RESEARCH ARTICLE

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

Dolores Bernenko[1], **Sang Hoon Lee**[2,3], **Per Stenberg**[4], **Ludvig Lizana**[1] *

1 Department of Physics, Integrated Science Lab, Umeå University, Umeå, Sweden, **2** Department of Physics and Research Institute of Natural Science, Gyeongsang National University, Jinju, Korea, **3** Future Convergence Technology Research Institute, Gyeongsang National University, Jinju, Korea, **4** Department of Ecology and Environmental Science, Umeå University, Umeå, Sweden

* ludvig.lizana@umu.se

Abstract

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 active 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.

Author summary

The 3D organization of mammalian DNA affects genetic processes, such as transcription, DNA repair, and epigenetics. To unravel the complexity of the 3D structure, researchers developed numerous experimental methods, the most advanced being Hi-C. This method enables scientists to create "contact maps" illustrating the 3D interactions among all pairs of DNA segments across the genome. These maps unveiled a multi-scale organization, ranging from megabase-pair compartments to short-range DNA loops. Common explanations for this organization rest on nested hierarchies, where DNA regions of similar sizes coalesce into larger structures. However, such a model is incompatible with competing molecular mechanisms, primarily loop extrusion and phase separation, that shape the chromosomes' 3D organization at different scales.

Our study aims to map out the actual chromosome folding relationships using Hi-C data sets. Treating the data as a weighted network of pairwise DNA segment interactions, we identified 3D communities across different network scales using the Generalized Louvain method, a standard community detection algorithm. By building a tree linking these communities, we discovered that chromosome organization is more intricate than a perfect hierarchy suggests. Instead, we found that chromosomes exhibit a mix of nested and non-nested community pairs alongside considerable randomness. The nested parts often associate with active chromatin. These results highlight that cross-scale relationships are critical for understanding the causal mechanisms of chromosome folding.

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 architectural proteins that reshape chromatin, 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 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 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 inevitably leads 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 actual folding by charting the cross-scale structural 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 networks' layers of organization. By mapping the hierarchical relationships between these assemblies, we find that some nest, others segregate, and still others 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.

Materials and methods

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 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 treat 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 to extract communities harbouring densely connected nodes that maintain fewer contacts with the rest of the network (the generalized Louvain method, see Methods: The GenLouvain algorithm—detecting 3D communities in Hi-C data).

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. In particular, we use the Knight-Ruiz (KR) matrix balancing [23] implemented in gcMapExplorer [24].

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 optimal community division of 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.

We chose GenLouvain based on principled and practical aspects. First, it is a generalized version of one of the most intensively tested algorithms: Louvain [25]. Second, for practicality, the developer's open-source codes are written in `MATLAB`, allowing us to modify essential parts such as the null-model term, which we will elaborate on in the forthcoming paragraphs.

One critical feature of GenLouvain is its resolution (or scale) parameter γ. This parameter allows us to sweep through the scales of the network and probe the network's community spectrum. Furthermore, this parameter is closely related to the parameters capturing the relative

tendency of intra- versus inter-group connections in the context of the stochastic block model [27]. Mathematically, γ is a part of the modularity function $\mathcal{M}$, defined as

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

where A_{ij} is the network's adjacency matrix representing the weight of the edge connecting nodes i and j, and the summand is counted only if i and j belong to the 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

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

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, 28, 29] that agrees with chromatin contact decay in Hi-C experiments, where the exponent is approximately -1.08 [3]. Note that we may use any decay exponent in Eq (2). For example, -0.75 matches better with within-TAD contacts, [30], but we kept -1 as our 3D communities are larger than TADs.

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 Results—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 [31, 32]. 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 ones. When completely segregated, the communities do not share any domains. In a perfect hierarchy, like a phylogenetic tree, the nestedness is either $+1$ or -1, indicating full nesting or complete segregation. But in a more complex multi-scale structure, 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
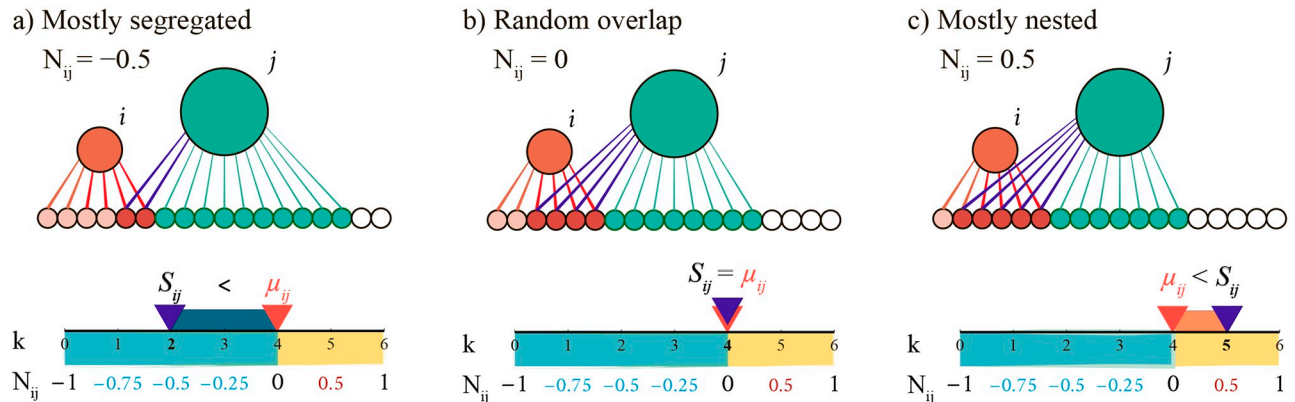
Fig 1. Three examples of nestedness (N_{ij}) in a simple bipartite network. The networks in panels (a)–(c) have the same number of nodes in each layer —18 domains (small circles) and two communities (i and j, large circles)—and the same number of links (18) but connected differently to achieve varying 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 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.

https://doi.org/10.1371/journal.pcbi.1011185.g001

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 ($N_{ij} = 0.5$).

We go through several steps to calculate N_{ij}. 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 [31]

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

where n is the total and k is the 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 $N_{ij} \in [-1, 1]$:

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

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

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

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

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

Significant community overlap and *p*-values. In addition to the expected 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 [31, 32]

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

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.

Chromatin states and folds of enrichment

In Results: Quantifying chromosome nestedness, we study cross-scale nestedness among communities associated with specific chromatin states. To calculate chromatin enrichment, we used published data integrating several resources (e.g., ChIP-seq and RNA-seq) to partition the genome into 15 chromatin types. [33] 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) [34]. The data is a genome-wide list of start-and-stop coordinates for each HMM state, where each instance is called a "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, 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 = S1, . . ., S15.

2. Calculate the expected number of X peaks $\overline{k}'_X$ given the community's total peak count n as $\overline{k}'_X = nK_X/N$, 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 $\overline{k}'_X$ under the hypergeometric null hypothesis. If less than the significance threshold of 0.05, we consider the community 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 [35] implemented in Python statsmodels [36]. 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 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 the 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).

