactions identified in (42, 45, 74) suggest this may be the case. Furthermore, functional characterization of perturbed regions may allow *a priori* prediction of resulting phenotypic changes which may then be experimentally validated; this is similar to the process used in discovery of *cis* elements controlling DNA replication (80).

Additionally, it should be possible to define the functional landscapes of many-body spatial chromatin interactions in different cells and tissues. With the identification of the participating functional elements and their principal loops, we will gain better understanding on how novel loop anchors can spatially coalesce multiple functional elements to form a coherent apparatus for transcription. A global understanding of many-body interactions is indeed feasible, as demonstrated by the large-scale study of a set of enhancer-rich loci (44).

Lastly, we have a new means to investigate how chromatin structure can provide the physical basis for cellular activity. We can asssess how cellular programs as defined by single-cell transcriptomes correspond to single-cell 3D chromatin conformation. The ability to quantify structural heterogeneity may allow us to delineate functional cellular subpopulations based on their shared chromatin folds. For example, we can assess correspondence between different structural clusters of chromatin and different types of cellular behavior, and whether certain 3D scaffolds are required for specific cellular states. The analysis of time-evolving patterns of chromatin

clusters may further shed light on embryonic development as seen in Drosophila (Fig 4e-4f) (45).

Conclusions

With powerful polymer models that can transform 2D Hi-C interaction maps into ensembles of single-cell 3D chromatin conformations, we expect to gain further understanding of the genome structure-function relationship. Emerging frontiers include: 1) relating ensembles of spatial chromatin structures to cellular phenotypes such as gene expression and gene usage; 2) establishing the structural basis and identifying spatial motifs for different cellular states that are consistent with experimentally measured single-cell transcriptomics; and 3) relating transcriptional heterogeneity to chromatin structural heterogeneity to improve understanding of embryogenesis and cellular reprogramming (81). We expect that integrated modeling and experimental studies will play important roles in investigating these important questions.

Acknowledgements

Summary:

- A minimalistic, self-avoiding 3D chromatin model with no adjustable parameters can transform population Hi-C into high-resolution, single-cell chromatin conformations;

- Chromatin folding at many loci is driven by a small number of specific interactions;

- TADs form as a byproduct of 3D chromatin folding driven by specific interactions;

- Cell subpopulations equipped with different chromatin structural scaffolds are developmental stage-dependent;

- Characterization of the functional landscape and epigenetic marks of many-body units which are simultaneously spatially co-interacting within enhancer-rich, euchromatic regions.

Bibliography

1. E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293, October 2009. doi: 10.1126/science.1181369.

2. Suhas S.P. Rao, Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D. Bochkov, James T. Robinson, Adrian L. Sanborn, Ido Machol, Arina D. Omer, Eric S. Lander, and Erez Lieberman Aiden. A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7):1665–1680, December 2014. doi: 10.1016/j.cell.2014.11.021.

3. Rieke Kempfer and Ana Pombo. Methods for mapping 3d chromosome architecture. *Nature Reviews Genetics*, 21(4):207–226, December 2019. doi: 10.1038/s41576-019-0195-2.

4. J. Dekker. Capturing chromosome conformation. *Science*, 295(5558):1306–1311, February 2002. doi: 10.1126/science.1067799.

5. Alistair N. Boettiger, Bogdan Bintu, Jeffrey R. Moffitt, Siyuan Wang, Brian J. Beliveau, Geoffrey Fudenberg, Maxim Imakaev, Leonid A. Mirny, Chao ting Wu, and Xiaowei Zhuang. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*, 529(7586):418–422, January 2016. doi: 10.1038/nature16496.

6. Bogdan Bintu, Leslie J. Mateo, Jun-Han Su, Nicholas A. Sinnott-Armstrong, Mirae Parker, Seon Kinrot, Kei Yamaya, Alistair N. Boettiger, and Xiaowei Zhuang. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, 362 (6413):eaau1783, October 2018. doi: 10.1126/science.aau1783.

7. Won-Ki Cho, Jan-Hendrik Spille, Micca Hecht, Choongman Lee, Charles Li, Valentin Grube, and Ibrahim I. Cisse. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science*, 361(6400):412–415, June 2018. doi: 10.1126/science.aar4199.

