Transient crosslinking kinetics optimize gene cluster interactions

Benjamin Walker[1], Dane Taylor[2,3], Josh Lawrimore[4], Caitlin Hult[5,6], David Adalsteinsson[1], Kerry Bloom[4], M. Gregory Forest[1,7] *

1 Department of Mathematics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, 2 Department of Mathematics, University at Buffalo, State University of New York, Buffalo, New York, United States of America, 3 Computational and Data Enabled Science and Engineering, University at Buffalo, State University of New York, Buffalo, New York, United States of America, 4 Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, 5 Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan, United States of America, 6 Department of Chemical Engineering, University of Michigan, Ann Arbor, Michigan, United States of America, 7 Departments of Applied Physical Sciences and Biomedical Engineering, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America

* forest@unc.edu

Abstract

Our understanding of how chromosomes structurally organize and dynamically interact has been revolutionized through the lens of long-chain polymer physics. Major protein contributors to chromosome structure and dynamics are condensin and cohesin that stochastically generate loops within and between chains, and entrap proximal strands of sister chromatids. In this paper, we explore the ability of transient, protein-mediated, gene-gene crosslinks to induce clusters of genes, thereby dynamic architecture, within the highly repeated ribosomal DNA that comprises the nucleolus of budding yeast. We implement three approaches: live cell microscopy; computational modeling of the full genome during G1 in budding yeast, exploring four decades of timescales for transient crosslinks between $5 kbp$ domains (genes) in the nucleolus on Chromosome XII; and, temporal network models with automated community (cluster) detection algorithms applied to the full range of 4D modeling datasets. The data analysis tools detect and track gene clusters, their size, number, persistence time, and their plasticity (deformation). Of biological significance, our analysis reveals an optimal mean crosslink lifetime that promotes pairwise and cluster gene interactions through "flexible" clustering. In this state, large gene clusters self-assemble yet frequently interact (merge and separate), marked by gene exchanges between clusters, which in turn maximizes global gene interactions in the nucleolus. This regime stands between two limiting cases each with far less global gene interactions: with shorter crosslink lifetimes, "rigid" clustering emerges with clusters that interact infrequently; with longer crosslink lifetimes, there is a dissolution of clusters. These observations are compared with imaging experiments on a normal yeast strain and two condensin-modified mutant cell strains. We apply the same image analysis pipeline to the experimental and simulated datasets, providing support for the modeling predictions.

Author summary

The spatiotemporal organization of the genome plays an important role in cellular processes involving DNA, but remains poorly understood, especially in the nucleolus, which does not facilitate conventional techniques. Polymer chain models have shown ability in recent years to make accurate predictions of the dynamics of the genome. We consider a polymer bead-chain model of the full yeast genome during the interphase portion of the cell cycle, featuring special dynamic crosslinking to model the effects of structural maintenance proteins in the nucleolus, and investigate how the kinetic timescale on which the crosslinks bind and unbind affects the resulting dynamics inside the nucleolus. It was previously known that when this timescale is sufficiently short, large, stable clusters appear, but when it is long, there is no resulting structure. We find that there additionally exists a range of timescales for which flexible clusters appear, in which beads frequently enter and leave clusters. Furthermore, we demonstrate that these flexible clusters maximize the cross-communication between beads in the nucleolus. Finally, we apply network temporal community detection algorithms to identify what beads are in what communities at what times, in a way that is more robust and objective than conventional visual-based methods.

Introduction

The 4D Nucleome Project [1] proposes the integration of diverse approaches: increasingly powerful chromosome conformation capture techniques including high-throughput chromosome conformation capture (Hi-C); statistical and topological analyses of these massive Hi-C datasets; 3-dimensional (3D) and 4D super-resolution imaging datasets; and computational modeling approaches, both constrained by and independent of Hi-C datasets. The project aims to gain mechanistic understanding of 3D structure and dynamics of the genome within the nucleus, and to learn how the active chromosome architecture facilitates nuclear functions. In this paper, we contribute to these aims by combining three approaches: (i) live cell microscopy for experiments studying the effect on gene clustering for normal and condensin-modified mutant cell strains; (ii) first-principles-based, computational modeling based on the statistical physics of chromosome polymers coupled with transient gene-gene crosslinks formed by condensin proteins; and (iii) analysis of the dynamic chromosome architecture with temporal network community detection algorithms applied to 4D modeling datasets across four decades of crosslinking timescales.

Our present understanding of basic principles that govern high-order genome organization can be attributed to incorporation of the physical properties of long-chain polymers [2–7]. The fluctuations of long-chain polymers, numerically simulated with Rouse-like bead-spring chain models of chromosomes confined to the nucleus, capture the tendency of chromosomes to self-associate and occupy territories [8–11] In addition, these models make predictions with regard to the spatial and dynamic timescales of inter-chromosomal interactions, a dynamic analog of topologically associated domains. The convergence of robust physical models with high-throughput biological data reveals the fractal nature of chromosome organization, namely an apparently self-similar cascade of loops within loops, or structure within structure, as one examines chromosomes at higher and higher resolution [12–14].

De novo stochastic bead-spring polymer models of the dynamics and conformation of "live" chromosomes, plus the action on top of the genome by transient binding interactions of structural maintenance of chromosome (*SMC*) proteins, e.g. condensin, provide complementary information to chromosome conformation capture (3C) techniques, genome-wide high-

throughput (Hi-C) techniques, and restraint-based modeling [12, 15–20]. 3C and Hi-C experiments rely on population averages of gene-gene proximity on all chromosomes over many thousands of dead cells whose chromosomes have been permanently crosslinked by formaldehyde; the restraint-based modeling approach then explores 3D chromosome architecture that optimizes agreement with the experimental data on gene-gene frequency and proximity across the genome. Many powerful inferences have been drawn from both Hi-C and polymer modeling approaches, using analyses of empirical and synthetic datasets encoding maps related to the pairwise distances between genes.

A common major limitation for existing polymer models and whole-genome contact maps in mammalian cells is in mapping two essential regions of the chromosome, namely the centromere and the nucleolus. The centromere, essential for chromosome segregation, and the nucleolus, the sub-nuclear domain of ribosomal DNA, are comprised of megabases of repeated DNA (centromere satellites and nucleolus rDNA). Furthermore, these regions are not captured in methods used for generating contact maps. We have used single live cell imaging of the nucleolus in budding yeast coupled with whole genome polymer modeling to explore the minimal requirements for sub-compartmentalization. Implementation of protein-mediated cross-linking within the nucleolus is sufficient to partition this region of the genome from the remaining chromosomes. Furthermore, stochastic polymer models reveal that the relative timescales of crosslinking kinetics and fluctuations of the chromosome chains have a profound influence on nucleolar morphology [21].

In single cells, the positional fluctuations of tagged DNA sequences on specific chromosomes [22–24] through the lac operator/lac repressor reporter system validate the bead-spring models. Chromosomes fluctuate as predicted for the conformational dynamics of idealized Rouse chains [25].

Polymer simulations over the entire genome have revealed the ability of relatively fast binding and unbinding, and thereby short-lived (fraction of a second) protein crosslinks to concentrate the rDNA chain sequence in a smaller volume and increase the simulated fluorescent signal intensity variance when the model datasets were convolved with a point spread function to create two-dimensional, maximum intensity projections [21]. Visualizations of the monomers in the simulations revealed that the fastest kinetics explored, or shortest-lived crosslinks (∼ .09s), generated several clusters of high polymer density, and overall compaction of the nucleolus. In contrast, much slower kinetics (decades longer-lived protein crosslinks (∼ 90s)) tended to homogenize the fluorescent signal intensity as evidenced in the decrease in simulated fluorescent signal intensity variance. These model visualizations were consistent with experimental results on live budding yeast.

There is a growing interest to analyze Hi-C datasets and model chromosome interactions using network models [26–28], which has opened the door to study chromosomal datasets using network-based algorithms including centrality analysis [29, 30] and community detection [31, 32]. In this context, a 'gene cluster' is a set of genes that are in close physical proximity, and it is represented in a network by a community of nodes (i.e., a set of nodes between which there is a prevalence of edges). These detection algorithms perform an unbiased search for robust structures (communities or clusters) at the scale they exist in an automated manner, quantifying how chromosome conformational changes can precede changes to transcription factors and gene expression [33, 34] and leading to new approaches for cellular reprogramming [29, 35].