Results

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 2A. 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 2B shows that some DNA stretches rarely break across a wide range of γ. We call these indivisible pieces "irreducible domains". We collect them by we collecting borders of intact DNA segments across many γ values into one list. We show the domains in the upper turquoise stripe in Fig 2B and their size distribution in S4 Fig. 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.
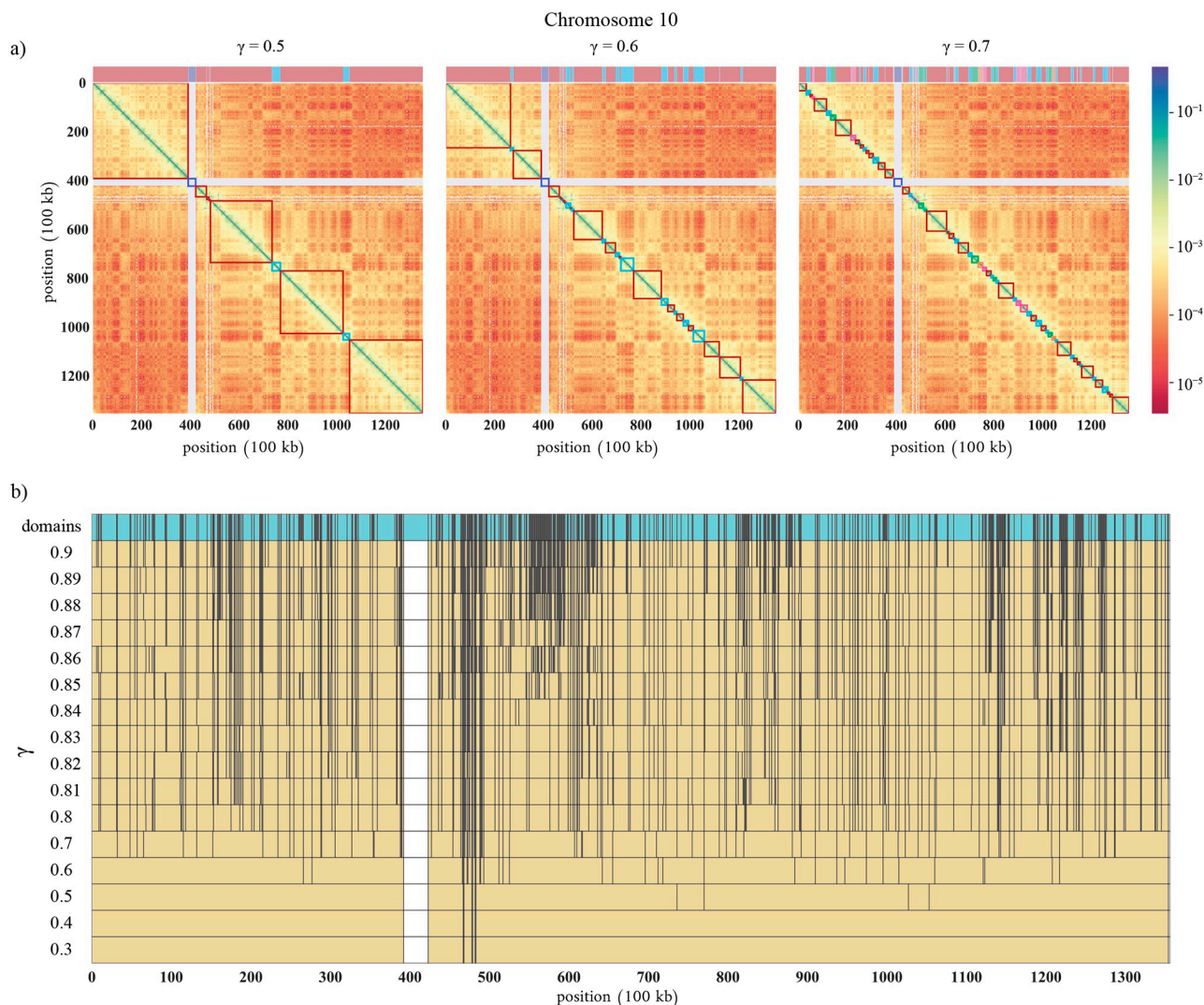
Fig 2. Hi-C maps, 3D communities, and domains. (a) Hi-C maps where the red-to-blue pixel colors are a proxy for short-to-long 3D distances. The squares decorating the map's diagonals represent GenLouvain-derived 3D communities for three γ values (0.5, 0.6, and 0.7). Above each map, we show the community 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 100 kb long, which is the resolution limit of the Hi-C data set we use.

https://doi.org/10.1371/journal.pcbi.1011185.g002

3D communities do not form perfect hierarchies

[Fig 2](#) suggests that 3D communities have 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 3A](#), 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
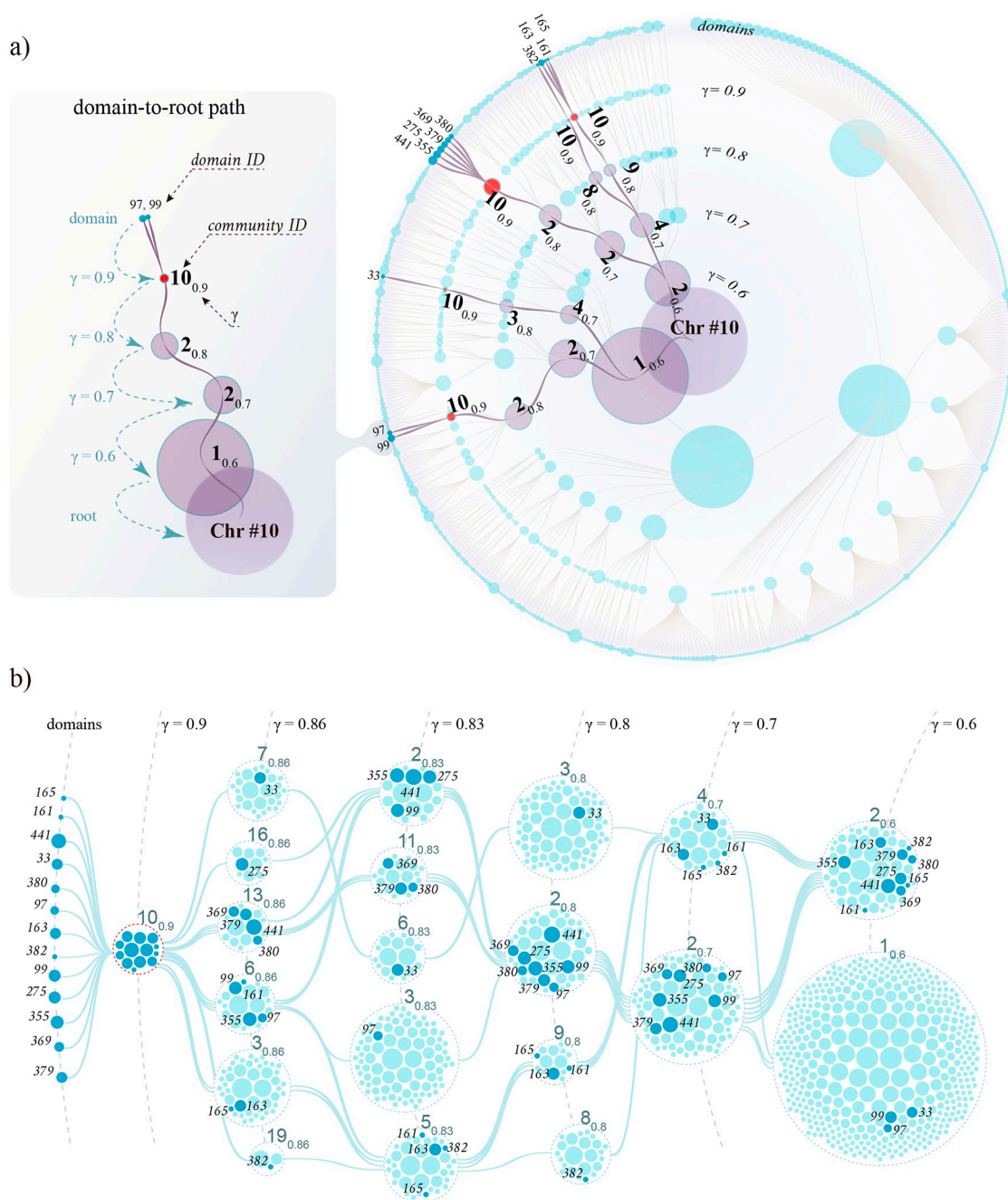
Fig 3. Cross-scale community organization in chromosome 10. (a) Circular tree showing how domains (filled circles on the outer rim) merge into larger and larger 3D structures (filled circles on the inner rings). Each ring represents one value of GenLouvain's resolution parameter γ, and the diameters 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 [37]. (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).