8. Alistair Boettiger and Sedona Murphy. Advances in chromatin imaging at kilobase-scale resolution. *Trends in Genetics*, 36(4):273–287, April 2020. doi: 10.1016/j.tig.2019.12.010.

9. Elphège P. Nora, Bryan R. Lajoie, Edda G. Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, Nynke L. van Berkum, Johannes Meisig, John Sedat, Joost Gribnau, Emmanuel Barillot, Nils Blüthgen, Job Dekker, and Edith Heard. Spatial partitioning of the regulatory landscape of the x-inactivation centre. *Nature*, 485(7398):381–385, April 2012. doi: 10.1038/nature11049.

10. Jesse R. Dixon, Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, and Bing Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380, April 2012. doi: 10.1038/nature11082.

11. Tom Sexton, Eitan Yaffe, Ephraim Kenigsberg, Frédéric Bantignies, Benjamin Leblanc, Michael Hoichman, Hugues Parrinello, Amos Tanay, and Giacomo Cavalli. Three-dimensional folding and functional organization principles of the drosophila genome. *Cell*, 148(3):458–472, February 2012. doi: 10.1016/j.cell.2012.01.010.

12. Quentin Szabo, Axelle Donjon, Ivana Jerković, Giorgio L. Papadopoulos, Thierry Cheutin, Boyan Bonev, Elphège P. Nora, Benoit G. Bruneau, Frédéric Bantignies, and Giacomo Cavalli. Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nature Genetics*, 52(11):1151–1157, October 2020. doi: 10.1038/s41588-020-00716-8.

13. ZhiZhuo Zhang, Guoliang Li, Kim-Chuan Toh, and Wing-Kin Sung. 3d chromosome modeling with semi-definite programming and hi-c data. *Journal of Computational Biology*, 20 (11):831–846, November 2013. doi: 10.1089/cmb.2013.0076.

14. Jonas Paulsen, Monika Sekelja, Anja R. Oldenburg, Alice Barateau, Nolwenn Briand, Erwan Delbarre, Akshay Shah, Anita L. Sørensen, Corinne Vigouroux, Brigitte Buendia, and Philippe Collas. Chrom3d: three-dimensional genome modeling from hi-c and nuclear lamin-genome contacts. *Genome Biology*, 18(1), January 2017. doi: 10.1186/s13059-016-1146-2.

15. Guangxiang Zhu, Wenxuan Deng, Hailin Hu, Rui Ma, Sai Zhang, Jinglin Yang, Jian Peng, Tommy Kaplan, and Jianyang Zeng. Reconstructing spatial organizations of chromosomes through manifold learning. *Nucleic Acids Research*, 46(8):e50–e50, February 2018. doi: 10.1093/nar/gky065.

16. Zhijun Duan, Mirela Andronescu, Kevin Schutz, Sean McIlwain, Yoo Jung Kim, Choli Lee, Jay Shendure, Stanley Fields, C. Anthony Blau, and William S. Noble. A three-dimensional model of the yeast genome. *Nature*, 465(7296):363–367, May 2010. doi: 10.1038/nature08973.

17. Tuan Trieu, Oluwatosin Oluwadare, and Jianlin Cheng. Hierarchical reconstruction of high-resolution 3d models of large chromosomes. *Scientific Reports*, 9(1), March 2019. doi: 10.1038/s41598-019-41369-w.

18. Nan Hua, Harianto Tjong, Hanjun Shin, Ke Gong, Xianghong Jasmine Zhou, and Frank Alber. Producing genome structure populations with the dynamic and automated PGS software. *Nature Protocols*, 13(5):915–926, April 2018. doi: 10.1038/nprot.2018.008.

19. Oluwatosin Oluwadare, Max Highsmith, and Jianlin Cheng. An overview of methods for reconstructing 3-d chromosome and genome structures from hi-c data. *Biological Procedures Online*, 21(1), April 2019. doi: 10.1186/s12575-019-0094-0.

20. Kimberly MacKay and Anthony Kusalik. Computational methods for predicting 3d genomic organization from high-resolution chromosome conformation capture data. *Briefings in Functional Genomics*, 19(4):292–308, April 2020. doi: 10.1093/bfgp/elaa046.