Here, we apply *temporal* community detection algorithms including multilayer modularity [36, 37] to simulated 4D datasets over four decades of *SMC*-binding kinetic timescales. This approach integrates both temporal and spatial information so that each community now represents a set of genes that are not only nearby one another, *but they remain in close proximity*

for some duration. This approach allows us to detect, track and label transient gene communities (clusters) in the nucleolus. Simultaneously, we record summary statistics on the sizes and numbers as well as persistence times of communities, and the frequencies of community interactions leading to gene exchanges. We likewise record standard bead-bead summary statistics. In doing so, we detect spatial and temporal organization at the scales they exist, beyond two-point (gene-gene) spatial proximity statistics. We identify the timescales over which spatial organization persists, linking the timescales to the cluster identification algorithm. Since clusters can deform through the flux of genes into and out of clusters, we further are able to identify crosslink timescales for which spatial clustering persists over extended timescales, and whether individual clusters are relatively permanent or experience frequent interactions and gene exchanges. Perhaps the most striking prediction of our modeling and data analysis is that specific gene organization tasks (amplified below) are optimized at a relatively short crosslink timescale, on the order of. 19 sec.

With these network tools applied to physics-based 4D nucleome simulated datasets, we explore the mechanistic basis for the experimentally observed variance in nucleolar morphology. *From a high-resolution sampling of the timescales for crosslinking of 5k base pair (bp) domains, 4D model simulations of the yeast genome reveal the nucleolus on Chromosome XII undergoes a stark transition in dynamics and structure, and does so within a narrow "mean on" crosslink timescale range of* .09 − 1.6 *sec.* A highly stable clustering regime exists with relatively short-lived crosslinks (.09 sec), with relatively few cluster interactions and gene exchanges, as reported previously in [21]. At slightly longer-lived (.19 sec) timescales, a novel "flexible" behavior is revealed. Gene clusters continue to self-organize, yet clusters are more mobile, frequently interact, and exchange genes. Indeed, *there is a peak timescale, marked by highly mobile gene clusters, at which both pairwise and community-scale gene interactions are maximized.* As the binding affinity of crosslinker proteins increases only slightly longer (1.6 sec), the community-scale structure has dissolved, with no identifiable nucleolar sub-substructure. See Fig 1.

From a methods perspective, our analysis of the 4D simulated datasets is based on network modeling and a temporal community-detection algorithm known as multilayer modularity [37]. From a biological perspective, this tunable dynamic self-organization reflects a powerful mechanism to coordinate gene regulation and the coalescence of non-contiguous genes into identifiable clusters (substructures). The transition shown in Fig 1 occurs *within* such a narrow crosslinker timescale regime (.09 − 1.6 sec), suggesting a relatively simple mechanism to control dynamic sub-organization of the genome; indeed we performed and report experiments below to support this prediction. Finally, we emphasize the counter-intuitive nature of this mechanism: clustering is most often associated with segregation, however we observe that the dynamic element of flexible clusters facilitates an overall increase in global gene interactions in the nucleolus.

Results

Our results are based on analysis of the model of the yeast genome originally presented in [21]. This model aims to have a sufficiently resolution to be able to qualitatively reproduce behaviors observed in the nucleolus (as is confirmed in [21]) while still being practical to computationally evaluate.

An overview of the model is provided in the Model subsection in the Materials and methods section. We begin by extending results from [21] in context with the extensions presented in this paper. In addition to simulated 4D datasets as in [21], here we compare wild-type and SMC protein-altered mutant experimental data to explore how the SMC-protein crosslinking timescale μ influences behavior of the nucleolus. In [21], three values of $\mu \in \{0.09, 0.9, 90\}$ were

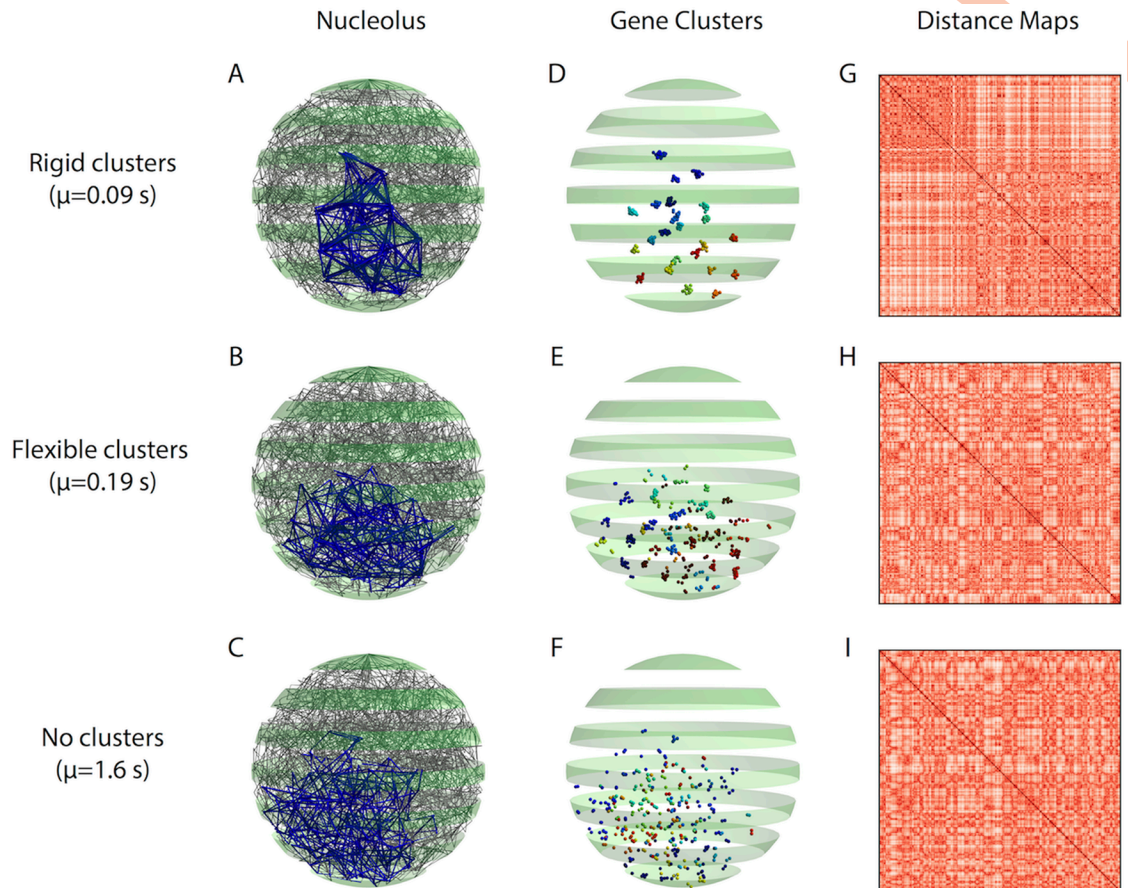| Nucleolus	Gene Clusters	Distance Maps

Fig 1. Interphase yeast genome (A-C: Full, D-I: Nucleolus) using the polymer bead-spring model of [21] (see Section: Model) that implements transient SMC-protein-mediated crosslinking of 5k bp domains within the nucleolus on Chromosome XII. The top, middle, and bottom rows depict chromosome conformations for three values of the kinetic timescale parameter μ: (top, $\mu = 0.09$) induces rigid clusters; (center, $\mu = 0.19$) induces flexible clusters; (bottom, $\mu = 1.6$) induces no clusters. We identify the timescales with the intra-nucleolar clustering behavior they induce. (A)–(C) 3D "snapshots" of all 16 yeast chromosomes during interphase. Blue beads and edges highlight the nucleolus. (D)–(F) Visualization of nucleolar beads ($5kbp$ chromosome domains) that self-organize into clusters. The beads' positions are identical to those in (A)–(C) and their colors indicate their cluster labels, identified using modularity optimization [38] as described in Methods Section: Cluster identification via network community detection. (G)–(I) Heatmaps of the pairwise distances between beads in the nucleolus from one snapshot of the 4D time series, which provide an analogue of Hi-C bead-bead proximity data (see Methods Section: Pairwise-distance maps for high-throughput chromosome conformation capture (Hi-C)). Note that it is difficult to predict the absence/presence of clusters from heat maps.

https://doi.org/10.1371/journal.pcbi.1007124.g001

studied. In this paper, we logarithmically sample across the full range of μ values between 0.09 and 90, providing a detailed investigation of consequences induced by the crosslinking timescale μ. We then construct simulated microscope images [21] for all 4D simulated datasets and compare them to experimental images of yeast nucleoli, including wild-type (WT), hmo1Δ and fob1Δ. We show that the mutations hmo1Δ and fob1Δ alter the nucleolus similarly to changes induced by varying μ in our model.