https://doi.org/10.1371/journal.pcbi.1011185.g003

the inner rings). Each ring corresponds to one γ value, i.e., one organization scale, and the filled circles' diameter symbolizes their DNA content.

The tree in Fig 3A 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 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 this way is a hallmark of 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 3B. 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.

Quantifying chromosome nestedness

Fig 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, denoted 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 Methods: Network nestedness and show a schematic in Fig 4A.

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); we choose these to mix large, intermediate, and small chromosomes. 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 4B). 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 an ideal hierarchy. This telltales that the 3D folding blends hierarchy-breaking contacts where some are possibly random.

To separate significant from random overlaps in Fig 4B, 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 Methods: Network nestedness. 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 4C. In orange, we highlight significant overlaps. In blue-green, we indicate overlaps that are indistinguishable from being random. From Fig 4C, 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
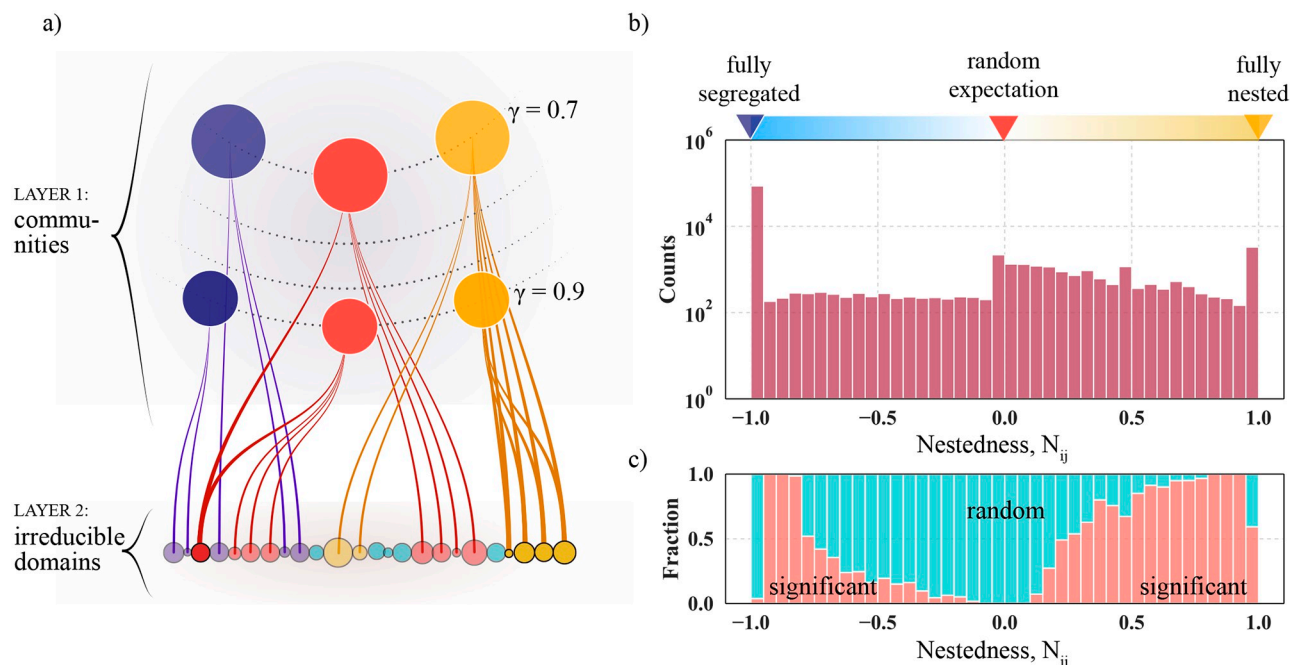
Fig 4. Nestedness of community pairs in human chromosomes 3, 5, 10, and 22. (a) Schematic community-domain overlap in three cases: fully segregated ($N_{ij} = -1$, violet), random ($N_{ij} = 0$, red), and fully nested ($N_{ij} = 1$, yellow). Layer 1 contains communities belonging to different γ values (dotted lines). The bottom layer shows the irreducible domains, and the edges indicate community memberships. (b) Nestedness histogram (N_{ij}) for chromosomes 3, 5, 10, and 22. For each chromosome, we derived communities from 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 Methods: Network nestedness. We show the nestedness of individual chromosomes in S3 Fig, and we visualize N_{ij} distributions for individual γ-pairs in Supplementary Material, S5 Fig (chromosome 10). (c) Significant versus random community nestedness. As outlined in Methods: Network nestedness, 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).

https://doi.org/10.1371/journal.pcbi.1011185.g004

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. Finally, in S5 Fig, we show community overlaps between specific γ pairs.

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 show 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 in 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 that represents the probability that two domains will change community memberships. Below, we outline the $Q = 0$ (perfect hierarchy) and $Q > 0$ limits separately.

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 consist 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 merge all domains into a giant superstructure, we use the domains' folding paths to organize the superstructures into a circular tree (Fig 5A). However, unlike the Hi-C derived tree in Fig 3, the rings in Fig 5A do not represent different γ values as this model does not use GenLouvain. 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 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 out that the $Q = 0$ limit is nearly identical to the so-called metaTAD algorithm [12].