21. Andrea M. Chiariello, Simona Bianco, Andrea Esposito, Luca Fiorillo, Mattia Conte, Ehsan Irani, Francesco Musella, Alex Abraham, Antonella Prisco, and Mario Nicodemi. Physical mechanisms of chromatin spatial organization. *The FEBS Journal*, February 2021. doi: 10.1111/febs.15762.

22. Chris A. Brackey, Davide Marenduzzo, and Nick Gilbert. Mechanistic modeling of chromatin folding to understand function. *Nature Methods*, 17(8):767–775, June 2020. doi: 10.1038/s41592-020-0852-6.

23. Bin Zhang and Peter G. Wolynes. Genomic energy landscapes. *Biophysical Journal*, 112 (3):427–433, February 2017. doi: 10.1016/j.bpj.2016.08.046.

24. M. Barbieri, M. Chotalia, J. Fraser, L.-M. Lavitas, J. Dostie, A. Pombo, and M. Nicodemi. Complexity of chromatin folding is captured by the strings and binders switch model. *Proceedings of the National Academy of Sciences*, 109(40):16173–16178, September 2012. doi: 10.1073/pnas.1204799109.

25. Daniel Jost, Pascal Carrivain, Giacomo Cavalli, and Cédric Vaillant. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Research*, 42(15):9553–9561, August 2014. doi: 10.1093/nar/gku698.

26. Bin Zhang and Peter G. Wolynes. Topology, structures, and energy landscapes of human chromosomes. *Proceedings of the National Academy of Sciences*, 112(19):6062–6067, April 2015. doi: 10.1073/pnas.1506257112.

27. Michele Di Pierro, Bin Zhang, Erez Lieberman Aiden, Peter G. Wolynes, and José N. Onuchic. Transferable model for chromosome architecture. *Proceedings of the National Academy of Sciences*, 113(43):12168–12173, September 2016. doi: 10.1073/pnas.1613607113.

28. Guang Shi, Lei Liu, Changbong Hyeon, and D. Thirumalai. Interphase human chromosome exhibits out of equilibrium glassy dynamics. *Nature Communications*, 9(1), August 2018. doi: 10.1038/s41467-018-05606-6.

29. Yifeng Qi and Bin Zhang. Predicting three-dimensional genome organization with chromatin states. *PLOS Computational Biology*, 15(6):e1007024, June 2019. doi: 10.1371/journal.pcbi.1007024.

30. Yifeng Qi, Alejandro Reyes, Sarah E. Johnstone, Martin J. Aryee, Bradley E. Bernstein, and Bin Zhang. Data-driven polymer model for mechanistic exploration of diploid genome organization. *Biophysical Journal*, 119(9):1905–1916, November 2020. doi:

10.1016/j.bpj.2020.09.009. • A full diploid genome study at 1 Mb resolution, where heteropolymer chromatin conformations governed by an energy function whose parameters were extracted from Hi-C data are generated by molecular dynamics. Both specific interchromosomal interactions and centromere clustering are found to be important for radial positioning of chromosomes.

31. Mattia Conte, Luca Fiorillo, Simona Bianco, Andrea M. Chiariello, Andrea Esposito, and Mario Nicodemi. Polymer physics indicates chromatin folding variability across single-cells results from state degeneracy in phase separation. *Nature Communications*, 11(1), July 2020. doi: 10.1038/s41467-020-17141-4. •• Authors use SBS, a block copolymer model, to simulate phase-separated globule formation within a locus. Binding types are assigned using PRISMR, a Bayesian Monte Carlo algorithm, and they primarily correspond to: i. CTCF/Cohesin, ii. H3K27ac/active marks, iii. H3K27me3/repressive marks, and iv. H4K16ac/other transcription factors. The molecular dynamics simulation among these binding types suggests a folding mechanism for phase separated globule formation and stability. Simulation results are validated against independent distance microscopy data.

32. A Kurz, S Lampel, J E Nickolenko, J Bradl, A Benner, R M Zirbel, T Cremer, and P Lichter. Active and inactive genes localize preferentially in the periphery of chromosome territories. *Journal of Cell Biology*, 135(5):1195–1205, December 1996. doi: 10.1083/jcb.135.5.1195.

