Mapping the semi-nested community structure of 3D chromosome contact networks

Dolores Bernenko*

Department of Physics, Integrated Science Lab, Umeå University, SE-901 87 Umeå, Sweden.

Sang Hoon Lee†

Department of Physics and Research Institute of Natural Science, Gyeongsang National University, Jinju 52828, Korea and
Future Convergence Technology Research Institute, Gyeongsang National University, Jinju 52849, Korea

Per Stenberg‡

EMG, Umeå University, SE-901 87 Umeå, Sweden

Ludvig Lizana§

Integrated Science Lab, Department of Physics, Umeå University, SE-901 87 Umeå, Sweden

(Dated: June 25, 2022)

Mammalian DNA folds into 3D structures that facilitate and regulate genetic processes such as transcription, DNA repair, and epigenetics. Several insights derive from chromosome capture methods, such as Hi-C, which allow researchers to construct contact maps depicting 3D interactions among all DNA segment pairs. These maps show a complex cross-scale organization spanning megabase-pair compartments to short-ranged DNA loops. To better understand the organizing principles, several groups analyzed Hi-C data assuming a Russian-doll-like nested hierarchy where DNA regions of similar sizes merge into larger and larger structures. Apart from being a simple and appealing description, this model explains, e.g., the omnipresent chequerboard pattern seen in Hi-C maps, known as A/B compartments, and foreshadows the co-localization of some functionally similar DNA regions. However, while successful, this model is incompatible with the two competing mechanisms that seem to shape a significant part of the chromosomes' 3D organization: loop extrusion and phase separation. This paper aims to map out the chromosome's actual folding hierarchy from empirical data. To this end, we take advantage of Hi-C experiments and treat the measured DNA-DNA interactions as a weighted network. From such a network, we extract 3D communities using the generalized Louvain algorithm. This algorithm has a resolution parameter that allows us to scan seamlessly through the community size spectrum, from A/B compartments to topologically associated domains (TADs). By constructing a hierarchical tree connecting these communities, we find that chromosomes are more complex than a perfect hierarchy. Analyzing how communities nest relative to a simple folding model, we found that chromosomes exhibit a significant portion of nested and non-nested community pairs alongside considerable randomness. In addition, by examining nesting and chromatin types, we discovered that nested parts are often associated with actively transcribed chromatin. These results highlight that cross-scale relationships will be essential components in models aiming to reach a deep understanding of the causal mechanisms of chromosome folding.

I. INTRODUCTION

Mammalian genomes fold into a network of 3D structures that facilitate and regulate genetic processes such as transcription, DNA repair, and epigenetics[1;2]. Most discoveries derive from chromosome capture methods, such as Hi-C, which measure the number of contacts between DNA segment pairs and allow researchers to construct genome-wide 3D contact maps[3–5]. These maps show that chromosomes comprise a spectrum of 3D structures spanning a range of scales: megabase-scale A/B compartments, sub-megabase-scale Topologically Associated Domains (TADs), and short-ranged loops. Some of these structures are associated with epigenetic marks, active genes, and chromatin remodelers, such as CCCTC-binding factors (CTCF), cohesin complexes, and CP190[6–9].

At first glance, Hi-C maps appear hierarchical where DNA regions sharing high contact counts fold into larger and larger structures. This scheme is appealing because it proposes a simple folding mechanism leading to a densely packed DNA without over-entanglement. It also predicts the existence of alternating megabase-sized 3D structures appearing in most Hi-C maps as plaid patterns[3;6;10]. More specifically, TADs tend to aggregate into sub-compartments (denoted A1, A2, B1, ..., B4)[11].

This folding scheme also posits that chromosomes form

*Electronic address: dolores.bernenko@umu.se
†Electronic address: lshlj82@gnu.ac.kr
‡Electronic address: per.stenberg@umu.se
§Electronic address: ludvig.lizana@umu.se

a perfect hierarchy. In other words, once two DNA regions join, such as two TADs, they remain in the same super-structure throughout the upstream folding hierarchy. This idea is the keystone assumption in several studies[12–15]. While it can explain how A/B compartments form and foreshadows the co-localization of some functionally similar DNA regions, critical observations question the basic idea.

First, 3D communities are not necessarily contiguous DNA segments[16]. Assembling such disconnected communities into larger and larger building blocks leads inevitably to a non-perfect hierarchy. Second, if the hierarchy is perfect, it suggests that similar folding mechanisms act across several scales. However, this conclusion is inconsistent with the competing mechanisms that seem to form TADs and A/B compartments: loop extrusion and phase-separation[17;18]. Third, in a recent paper[19], researchers fit a Gaussian polymer model to Hi-C data and recovered several established sub-structures—TADs, subTADs, A/B compartments etc.—showing that they were not perfectly hierarchical.

This paper aims to unveil the chromosomes' hierarchical folding by charting the actual cross-scale folding relationships from empirical data. In particular, we use data from Hi-C experiments that we recast into a weighted network of 3D interactions and use tools from network science to find the optimal community assembly while scanning through the network's layers of organization. By mapping the hierarchical relationships between these assemblies, we find that some nest, others segregate, and some are not significantly different from random. To better understand these results, we propose a minimal folding model mixing perfect and random nesting. We also relate community nesting to established chromatin states. We discovered that communities associated with active transcription are more distinct and show significant nesting relative to the chromosome-wide average.

II. METHOD

A. Hi-C data treatment

We used Hi-C data for the human cell line GM12878 (B-lymphoblastoid)[4], downloaded from the GEO database[20]. We used the MAPQG0 data set in our analysis at 100 kilobase-pair (kb) resolution. Stored in matrix form, the Hi-C data contains the pairwise contact counts between DNA loci. We omit inter-chromosome contacts due to their low signal-to-noise ratio[21;22].

We consider the Hi-C data as a DNA contact network. Each network node represents a 100 kb DNA segment and the link weights are proportional to the number of measured Hi-C contacts. We use network methods (the generalized Louvain method, see Sec. II.B), to extract communities harbouring densely connected nodes that maintain less contacts with the rest of the network.

Before investigating the community structure, we normalized the raw Hi-C counts to reduce biases and to make fair comparisons between chromosomes that may vary in size up to one order of magnitude. To normalize the Hi-C data, we use the Knight-Ruiz (KR) matrix balancing[23] implemented in gcMapExplorer[24].

B. The GenLouvain algorithm—detecting 3D communities in Hi-C data

To find communities in the Hi-C contact network, we use the generalized Louvain method (GenLouvain)[25;26]. It is a community detection method that takes advantage of so-called modularity-maximization to find the nodes' optimal community division in a network. By "optimal," we mean the community assembly that maximizes the number of internal contacts, measured by a so-called modularity function, with respect to a null hypothesis. A common null model is random rewiring keeping the node degree fixed. GenLouvain is a greedy optimization algorithm that starts with single-node communities and then searches for the optimal solution by generating trial node agglomerates and evaluating the modularity function.

One critical feature of GenLouvain is its resolution (or scale) parameter γ. This parameter allows us to sweep through scales of the network and probe the network's community spectrum. Mathematically, γ is a part of the modularity function $\mathcal{M}$, defined as

$$\mathcal{M} = \frac{1}{2m} \sum_{i \neq j} \left[\left(A_{ij} - \gamma P_{ij}^{\mathrm{FG}} \right) \delta(g_i, g_j) \right] , \qquad (1)$$

where, A_{ij} is the adjacency matrix representing the weight of the edge between nodes i and j, and the summand is counted only if i and j belong to a same community, thus the Kronecker delta $\delta(g_i, g_j)$. In our case, A_{ij} corresponds to the KR-normalized Hi-C matrix.

Furthermore, the second term of the summand in Eq. (1) represents our null hypothesis for the network's "background" connectivity. Building on previous work[16], we use the so-called fractal-globule (FG) null-model term P_{ij}^{FG} that assumes that the average interaction strength between two DNA segments, i and j, decays as a power-law with exponent -1. The FG null model term is

$$P_{ij}^{\mathrm{FG}} = \frac{2m k_i k_j |i-j|^{-1}}{\sum_{i' \neq j'} k_{i'} k_{j'} |i'-j'|^{-1}} , \qquad (2)$$

where the strength k_i represents the sum of weights around node i, $2m = \sum_i k_i$ is a normalization constant, and $\propto 1/|i-j|$ is the expected amount of reduced interaction as a function of the one-dimensional distance separating nodes i and j. This decay follows the fractal-globule scaling[16;27;28] that matches with chromatin contact decay in Hi-C experiments[3].