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 2B, illustrating the cross-scale folding paths of 12 domains in chromosome 10. Following these paths, we note they do not perfectly correlate as they would in a perfect hierarchy. While some domains often stay together, 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 5A) 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 exhaust 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).

By varying the parameter Q, we retrieved several N_{ij} distributions. To find the optimal $Q_{\text{opt.}}$ —the Q that produces the N_{ij} distribution that is most similar to the real data—we utilized a Kolmogorov-Smirnov test. This test gave $Q_{\text{opt.}} \approx 0.3$ (Supplementary Material, S8 Fig and S7 Text). Fig 5B shows the associated domain folding paths.

In contrast to $Q = 0$ in Fig 5A and 5B shows that the hierarchy breaks when $Q > 0$. We highlighted the domains forming the same superstructure we studied in (a) ('453', dark violet) to better see the difference. Like in (a), this superstructure has 12 domains (11 out of 12 are the same). But unlike (a), superstructure 453 appears in different tree branches. This better
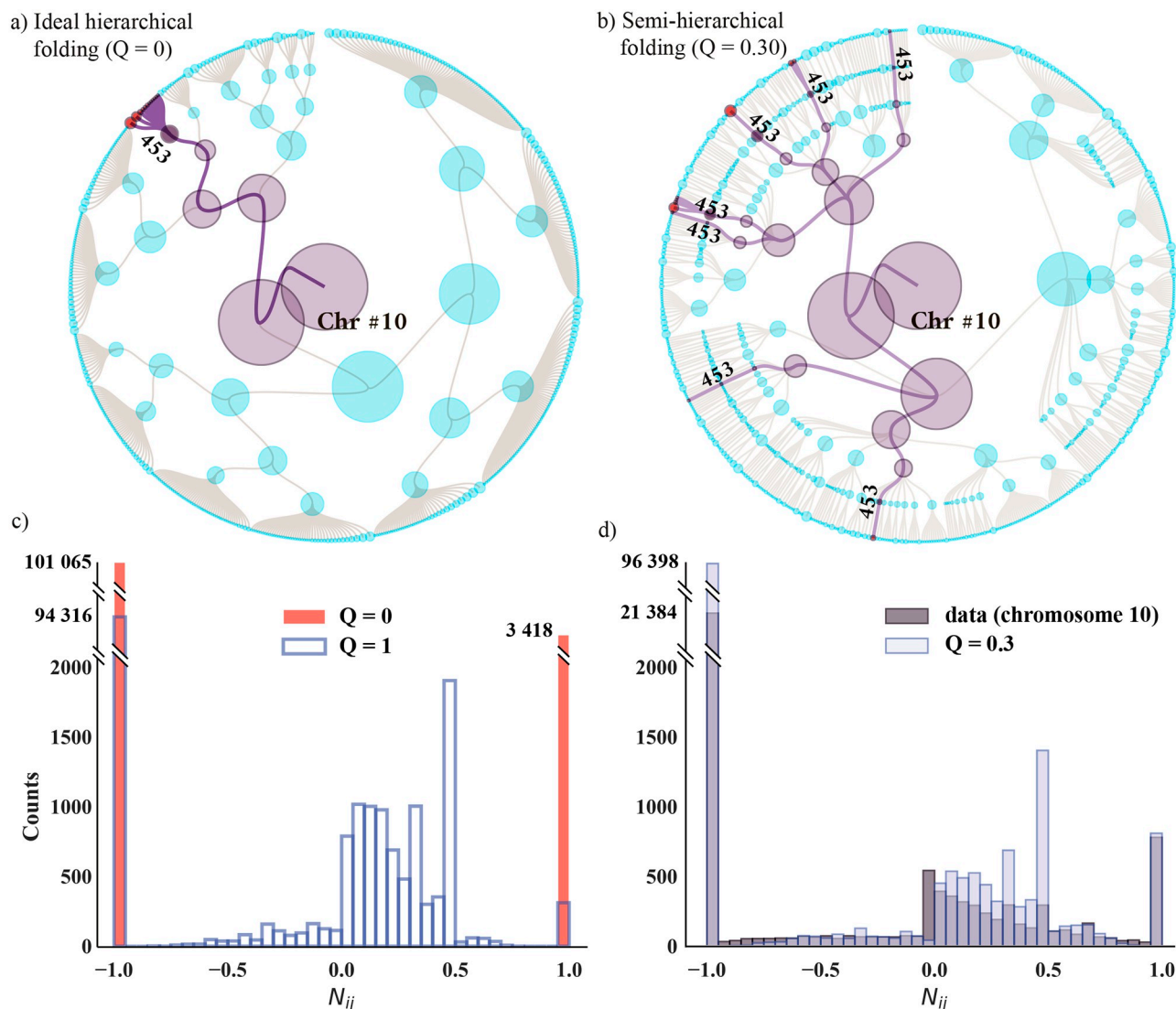
Fig 5. Hierarchical and semi-hierarchical models of chromatin folding for human chromosome 10. (a) Ideal hierarchical folding ($Q = 0$). Filled circles on the outer rim represent domains; the root symbolizes the entire chromosome. We align the domain aggregates (superstructures) with the inner tree rings, each defining a scale of organization. We select a few domains (red-filled circles) and show their domain-to-root paths with thick edges. These domains assemble into yet larger structures (violet) at every inner ring. As soon as the domains merge into a superstructure labeled '453' (dark violet), they never split apart. (b) Semi-hierarchical folding ($Q = 0.30$). 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 the outer rim to the root. Unlike (a), node '453' is scattered across seven 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 Methods: Network nestedness), there is also partial overlap for $-0.8 < N_{ij} < 0.8$. (d) Nestedness histogram with some randomness ($Q = 0.30$, light-blue bars) overlaying the actual GenLouvain-derived data for chromosome 10 (dark-grey bars). We produced (a) and (b) using RAW Graphs [37].

reassembles the Hi-C derived tree in Fig 3B, 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.

Nestedness for hierarchical and semi-hierarchical folding. To study how the Q parameter in the model affects superstructure nestedness, we calculated and studied the N_{ij} histograms. Just as in Results: Quantifying chromosome nestedness, 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 (Methods: Network nestedness). We show three cases in Fig 5C and 5D: perfect hierarchy ($Q = 0$), full randomness ($Q = 1$), and intermediate randomness ($Q = 0.30$). All three cases build on domains derived from chromosome 10. Below, we discuss each case separately.

As expected for the ideal hierarchy, Fig 5C 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 actual data. We interpret this as the domain reshuffling split several nested communities while keeping the segregation primarily intact.

To better mimic the real data, we tweaked Q to reassemble the actual N_{ij} histogram. In Fig 5D, we show the $Q = 0.30$ 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 must not be large for the model to reproduce the nestedness data in Fig 4B. About 30% domain redistribution seems enough.