In [21], 4D imaging of the simulated nucleolus beads investigated clusters, sets of beads such that the distances between genes in the same cluster are smaller than those between genes in different clusters, providing visual evidence of clusters at the shortest crosslinking timescale ($\mu \approx .09$ sec) and a lack of clusters at the longest timescales ($\mu \approx 90$ sec). Herein, we simulated across four decades of μ, revealing transitions in dynamic architecture that were not explored in [21] since we did not have the tools to automate detection apart from obvious visual images

at polar extremes, robust clusters and no clusters. Depending on the crosslinking timescale μ, the nucleolar beads can aggregate into self-organized clusters, which may or may not change over time via mergers/divisions of clusters with bead exchanges between clusters. The choice $\mu \approx 1.6$ gives rise to plasticity in the dynamic architecture of the nucleolus, whereby beads aggregate into clusters, yet the clusters frequently come into contact, merge into a super cluster, then rapidly divide with several exchanges of beads. This process of frequent cluster interactions optimizes the traditional statistics of the frequency of pairwise bead-bead interactions within the nucleolus, while furthermore generalizing pairwise interactions to gene cluster interactions and the frequency of pairs of beads to share the same cluster. By a fine sampling of the crosslinking timescales, coupled with the automated dynamic community detection algorithms, we discover the novel non-monotone behavior in the interaction statistics, at both gene-gene and gene cluster scales.

Motivated by the need to identify, label, and track clusters over time, we adapt a method based on network modeling and the temporal community-detection algorithm known as multilayer modularity [37]. We identify communities in temporal networks that represent clusters of genes that remain in close proximity for a duration. The appropriate size and duration of communities/clusters are revealed as part of the algorithm, rather than specified *a priori*. Our methodology not only confirms the striking visual evidence of robust clustering at $\mu = .09$ in [21], but allows us to track and label whether clusters of genes exist, cluster interactions (merger and division), their frequency, and gene exchanges per interaction. If there are no clusters, the algorithm reports that all clusters are essentially single genes with doublets and triplets that interact and deform. Our methods therefore reveal dynamic sub-organization features at the scales they exist, automatically and unbiased, as we scan 4D datasets across 4 decades of crosslinking timescales. We show our algorithm efficiently identifies robust as well as flexible communities using a cost function that factors in both physical proximity and temporal coherence. Finally, we demonstrate that the identification and analysis of time-evolving communities reveals a larger scale explanation for the non-monotone behavior in bead-bead interaction statistics versus crosslinking timescales. The explanation lies precisely in the transition from robust, non-interacting clusters for $\mu = .09$ sec, to flexible clusters for $\mu = .19$ sec, and then a slow dissolution of clusters as μ increases, with essentially no clustering by $\mu = 1.6$.

Transient crosslinking timescale influences nucleolus clustering

We first focus in the relatively short crosslink timescale regime, extending the simulations of [21] at discrete values $\mu = 0.09, 0.9, 90$. These will establish a basic understanding of how the kinetic timescale μ for crosslinking sensitively affects the organization of the nucleolus and the dynamics of the architecture. From our refined simulations across the above four decades, the essence of the story can be told with results for three selected values $\mu \in \{0.09, 0.19, 1.6\}$. In Fig 1(A)–1(C), we present visualizations, i.e., "snapshots," of the beads' 3D positions during the simulations. The nucleolus on Chromosome XII is highlighted in blue and all remaining chromosome arms are colored gray. In Fig 1(D)–1(F), we show only the nucleolar beads, which are colored according to the network community detection analyses that we describe in the following sections. We also show videos of the time evolution of the beads, along with a simulated microscope projection, for each timescale in S1, S2 and S3 Videos.

Based on Fig 1(A)–1(F) and the videos, we identify three qualitative regimes for nucleolus clustering:

1. *rigid clustering* whereby strong, stable clusters arise, e.g., with $\mu = 0.09$.

2. *flexible clustering, or cluster plasticity* with slightly weaker clustering whereby communities of genes form and persist, however the clusters frequently interact, merge, and divide, swapping genes per interaction, e.g., with $\mu = 0.19$.

3. *non-clustering* with a lack of robust communities in which the beads act as lone units, pairs or triplets, e.g., with $\mu = 1.6$.

We will continue to use this terminology when referring to these three clustering regimes. We note that [21] discovered the two extreme regimes: robust clusters for $\mu = 0.09$, and the lack of clusters for $\mu = 90$. As they did not finely sample the decades of timescales in between, they did not discover that the transition from robust to no clustering is in fact non-monotone with respect to gene-gene interactions, nor that the transition is essentially complete already at $\mu = 1.6$, and that the most biologically interesting and relevant regime occurs at $\mu = 0.19$. Furthermore, without automated structure detection algorithms, they would not have been able to detect and dynamically track clusters of genes and their interactions that explain the peak in gene-gene interactions at $\mu = 0.19$. This transition behavior and the optimal properties that arise will be the focus of several sections to follow.

In Fig 1(G)–1(I), we show heatmaps of the bead-bead distances associated with the bead positions of the snapshots in (D)–(F), identical to those in (A)–(C); construction of heatmaps is described in the Methods Section: Pairwise-distance maps for high-throughput chromosome conformation capture (Hi-C). Heatmaps are widely used in Hi-C to depict population averages of pairwise gene-gene proximity data [16, 39–43] and in simulated data from polymer bead-spring models, both from 3D snapshots and time averages [9, 11, 21]. Comparing the second and third columns of Fig 1, we note the difficulty (false negatives and false positives) in detecting the presence of structure and sub-organization in column 2 from visual examination of heatmaps in column 3.

As shown in [21], the time average of 4D simulated datasets, even in the strong clustering regime, wipes out the sub-structure of snapshots when averaging over the entire G1 phase. An alternative approach has been to use polymer modeling to generate chromosome conformations, and to select those conformations that best match Hi-C data, so-called restraint-based polymer modeling [1]. Simultaneously, there have been efforts to develop methodologies to identify gene clusters in a rigorous and automated way from Hi-C data [26–28]. Our conclusion is that there is a need for a more reliable and objective method to study the clustering of chromosome domains in the nucleolus, especially spatio-temporal methods that take into account how bead positions and sub-organization change with time, weighing both spatial proximity and temporal coherence in the detection method. In the following sections, we present a scalable and automated technique to identify and track the dynamics of clusters. First, however, we will present new experiments that provide empirical evidence for clustering in the nucleolus.

Evidence of nucleolus clustering in experimental and simulated microscopy images

We conducted experiments to qualitatively compare image-based cluster analysis between our model and empirical measurements obtained from live cell microscopy and demonstrate the effect that SMC protein mutation can have on clustering in the nucleolus, extending the results previously reported in [21]. Here, we study three yeast strains: wild-type (WT), fob1 and hmo1. Importantly, fob1Δ and hmo1Δ are mutations that lack key proteins reported to crosslink or loop segments of rDNA within the nucleolus. Fob1Δ is required for maintenance of the rDNA copy number and regulates the association of condensin with rDNA repeats
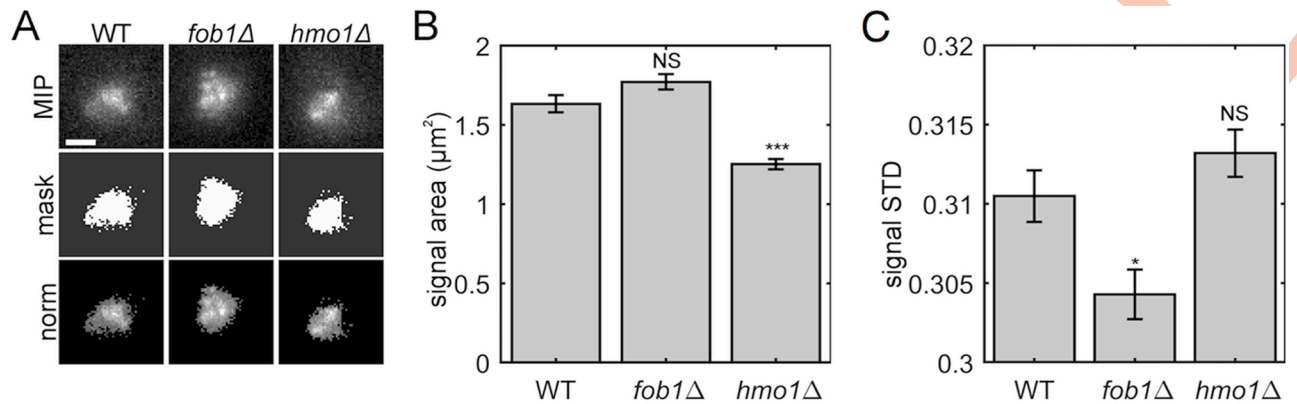
Fig 2. Experimental images of the nucleolus. Algorithmic thresholding of CDC14-GFP reveals alterations in nucleolar area and variance of signal intensity. (A) Top row depicts images of maximum intensity projection of WT, fob1Δ, and hmo1Δ yeast cells containing CDC14-GFP to label the nucleolus. Scale bar is 1 μm. The middle row are images of the binary mask generated by thresholding the projections using Otsu's method. The bottom row are images of the projections where the intensities of pixels below the threshold were set to zero. See Methods Section: Image acquisition and baseline processing for more detail. (B) Bar chart of the nucleolar signal are calculated by the area of the binary mask. (C) Bar chart of the standard deviation of normalized nucleolar signal. In panels (B) and (C), the non-significant changes are labeled 'NS' (see text), the bars represent an average value across n cells, and error bars indicate standard error.

https://doi.org/10.1371/journal.pcbi.1007124.g002

[44, 45]. The replication fork barrier within the rDNA is a binding site for Fob1Δ that, together with several other components (Tof1, Csm1 and Lrs4), are responsible for the concentration of condensin within the nucleolus [44]. Strain hmo1 is an abundant high mobility group protein that localizes to the nucleolus and has been proposed to share functions with UBF1, which is involved in rDNA transcriptional regulation within the nucleolus [46, 47]. Fob1 and Hmo1 are non-essential genes and were deleted from the genome to allow us to study their effect on nucleolus morphology due to functional modifications of crosslinking. See Methods Section: Yeast strains for experiment for further details.