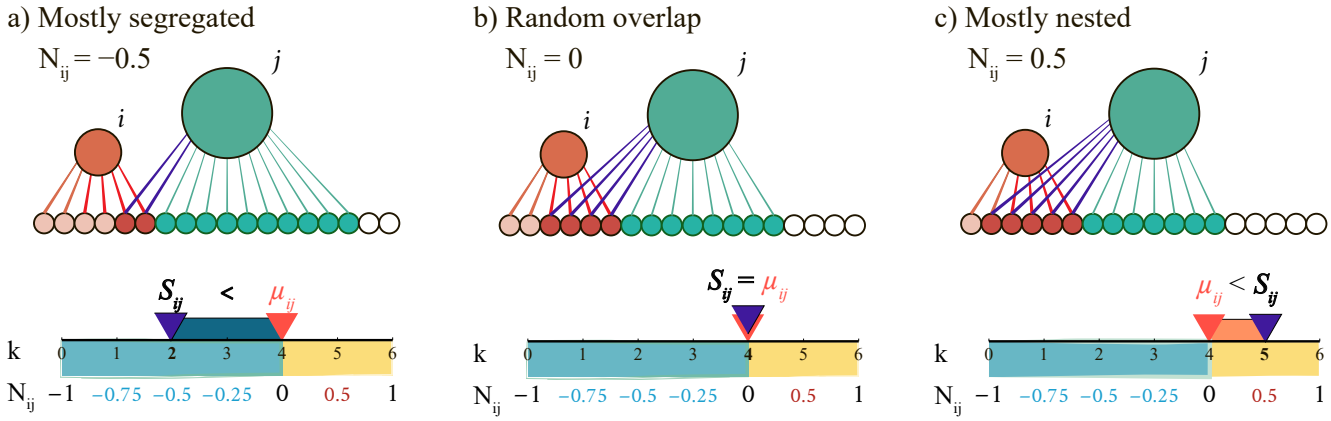

FIG. 1 Three examples of nestedness (N_{ij}) in a simple bipartite network. The networks in panels (a)–(c) have the same number of nodes in each layer—18 domains (small circles) and two communities (i and j, large circles)—and the same number of links (18) but connected differently to achieve varying nestedness. Below each network, we illustrate how we calculate N_{ij} using Eqs. (4)–(7). On the horizontal k-axis, we indicate the number of shared nodes S_{ij} for the community pair and the expected overlap μ_{ij} calculated from Eq. (3). The panels (a)–(c) show three essential N_{ij} regimes ($\mu_{ij} = 4$ for all of the cases). (a) Mostly segregated ($S_{ij} < \mu_{ij}$, $N_{ij} = -0.5$). Because $S_{ij} = 2$ and $\mu_{ij} = 4$, the i and j communities are half-way to full segregation. We illustrate this with a dark-blue stripe covering half of the $0 \leq k \leq 4 = \mu_{ij}$ range. (b) Random overlap ($S_{ij} = \mu_{ij}$, $N_{ij} = 0$). The number of shared nodes equals the random expectation. (c) Mostly nested ($S_{ij} > \mu_{ij}$, $N_{ij} = 0.5$). Here i and j share one domain more than expected ($S_{ij} = 5$). This yields $N_{ij} = 0.5$ because their overlap is at the midpoint between the random and the maximum overlap ($S_{ij} = 6$) that would result in ideal nesting ($N_{ij} = 1$). We illustrate this with the orange stripe spanning half of the range $\mu_{ij}(= 4) \leq k \leq 6$. This example shows that N_{ij} measures the relative overlap compared to what is achievable given the link density rather than absolute numbers.

C. Network nestedness

By varying the resolution parameter γ embedded in the GenLouvain algorithm, we scan through the scales of chromosomes' 3D organization. While scanning, we keep track of uninterrupted DNA segments—we will refer to these segments as *domains* in Sec. III—and how they distribute between the communities as the scale changes. This allows us to chart cross-scale folding relationships.

To better understand these relationships and quantify the deviations from a perfect hierarchy, we use an approach developed for ecological networks[29;30]. Designed for interacting species pairs, say plants and pollinators, this approach rests on a nestedness metric, called N_{ij}, measuring how many plants two pollinators have in common compared to a random benchmark. In our case, we track how many DNA segments (domains) two 3D communities share, given that they appear at different hierarchical levels (different γ values)—by construction, two communities at the same hierarchical level do not share any domains. We illustrate the philosophy behind N_{ij} in Fig. 1.

The N_{ij} metric benchmarks the community overlaps to a combinatorial link redistribution, normalized to vary between -1 and $+1$. These endpoints indicate complete network segregation ($N_{ij} = -1$) or perfect nesting ($N_{ij} = +1$). When perfectly nested, the larger community engulfs the domains in the smaller one. When completely segregated, the communities do not share any domains[1]. In a perfect hierarchy, like a phylogenetic tree, the nestedness is either $+1$ or -1, indicating either full nesting or complete segregation. But in more complex hierarchies, N_{ij} takes any value between these two extremes because the communities may share more or fewer domains relative to a random overlap. We note that N_{ij} is normalized so that the midpoint $N_{ij} = 0$ represents random overlap and that $N_{ij} = \pm x$ indicates the same relative proportion x of segregation or nesting. We exemplify this property in Fig. 1 using a small bipartite network having varying nestedness: (a) mostly segregated ($N_{ij} = -0.5$), (b) random overlap ($N_{ij} = 0$), and (c) mostly nested ($N_{ij} = 0.5$).

To calculate N_{ij}, we go through several steps. First, we extract the overlap S_{ij} between two communities i and j from data—we study nestedness in empirical (Hi-C-derived) and simulated data. Second, we calculate the expected overlap μ_{ij} assuming a random arrangement. Denoting d_i as community i's internal number of domains, μ_{ij} is[29]

$$\mu_{ij} = \sum_{k=1}^{\min(d_i,d_j)} \frac{k\binom{n}{k}\binom{n-k}{d_j-k}\binom{n-d_j}{d_i-k}}{\binom{n}{d_j}\binom{n}{d_i}}, \qquad (3)$$

[1] However, we can get $N_{ij} = -1$ in rare cases even if there is some domain overlap. This happens when $d_i + d_j > n$ in Eqs. (3)-(7).

where n and k are the total and shared number of domains. Next, we shift S_{ij} by μ_{ij} to center the expected overlap for random arrangement at zero and normalize so that $N_{ij} \in [-1, 1]$:

$$N_{ij} = \frac{S_{ij} - \mu_{ij}}{\Omega_{ij} \min(d_i, d_j)}\,, \tag{4}$$

where Ω_{ij} is the maximum or minimum achievable overlap, depending on if $S_{ij} > \mu_{ij}$ or $S_{ij} < \mu_{ij}$. In these cases, we calculate Ω_{ij} as

(a) $S_{ij} > \mu_{ij}$

$$\Omega_{ij} = \frac{\min(d_i, d_j) - \mu_{ij}}{\min(d_i, d_j)} \tag{5}$$

(b) $S_{ij} < \mu_{ij}$, which is further classified into two cases:

(*i*) $d_i + d_j - n < 0$

$$\Omega_{ij} = \frac{\mu_{ij}}{\min(d_i, d_j)} \tag{6}$$

(*ii*) $d_i + d_j - n \geq 0$

$$\Omega_{ij} = \frac{\mu_{ij} - (d_i + d_j - n)}{\min(d_i, d_j)} \tag{7}$$

1. Significant community overlap and p-values

In addition to expected the overlap μ_{ij}, we calculate the likelihood that two communities share S_{ij} domains given the random null hypothesis. Under this hypothesis, the probability that $S_{ij} = k$ is[29;30]

$$\mathcal{P}(S_{ij} = k) = \frac{\binom{n}{k}\binom{n-k}{d_j-k}\binom{n-d_j}{d_i-k}}{\binom{n}{d_j}\binom{n}{d_i}} \tag{8}$$

To assign p-values to the observations, we sum $\mathcal{P}(S_{ij} = k)$ over k. However, depending on if k is smaller or larger than S_{ij}, we must separate two cases

(*i*) $k_{\text{obs.}} \leq S_{ij}$,

$$p = \sum_{k=0}^{k_{\text{obs.}}} \mathcal{P}(S_{ij} = k) \tag{9}$$

(*ii*) $k_{\text{obs.}} \geq S_{ij}$

$$p = \sum_{k=k_{\text{obs.}}}^{\min(d_i,d_j)} \mathcal{P}(S_{ij} = k) \tag{10}$$

In our analyses, we set the significance threshold to $p \leq 0.025$ to distinguish significant from random overlap.

D. Chromatin states and folds of enrichment

In Sec. III.C, we study cross-scale nestedness among communities associated with specific chromatin states. To calculate chromatin enrichment, we used published data that integrates several resources (e.g., ChIP-seq and RNA-seq) to partition the genome into 15 chromatin types.[31] Derived from a multivariate Hidden Markov Model (HMM), these states are (S1–S15): Active Promoter (S1), Weak Promoter (S2), Inactive/poised Promoter (S3), Strong Enhancer (S4, S5), Weak/poised Enhancer (S6, S7), Insulator (S8), Transcriptional transition (S9), Transcriptional elongation (S10), Weakly transcribed (S11), Polycomb-repressed (S12), Heterochromatin (S13), and Repetitive/Copy number variation (S14, S15).

We downloaded the HMM data from ENCODE (human cell line GM12878)[32]. The data file contains a genome-wide list of start and stop coordinates for each HMM state, where each instance is called peak. To determine the HMM content in a long DNA stretch, say a community, we count the number of peaks belonging to each of the 15 states. Because some HMM peaks may cross community borders, in practise we count the number of peak starts.