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 [33]. 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 Methods: 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 Methods: Chromatin states and folds of enrichment 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 6B, 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 nest more than the rest of the chromosome as the N_{ij} distributions skew to the right. This observation differs from DD, which seems to follow the chromosome's overall nestedness distribution. To lend quantitative support to these observations, we performed a Kolmogorov-Smirnov test, which compares the cumulative distribution functions of the histograms AA, BB, CC, and DD (Supplementary Material, S6 Fig).
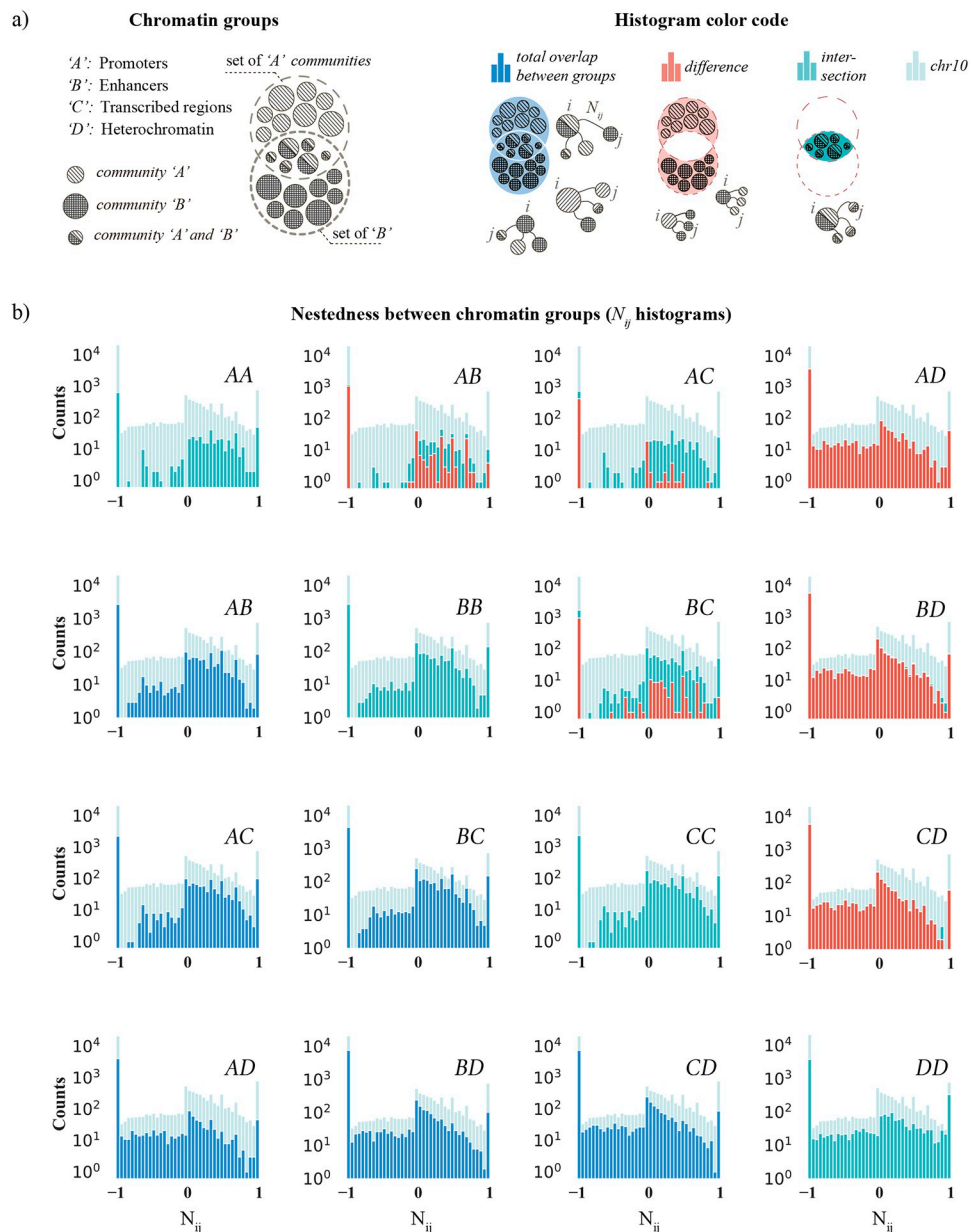
Fig 6. Chromatin type and cross-scale nestedness between community pairs in chromosome 10. (a, left) Each chromatin cross pair (AB, AC, AD, etc.) has three community types. For example, these may enrich A ("Promoters"), B ("Enhancers"), or A and B. The large dashed circles represent the Venn diagram of all A or B community types (small filled circles). The set illustrates that one community can be in one of three categories: enriched with A (upper), B (lower), or both (intersection). (a, right) Schematic illustrating the color codes used in the nestedness histograms associated: complete overlap (dark blue), the difference (dark red), and the intersection (light blue). In pale blue, we indicate the chromosome-wide average of chromosome 10. (b) Nestedness distributions (N_{ij}) for 10 combinations of chromatin types A–D. 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 (a) for detailed descriptions. The faint pale blue background in all histograms portrays the complete nestedness histogram from chromosome 10 (like Fig 4).

https://doi.org/10.1371/journal.pcbi.1011185.g006

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 Fig 6B 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 Fig 6A, where the large dashed circles encompass all communities flagged as A or B. At the circles' intersection, communities are enriched in A and B (half-filled circles).

The blue histograms in the lower triangular part of Fig 6B 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 belongs 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 in 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 AD, BD, and CD panels).

In the upper triangular part in Fig 6B, 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 B). In blue-green, we represent the intersection (e.g., A and B), and orange symbolizes the difference (e.g., only A or only B). Admitting that the sample size is relatively small, we note that the AB, AC, and BC histograms have a more significant fraction of data points for positive N_{ij} values than those in the lower triangle, indicating more nesting. However, the AD, BD, and CD histograms remained almost identical.

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.

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 the modularity scores high, the internal nodes interconnect more than the background. If scoring low, they connect less (see Methods: The GenLouvain algorithm—detecting 3D communities in Hi-C data for an explanation). We recover the global modularity $\mathcal{M}$ in Eq (1) by summing over all communities, $\mathcal{M} = \sum_c \mathcal{M}_c$.
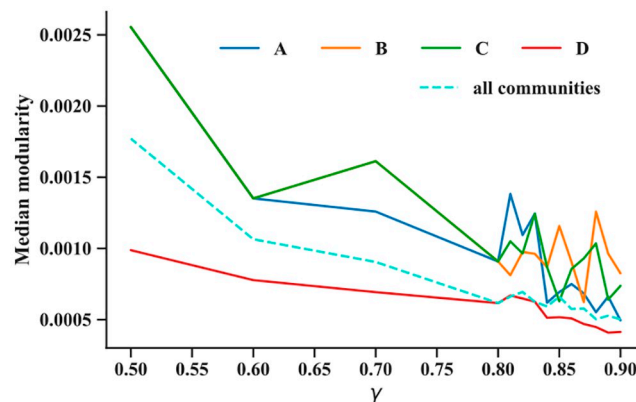
Fig 7. Median community modularity (rescaled with community size) for four chromatin groups A—D (solid lines) across different scale parameters γ. The dashed line shows the median for the entire network. A, B, and C communities have higher modularity than D and the whole network.