33. Ryan R Cheng, Vinicius G Contessoto, Erez Lieberman Aiden, Peter G Wolynes, Michele Di Pierro, and Jose N Onuchic. Exploring chromosomal structural heterogeneity across multiple cell lines. *eLife*, 9, October 2020. doi: 10.7554/elife.60312. ••Authors extend an earlier study of predicting chromatin structure using the MiChroM heteropolymer chromatin method beyond lymphoblastoid GM12878 cells to six other cell types. Results from analysis of DNA paint imaging and simulated chromatin conformations show that chromosome structures are fluid-like and heterogeneous. The authors conclude that there is no single native chromatin structure and that descriptions by a statistical ensemble are required.

34. Antonio B. Oliveira Junior, Vinícius G. Contessoto, Matheus F. Mello, and José N. Onuchic. A scalable computational approach for simulating complexes of multiple chromosomes. *Journal of Molecular Biology*, 433(6):166700, March 2021. doi: 10.1016/j.jmb.2020.10.034.

35. Luca Fiorillo, Simona Bianco, Andrea M. Chiariello, Mariano Barbieri, Andrea Esposito, Carlo Annunziatella, Mattia Conte, Alfonso Corrado, Antonella Prisco, Ana Pombo, and Mario Nicodemi. Inference of chromosome 3d structures from GAM data by a physics computational approach. *Methods*, 181-182:70–79, October 2020. doi: 10.1016/j.ymeth.2019.09.018.

36. Luca Fiorillo, Francesco Musella, Mattia Conte, Rieke Kempfer, Andrea M. Chiariello, Simona Bianco, Alexander Kukalev, Ibai Irastorza-Azcarate, Andrea Esposito, Alex Abraham, Antonella Prisco, Ana Pombo, and Mario Nicodemi. Comparison of the hi-c, GAM and SPRITE methods using polymer models of chromatin. *Nature Methods*, 18(5):482–490, May 2021. doi: 10.1038/s41592-021-01135-1.

37. M Madan Babu, Nicholas M Luscombe, L Aravind, Mark Gerstein, and Sarah A Teichmann. Structure and evolution of transcriptional regulatory networks. *Current Opinion in Structural Biology*, 14(3):283–291, June 2004. doi: 10.1016/j.sbi.2004.05.004.

38. Gamze Gürsoy, Yun Xu, Amy L. Kenter, and Jie Liang. Spatial confinement is a major determinant of the folding landscape of human chromosomes. *Nucleic Acids Research*, 42 (13):8223–8230, July 2014. doi: 10.1093/nar/gku462.

39. Elizabeth H. Finn, Gianluca Pegoraro, Hugo B. Brandão, Anne-Laure Valton, Marlies E. Oomen, Job Dekker, Leonid Mirny, and Tom Misteli. Extensive heterogeneity and intrinsic variation in spatial genome organization. *Cell*, 176(6):1502–1515.e10, March 2019. doi: 10.1016/j.cell.2019.01.020.

40. Kit Fun Lau and Ken A. Dill. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules*, 22(10):3986–3997, October 1989. doi: 10.1021/ma00200a030.

41. Ken A. Dill, Sarina Bromberg, Kaizhi Yue, Hue Sun Chan, Klaus M. Ftebig, David P. Yee, and Paul D. Thomas. Principles of protein folding - a perspective from simple exact models. *Protein Science*, 4(4):561–602, December 2008. doi: 10.1002/pro.5560040401.

42. Gamze Gürsoy, Yun Xu, Amy L. Kenter, and Jie Liang. Computational construction of 3d chromatin ensembles and prediction of functional interactions of alpha-globin locus from 5c data. *Nucleic Acids Research*, 45(20):11547–11558, September 2017. doi: 10.1093/nar/gkx784.

43. Gamze Gürsoy, Yun Xu, and Jie Liang. Spatial organization of the budding yeast genome in the cell nucleus and identification of specific chromatin interactions from multi-chromosome constrained chromatin model. *PLOS Computational Biology*, 13(7):e1005658, July 2017. doi: 10.1371/journal.pcbi.1005658.