Next, to calculate the enrichment, we use a hypergeometric test that benchmarks the HMM content in a community to the chromosome-wide random expectation (sampling without replacement). The test goes through the following three steps.

1. Get community content from HMM data. We denote the number of peaks for each state as k_X, $X = \text{S1}, \ldots, \text{S15}$.

2. Calculate the expected number of X peaks $\bar{k}'_X$ given the community's total peak count n as $\bar{k}'_X = nK_X/N$, where N is the total number of peaks in the chromosome (including all HMM states), and K_X is the number of X peaks in the chromosome.

3. Calculate the p-value for $\bar{k}'_X$ under the hypergeometric null hypothesis. If less than the significance threshold 0.05, we consider the community as enriched or depleted in HMM state X (two-sided test). However, because we make multiple comparisons, one for each HMM state, we correct the p-value to reduce the false discovery rate. We do this using the Benjamini-Hochberg procedure[33] implemented in Python statsmodels[34]. We set the false discovery rate to 0.05.

After going through all communities using this procedure, they get labeled as "enriched" or "depleted" in each one of the 15 chromatin states. We point out that one community can be enriched in several HMM states.

In addition, to make our analysis more tractable when studying the nestedness of different chromatin types, we make a coarser classification and partition the communities into four large groups A—D. These groups reflect

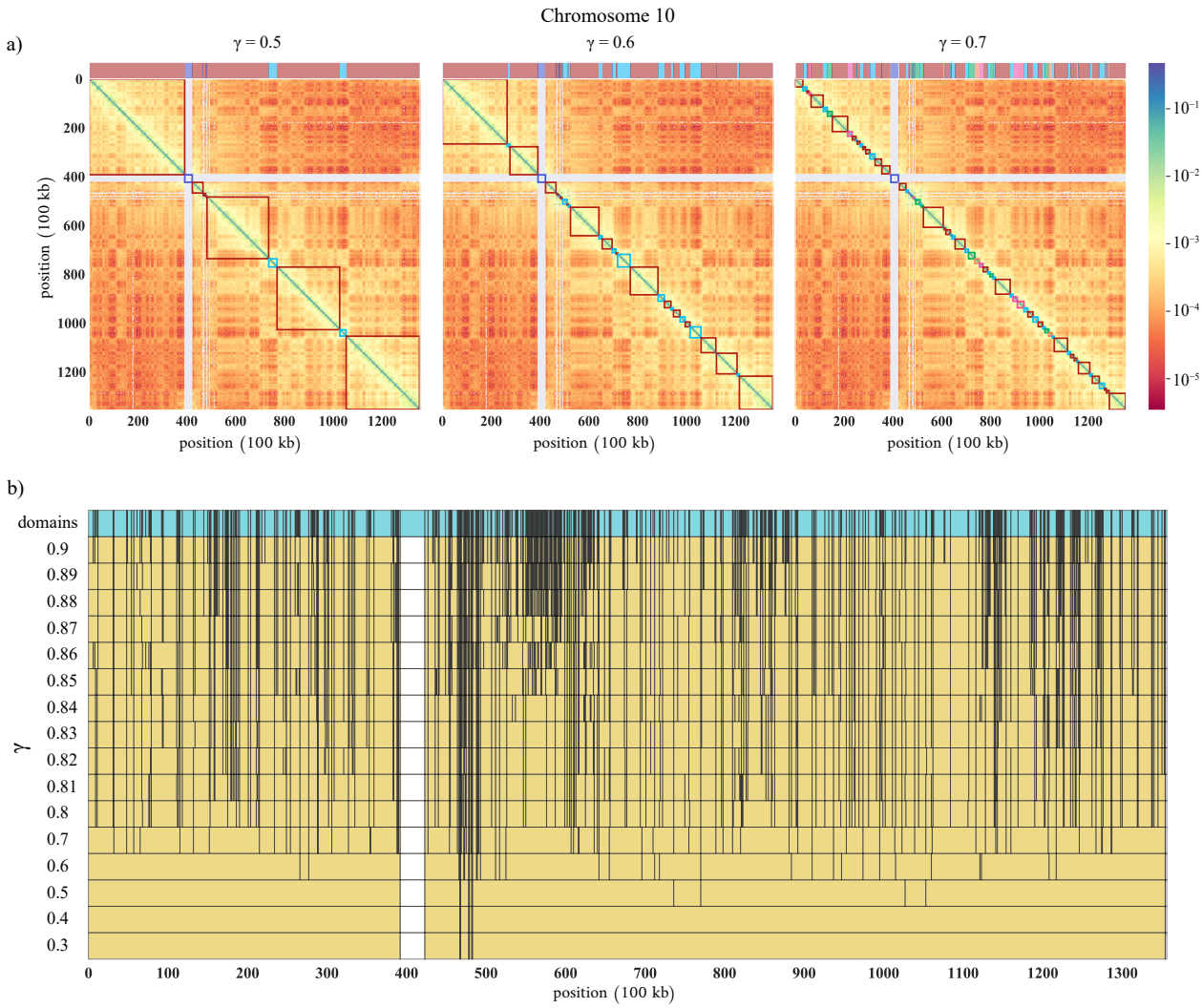

FIG. 2 Hi-C maps, 3D communities, and domains. (a) Hi-C maps where the red-to-blue pixel colors is a proxy for short-to-long 3D distances. The squares decorating the map's diagonals represent GenLouvain-derived 3D communities for three γ values (0.5, 0.6, and 0.7). Above each map, we show the community coverage as a colored stripe. Having unique colors, we observe that the communities comprise scattered linear DNA segments. The white cross shows the centromere. (b) Community borders and coverage across chromosome 10 for 16 γ values. The upper turquoise stripe shows DNA stretches that never split for $0 < \gamma \lesssim 1$. We refer to these indivisible regions as domains. The smallest 3D domain is one 100 kb long, which is the resolution-limit of the Hi-C data set we use.

the communities' overall HMM state enrichment: A: Active promoters (S1–S2), B: Enhancers (S4–S7), C: Transcribed regions (S9–S11), and D: Heterochromatin (S3 and S12–S15).

III. RESULTS & DISCUSSION

A. Distant domains aggregate into 3D communities spanning a range of scales

To illustrate how the 3D communities partition the chromosome, we superimpose GenLouvain-derived communities as squares along the Hi-C map's diagonal in Fig. 2(a). By assigning each community a unique color, we see that some 3D communities contain distant DNA segments. This community type—a distributed assembly of DNA segments—widens commonly used 3D partitions, like TADs, that assumes contiguous DNA stretches.

Furthermore, Fig. 2(b) shows that some DNA stretches rarely break across a wide range of γ. We call these indivisible pieces domains. To extract these domains, we collect the borders of intact DNA segments across many γ values and put them in one list. We show the domains in the upper turquoise stripe in Fig. 2(b). Admittedly, making γ large enough, we break even the domains into smaller linear DNA pieces so that eventually, every Hi-C bin (100 kb) represents one domain. However, we do not cover this extreme limit here.