The community modularity varies significantly within and between chromatin groups A–D (S9 Fig). We also found that $\mathcal{M}_c$ grows linearly with the community sizes (number of domains) (S9 Fig). Therefore, to make a fair comparison, we plotted the median modularity rescaled with the community sizes (Fig 7). The solid lines represent each chromatin group, including the global median modularity as a reference curve (dashed). These curves show that A–C communities ("active chromatin") have higher modularity than chromatin group D ("inactive chromatin") and the entire Hi-C network. This implies that A–C communities partition the network better than the D communities. To complement Fig 7, we show the $\mathcal{M}$ variability in S10 Fig.

In addition to forming tighter node clusters, A–C communities tend to nest with each other (as shown in Fig 4 and exemplified on a subset of domains in Supplementary Material, S7 Fig). 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 (Results: Modeling non-nested chromosome folding), random nesting breaks ideal hierarchies.

Conclusion

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 substantial 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 chromatin 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.

Our results derive from 100 kb Hi-C data. However, our approach is not restricted to any specific resolution or interaction matrix. It can efficiently analyze various chromatin interaction matrices such as single-cell Hi-C (scHi-C [38]), HiCap [39], HiChIP [40], and distance matrices [41]. Nevertheless, modifications to the GenLouvain null model may be necessary for

some of these scenarios. For instance, if using Hi-C at higher resolution, e.g., 1 kb, numerous contacts appear inside TADs where the contacts decay as a power-law with an exponent of −0.75, rather than approximately −1 [42].

Next, we use the GenLouvain method to extract 3D communities. However, generating communities 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 where multi-scale interactions govern the organization [43]. 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 mainly 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 [44] 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). One of the authors of this paper tried the Leiden algorithm to study general community inconsistency problems [43]. That experience showed that the specific property of reducing the inconsistency does hamper proper detection of scale-dependent inconsistency. 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.

There are several TAD-finding methods [45]. These can be broadly categorized into feature-based algorithms, clustering methods, and graph-partitioning tools [46]. In this paper, we use a technique that falls into the graph-partitioning category, which includes perhaps the most popular community-detection algorithm based on modularity maximization. For example, Ref. [47] finds TADs using Louvain but assumes that the background connectivity is a random network under given node degrees (the Newman-Girvan model). In contrast, we use the fractal-globule null model, which better agrees with the empirical distance decay in contact probability in human Hi-C maps. Although the approach is similar to ours—varying γ and extracting communities—there are meaningful quantitative differences. For example, the fractal globule model tends to capture widely spread and delocalized 3D communities, while the Newman-Girvan model typically groups contiguous DNA stretches into local communities, like TADs. In addition, another reference combines maximum modularity and Hi-C-like distance decay and extracts communities for different γ values [48]. However, they treat TADs as unbroken DNA stretches, not delocalized as we do here. Using polymer simulations [16], we demonstrated that this generalization partitions spatially close monomers into meaningful 3D communities.

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 in 3D proximity to regulate transcription. These interactions are not random, so they could contribute to shaping the 3D structure toward a perfect hierarchy. While we lack data to validate this hypothesis, we note that our simple folding model requires fine-tuned interactions to create an ideal hierarchical order and that 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, epigenetic 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 manifesting as a broad nestedness distribution, as in Fig 4, and a less hierarchical structure than active chromatin.

Several lines of evidence indicate that compartments and TADs emerge from distinct mechanisms, like loop extrusion and phase separation, and are not a hierarchy that stems from identical phenomena operating on different scales. Our paper sheds new light on the hierarchical organization resulting from these phenomena. 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 further open 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.

Supporting information

S1 Fig. Division of the Hi-C network into 3D communities and domains across 16 γ values for human chromosomes 22(a), 5(b), and 3(c). Each stripe represents a community partition for a single γ value. Within each stripe, vertical lines separate two adjacent DNA segments that belong to different communities. The white area shows the centromere. The top turquoise stripes in (a)—(c) show the domains. These are DNA segments that did not break across the shown γ range. In (a), we label two domains '1' and '96,' representing the first a last domain along chromosome 22. The domain IDs follow sequential order along the chromosome so that domain 2 is a linear neighbor of domains 1 and 3.
(TIF)

S2 Fig. Community-domain network. (a) We illustrate the domains as a chain of circles starting from '1' and ending with '96.' With colored links, we show domain memberships in two communities: $7_{0.9}$ and $4_{0.7}$. (b) Example of community pair overlap in a bipartite graph. The domains in the bottom layer connect to communities in the upper layer. White circles show domains that do not belong to any community (excluded from the nestedness analysis).
(TIF)

S3 Fig. Individual chromosome nestedness. (top) Nestedness histograms for chromosomes 3, 5, 10, and 22. (bottom) The fraction of N_{ij} scores that are significant (orange) or random (blue-green).
(TIF)

S4 Fig. Characterization of domains and communities (human chromosome 10). (a) Letter-value plot showing the size distribution of irreducible domains. The domain sizes vary between $\sim 1 - 30$ Hi-C bins. The median domain size is 1 Hi-C bin (100 kb). (b) The scale-dependent number of communities (defined by γ). The number of communities grows exponentially with γ (red).
(TIF)

S5 Fig. N_{ij} histograms for individual γ-pairs (stacked bars with colors specified in legend). To better illustrate range $-1 < N_{ij} < 1$, we truncate the bars at $N_{ij} \pm 1$ if their counts exceed 100. a) pairs between $\gamma = 0.6$ and all other γ. Cross-scale interaction between communities found at $\gamma = 0.6$ and all other communities are fully nested, fully segregated, but also their nestedness is close to random ($-0.5 < N_{ij} < 0.5$); b) for $\gamma = 0.7$ and all other γ. The distribution changes for $N_{ij} \sim 0$), showing that communities tend to be more nested (more counts in the

range: $0 < N_{ij} < 0.5$); c) and d) for $\gamma = 0.8$ and 0.9 versus all other γ. We observe that distribution peaks near $N_{ij} = 0$ when comparing to other distributions.
(TIF)

S6 Fig. Nestedness and chromatin type. (a) Nestedness distributions (N_{ij}) between communities of the same type (either A, B, C, or D). (b) CDF of N_{ij} distributions. The vertical bars indicate the Kolmogorov-Smirnov distance between 'DD' distribution (thick pink line) and all others.
(TIF)

S7 Fig. Rearrangements of chromatin domains across structural scales. Color codes represent their dominant chromatin types. Red domains are enriched in A/B/C states, blue domains in the D-group, green domains in a combination of A/B/C and D-group states, and gray domains show no significant enrichment (p-value set at 0.025). The domain content of communities varies, with some dominated by active chromatin types and others suppressed, similar to the well-known A/B compartmentalization. As resolution increases from $\gamma = 0.6$ to $\gamma = 0.9$, domains reorganize into a community dominated by A/B/C chromatin. At intermediate scales, domains preferentially exchange between structural scales while maintaining biological similarity.
(TIF)