44. Alan Perez-Rathke, Qiu Sun, Boshen Wang, Valentina Boeva, Zhifeng Shao, and Jie Liang. CHROMATIX: computing the functional landscape of many-body chromatin interactions in transcriptionally active loci from deconvolved single cells. *Genome Biology*, 21(1), January 2020. doi: 10.1186/s13059-019-1904-z. ••Authors develop CHROMATIX, a Bayesian generative model incorporating extensive polymer simulations, to deconvolve population Hi-C into ensembles of individual 3D chromatin conformations. This approach is applied to 39 super-enhancer (SE) loci in GM12878 cells. Individual conformations are consistent with independent Dip-C data, and simulated ensembles are consistent with Hi-C and independent SPRITE data. Authors use the modeled, single-locus data to characterize the functional landscape of many-body units, higher-order interactions co-occurring within a single nucleus, and find increased participation of SE, enhancers, and promoters in the principle loops, the longest spanning loops in base pairs, of specific many-body units.

45. Qiu Sun, Alan Perez-Rathke, Daniel M. Czajkowsky, Zhifeng Shao, and Jie Liang. High-resolution single-cell 3d-models of chromatin ensembles during drosophila embryogenesis. *Nature Communications*, 12(1), January 2021. doi: 10.1038/s41467-020-20490-9. ••Authors use Hi-C and minimalistic polymer modeling to reconstruct the 3D chromatin fold in individual cells during *Drosophila* embryogenesis. Using the specific interactions consisting of 5-7% of Hi-C interactions, the modeled single-cell chromatin conformations reproduce single-cell imaging experiments and FISH-measured distance distributions. The single-cell conformations also quantify chromatin structural heterogeneity, and the relative representation of distinct structural subpopulations is likely important for developmental progression. Results also indicate that TADs are a consequence of 3D chromatin folding driven by a small number of specific interactions.

46. J. Mateos-Langerak, M. Bohn, W. de Leeuw, O. Giromus, E. M. M. Manders, P. J. Verschure, M. H. G. Indemans, H. J. Gierman, D. W. Heermann, R. van Driel, and S. Goetze. Spatially confined folding of chromatin in the interphase nucleus. *Proceedings of the National Academy of Sciences*, 106(10):3812–3817, February 2009. doi: 10.1073/pnas.0809501106.

47. Adrian L. Sanborn, Suhas S. P. Rao, Su-Chen Huang, Neva C. Durand, Miriam H. Huntley, Andrew I. Jewett, Ivan D. Bochkov, Dharmaraj Chinnappan, Ashok Cutkosky, Jian Li, Kristopher P. Geeting, Andreas Gnirke, Alexandre Melnikov, Doug McKenna, Elena K. Stamenova, Eric S. Lander, and Erez Lieberman Aiden. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences*, 112(47):E6456–E6465, October 2015. doi: 10.1073/pnas.1518552112.

48. Predrag Jevtić, Lisa J Edens, Lidija D Vuković, and Daniel L Levy. Sizing and shaping the nucleus: mechanisms and significance. *Current Opinion in Cell Biology*, 28:16–27, June 2014. doi: 10.1016/j.ceb.2014.01.003.

49. H. Tjong, K. Gong, L. Chen, and F. Alber. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Research*, 22(7):1295–1305, May 2012. doi: 10.1101/gr.129437.111.

50. Hua Wong, Hervé Marie-Nelly, Sébastien Herbert, Pascal Carrivain, Hervé Blanc, Romain Koszul, Emmanuelle Fabre, and Christophe Zimmer. A predictive computational model of the dynamic 3d interphase yeast nucleus. *Current Biology*, 22(20):1881–1890, October 2012. doi: 10.1016/j.cub.2012.07.069.

51. Irina Solovei, Audrey S. Wang, Katharina Thanisch, Christine S. Schmidt, Stefan Krebs, Monika Zwerger, Tatiana V. Cohen, Didier Devys, Roland Foisner, Leo Peichl, Harald Herrmann, Helmut Blum, Dieter Engelkamp, Colin L. Stewart, Heinrich Leonhardt, and Boris Joffe. LBR and lamin a/c sequentially tether peripheral heterochromatin and inversely regulate differentiation. *Cell*, 152(3):584–598, January 2013. doi: 10.1016/j.cell.2013.01.009.

52. Bas van Steensel and Andrew S. Belmont. Lamina-associated domains: Links with chromosome architecture, heterochromatin, and gene repression. *Cell*, 169(5):780–791, May 2017. doi: 10.1016/j.cell.2017.04.022.