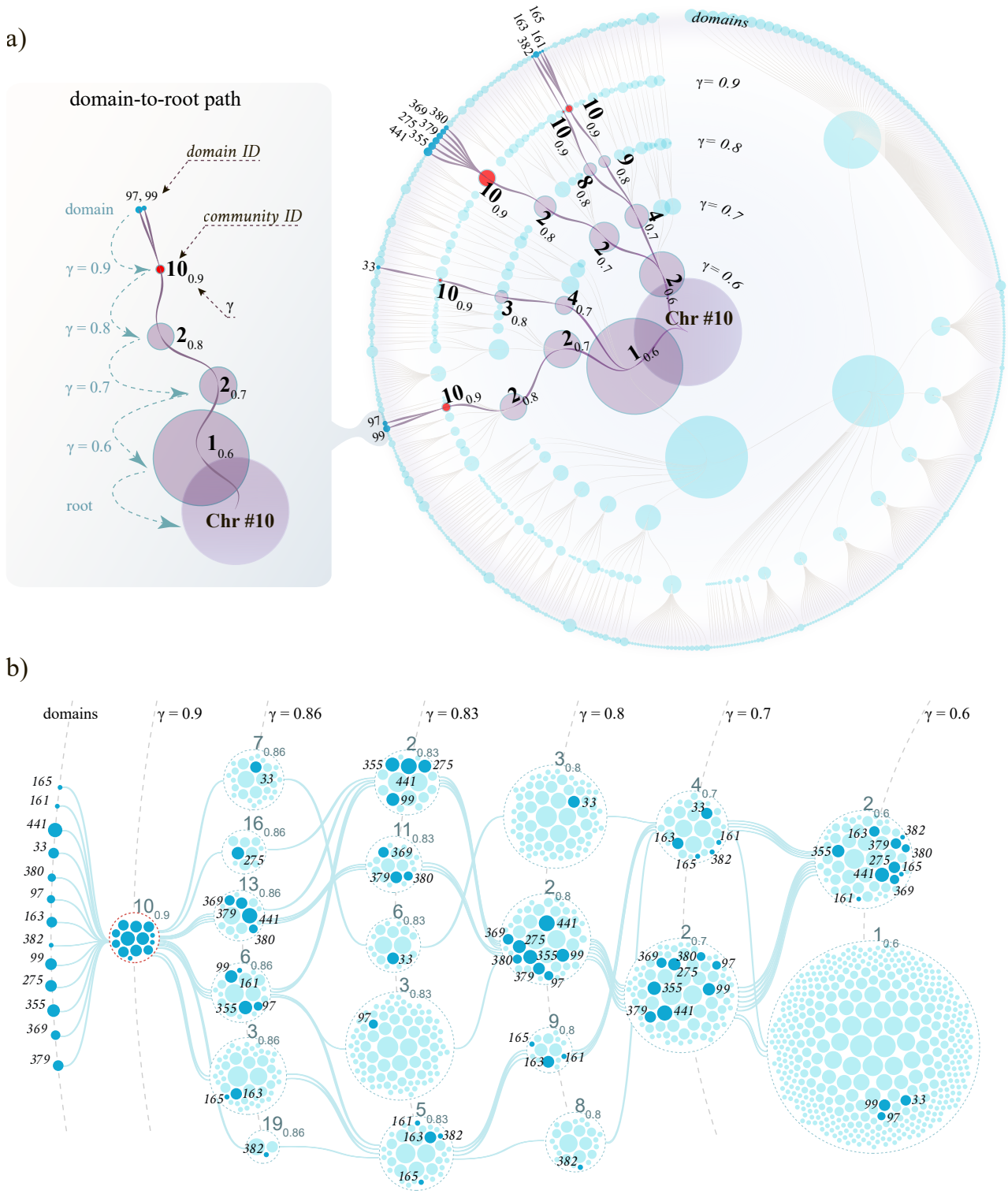

FIG. 3 Cross-scale community organization in chromosome 10. (a) Circular tree showing how domains (filled circles on the outer rim) merge into larger and larger 3D structures (filled circles on the inner rings). Each ring represents one value of GenLouvain's resolution parameter γ, and the diameters of the filled circles are associated with their DNA length (measured by the number of Hi-C bins). The red circles mark delocalized 3D structures forming a single 3D community at $\gamma = 0.9$ (denoted $10_{0.9}$). The dark links show folding trajectories for the domains passing through $10_{0.9}$ towards the root. The left panel shows a two-domain folding path and defines our label convention. We plotted the tree using RAW Graphs[35]. (b) Joining and splitting of the 13 domains belonging to the community $10_{0.9}$. These domains (filled dark-blue circles) pass through the 3D communities (open circles), joining other domains (filled light-blue circles). The edges connect 3D communities with dark-blue domains. We also highlighted these folding pathways in (a) (dark links).

B. 3D communities do not form perfect hierarchies

Figure 2 suggests that 3D communities have a complex cross-scale relationships. To better visualize such relations, we constructed a hierarchical tree from the same Hi-C data set (chromosome 10), showing how domains, the least divisible DNA regions, join into large 3D structures that, in turn, make up 3D communities (Fig. 3).

To construct the tree in Fig. 3, we first extracted chromosome 10's domain list and calculated the optimal community division associated with a few γ values. Next, we stored the folding pathways of all domains by tracing how their community memberships change with γ (Fig. 3(a), left). The circular tree illustrates the collection of all these pathways, where the links indicate how domains (filled circles on the outer rim) assemble into 3D structures (filled circles on the inner rings). Each ring corresponds to one γ value, i.e., one organization scale, and the filled circles' diameter symbolizes their DNA length.

The tree in Fig. 3(a) looks hierarchical. But a more complex pattern emerges if factoring in the 3D communities. To this end, we print the community ID next to a few filled circles (black numbers). In addition to the ID, we add a subscript to indicate the γ value (e.g., ID_γ). However, we omit ID numbers for the domains on the outer rim. Instead, their numbering denotes sequential ordering along the chromosome (e.g., domain 2 is next to domains 1 and 3, etc.).

Interestingly, the same community ID appears several times within one γ ring. One example is the community $10_{0.9}$, highlighted in red, that appears five times on the $\gamma = 0.9$ ring. This community contains 13 domains, scattered all over the chromosome as seen from their non-consecutive ID numbers. But even if scattered, they belong to the same 3D community that is a part of the optimal network division (according to GenLouvain). Delocalized domains forming communities in this way is a hallmark for an imperfect hierarchical folding.

To further exemplify this observation, we depict the folding pathways of the 13 individual domains belonging to the community $10_{0.9}$ in Fig. 3(b). By following the folding paths (edges) from left to right, we see that these domains (filled dark-blue circles) start in the same community and then split apart to become members of other 3D communities having different domain content (light-blue circles). Going even further to the right, 10 out of 13 dark-blue domains join yet again into one huge community (at $\gamma \gtrsim 0.6$). Again, this complex merging-and-splitting behavior is far from a perfect hierarchy. For clarity, we highlighted community $10_{0.9}$'s folding pathways as dark lines in (a) connecting the violet circles.

C. Quantifying chromosome nestedness

Figure 3 shows that domains mix between 3D communities as they approach the tree's root. This finding suggests that the folding mechanics is not perfectly hierarchical. To quantify deviations from being perfect, we calculate the pairwise community-domain overlap relative to random chance between two communities, i and j, belonging to different tree rings. To this end, we use a normalized nestedness metric, which we denote N_{ij}, that varies from -1 to $+1$. These two extreme points indicate complete segregation ($N_{ij} = -1$) and perfect nesting ($N_{ij} = 1$). When $N_{ij} = 0$, the overlap is not different from being random. We outline the explicit calculations and some of N_{ij}'s critical properties in Sec. II.C and show a schematic in Fig. 4(a).

To study the cross-scale nestedness in Hi-C-derived trees, like Fig. 3, we calculated N_{ij} across several γ values in four chromosomes (3, 5, 10, and 22). Plotting the N_{ij} histogram for all chromosomes in one graph, we find that the distribution has two pronounced peaks at ± 1 and a flat but slightly right-skewed intermediate region [Fig. 4(b)]. These two peaks indicate that some communities segregate (-1) while others nest ($+1$), just like in a perfect hierarchy that is either completely segregated or fully nested. However, the histogram's intermediate N_{ij} region is not zero and thus differs from a perfect hierarchy. This telltales that the 3D folding blends hierarchy-breaking contacts where some are possibly random.

To separate significant from random overlaps in Fig. 4(b), we calculated the probability that two 3D communities, having sizes d_i and d_j, share k domains in a random assignment—we defer all details to Sec. II.C. Based on this probability, we associate p-values to each N_{ij} observation. Setting the threshold to $p \leq 0.025$, we count the fraction of significant observations and illustrate the relative proportions in Fig. 4(c). In orange, we highlight significant overlaps. In blue-green, we indicate overlaps that are indistinguishable from being random. From panel (c), we make three key observations. First, the most segregated part (-1) is almost entirely blue-green and thus classified as insignificant. Second, roughly half of the perfectly nested communities ($+1$) share significant domain overlaps. Third, two regions show substantial overlaps that err on the side of segregation ($-0.95 < N_{ij} < -0.8$) and nesting ($0.65 < N_{ij} < 0.95$). The remaining data points appear random, particularly those surrounding $N_{ij} = 0$.

From Figs. 2–4, we conclude that chromosomes fold into complex hierarchies that mix nested and segregated parts. On average, however, the nesting is close to being random $\overline{N_{ij}} \approx 0$ if neglecting the -1 peak that skews the average (if included, it is $\overline{N_{ij}} \approx -0.8$). But since the distribution is so broad, community pairs show substantial differences where some are completely segregated, others are perfectly nested, and the rest is somewhere in between. This finding sheds new light on the hierarchical chromosome paradigm underlying several papers.