S8 Fig. Optimal Q analysis for chromosome 10. (a) CDF for nestedness score for the real data (black thick line) and models associated with varying Q. In all CDFs, the extreme $N_{ij} = \pm 1$ values are excluded. (b) Kolmogorov-Smirnov distance, D, versus the reshuffling parameter Q. We find the minimal distance when $Q = 0.3$. This represents the optimal $Q_{rmopt.}$. (c) Normalized N_{ij} distributions for chromosome 10 and $Q = 0.3$ with $N_{ij} = \pm 1$ removed from the plot. (d) Same as panel c), but including all N_{ij} counts. We use log-scaled axes to fit the ± 1 peaks.
(TIF)

S9 Fig. Community modularity for four γ values: (a) $\gamma = 0.9$, (b) $\gamma = 0.8$, (c) $\gamma = 0.7$, (d) $\gamma = 0.6$. The top panels show a linear regression fit between community size (number of domains) and modularity (Eq (1), main text); We observe a nearly linear relationship. The bottom panels show the community modularity rescaled with the community size. We define the A—D chromatin groups in Methods: Chromatin states and folds of enrichment (main text). The 'NA' group represents communities that are not enriched in any chromatin group.
(TIF)

S10 Fig. Community modularity (rescaled with community size) for chromatin groups A—D across different scale parameters γ. The top four plots visualize community modularity as bar plots, where values of median modularity are connected across γ. In bar plots, we compare groups A (blue), B (yellow), and C (green) with group D (red bar plot). However, the D group we compare with the modularity of all communities (light blue plot).
(TIF)

S1 Text. Communities and domains for chromosomes 3, 5, and 22.
(DOCX)

S2 Text. Community nestedness for chromosomes 3, 5, 10, and 22.
(DOCX)

S3 Text. Characterization of irreducible domains and structural scales.
(DOCX)

S4 Text. Nestedness distribution for specific γ-pairs.
(DOCX)

S5 Text. Kolmogorov-Smirnov test for nestedness and chromatin types.
(DOCX)

S6 Text. Folding pathways and chromatin types.
(DOCX)

S7 Text. Optimal Q-parameter.
(DOCX)

S8 Text. Community modularity and chromatin type.
(DOCX)

Author Contributions

Conceptualization: Dolores Bernenko, Sang Hoon Lee, Per Stenberg, Ludvig Lizana.

Data curation: Dolores Bernenko, Sang Hoon Lee, Ludvig Lizana.

Formal analysis: Dolores Bernenko, Sang Hoon Lee, Per Stenberg, Ludvig Lizana.

Funding acquisition: Ludvig Lizana.

Investigation: Dolores Bernenko, Sang Hoon Lee, Per Stenberg, Ludvig Lizana.

Methodology: Dolores Bernenko, Sang Hoon Lee, Per Stenberg, Ludvig Lizana.

Project administration: Ludvig Lizana.

Resources: Ludvig Lizana.

Software: Dolores Bernenko, Sang Hoon Lee, Ludvig Lizana.

Supervision: Sang Hoon Lee, Per Stenberg, Ludvig Lizana.

Validation: Dolores Bernenko, Sang Hoon Lee, Per Stenberg, Ludvig Lizana.

Visualization: Dolores Bernenko, Sang Hoon Lee, Ludvig Lizana.

Writing – original draft: Dolores Bernenko, Sang Hoon Lee, Ludvig Lizana.

Writing – review & editing: Dolores Bernenko, Sang Hoon Lee, Per Stenberg, Ludvig Lizana.

References

1. Schwartz YB, Cavalli G. Three-dimensional genome organization and function in Drosophila. Genetics. 2017; 205(1):5–24. https://doi.org/10.1534/genetics.115.185132 PMID: 28049701

2. Bonev B, Cavalli G. Organization and function of the 3D genome. Nature Reviews Genetics. 2016; 17 (11):661–678. https://doi.org/10.1038/nrg.2016.112 PMID: 27739532

3. Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. science. 2009; 326(5950):289–293. https://doi.org/10.1126/science.1181369 PMID: 19815776

4. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014; 159(7):1665–1680. https://doi.org/10.1016/j.cell.2014.11.021 PMID: 25497547

5. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, et al. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell. 2012; 148(3):458–472. https://doi.org/10.1016/j.cell.2012.01.010 PMID: 22265598

6. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012; 485(7398):376–380. https://doi.org/10.1038/nature11082 PMID: 22495300

7. Kaushal A, Mohana G, Dorier J, Özdemir I, Omer A, Cousin P, et al. CTCF loss has limited effects on global genome architecture in Drosophila despite critical regulatory functions. Nature communications. 2021; 12(1):1–16. https://doi.org/10.1038/s41467-021-21366-2 PMID: 33579945

8. Remeseiro S, Hörnblad A, Spitz F. Gene regulation during development in the light of topologically associating domains. Wiley Interdisciplinary Reviews: Developmental Biology. 2016; 5(2):169–185. https://doi.org/10.1002/wdev.218 PMID: 26558551

9. Szabo Q, Bantignies F, Cavalli G. Principles of genome folding into topologically associating domains. Science advances. 2019; 5(4):eaaw1668. https://doi.org/10.1126/sciadv.aaw1668 PMID: 30989119

10. Kumar R, Lizana L, Stenberg P. Genomic 3D compartments emerge from unfolding mitotic chromosomes. Chromosoma. 2019; 128(1):15–20. https://doi.org/10.1007/s00412-018-0684-7 PMID: 30357462

11. Sarnataro S, Chiariello AM, Esposito A, Prisco A, Nicodemi M. Structure of the human chromosome interaction network. PLoS One. 2017; 12(11):e0188201. https://doi.org/10.1371/journal.pone.0188201 PMID: 29141034

12. Fraser J, Ferrai C, Chiariello AM, Schueler M, Rito T, Laudanno G, et al. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. Molecular systems biology. 2015; 11(12):852. https://doi.org/10.15252/msb.20156492 PMID: 26700852

13. An L, Yang T, Yang J, Nuebler J, Xiang G, Hardison RC, et al. OnTAD: hierarchical domain structure reveals the divergence of activity among TADs and boundaries. Genome biology. 2019; 20(1):1–16. https://doi.org/10.1186/s13059-019-1893-y PMID: 31847870

14. Zhang YW, Wang MB, Li SC. SuperTAD: robust detection of hierarchical topologically associated domains with optimized structural information. Genome biology. 2021; 22(1):1–20. https://doi.org/10.1186/s13059-020-02234-6 PMID: 33494803

15. Zhan Y, Mariani L, Barozzi I, Schulz EG, Blüthgen N, Stadler M, et al. 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. 2017; 27(3):479–490. https://doi.org/10.1101/gr.212803.116 PMID: 28057745