In Fig 2, we present images and analyses of nucleoli of these strains using fluorescent, live-cell microscopy. To visualize nucleoli, we fused Cdc14 protein phosphatase to green fluorescent protein (GFP) [21]. Nucleolar protein fusions occupy a distinct region of the nucleus that is adjacent to the nuclear envelope and (typically) opposed to the spindle pole body. We describe the image acquisition and processing steps in Methods Section: Image acquisition and baseline processing, and we highlight a few details here. Following image acquisition, we construct *maximum intensity projections* (MIP) centered on the nucleolus. See top row of Fig 2(A). Due to potential variation in CDC14-GFP protein copy number and nucleolar/rDNA size from cell to cell, we normalized the nucleolar CDC14-GFP signal after excluding all intensity values below a intensity threshold. To this end, we first selected a threshold using Otsu's method [48], which we implemented using the MATLAB function multithresh. One can interpret the threshold as a binary mask, as shown in the second row of Fig 2(A). After applying the mask, we normalized the nucleolar signal by subtracting all intensities by the minimum value and then dividing them by the new maximum intensity that is obtained after subtraction. The third row of Fig 2(A) depicts normalized images. Mutations of Hmo1 and Fob1 were found to alter the area and signal intensity of nucleoli labeled with CDC14-GFP across a range of intensity thresholds, which we surmise is due to alterations in the architecture of, i.e., clustering within, the nucleolus.

In Fig 2(B) and 2(C), we provide results for an analysis of nucleolar morphology: (B) the area of nucleolar signal; and (C) the standard deviation of the normalized signal. This analysis

was implemented using the numerical algorithms presented in [21], which we further describe in Methods Section: Image analysis.

As shown in Fig 2(B), null mutations of hmo1Δ significantly altered the area of the nucleolar signal, whereas null mutations of fob1Δ did not. This was assessed by a Student's two-tailed T-test, which yielded $p = 3 \times 10^{-8}$ for the former and $p = 0.07$ for the latter. As shown in Fig 2(C), the standard deviations of the normalized images were significantly lowered for the fob1Δ null mutation, but this did not occur for the hmo1Δ null mutation $p = 0.01$ versus $p = 0.2$). The non-significant changes are labeled 'NS' in the figure. The error bars indicate standard errors across n cells, where $n = 84$, 70 and 77 for the WT, fob1Δ and hmo1Δ strains, respectively.

We note that [21] also studied the area and variance of the nucleolus using experimental and simulated images. They found, for example, that the distribution of areas occupied by the nucleolus displays a lognormal distribution for WT cells in G1. Also, recall that we implemented thresholding based on Otsu's method; in contrast, [21] explored a range of threshold values and found qualitatively similar results to be consistent across a range of threshold values. They did not, however, explore the area and variance for simulated images for a wide range of μ, which is the focus of our next experiment.

To explore whether varying the kinetic timescale μ for our simulations yields similar changes as those arising under the fob1Δ and hmo1Δ mutations, we applied the microscope simulator of [21] to our 4D simulated data and analyzed the images using the same image analyses as described in Fig 2. First, we converted our 4D simulated data into a timelapse sequence with 22 time points, i.e., snapshots. Each nucleolus bead was convolved with a point spread function and a maximum intensity projection was created for each timepoint. We depict 11 such images in Fig 3(A). In panel (B), we plot the area of the nucleolar signal (computed using Otsu's threshold) versus μ. Note that the nucleolus area increases as μ increases. In panel (C), we plot the standard deviation of nucleolar signal versus μ, which has the opposite trend. In Fig 3(D), we plot the standard deviation of images obtained after a normalization step that is identical to that implemented for the experimental images (see discussion for Fig 2. Interestingly, the dependence on μ of the signal's standard deviation drastically changes depending on whether or not it is normalized. Given that normalization is required to control for cell-to-cell differences in CDC14-GFP and in nucleolar/rDNA size, we sought develop a metric to measure clustering in the CDC14-GFP signal that was independent of the absolute values of the intensities.

Our final experiment studies cluster formation in the nucleolus and compares clustering observed in the experimental and simulated microscopy images. We developed a cluster detection algorithm written with MATLAB (see Methods Section: Image analysis) and applied it to both the experimental and simulated images. We have made the code available at [49]. In Fig 4(A), we depict images of maximum intensity projections for WT, fob1Δ and hmo1Δ strains with (top row) and without (bottom row) visualizations of detected clusters, which are represented by green circles. In panel (B), we depict identical information as in panel (A) except we show simulated images for three values of μ. In Fig 4(C) and 4(D), we show the number of clusters for the experimental and simulated images, respectively. For the experimental images, we give results for WT, fob1Δ and hmo1Δ, whereas for the simulated images we present results for $\mu \in [.09, 90]$. We observe that the number of clusters was significantly decreased in the hmo1Δ null mutation, but not the fob1Δ null mutation ($p = 0.04$ for hmo1Δ versus $p = 0.3$ for fob1Δ). We also observe that increasing μ yielded a general trend in which there were fewer clusters. Taken together, these data suggest that gene clustering can directly impact the size and shape of the nucleolus. This underscores the need for robust and objective tools for identifying gene clusters.
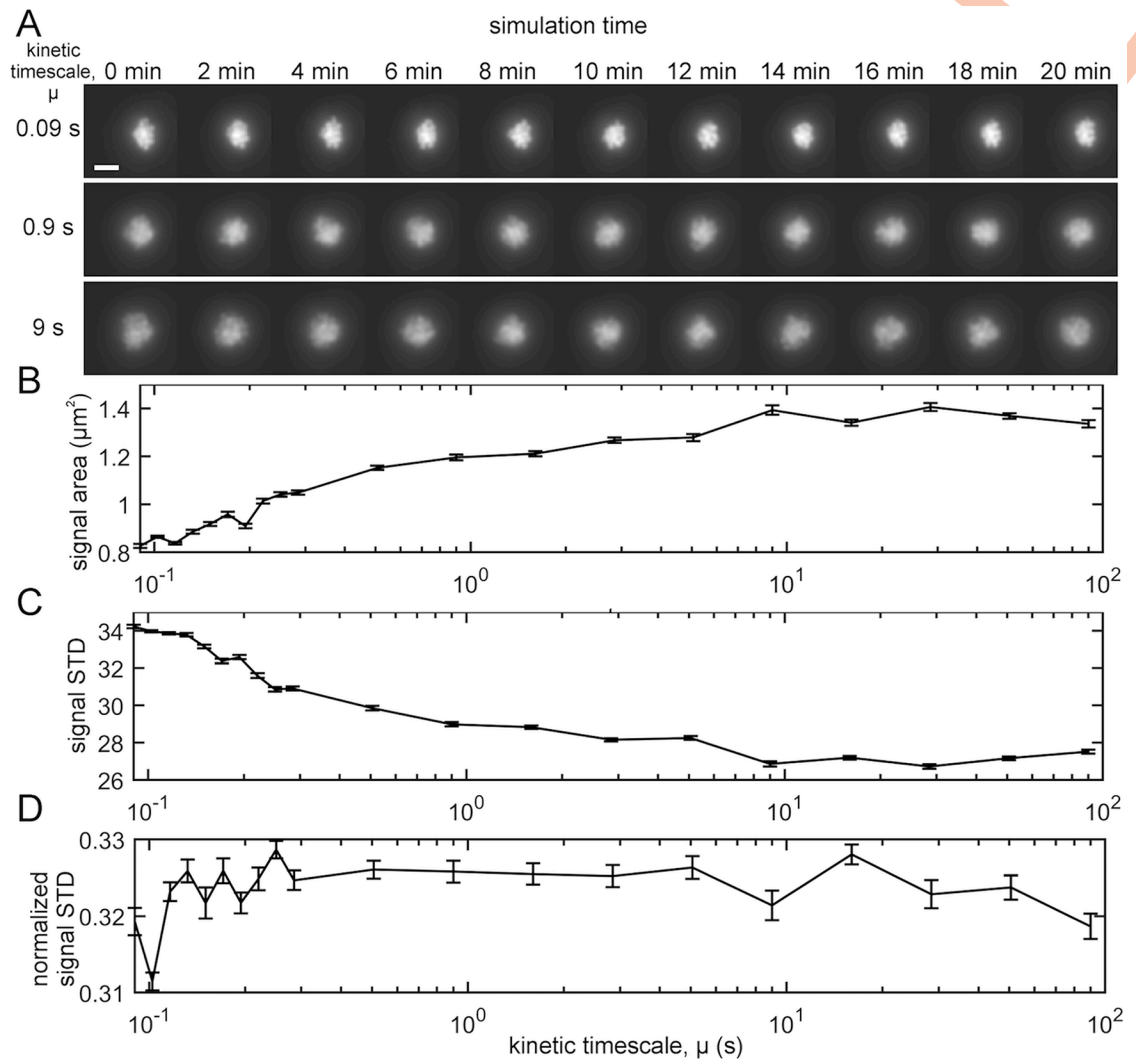
Fig 3. Simulated microscope images for polymer bead-spring model with transient crosslinking and varying kinetic timescale μ. Varying μ alters the area and variance of the intensity for the nucleolar signal, which models the affect of the fob1Δ and hmo1Δ mutations. (A) Timelapse montage of simulated microscope images for $\mu \in \{0.09, 0.9, 0\}$ (seconds). Scale bar is μm. (B) Area of nucleolus signal (μm^2) as a function of $\mu \in [10^{-1}, 10^2]$. Areas were calculated by measuring the area of the binary mask generated by applying Otsu's threshold to each image. (C) Standard deviation of the (non-normalized) nucleolus signal versus μ. (D) Standard deviation of the normalized nucleolus signals versus μ; these we normalized identically to the normalization of the experimental microscope images. In panels (B)-(D), error bars indicate standard errors observed using 22 time points for each value of μ.