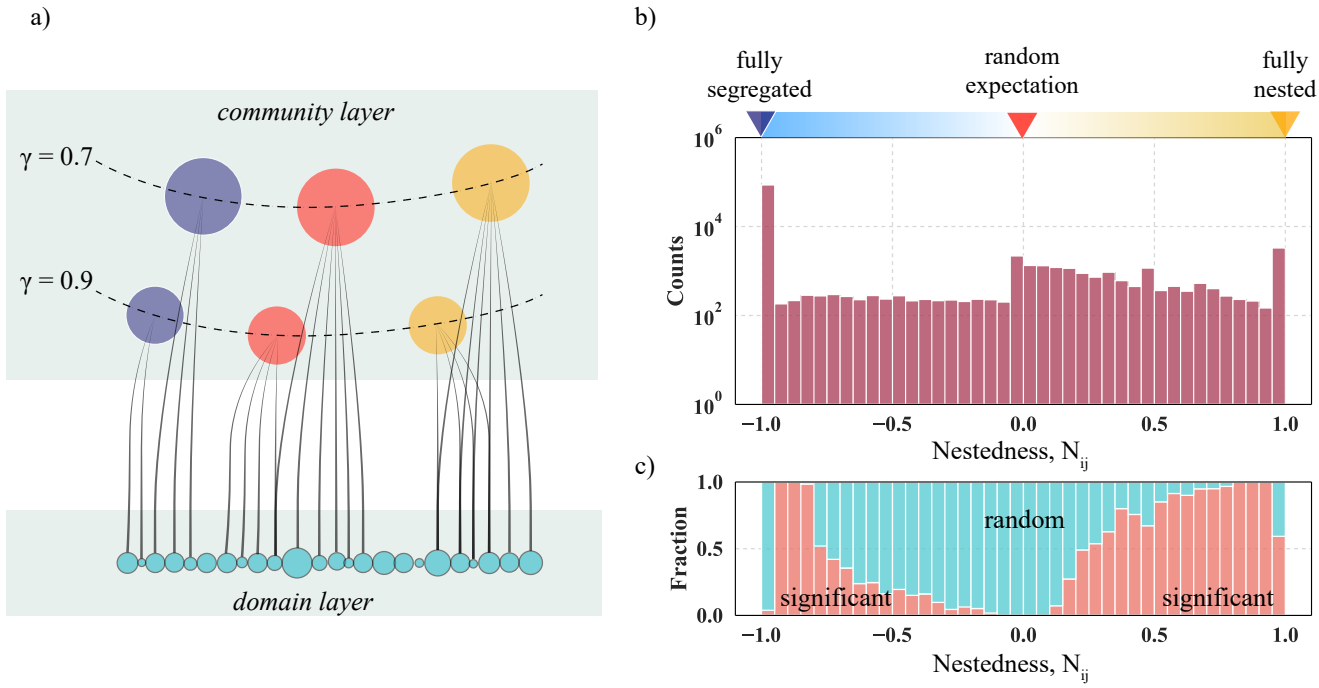

FIG. 4 Nestedness of community pairs in human chromosomes 3, 5, 10, and 22. (a) Schematic community-domain overlap in three cases: fully segregated ($N_{ij} = -1$, violet), random ($N_{ij} = 0$, red), and fully nested ($N_{ij} = 1$, yellow). The two upper layers contains communities belonging to different γ values (dashed lines). The bottom layer shows the domains and the edges indicate community memberships. (b) Nestedness histogram (N_{ij}) for chromosomes 3, 5, 10, and 22 across 16 γ values. The peaks at ± 1 suggest that several communities are segregated (-1) and nested ($+1$). However, there also exist significant intermediate levels of nestedness. The stripe overlaying the histogram indicates what we classify as fully segregated or nested according to the nestedness metric outlined in Sec. II.C. We show the nestedness of individual chromosome in Fig. S2. (c) Significant versus random community nestedness. As outlined in Sec. II.C, we filter community overlaps having p-values ≤ 0.025 and show the relative proportions of significant and random overlap associated with the N_{ij} histogram in (a). The colors indicate significant (orange) and random overlaps (blue-green).

D. Modeling non-nested chromosome folding

Several papers assume that linear DNA regions, like TADs, form higher-order structures by folding into each other in a perfect hierarchy (e.g.,[12–15]). However, our data shows that the nesting is more complex (Figs. 2–4). To better understand this disconnect, we propose a model for semi-nested chromosome folding. At the core, the model assembles ideally nested domain groups, consistent with the significant nestedness seen the N_{ij} histograms (Fig. 4). Then we break this pattern by reshuffling some domains among the communities. We denote the critical reshuffling parameter Q. Below, we outline the $Q = 0$ (perfect hierarchy) and $Q > 0$ limits separately.

1. Perfect hierarchical folding ($Q = 0$)

To achieve ideal hierarchical folding, we agglomerate domains into superstructures and superstructures into yet larger superstructures, following a few simple steps. First, we calculate the pairwise domain-domain interaction strength from their average Hi-C contact frequency (domains typically consists of several Hi-C 100 kb bins). Second, we select the domain pair having the strongest interaction and merge them into a superstructure. Then we replace the two merged domains in the list of pairwise interactions with the new superstructure and join the next most interacting pair. Regardless of choice, this scheme yields a new superstructure at each iteration. Notably, the algorithm does not only merge linearly adjacent domains.

Once we merged all domains into a giant superstructure, we use the domains' folding paths to organize the superstructures into a circular tree [Fig. 5(a)]. However, unlike the Hi-C derived tree in Fig. 3, the rings in Fig. 5(a) do not represent different γ values. Instead, they show consecutive mergers of the superstructures. Because some branches are so deep (> 10 steps), we show only the last five merging events and put all the domains on the outer (sixth) rim.

To illustrate that this scheme produces an ideal hierarchy, we select a group of 12 domains (red filled circles, outer rim) and highlight their folding paths across the tree with heavy links. After forming one super structure ('453', dark violet), this domain group stays intact

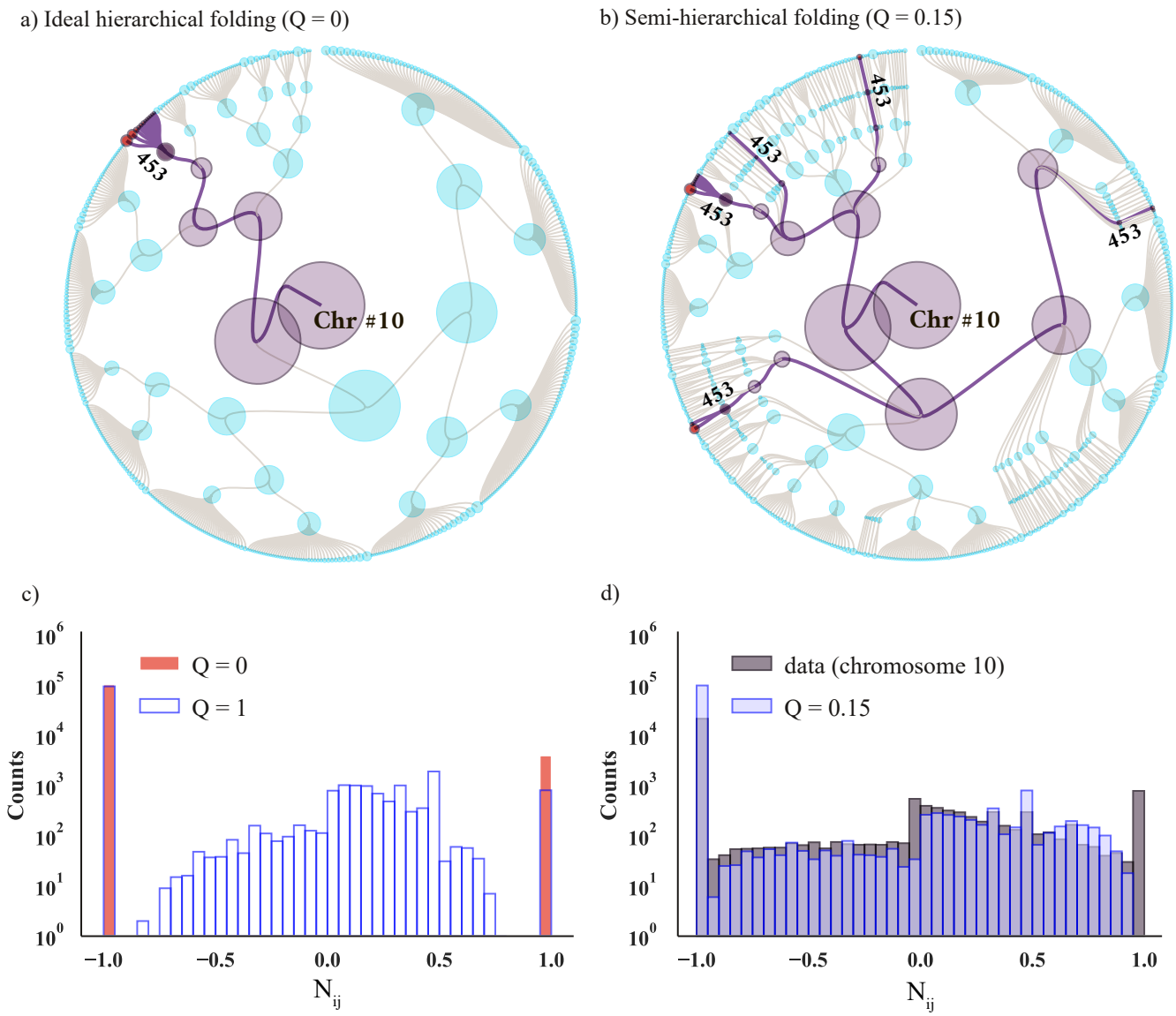

FIG. 5 Hierarchical and semi-hierarchical models of chromatin folding for human chromosome 10. (a) Ideal hierarchical folding ($Q = 0$). Filled circles on the outer rim represent domains and the root symbolizes the entire chromosome. We align the domain aggregates (superstructures) with the inner tree rings, each defining a scale of organization. We select a few domains (red filled circles) and show their domain-to-root paths with thick edges. At every inner ring, these domains assemble into yet larger structures (violet). As soon the domains merge into superstructure labelled '453' (dark violet), they never split apart. (b) Semi-hierarchical folding ($Q = 0.15$). As in (a), we color the domains in red that merge into a superstructure '453' and highlight their folding paths with thick edges going from from the outer rim to the root. Contrary to (a), node '453' is scattered across five tree branches. Thus, '453' only partially nests into larger structures, and the domains split and reunite when approaching the root. (c) Nestedness histogram when $Q = 0$ (ideal hierarchy, red bars) and $Q = 1$ (random nesting, open bars). When $Q = 0$, we see two peaks at $N_{ij} \pm 1$ indicating complete segregation and full nestedness. When $Q = 1$, the domains are fully randomized between the superstructures. While there is still perfect nesting and segregation (as we expect from the random null hypothesis in Sec. II.C), there is also partial overlap for $-0.8 < N_{ij} < 0.8$. (d) Nestedness histogram with some randomness ($Q = 0.15$, light-blue bars) overlaying the actual GenLouvain-derived data for chromosome 10 (dark-grey bars). We produced (a) and (b) using RAW Graphs[35].

as it joins more and more domains forming increasingly larger superstructures (violet circles). This exemplifies that domains never split apart once they end up in the same superstructure. However, this behaviour contrasts with what we observed in Fig. 3, where domains split and merge as they form communities. Therefore, this simple description cannot explain actual chromosome folding. We point point out that the $Q = 0$ limit is nearly identi-

cal to the so-called metaTAD algorithm[12].