16. Lee SH, Kim Y, Lee S, Durang X, Stenberg P, Jeon JH, et al. Mapping the spectrum of 3D communities in human chromosome conformation capture data. Scientific reports. 2019; 9(1):1–7. https://doi.org/10.1038/s41598-019-42212-y PMID: 31048738

17. Nuebler J, Fudenberg G, Imakaev M, Abdennur N, Mirny LA. Chromatin organization by an interplay of loop extrusion and compartmental segregation. Proceedings of the National Academy of Sciences. 2018; 115(29):E6697–E6706. https://doi.org/10.1073/pnas.1717730115 PMID: 29967174

18. Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, et al. Two independent modes of chromatin organization revealed by cohesin removal. Nature. 2017; 551(7678):51–56. https://doi.org/10.1038/nature24281 PMID: 29094699

19. Bak JH, Kim MH, Liu L, Hyeon C. A unified framework for inferring the multi-scale organization of chromatin domains from Hi-C. PLoS computational biology. 2021; 17(3):e1008834. https://doi.org/10.1371/journal.pcbi.1008834 PMID: 33724986

20. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic acids research. 2002; 30(1):207–210. https://doi.org/10.1093/nar/30.1.207 PMID: 11752295

21. Nyberg M, Ambjörnsson T, Stenberg P, Lizana L. Modeling protein target search in human chromosomes. Physical Review Research. 2021; 3(1):013055. https://doi.org/10.1103/PhysRevResearch.3.013055

22. Kaufmann S, Fuchs C, Gonik M, Khrameeva EE, Mironov AA, Frishman D. Inter-chromosomal contact networks provide insights into Mammalian chromatin organization. PloS one. 2015; 10(5):e0126125. https://doi.org/10.1371/journal.pone.0126125 PMID: 25961318

23. Knight PA, Ruiz D. A fast algorithm for matrix balancing. IMA Journal of Numerical Analysis. 2013; 33(3):1029–1047. https://doi.org/10.1093/imanum/drs019

24. Kumar R, Sobhy H, Stenberg P, Lizana L. Genome contact map explorer: a platform for the comparison, interactive visualization and analysis of genome contact maps. Nucleic Acids Research. 2017; 45(17):e152. https://doi.org/10.1093/nar/gkx644 PMID: 28973466

25. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. 2008; 2008(10):P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

26. Jeub LGS, Bazzi M, Jutla IS, Mucha PJ. A generalized Louvain method for community detection implemented in MATLAB; 2011-2019. Available from: https://github.com/GenLouvain/GenLouvain.

27. Newman MEJ. Equivalence between modularity optimization and maximum likelihood methods for community detection. Phys Rev E. 2016; 94:052315. https://doi.org/10.1103/PhysRevE.94.052315 PMID: 27967199

28. Grosberg A, Rabin Y, Havlin S, Neer A. Crumpled globule model of the three-dimensional structure of DNA. EPL (Europhysics Letters). 1993; 23(5):373. https://doi.org/10.1209/0295-5075/23/5/012

29. Mirny LA. The fractal globule as a model of chromatin architecture in the cell. Chromosome research. 2011; 19(1):37–51. https://doi.org/10.1007/s10577-010-9177-0 PMID: 21274616

30. Sanborn AL, Rao SS, Huang SC, Durand NC, Huntley MH, Jewett AI, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proceedings of the National Academy of Sciences. 2015; 112(47):E6456–E6465. https://doi.org/10.1073/pnas.1518552112 PMID: 26499245

31. Strona G, Veech JA. A new measure of ecological network structure based on node overlap and segregation. Methods in Ecology and Evolution. 2015; 6(8):907–915. https://doi.org/10.1111/2041-210X.12395

32. Veech JA. A probabilistic model for analysing species co-occurrence. Global Ecology and Biogeography. 2013; 22(2):252–260. https://doi.org/10.1111/j.1466-8238.2012.00789.x

33. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Systematic analysis of chromatin state dynamics in nine human cell types. Nature. 2011; 473(7345):43–49.

34. GM12878 Chromatin State Segmentation by HMM from ENCODE/Broad;. Available from: https://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=1295125293_1uAxm5NGeRepzfvVCPEKgWcUZura&db=hg19&g=wgEncodeBroadHmmGm12878HMM.

35. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological). 1995; 57(1):289–300.

36. Seabold S, Perktold J. statsmodels: Econometric and statistical modeling with python. In: 9th Python in Science Conference; 2010.

37. Mauri M, Elli T, Caviglia G, Uboldi G, Azzi M. RAWGraphs: a visualisation platform to create open outputs. In: Proceedings of the 12th biannual conference on Italian SIGCHI chapter; 2017. p. 1–5.

38. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature. 2013; 502(7469):59–64. https://doi.org/10.1038/nature12593 PMID: 24067610

39. Zhigulev A, Sahlén P. Targeted Chromosome Conformation Capture (HiCap). In: Spatial Genome Organization: Methods and Protocols. Springer; 2022. p. 75–94.

40. Chakraborty C, Nissen I, Vincent CA, Hagglund AC, Hornblad A, Remeseiro S. Rewiring of the promoter-enhancer interactome and regulatory landscape in glioblastoma orchestrates gene expression underlying neurogliomal synaptic communication. bioRxiv. 2022; p. 2022–11.

41. Bintu B, Mateo LJ, Su JH, Sinnott-Armstrong NA, Parker M, Kinrot S, et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science. 2018; 362(6413): eaau1783. https://doi.org/10.1126/science.aau1783 PMID: 30361340

42. Holmgren A, Bernenko D, Lizana L. Mapping robust multiscale communities in chromosome contact networks. arXiv preprint arXiv:221208456. 2022;.

43. Lee D, Lee SH, Kim BJ, Kim H, et al. Consistency landscape of network communities. Physical Review E. 2021; 103(5):052306. https://doi.org/10.1103/PhysRevE.103.052306 PMID: 34134219

44. Traag VA, Waltman L, Van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Scientific reports. 2019; 9(1):1–12. https://doi.org/10.1038/s41598-019-41695-z PMID: 30914743

45. Zufferey M, Tavernari D, Oricchio E, Ciriello G. Comparison of computational methods for the identification of topologically associating domains. Genome biology. 2018; 19(1):1–18. https://doi.org/10.1186/s13059-018-1596-9 PMID: 30526631

46. Sefer E. A comparison of topologically associating domain callers over mammals at high resolution. BMC bioinformatics. 2022; 23(1):127. https://doi.org/10.1186/s12859-022-04674-2 PMID: 35413815

47. Norton HK, Emerson DJ, Huang H, Kim J, Titus KR, Gu S, et al. Detecting hierarchical genome folding with network modularity. Nature methods. 2018; 15(2):119–122. https://doi.org/10.1038/nmeth.4560 PMID: 29334377

48. Yan KK, Lou S, Gerstein M. MrTADFinder: A network modularity based approach to identify topologically associating domains in multiple resolutions. PLoS computational biology. 2017; 13(7):e1005647. https://doi.org/10.1371/journal.pcbi.1005647 PMID: 28742097