https://doi.org/10.1371/journal.pcbi.1007124.g003

Histograms of 2-point pairwise distances between nucleolar beads

A simple and previously used method for analyzing distances between beads is to create a histogram of all bead-bead pairwise distances. As explored in [21], this two-point statistic can provide evidence of clustering and can be used to query simple properties such as whether or not the clusters change over time. In this section, we repeat this analysis on our data and extend it by showing what effects averaging over time and averaging over populations has on the results. We show that averaging one cell over time prevents observing the flexible clustering through pairwise distances, and averaging over populations prevents observing any sort of clustering. We provide further details on the computation of these distances in Methods Section: Pairwise-distance maps for high-throughput chromosome conformation capture (Hi-C).
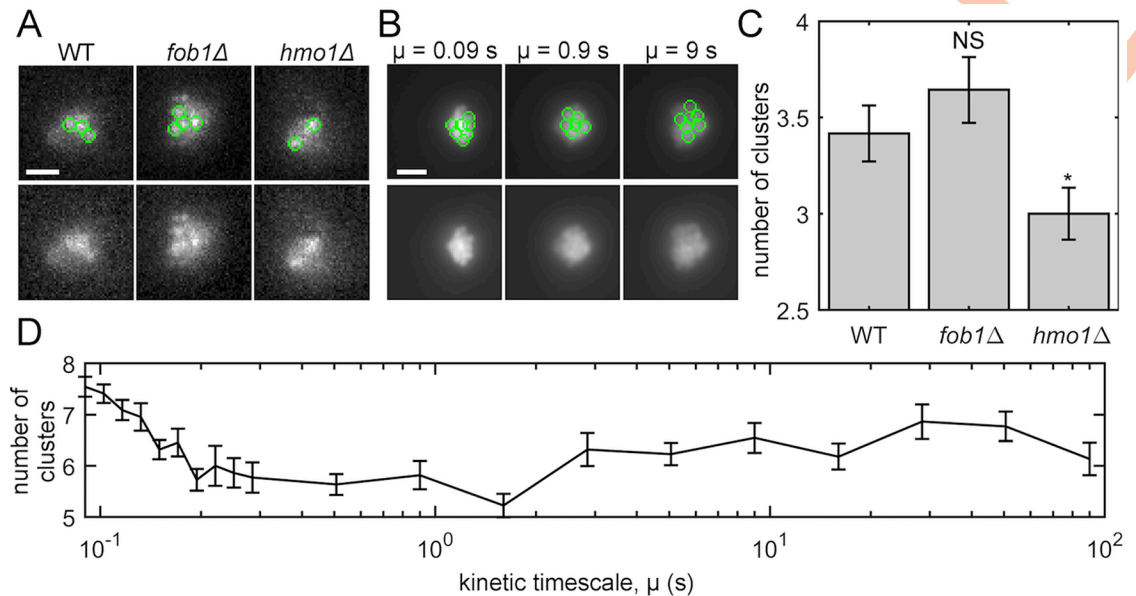
Fig 4. Comparison of experimental and simulated microscope images. Varying μ in the polymer bead-spring model captures the effect on nucleolus clustering incurred by the crosslink-altering mutations. (A) Maximum intensity projections of CDC14-GFP in WT, fob1Δ, and hmo1Δ cells with identified clusters marked by green circles. The lower row depicts the same images as the top row, except the circles are removed. (Images are the same as Fig 2(A)). (B) Same information as panel (A), except the results depict simulated images based on 4D simulation data for three values of μ. Scale bars in (A) and (B) are 1 μm. (C) Bar chart showing the number of clusters for each strain. The bars and error bars indicate the average and standard error across n cells, where $n = 84, 70$, and 77 for WT, fob1Δ, and hmo1Δ, respectively. Significance was assessed via a Student's two-tailed T-test: $p = 0.3$ for WT versus fob1Δ and $p = 0.04$ for WT versus hmo1Δ. (D) Average number of clusters for simulated images as a function of μ. The averages and standard errors (see error bars) were calculated using 22 time points for each μ.

In Fig 5(A), we plot the distribution of all pairwise distances $\{d_{ij}^{(t)}\}$ at a single time t for three kinetic timescales, given by the same values $\mu = 0.09, 0.19, 1.6$ as shown in Fig 1. For $\mu = 0.09$, the pairwise distance distribution is clearly a multimodal distribution [21]. The peak near $d \approx 50$ represents a large number of very short pairwise distances between beads in the same cluster. For slightly larger d, the density drops to zeros, indicating a separation distance between clusters. Interestingly, we observe two more peaks near $d \approx 300$ and $d \approx 600$. The clarity of these peaks suggests that the clusters themselves are regularly spaced from one another, reminiscent of a lattice structure. This shows the three layers of the multiscale structure of the nucleolus for $\mu = 0.09$: its existence as a dense, secluded section of the nucleus, the self-organization of intra-nucleolar clusters, and the individual beads within each cluster. For $\mu = 0.19$, one can also observe in Fig 5(A) three peaks in the empirical probability density for bead-bead distances, but these peaks are much less pronounced. This shows a gradual transition in the degree of clustering as we increase μ. There is also a smaller gap between peaks. Together, these observations recapitulate our observations in Fig 1(E), wherein the clusters can be observed to be less compact. Finally, for $\mu = 1.6$, there is no multimodal structure in the bead-bead distance plot. This is consistent with our expectation that there is no clustering structure present for this range of μ.

The rigid and flexible clustering cases differ not only in how strong the clustering is at any given time, but also in how stable the structure is in time. We investigate this by considering how averaging pairwise-distances either across across time (Fig 5(B)) or over multiple simulations (Fig 5(C)) influences pairwise-distance probability densities.
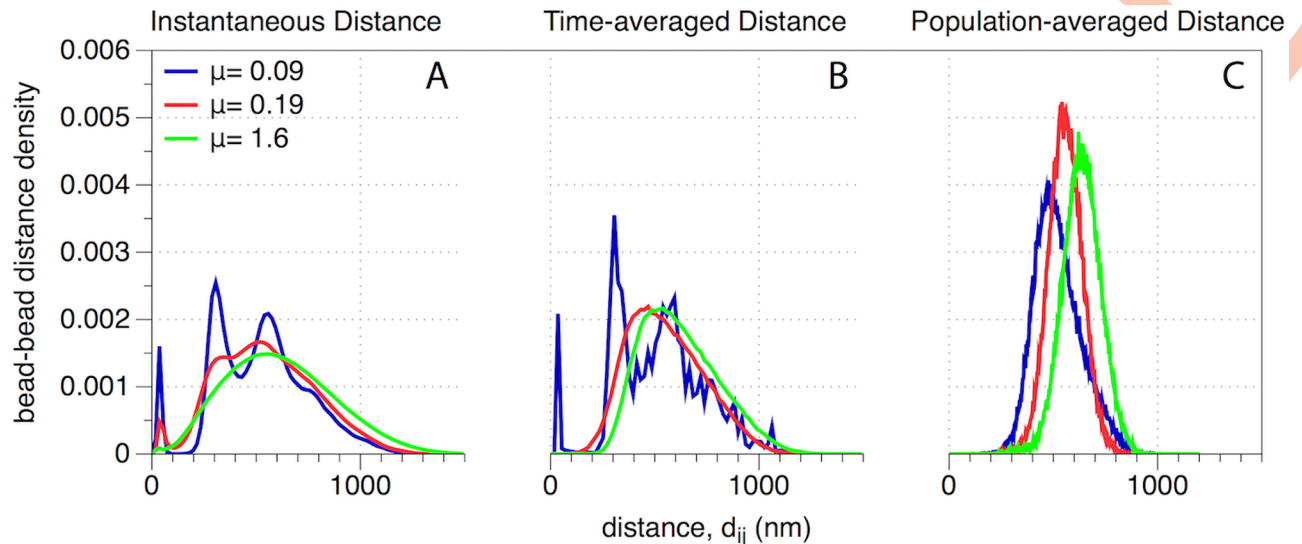
Fig 5. (A) Instantaneous distances, (B) time-average distances, and (C) population-averaged distances. In each panel, we show results for the three kinetic regimes illustrated in Fig 1: $\mu = 0.09$, $\mu = 0.19$), and $\mu = 1.6$. As reported in [21], multimodal histograms are a "signal" for the presence of clusters. This signal is strongest for $\mu = 0.09$, is nonexistent for $\mu = 1.6$, and we observe a new regime for $\mu = 0.19$ (flexible clustering or cluster plasticity), whereby 'soft' clusters form but deform over time through cluster interactions.