53. Nolwenn Briand and Philippe Collas. Lamina-associated domains: peripheral matters and internal affairs. *Genome Biology*, 21(1), April 2020. doi: 10.1186/s13059-020-02003-5.

54. Sofia A. Quinodoz, Noah Ollikainen, Barbara Tabak, Ali Palla, Jan Marten Schmidt, Elizabeth Detmer, Mason M. Lai, Alexander A. Shishkin, Prashant Bhat, Yodai Takei, Vickie Trinh, Erik Aznauryan, Pamela Russell, Christine Cheng, Marko Jovanovic, Amy Chow, Long Cai, Patrick McDonel, Manuel Garber, and Mitchell Guttman. Higher-order inter-chromosomal hubs shape 3d genome organization in the nucleus. *Cell*, 174(3):744–757.e24, July 2018. doi: 10.1016/j.cell.2018.05.024.

55. Andrew S Belmont. Large-scale chromatin organization: the good, the surprising, and the still perplexing. *Current Opinion in Cell Biology*, 26:69–78, February 2014. doi: 10.1016/j.ceb.2013.10.002.

56. Luca Giorgetti, Rafael Galupa, Elphège P. Nora, Tristan Piolot, France Lam, Job Dekker, Guido Tiana, and Edith Heard. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*, 157(4):950–963, May 2014. doi: 10.1016/j.cell.2014.03.025.

57. Wibke Schwarzer, Nezar Abdennur, Anton Goloborodko, Aleksandra Pekowska, Geoffrey Fudenberg, Yann Loe-Mie, Nuno A Fonseca, Wolfgang Huber, Christian H. Haering, Leonid Mirny, and Francois Spitz. Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, 551(7678):51–56, September 2017. doi: 10.1038/nature24281.

58. Suhas S.P. Rao, Su-Chen Huang, Brian Glenn St Hilaire, Jesse M. Engreitz, Elizabeth M. Perez, Kyong-Rim Kieffer-Kwon, Adrian L. Sanborn, Sarah E. Johnstone, Gavin D. Bascom, Ivan D. Bochkov, Xingfan Huang, Muhammad S. Shamim, Jaewoon Shin, Douglass Turner, Ziyi Ye, Arina D. Omer, James T. Robinson, Tamar Schlick, Bradley E. Bernstein, Rafael Casellas, Eric S. Lander, and Erez Lieberman Aiden. Cohesin loss eliminates all loop domains. *Cell*, 171(2):305–320.e24, October 2017. doi: 10.1016/j.cell.2017.09.026.

59. Yad Ghavi-Helm, Aleksander Jankowski, Sascha Meiers, Rebecca R. Viales, Jan O. Korbel, and Eileen E. M. Furlong. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nature Genetics*, 51(8):1272–1282, July 2019. doi: 10.1038/s41588-019-0462-3. •• The authors probed the genome-wide structure-function relationship of Drosophila genomes using balancer chromosomes containing extensive rearrangements, and found that architectural features of TADs and loops are generally disconnected with gene activity, affecting only 10-12% of tested genes. Their results suggest that TADs may arise as a consequence of sequence features and gene activity, and multiple enhancers may coordinate to act on a promoter,.

60. Yad Ghavi-Helm. Functional consequences of chromosomal rearrangements on gene expression: Not so deleterious after all? *Journal of Molecular Biology*, 432(3):665–675, February 2020. doi: 10.1016/j.jmb.2019.09.010.

61. Mustafa Mir, Wendy Bickmore, Eileen E. M. Furlong, and Geeta Narlikar. Chromatin topology, condensates and gene regulation: shifting paradigms or just a phase? *Development*, 146(19), September 2019. doi: 10.1242/dev.182766.

62. Paula A. Vasquez, Caitlin Hult, David Adalsteinsson, Josh Lawrimore, Mark G. Forest, and Kerry Bloom. Entropy gives rise to topologically associating domains. *Nucleic Acids Research*, 44(12):5540–5549, June 2016. doi: 10.1093/nar/gkw510.

63. Leslie J. Mateo, Sedona E. Murphy, Antonina Hafner, Isaac S. Cinquini, Carly A. Walker, and Alistair N. Boettiger. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature*, 568(7750):49–54, March 2019. doi: 10.1038/s41586-019-1035-4.