2. Hierarchical folding with randomness ($Q > 0$)

The model yields a perfect domain hierarchy when the reshuffling parameter Q is zero. However, the actual folding patterns appear more complex. We exemplified this in Fig. 2(b), illustrating the cross-scale folding paths of 13 domains in chromosome 10. Following these paths, we note they do not perfectly correlate as they otherwise would in a perfect hierarchy. While some domains often stay together (e.g., 369, 379, and 380), others split only to reunite later. This represents the feature we aim to mimic by considering $Q > 0$.

To this end, we reshuffle a fraction of domains between the communities, restricting the reshuffling to communities within the same level of organization. The number of domains we interchange is proportional to Q. Algorithmically, we follow these three steps. (1) Go through all superstructures in the same organizational level [one ring in Fig. 5(a)] and identify the domain IDs and superstructure memberships. We exclude domains that do not yet belong to any community. (2) Select two of these domains randomly and swap their superstructure memberships with probability Q. (3) Repeat (1)–(2) until we exhausted all domain pairs, excluding those we already interchanged. If one domain remains without a pair, we keep its superstructure membership. Next, we pick another tree ring and repeat steps (1)–(3). In Fig. 5(b), we show the domains' folding paths when $Q = 0.15$.

In contrast to $Q = 0$ in Fig. 5(a), Fig. 5(b) shows that the hierarchy breaks when $Q > 0$. To better see the difference, we highlighted the domains forming the same superstructure we studied in (a) ('453', dark violet). Like in (a), this superstructure has 12 domains (10 out of 12 are the same). But unlike (a), superstructure 453 appears in different tree branches. This better reassembles the Hi-C derived tree in Fig. 3(b), where domains merge that do not have identical domain-to-root folding paths.

We point out that the tree's backbone formed in this way is identical to the $Q = 0$ case, but the domain memberships differ. Therefore, we foreshadow that this model is valid for small Q. But as we show in the following section, this is enough to reproduce the actual nestedness distribution in Fig. 4.

3. Nestedness for hierarchical and semi-hierarchical folding models

To study how the Q parameter in the model affects superstructure nestedness, we calculated and studied the N_{ij} histograms. Just as in Sec. III.C, we calculate these histograms by going through all superstructure pairs, omitting those belonging to the same tree ring, and counting the number of shared domains (Sec. II.C). We show three cases in Figs. 5(c) and (d): perfect hierarchy ($Q = 0$), full randomness ($Q = 1$), and intermediate randomness ($Q = 0.15$). All three cases build on domains derived from chromosome 10. Below, we discuss each case separately.

As expected for the ideal hierarchy, Fig. 5(c) has two isolated peaks at ± 1 (red bars), indicating that the communities are either fully nested or fully segregated. Put differently, the structure is "modular." These peaks also appear in the complete randomness limit ($Q = 1$). However, the $+1$ bar is lower relative to the $Q = 0$ case, and there is a distribution of N_{ij} values surrounding $N_{ij} = 0$, albeit not as wide as the real data. We interpret this as the domain reshuffling split several nested communities while keeping the segregation mostly intact.

To better mimic the real data, we tweaked Q to reassemble the actual N_{ij} histogram. In panel (d), we show the $Q = 0.15$ case overlaying the empirical data for chromosome 10. Apart from underestimating the histogram for negative N_{ij} values and overestimating it for large values, the two histogram lies on top of each other for the most part. This shows that the reshuffling parameter does not have to be large for the model to reproduce the nestedness data in Fig. 4(b). About 15% domain redistribution seems enough.

E. Nestedness and chromatin states

In Fig. 4, we found that some communities nest and others segregate. Also, Fig. 5 showed that we could reproduce the chromosome-wide nestedness distribution by slightly breaking an otherwise perfect folding hierarchy. This section analyzes if this behavior is associated with specific chromatin types.

To this end, we take advantage of published data that partition the genome into 15 chromatin states[31]. However, to make the analysis more tractable, we aggregate these states into four groups A–D, and study their pairwise nestedness. The groups are: promoters (A), enhancers (B), transcribed regions (C), and heterochromatin (D) (see Sec. II.D for complete definitions). To assign communities to these groups, we calculate folds of enrichment for each of the 15 chromatin states relative to the chromosome-wide average. We then use the hypergeometric statistical test to judge the enrichment significance (see Sec. II.D for details). Notably, because one community may enrich several chromatin states, it can belong to several A–D categories..

Next, we go through all community pairs to extract their nestedness N_{ij} and chromatin group (A–D). Then we plot N_{ij} histograms for all paired combinations—AA, AB, AC, AD, BB, etc. We show these histograms as panels in Fig. 6(a), where the light blue background portrays the entire chromosome's nestedness (we use data from chromosome 10). The diagonal panels represent community pairs having the same chromatin type (AA, BB, CC, and DD). These pairs seem to nest more than the rest of the chromosome as the N_{ij} distributions skews to the right. This observation differs from DD that seems to