https://doi.org/10.1371/journal.pcbi.1007124.g005

In Fig 5(B), we plot the empirical probability densities for pairwise distances averaged across our 20 minute simulations. Note for $\mu = 0.19$ that the density is no longer multimodal, implying that aggregating the data across a large time range inhibits the detection of flexible clusters, which by definition change with time. Note that the rigid clusters, which are very stable across time, remain discernable as the pairwise probability density remains multimodal. Unsurprisingly, the slow crosslinking appears qualitatively very similar in the long time average, as there was no apparent structure in the first place.

In Fig 5(C), we plot the empirical probability densities for pairwise distances at a single time but averaged across 10 simulations with different random initial conditions. Note for all μ that there is no longer any multimodal structure for these densities, highlighting that averaging across heterogeneous cell populations obscures the detection of clusters.

Flexible clustering regime maximizes bead mixing

Next, we study how the kinetic time scale μ (i.e., and thus the presence of clusters) affects the properties of pairwise *gene interactions*. A pair of beads is said to be interacting if they are in very close proximity and the distance between them drops below d^*. As discussed in Fig 5, we choose $d^* = 100nm$ unless otherwise noted. In the following experiment, we show that increasing μ not only inhibits the formation of clusters, but that there exists a particular range of μ that optimizes *gene mixing*, or the overall interaction frequency of all pairs of genes. These experiments illustrate how clustering—which inherently describes multi-way relationships—can be studied through pairwise distances—which inherently describe two-way relationships —, and how there remain important open problems related to the time series signal processing of 4D chromosome conformation datasets.

We study the following summary statistics for gene mixing:

(A). The *interaction fraction* indicates the fraction of possible unique bead pairs that interact at least once during an interphase simulation.

(B). The *mean interaction number* indicates the number of simultaneous interactions (i.e., beads within distance d^*) for a nucleolus bead, averaged across time and across beads.

(C). The *mean waiting time* indicates for any two beads, selected at random, the average time that passes between their i-th and $(i + 1)$-th interactions.

(D). The *mean interaction duration* indicates the amount of time beads enter and reside within the interaction distance.

In Fig 6(A)–6(D), we plot these summary statistics across a wide range of μ. We identify three regimes that optimize different attributes. $\mu \approx 0.1$ yields a self-organized structure that maximizes the number and the duration of gene-gene interactions (see panels (B) and (D)). Recall from Video 1 that $\mu = 0.09$ yields many large clusters that are stable (i.e., do not change) over time. This is reflected in a high number of interactions with beads in the same cluster and low number of interactions with beads not in the same cluster.

With $0.15 \lesssim \mu \lesssim 1$, we see flexible clustering behavior from Video 2. Notably, we find here that this flexible clustering has interesting properties beyond simply being a weaker version of the strong clustering from the rigid clustering regime. Namely, Fig 6(A) shows that these μ values maximize the fraction of pairs of beads that interact at least once over the simulation, and Fig 6(C) shows that these values minimize the waiting times between subsequent interactions. Thus, we can say that *flexible clustering promotes the number of both simultaneous and overall distinct pairwise gene interactions in the nucleolus*. This behavior arises from a balance between
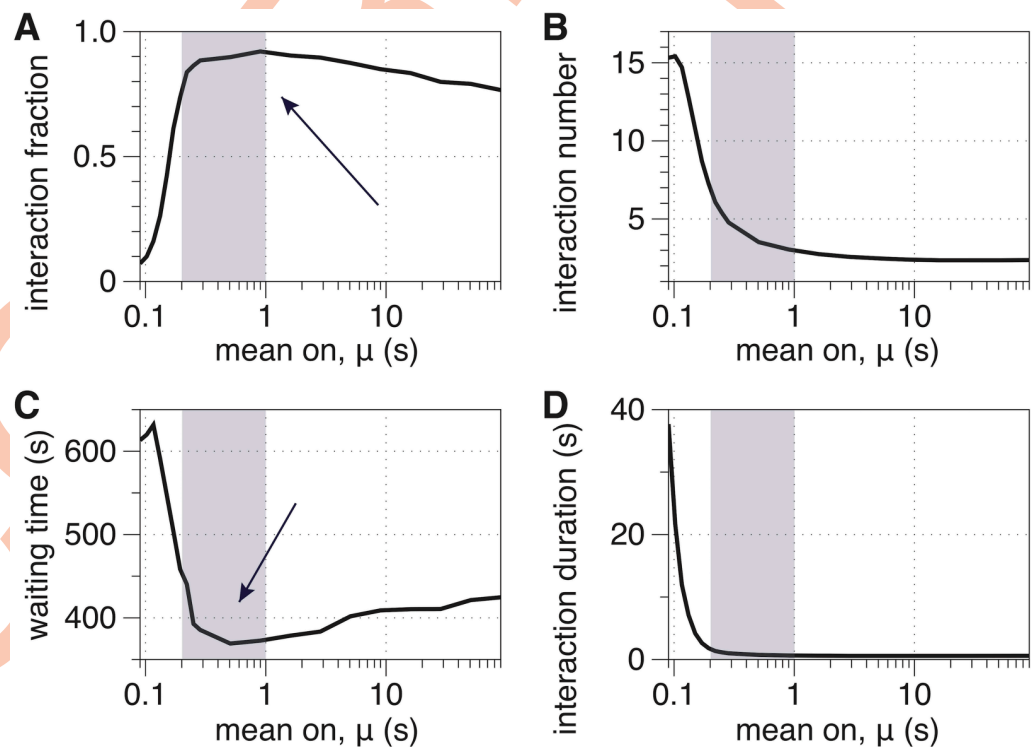
Fig 6. Mixing statistics for gene interactions: (A) mixing fraction; (B) mean interaction number; (C) mean waiting time; and (D) mean interaction duration, which we plot as versus μ. The shaded regions indicate the regime of flexible clustering, $\mu \in (0.19, 1)$. The arrows in panels (A) and (C) highlight that the interaction fraction and mean waiting time are both optimized for this range of μ. In contrast, this regime is associated with sharp transitions for the mean interaction number and duration, which both monotonically decrease with μ.

https://doi.org/10.1371/journal.pcbi.1007124.g006

the number of intra-cluster gene-gen interactions, which is still elevated due to the moderate clustering as shown in (B), and the ability for genes to frequently switch between clusters during cluster interactions, as indicated by the reduced waiting time in (C). SMC proteins with such crosslinking timescales will thereby promote collective interactions among all active genes. These circumstances could accelerate a homology search, for example, to facilitate DNA repair, if the sister chromosomes were suddenly activated by a family of SMC proteins whose binding affinity was near this "sweet spot".

Finally, $\mu \gtrsim 1.5$ is associated with a non-clustering regime, as shown in Video 3. The lack of clustering is reflected by a low number of gene-gene interactions, and the freely diffusing nature of the beads is reflected by short interaction duration and high interaction fraction.

Identifying cluster membership with network community detection algorithms

Having found that flexible clustering maximizes interesting properties of gene interaction, we seek to develop tools to identify and label the spatiotemporal clusters. In the rigid clustering regime, the clusters are so well-defined that any reasonable algorithm will detect them, but this is not the case for the flexible clustering. To detect and track flexible clustering, we utilize both spatial and temporal information to identify and track clusters.

While we have access to 4D bead position time series data, we begin by transforming this into a multilayer network problem as described in Methods Section: Gene-interaction networks from pairwise-distance data. This is motivated by the fact that the most similar data available in biology, the Hi-C dataset, does not measure true distances between genome regions, but rather a notion of similarity based on average proximity [12]. The result of this transformation is a time sequence of weighted, undirected networks whose edge weights represent how near two beads tend to be to each other at that point in time. We refer to this sequence of networks as a temporal network.