64. Clemens B. Hug, Alexis G. Grimaldi, Kai Kruse, and Juan M. Vaquerizas. Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell*, 169(2):216–228.e19, April 2017. doi: 10.1016/j.cell.2017.03.024.

65. Benjamin R. Sabari, Alessandra Dall'Agnese, Ann Boija, Isaac A. Klein, Eliot L. Coffey, Krishna Shrinivas, Brian J. Abraham, Nancy M. Hannett, Alicia V. Zamudio, John C. Manteiga, Charles H. Li, Yang E. Guo, Daniel S. Day, Jurian Schuijers, Eliza Vasile, Sohail Malik, Denes Hnisz, Tong Ihn Lee, Ibrahim I. Cisse, Robert G. Roeder, Phillip A. Sharp, Arup K. Chakraborty, and Richard A. Young. Coactivator condensation at super-enhancers links phase separation and gene control. *Science*, 361(6400):eaar3958, June 2018. doi: 10.1126/science.aar3958.

66. Denes Hnisz, Krishna Shrinivas, Richard A. Young, Arup K. Chakraborty, and Phillip A.

67. Sharp. A phase separation model for transcriptional control. *Cell*, 169(1):13–23, March 2017. doi: 10.1016/j.cell.2017.02.007.

67. Derek M Bickhart, Benjamin D Rosen, Sergey Koren, Brian L Sayre, Alex R Hastie, Saki Chan, Joyce Lee, Ernest T Lam, Ivan Liachko, Shawn T Sullivan, Joshua N Burton, Heather J Huson, John C Nystrom, Christy M Kelley, Jana L Hutchison, Yang Zhou, Jiajie Sun, Alessandra Crisà, F Abel Ponce de León, John C Schwartz, John A Hammond, Geoffrey C Waldbieser, Steven G Schroeder, George E Liu, Maitreya J Dunham, Jay Shendure, Tad S Sonstegard, Adam M Phillippy, Curtis P Van Tassell, and Timothy P L Smith. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature Genetics*, 49(4):643–650, March 2017. doi: 10.1038/ng.3802.

68. Longzhi Tan, Dong Xing, Chi-Han Chang, Heng Li, and X. Sunney Xie. Three-dimensional genome structures of single diploid human cells. *Science*, 361(6405):924–928, August 2018. doi: 10.1126/science.aat5641.

69. Geoffrey Fudenberg and Maxim Imakaev. FISH-ing for captured contacts: towards reconciling FISH and 3c. *Nature Methods*, 14(7):673–678, June 2017. doi: 10.1038/nmeth.4329.

70. Takashi Nagano, Yaniv Lubling, Tim J. Stevens, Stefan Schoenfelder, Eitan Yaffe, Wendy Dean, Ernest D. Laue, Amos Tanay, and Peter Fraser. Single-cell hi-c reveals cell-to-cell variability in chromosome structure. *Nature*, 502(7469):59–64, September 2013. doi: 10.1038/nature12593.

71. A. Marieke Oudelaar, James O. J. Davies, Lars L. P. Hanssen, Jelena M. Telenius, Ron Schwessinger, Yu Liu, Jill M. Brown, Damien J. Downes, Andrea M. Chiariello, Simona Bianco, Mario Nicodemi, Veronica J. Buckle, Job Dekker, Douglas R. Higgs, and Jim R. Hughes. Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nature Genetics*, 50(12):1744–1751, October 2018. doi: 10.1038/s41588-018-0253-2.

72. Carlo Vermeulen, Amin Allahyar, Britta A. M. Bouwman, Peter H. L. Krijger, Marjon J. A. M. Verstegen, Geert Geeven, Christian Valdes-Quezada, Ivo Renkens, Roy Straver, Wigard P. Kloosterman, Jeroen de Ridder, and Wouter de Laat. Multi-contact 4c: long-molecule sequencing of complex proximity ligation products to uncover local cooperative and competitive chromatin topologies. *Nature Protocols*, 15(2):364–397, January 2020. doi: 10.1038/s41596-019-0242-7.