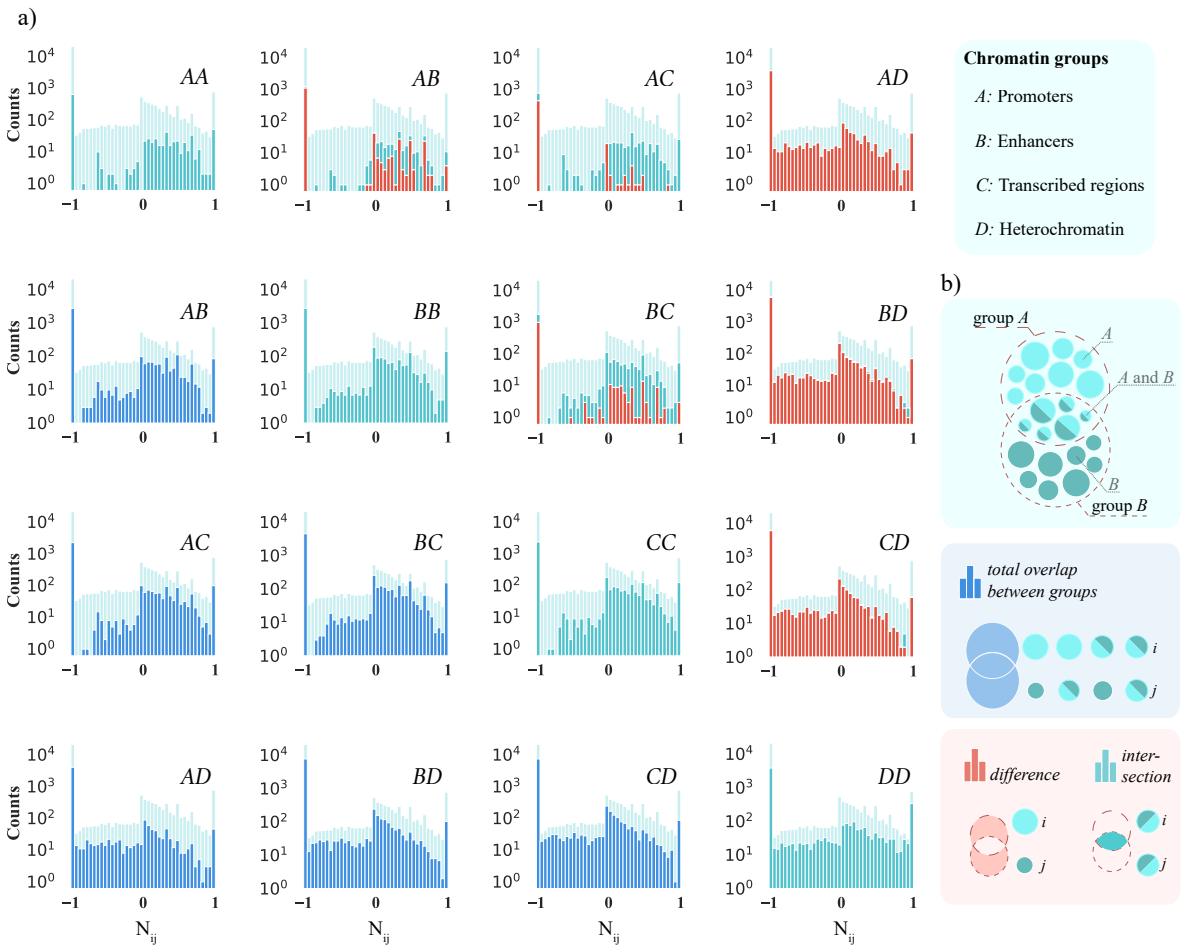

FIG. 6 Chromatin type and cross-scale nestedness between community pairs in chromosome 10. (a) Nestedness distributions (N_{ij}) for 10 combinations of chromatin types A–D (see the right legend). The diagonal panels show the nestedness histograms for community pairs belonging to the same chromatin type (AA, BB, etc.). The off diagonal panels show the other six paired combinations (AB, AC, AD, etc.); see panel (b) for detailed descriptions. The faint background in all histograms portrays the complete nestedness histogram from chromosome 10 (like Fig. 4). (b) Each cross pair (AB, AC, AD, etc.) has three community types, e.g., enriched in A, enriched B, or enriched in both. (top) Large dashed circles encompassing all A and B communities (small filled circles). One community can be enriched with A (light blue), B (dark blue), or both (half-filled). (middle) Schematic illustrating the community types used to calculate the nestedness histograms in the lower triangle in (a). For example, the AB histogram represents community pairs where one is enriched with A or "A and B," and the another is enriched with B or "A and B." The other panels follow the same pattern. (bottom) Schematic illustrating the community types used to calculate the nestedness histogram in the upper triangle in (a). These show a smaller subset than in (a): intersection (orange) and difference (green-blue)

follow the chromosome's overall nestedness distribution.

However, we could argue that the folding structure is segregated rather than nested because the average N_{ij} is negative in all diagonal panels; It becomes negative due to the large peak at -1 skewing the average. But as we showed in Fig. 4, this peak represents mostly random segregation (admittedly, roughly half of the $+1$ peak is also random) and the significant overlaps mostly appear for $N_{ij} > 0$ where AA–CC histograms carry heavy weight. Therefore, we conclude that these chromatin groups nest more than the chromosome average and that the nesting is significant.

Furthermore, the off-diagonal panels in (a) show the N_{ij} histograms for the six cross pairs, AB, AC, etc. But as noted above, some communities may enrich two groups simultaneously, say A and B. So when studying the AB cross-pair, it is natural to analyze separately communities enriched in A, B, or both, those enriched in A and B simultaneously, or those enriched in only A or only B. We depict these combinations and the color coding in panel (b), where the large dashed circles encompass all communities flagged as A or B. At the circles' intersection, there are communities enriched in A and B (half-filled circles).

The blue histograms in the lower triangular part show the nestedness among communities belonging to the broadest class (e.g., A, B, or both). These off-diagonal

histograms show that A, B, and C types tend to nest with each other (panels AB, AC, and BC), similar to AA–CC along the diagonal. In contrast, their overlap with D shows a wider variability reassembling the chromosome-wide N_{ij} distribution, apart from the dip close to $N_{ij} = 1$ hinting that A—C nest less with D than is expected. This observation likely reflects that A–C broadly belong to what is commonly referred to as "active chromatin" and D is "inactive chromatin" (e.g., measured by low or high RNA expression levels). In addition, a more granular study analyzing all 15 chromatin states showed that the five states making up group D rarely enrich more than random alongside the others making up A, B, and C. This differs from the A–C communities, where the internal chromatin states often co-appear. This explains why the significant nesting with group D is relatively scarce (see panels AD, BD, and CD).

In the upper triangular part in (a), we show stacked N_{ij} histograms for the other two more restricted cross pairs (e.g., communities simultaneously enriched in A and B, or only A or only B). In blue-green, we represent the intersection (e.g., A and B), and the orange symbolizes the difference (e.g., only A or only B). Compared to the histograms in the lower triangle, we note that AB, AC, and BC nest even more as the histograms get even more right-skewed. However, the domain overlaps between A–C with D remained almost the same.

In summary, when studying the cross-scale community nestedness, our data suggest that 3D communities belonging to "active chromatin" tend to nest more than the chromosome-wide average and appear to segregate from "inactive chromatin." The data also indicates that communities embedded in inactive chromatin seem to have substantial random cross-scale overlaps.

F. Active chromatin appears more hierarchical than inactive

To better understand the implications of the results in Fig. 6 regarding the chromosome's 3D organization, we quantified how well different 3D communities partition the Hi-C network and if solid or weak divisions are associated with the chromatin groups A–D. To this end, we calculated the modularity associated with the GenLouvain-derived communities ($\mathcal{M}_c$). To calculate $\mathcal{M}_c$, we use Eq. (1) and sum only those terms belonging to the same community. If this community modularity scores high, the internal nodes interconnect more than the background. If scoring low, they connect less (See Sec. B for explanation). We recover the global modularity $\mathcal{M}$ in Eq. (1) by summing over all communities, $\mathcal{M} = \sum_c \mathcal{M}_c$.

The community modularity varies significantly within and between chromatin groups A—D (Fig. S3). We also found that $\mathcal{M}_c$ grows linearly with the community sizes (number of domains) (Fig. S3). Therefore, to make a fair comparison, we plotted the median modularity rescaled with the community sizes (Fig. 7). The panels show

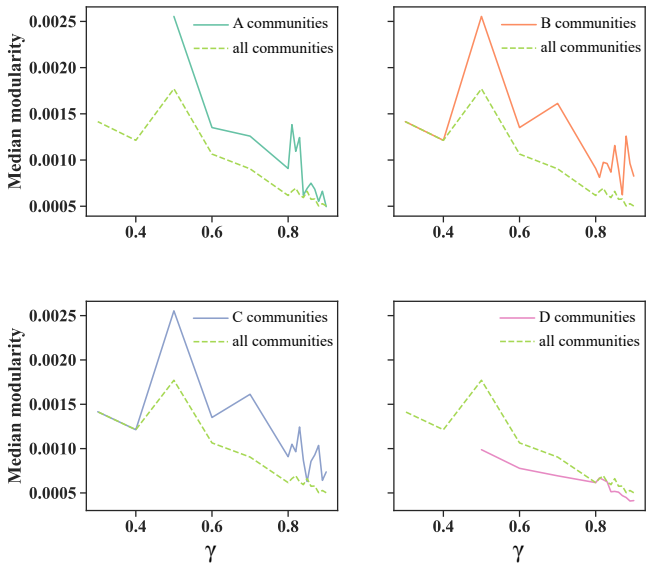

FIG. 7 Median community modularity (rescaled with community size) for four chromatin groups A—D (solid lines) across different scale parameters gamma. The dashed line shows the median for the entire network. A, B, and C communities have higher modularity than D and the whole network.

each chromatin group, including the global median modularity as a reference curve (dashed). From these panels, we observe that A—C communities ("active chromatin") have higher modularity than chromatin group D ("inactive chromatin") as well as the entire Hi-C network. This implies that A—C communities partition the network better than the D communities.

In addition to forming tighter node clusters, A—C communities tend to nest with each other (as shown in Fig. 4). These findings argue that active chromatin is hierarchical. At least it is more hierarchical than the D communities that form less convincing communities with substantial random nestedness N_{ij}. As we concluded from our simple folding model (Sec. III.D), random nesting breaks ideal hierarchies.

IV. CONCLUSIONS

In this paper, we have mapped out the semi-hierarchical organization of chromatin in human cells. Viewing the Hi-C data as a DNA contact network, we extracted significant 3D structures using the GenLouvain community detection algorithm that allows us to scan seamlessly through different organization scales. Contrasting common assumptions, the communities form non-hierarchical structures, where some organizational levels show a substantial degree of randomness. To better understand this result, we developed a model blending hierarchical folding and random contacts. This model reproduces the degree of nestedness we observe in actual data. We also study the nestedness in terms of chro-

matin states. We uncover that transcriptionally active states tend to nest more with each other and form more distinct 3D communities relative to the chromosome-wide average and inactive or repressed chromatin.

We use the GenLouvain method to extract 3D communities from Hi-C data. However, generating communities in this way is a random process, meaning that the nodes' community membership will differ between realizations even if the scale parameter γ is the same. Interestingly, the community–node correlation varies with γ, indicating that some communities are more stable than others. This problem exists in most complex networks whenever there are multi-scale interactions governing the organization[36]. Therefore, depending on algorithm design, two community detection methods focusing on slightly different connectivity features may disagree on the optimal node assembly. This aspect is yet unexplored for chromosome organization and worth pursuing in future work.

Furthermore, different community detection algorithms may disagree on the hierarchical levels. For instance, the Leiden algorithm[37] could yield a more strict community hierarchy than GenLouvain (the Leiden algorithm was developed to "solve" the non-hierarchical features of the original Louvain method). While studying the effects of other algorithms is a reasonable research direction, it does not invalidate our approach, which is agnostic to the specific algorithm choice.