Given a gene-interaction network, we identify communities using an approach based on multilayer modularity [37]. See [31, 32] for examples where community detection was applied to network models derived from Hi-C data. We present the algorithm in detail in Methods Section: Spatiotemporal gene clusters revealed by community detection in temporal networks, and we briefly describe it here.

The *modularity measure* was originally introduced [36] to detect communities in a single, non-temporal network; it is a scalar that quantifies—as compared to a null-model lacking communities—the extent to which a network's nodes can be partitioned into disjoint sets (i.e. communities) so that there is a prevalence of edges between nodes in the same community and relatively few edges between nodes in different communities. By searching over different possible ways to partition nodes into communities, one seeks to find an optimal partition that maximizes the modularity score [38]. Because each community contains a prevalence of edges, and edges only exist between pairs of genes that are in close proximity, a modularity-optimizing partition equivalently assigns genes into disjoint clusters so that each gene is nearer to genes in its cluster than to genes in other clusters.

Our analysis is primarily based on an extended version of modularity that allows one to detect time-varying communities in temporal networks and is called the multilayer modularity measure [37]. In contrast to a community in a time-independent network (which is defined by a set of nodes), to specify a time-varying community one must also identify for each node the time-steps for which it is in the community. Using a variational technique [38], we optimize the multilayer modularity measure to simultaneously assign every node to a community at every time step. Each time-varying community in the network corresponds

to a time-varying gene cluster, which is a set of genes that remain in close proximity for some duration.

A key feature of the multilayer-modularity approach for community detection is that the framework involves two parameters, γ and ω, which provide "tuning knobs" [37, 50] to identify, respectively, the appropriate sizes and temporal coherence of communities/clusters. Parameter γ is a *resolution parameter* [38] and allows one to select whether modularity-optimizing partitions involve many small communities or just a few, very large communities. Similarly, ω is a *coupling parameter* and allows one to choose if the communities can change drastically from one time step to the next or if they are restricted to changing slowly over time. We explored a range of choices to select appropriate values. Finally, we highlight that this approach significantly contrasts traditional clustering algorithms such as k-means clustering [51], which specifies the number of clusters *a priori* (i.e., rather than allow the appropriate resolution to be dictated by the data) and which does not naturally extend to time-varying data.

In S4, S5 and S6 Videos, we present equivalent videos to S1, S2 and S3 Videos, respectively, except in the new videos we color the beads according to the community labels that we detect. These new videos provided qualitative evidence supporting the ability of our algorithm to find clusters that have appropriate spatial and temporal scales. Looking at the rigid clustering in S4 Video, we see the coloring strongly agrees with our visual perception in the clusters. A similar but less decisive conclusion can be made from observing S5 Video.

These videos indicate good agreement between visual perception of clusters and the clusters that are detected by the multilayer modularity algorithm—when beads visually appear to be clumped together, they tend to also be the same color in the videos, which reaffirms the validity of our choice of clustering algorithm. However, especially when looking at S6 Video depicting the non-clustering regime with slow crosslinking, we identify a key and common issue with clustering algorithms—they typically identify the "best" clusters, even when no clusters actually exist.

Gene mixing at the community level further supports flexible clustering as the mechanism for optimality

In this section, we provide further evidence that flexible clustering is the mechanism that is responsible for the optimality observed in Fig 6. To this end, we will revisit and modify our definitions for *gene interactions* and *gene mixing*, which were defined at the "bead level" (i.e., for pairs of beads). We now define similar, but slightly different, concepts that are defined at the "community-level" in that they reflect only community-membership information and do not require the precise bead locations. We say that two beads are "communicating" if they are in the same community. That is, all beads in the same cluster are communicating with each other, and beads in different clusters are not communicating. With this modified definition in hand, we define summary statistics for gene mixing at the community level that are analogous to the 2-point summary statistics for pairwise gene interactions that we previously defined in Section: Flexible clustering regime maximizes bead mixing. Analogous to gene mixing at the bead level, we now define "cross communication" at the community level.

(A). The *communicating fraction* indicates the fraction of bead pairs that are in the same community at least once during a simulation.

(B). The *average beads per community* indicates, for a nucleolus bead, the average number of beads in the same community at the same time, averaged across time and across beads.

(C). The *mean waiting time* indicates for any two beads, selected at random, the average time that passes between when they are no longer in the same community and when they are next in the same community.

(D). The *mean interaction duration* indicates the amount of time between beads when they are first in the same community and when they are no longer in the same community.

In Fig 7, we present summary statistics for cross communication that are analogous to our results in Fig 6 for pairwise gene interactions. Note that the results in Fig 7 are qualitatively identical to those in Fig 6, supporting our hypothesis that gene-level mixing is determined by community-level cross communication. That is, the formation of gene clusters and exchanges of genes between them governs the timescale at which nucleolar domains come in close proximity of one another.

Temporal stability of clusters

Using temporal community detection, we are now finally able to quantitatively support our first observations made in Section: Transient crosslinking timescale influences nucleolus clustering that there are three distinct clustering regimes: rigid clustering; flexible clustering; and no clustering. We support these observations by studying the properties of the detected clusters.
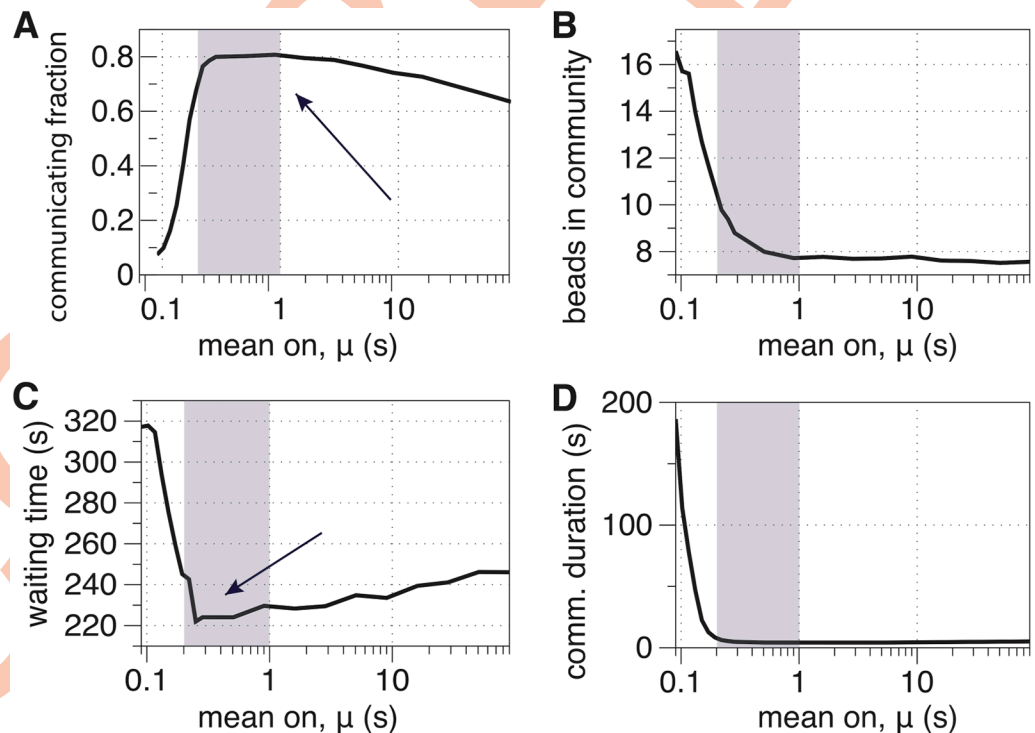
Fig 7. Cross communication describes the dynamics of community memberships of beads through community-level mixing. We study cross-communication for a large range of μ by plotting four summary statistics: (A) communicating fraction; (B) mean interaction number; (C) mean waiting time; and (D) mean interaction duration (see text). The shaded regions indicate the regime of flexible clustering, $\mu \in (0.19, 1)$, and the arrows in panels (A) and (C) highlight that the interaction fraction and mean waiting time are both optimized for this range of μ. These results are qualitatively identical to the results in Fig 6 for gene mixing, illustrating that cluster formation and the exchanges of beads between clusters determines the timescale of mixing.