73. Robert A. Beagrie, Antonio Scialdone, Markus Schueler, Dorothee C. A. Kraemer, Mita Chotalia, Sheila Q. Xie, Mariano Barbieri, Inês de Santiago, Liron-Mark Lavitas, Miguel R. Branco, James Fraser, Josée Dostie, Laurence Game, Niall Dillon, Paul A. W. Edwards, Mario Nicodemi, and Ana Pombo. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature*, 543(7646):519–524, March 2017. doi: 10.1038/nature21411.

74. Alan Perez-Rathke, Samira Mali, Lin Du, and Jie Liang. Alterations in chromatin folding patterns in cancer variant-enriched loci. In *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE, May 2019. doi: 10.1109/bhi.2019.8834565.

75. Sergio Martin Espinola, Markus Götz, Maelle Bellec, Olivier Messina, Jean-Bernard Fiche, Christophe Houbron, Matthieu Dejean, Ingolf Reim, Andrés M. Cardozo Gizzi, Mounia Lagha, and Marcelo Nollmann. Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early drosophila development. *Nature Genetics*, 53(4):477–486, April 2021. doi: 10.1038/s41588-021-00816-z. •• Using microscopy to capture a time course of spatial gene expression patterns in Drosophila, authors show *cis*-regulatory modules (CRMs) such as enhancers can spatially coalesce to form multi-way hubs. The formation of these CRM hubs occurs in early development and precedes gene transcription and TAD formation.

76. Lennart Hilbert, Yuko Sato, Ksenia Kuznetsova, Tommaso Bianucci, Hiroshi Kimura, Frank Jülicher, Alf Honigmann, Vasily Zaburdaev, and Nadine L. Vastenhouw. Transcription organizes euchromatin via microphase separation. *Nature Communications*, 12(1), March 2021. doi: 10.1038/s41467-021-21589-3. • Authors use a combination of super-resolution microscopy and lattice-based computational simulation to propose RNA accumulation as a driver for formation of transcriptional pockets in actively transcribed euchromatin. The formation of these pockets results in spatial displacement of transcriptionally inactive chromatin. Authors characterize this process as microphase separation and propose full phase separation is likely inhibited by RNA pol tethering of RNA to actively transcribed euchromatin.

77. Geoffrey Fudenberg, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A. Mirny. Formation of chromosomal domains by loop extrusion. *Cell Reports*, 15(9):2038–2049, May 2016. doi: 10.1016/j.celrep.2016.04.085.

78. Jennifer E. Phillips-Cremins, Michael E.G. Sauria, Amartya Sanyal, Tatiana I. Gerasimova, Bryan R. Lajoie, Joshua S.K. Bell, Chin-Tong Ong, Tracy A. Hookway, Changying Guo, Yuhua Sun, Michael J. Bland, William Wagstaff, Stephen Dalton, Todd C. McDevitt, Ranjan Sen, Job Dekker, James Taylor, and Victor G. Corces. Architectural protein subclasses shape 3d organization of genomes during lineage commitment. *Cell*, 153(6):1281–1295, June 2013. doi: 10.1016/j.cell.2013.04.053.

79. Jonathan A. Beagan and Jennifer E. Phillips-Cremins. On the existence and functionality of topologically associating domains. *Nature Genetics*, 52(1):8–16, January 2020. doi: 10.1038/s41588-019-0561-1.

80. Jiao Sima, Abhijit Chakraborty, Vishnu Dileep, Marco Michalski, Kyle N. Klein, Nicolas P. Holcomb, Jesse L. Turner, Michelle T. Paulsen, Juan Carlos Rivera-Mulia, Claudia Trevilla-Garcia, Daniel A. Bartlett, Peiyao A. Zhao, Brian K. Washburn, Elphège P. Nora, Katerina Kraft, Stefan Mundlos, Benoit G. Bruneau, Mats Ljungman, Peter Fraser, Ferhat Ay, and David M. Gilbert. Identifying cis elements for spatiotemporal control of mammalian DNA replication. *Cell*, 176(4):816–830.e18, February 2019. doi: 10.1016/j.cell.2018.11.036.

81. Constantinos Chronis, Petko Fiziev, Bernadett Papp, Stefan Butz, Giancarlo Bonora, Shan Sabri, Jason Ernst, and Kathrin Plath. Cooperative binding of transcription factors orchestrates reprogramming. *Cell*, 168(3):442–459.e20, January 2017. doi: 10.1016/j.cell.2016.12.016.