We interpret our data as active chromatin being more hierarchical than inactive chromatin. From a biological standpoint, this has exciting implications. In active chromatin, there is a menagerie of specific proteins, like transcription factors, that coordinate transcription regulation. These proteins interact with chromatin elements at all distances, bringing some of them in 3D proximity to regulate transcription. These interactions are not random, so they could help shape the 3D structure toward a perfect hierarchy. While we lack data to validate this hypothesis, our simple folding model requires fine-tuned interactions to create an ideal hierarchical order, where a slight degree of arbitrary nesting causes noticeable deviations.

While specific proteins regulate transcription in active chromatin, inactive chromatin is often epigenetically repressed. The proteins managing epigenetic repression decorate chromatin with chemical tags over large DNA regions (e.g., methylation of specific histone sites). In this respect, this type of repression is not relying on characteristic long-range attractions. It is enough if the right chromatin type is close. If so, this idea foreshadows many random contacts that would manifest as a broad nestedness distribution, as in Fig. 5, and a less hierarchical structure than active chromatin.

Taken together, our paper shed new light on the hierarchical chromosome 3D organization. While large sections nest, others segregate, and there is a significant portion of randomness. We anticipate that cross-scale relationships capturing these features will be essential components in future models aiming to reach a deep understanding of the causal mechanisms of chromosome folding. In the short perspective, our results open further questions worthy of research, including the reliability of 3D communities or what biological factors govern the different organization scales such as DNA-binding proteins, epigenetic marks, or general chromatin types.

Acknowledgments

We acknowledge financial support from the National Research Foundation (NRF) of Korea (S.H.L. grant Nos. NRF-2021R1C1C1004132 and NRF-2022R1A4A1030660), the Swedish Research Council (L.L. grant Nos. 2017-03848 and 2021-04080), and the Knut and Alice Wallenberg foundation (grant No. 2014-0018, co-PI: P.S.)

References

[1] Yuri B Schwartz and Giacomo Cavalli. Three-dimensional genome organization and function in drosophila. *Genetics*, 205(1):5–24, 2017.

[2] Boyan Bonev and Giacomo Cavalli. Organization and function of the 3d genome. *Nature Reviews Genetics*, 17(11):661–678, 2016.

[3] Erez Lieberman-Aiden, Nynke L Van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*, 326(5950):289–293, 2009.

[4] Suhas SP 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, et al. A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7):1665–1680, 2014.

[5] 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, 2012.

[6] 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, 2012.

[7] Anjali Kaushal, Giriram Mohana, Julien Dorier, Isa Özdemir, Arina Omer, Pascal Cousin, Anastasiia Semenova, Michael Taschner, Oleksandr Dergai, Flavia Marzetta, et al. Ctcf loss has limited effects on global genome architecture in drosophila despite critical regulatory functions. *Nature communications*, 12(1):1–16, 2021.

[8] Silvia Remeseiro, Andreas Hörnblad, and François Spitz. Gene regulation during development in the light of topologically associating domains. *Wiley Interdisciplinary Reviews: Developmental Biology*, 5(2):169–185, 2016.

[9] Quentin Szabo, Frédéric Bantignies, and Giacomo Cavalli. Principles of genome folding into topologically associating domains. *Science advances*, 5(4):eaaw1668, 2019.

[10] Rajendra Kumar, Ludvig Lizana, and Per Stenberg. Genomic 3d compartments emerge from unfolding mitotic chromosomes. *Chromosoma*, 128(1):15–20, 2019.

[11] Sergio Sarnataro, Andrea M Chiariello, Andrea Esposito, Antonella Prisco, and Mario Nicodemi. Structure of the human chromosome interaction network. *PLoS One*, 12(11):e0188201, 2017.

[12] James Fraser, Carmelo Ferrai, Andrea M Chiariello, Markus Schueler, Tiago Rito, Giovanni Laudanno, Mariano Barbieri, Benjamin L Moore, Dorothee CA Kraemer, Stuart Aitken, et al. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Molecular systems biology*, 11(12):852, 2015.

[13] Lin An, Tao Yang, Jiahao Yang, Johannes Nuebler, Guanjue Xiang, Ross C Hardison, Qunhua Li, and Yu Zhang. Ontad: hierarchical domain structure reveals the divergence of activity among tads and boundaries. *Genome biology*, 20(1):1–16, 2019.

[14] Yu Wei Zhang, Meng Bo Wang, and Shuai Cheng Li. Supertad: robust detection of hierarchical topologically associated domains with optimized structural information. *Genome biology*, 22(1):1–20, 2021.

[15] Yinxiu Zhan, Luca Mariani, Iros Barozzi, Edda G Schulz, Nils Blüthgen, Michael Stadler, Guido Tiana, and Luca Giorgetti. Reciprocal insulation analysis of hi-c data shows that tads represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. *Genome research*, 27(3):479–490, 2017.

[16] Sang Hoon Lee, Yeonghoon Kim, Sungmin Lee, Xavier Durang, Per Stenberg, Jae-Hyung Jeon, and Ludvig Lizana. Mapping the spectrum of 3d communities in human chromosome conformation capture data. *Scientific reports*, 9(1):1–7, 2019.

[17] Johannes Nuebler, Geoffrey Fudenberg, Maxim Imakaev, Nezar Abdennur, and Leonid A Mirny. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proceedings of the National Academy of Sciences*, 115(29):E6697–E6706, 2018.

[18] Wibke Schwarzer, Nezar Abdennur, Anton Goloborodko, Aleksandra Pekowska, Geoffrey Fudenberg, Yann Loe-Mie, Nuno A Fonseca, Wolfgang Huber, Christian H Haering, Leonid Mirny, et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, 551(7678):51–56, 2017.

[19] Ji Hyun Bak, Min Hyeok Kim, Lei Liu, and Changbong Hyeon. A unified framework for inferring the multi-scale organization of chromatin domains from hi-c. *PLoS computational biology*, 17(3):e1008834, 2021.

[20] Ron Edgar, Michael Domrachev, and Alex E Lash. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic acids research*, 30(1):207–210, 2002.

[21] Markus Nyberg, Tobias Ambjörnsson, Per Stenberg, and Ludvig Lizana. Modeling protein target search in human chromosomes. *Physical Review Research*, 3(1):013055, 2021.

[22] Stefanie Kaufmann, Christiane Fuchs, Mariya Gonik, Ekaterina E Khrameeva, Andrey A Mironov, and Dmitrij Frishman. Inter-chromosomal contact networks provide insights into mammalian chromatin organization. *PloS one*, 10(5):e0126125, 2015.

[23] Philip A. Knight and Daniel Ruiz. A fast algorithm for matrix balancing. *IMA Journal of Numerical Analysis*, 33(3):1029–1047, July 2013.

[24] Rajendra Kumar, Haitham Sobhy, Per Stenberg, and Ludvig Lizana. Genome contact map explorer: a platform for the comparison, interactive visualization and analysis of genome contact maps. *Nucleic Acids Research*, 45(17):e152, September 2017.

[25] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, oct 2008.

[26] Lucas G. S. Jeub, Marya Bazzi, Inderjit S. Jutla, and Peter J. Mucha. A generalized louvain method for community detection implemented in matlab, 2011-2019.

[27] A Grosberg, Y Rabin, S Havlin, and A Neer. Crumpled globule model of the three-dimensional structure of dna. *EPL (Europhysics Letters)*, 23(5):373, 1993.

[28] Leonid A Mirny. The fractal globule as a model of chromatin architecture in the cell. *Chromosome research*, 19(1):37–51, 2011.

[29] Giovanni Strona and Joseph A. Veech. A new measure of ecological network structure based on node overlap and segregation. *Methods in Ecology and Evolution*, 6(8):907–915, 2015.

[30] Joseph A. Veech. A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography*, 22(2):252–260, 2013.

[31] Jason Ernst, Pouya Kheradpour, Tarjei S. Mikkelsen, Noam Shoresh, Lucas D. Ward, Charles B. Epstein, Xiaolan Zhang, Li Wang, Robbyn Issner, Michael Coyne, Manching Ku, Timothy Durham, Manolis Kellis, and Bradley E. Bernstein. Systematic analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345):43–49, May 2011.

[32] GM12878 Chromatin State Segmentation by HMM from ENCODE/Broad. url = `https://genome.ucsc.edu/cgi-bin/hgtrackui?hgsid=1295125293_1uaxm5ngerepzfvvcpekgwcuzura&db=hg19&g=wgencodebroadhmmgm12878hmm#track_html`,.

[33] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. 57(1):289–300.

[34] Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

[35] Michele Mauri, Tommaso Elli, Giorgio Caviglia, Giorgio Uboldi, and Matteo Azzi. RAWGraphs: A visualisation platform to create open outputs. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter*, CHItaly '17, pages 1–5. Association for Computing Machinery.

[36] Daekyung Lee, Sang Hoon Lee, Beom Jun Kim, Heetae Kim, et al. Consistency landscape of network communities. *Physical Review E*, 103(5):052306, 2021.

[37] Vincent A Traag, Ludo Waltman, and Nees Jan Van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):1–12, 2019.