Fig 8. The birth and death of gene clusters identified using modularity-based community detection in temporal networks [37]. Panels (A), (B) and (C) depict the lifetime (i.e., duration) of each gene cluster versus the average cluster size for $\mu = 0.09$, 0.19, and 1.6, respectively. The points' colors have been chosen to highlight the density of points (with red indicating where there are many points close to one anot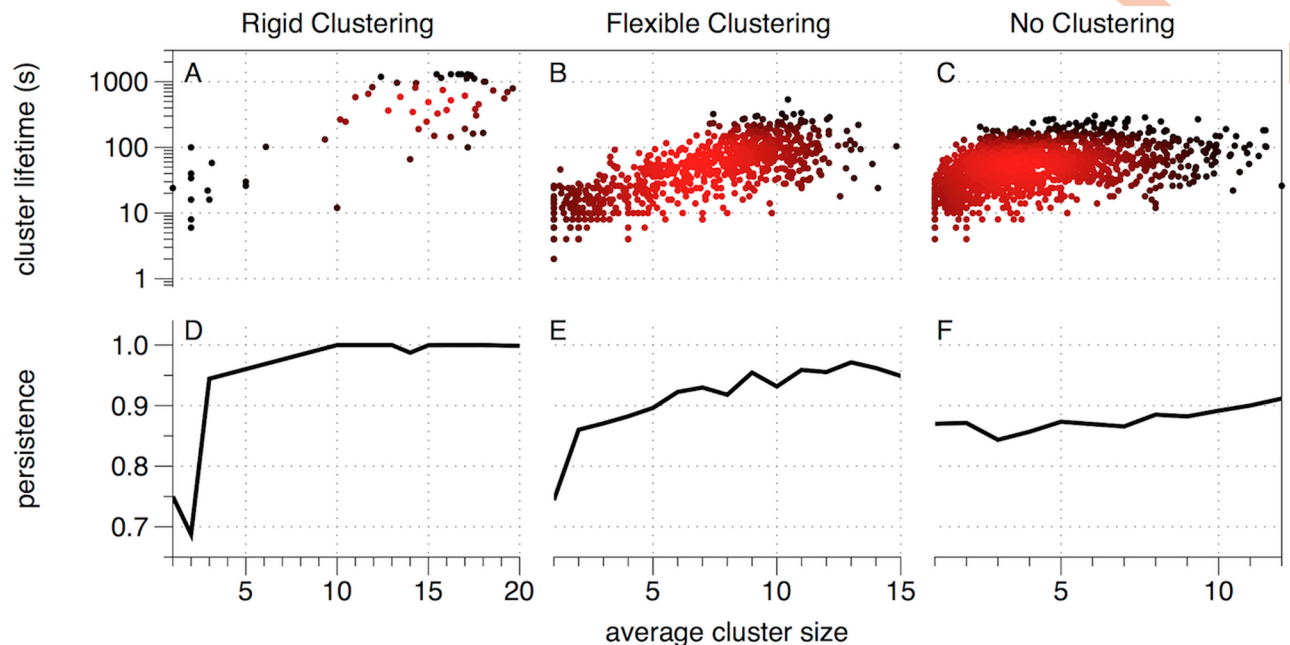her). (D)–(F): The persistence (i.e., temporal coherence) of clusters is indicated by the probability that a randomly selected bead remains in the same clusters in the next time window, again plotted versus the average number of beads in that clusters. Results reflect $d^* = 325$, $\gamma = 10$, $\omega = 1$.

https://doi.org/10.1371/journal.pcbi.1007124.g008

In Fig 8(A)–8(C), we plot the average lifetime of clusters as a function of their average size (averaged over time). Panels (A)–(C) indicate the three clustering regimes with $\mu \in \{0.09, 0.19, 1.6\}$, respectively.

For the rigid clustering regime, in Fig 8(A), we see that most of the clusters are large, with an average size of 10 or more beads, and also have a long lifetime of over 100 seconds. This is consistent with our prior observations (e.g. Video 1) that showed large clusters that appeared very stable in time. We also see that the large clusters survive for much longer than the small clusters.

For the flexible clustering regime, in Fig 8(B), we can see the same general trend that larger clusters tend to have a longer lifetime than smaller clusters, but there is a much wider spread of cluster sizes, with only a moderate number of large clusters.

For the non-clustering regime, there appears to be little relationship between cluster size and stability beyond an average size of approximately 3 beads. The clusters also tend to be much smaller, with almost no clusters with an average size over 10 beads.

In Fig 8(D)–8(E), we plot the probability that a bead remains in the same community upon the next timestep, again as a function of cluster size. Panels (D)–(E) indicate results for $\mu \in \{0.09, 0.19, 1.6\}$, respectively. In agreement with panels (A)–(C), one can observe that larger clusters are more stable. Note also that the communities exhibit more plasticity for $\mu = 0.19$ than for $\mu = 0.09$ since beads have a higher average probability for changing the community to which they belong.

Discussion

The dynamic self-organization of the eukaryote genome is fundamental to the understanding of life at the cellular level. The last quarter century has witnessed remarkable technological

advances that provide massive datasets of both the spatial conformation of chromosomal DNA from cell populations (3C and Hi-C generalizations) and the dynamic motion of domains in living cells (GFP tagging and tracking of specific DNA sequences), from the yeast to the human genome. Data mining of this massive data has likewise witnessed remarkable advances in understanding the hierarchical packaging mechanisms of DNA that act on top of the genome, e.g., histones and structural maintenance of chromosome (SMC) proteins, the topology of individual chromosome fibers, their topologically associated domains, and the territories they occupy in the nucleus. The third wave of advances has come from 4D modeling of chromosomes based on stochastic models of entropic, confined polymers, and the coupling of SMC proteins that either bind and crosslink genes on chromosomes or generate loops on individual chromosomes. As these three approaches continue to mature and inform one another, at an ever-increasing pace, insights into the structure and dynamics of the genome continue to deepen.

The motivation for this paper lies in the information that can be inferred from these massive datasets, from Hi-C, live cell imaging experiments, and polymer physics modeling. In previous studies, cf. [21] and references therein, we showed that heterogeneity in experimental images derive from substructures that are formed within the nucleolus. We decided to use the array of tools from network community detection analysis, modeling, and associated fast algorithms, to automate the search for dynamic architectures in the 4D datasets generated by polymer modeling. We note a similar network analysis approach has been applied to Hi-C datasets [26–28], whereas our datasets have the added feature of highly resolved temporal information. Our aim was to infer organization beyond 2-point, time-averaged or population-averaged, gene-gene proximity statistics and heat maps generated from the statistics, and to remove the bias of an individual's visual determination of structure. To do so, we used the advances in network-based models, their temporal generalization, data analysis, and algorithms, and applied this arsenal of tools to 4D datasets across four decades of crosslinking timescales to: (i) robustly identify clusters, or communities, of genes (5k bp domains, or beads, in our model), i.e., to directly detect gene sub-organization at the scale it exists rather than attempt to seek larger scale organization from gene-gene statistics and heat maps; (ii) determine the size distribution (number of genes) in such sub-structures; (iii) determine the persistence times of communities; and, (iv) determine the interaction frequency of communities and corresponding gene exchanges, which are the drivers of gene-gene interaction statistics. In this way, network algorithms automate gene community detection and persistence, with robustness built in by enforcing insensitivity to algorithm tuning parameters.

We elected to build and implement these network tools on the 4D datasets generated in house, from simulations of our recent polymer modeling of interphase budding yeast [21]. In this model, a pool of SMC proteins transiently and indiscriminantly crosslink 5k bp domains within the nucleolus on Chromosome XII. The kinetics of the cross-linking anchors relative to the substrate is a major driver of sub-nuclear organization. If the crosslinkers bind and release more rapidly than the chains can relocate, the non-intuitive consequence is that the chains explore less space. When dense clusters of crosslinker/binding sites arise, they persist for extended time periods when the crosslinking kinetics is sufficiently fast.

We previously showed in [21] that very short-lived ($\mu = 0.09sec$) binding kinetics provided closer agreement with experimental results (highest degree of compaction of the nucleolus into a crescent shape against the nuclear wall). It was also shown via visualization of the simulated 4D datasets that this timescale induces a decomposition of the nucleolus into a large number of clusters each consisting of many 5k bp domains, and these clusters were persistent over time. On the other hand, with long-lived crosslinks ($\mu = 90sec$) the clusters disappeared. These results reveal that the timescales of the crosslinkers relative to

entropic fluctuations of the chromosome polymer chains are a fundamental contributor to genome organization.

For the present paper, the sample set of binding kinetics in [21] was expanded to 4D datasets of interphase, sampling over four decades of crosslinking duration timescales. We applied standard, distance-based, 2-point statistical metrics and visualization tools, and then analyzed the full range of 4D datasets with the fast, automated network models and tools. The network algorithms search for and detect time-varying communities (clusters, sub-structures) at the scales they exist, not at a prescribed scale of two or more genes; the spatial and temporal scales are identified with the criteria that they are robust to the algorithm parameters. We then use this information to label and color-code communities, using community-level description to understand the persistence and interactions (merger and division marked by gene exchanges) between communities.

With the above community-scale information and statistics, we generalize standard gene-gene interaction statistics across the four decades of bond duration timescale. As a generalization of waiting times for 2 distant genes beads to come within a specific distance of one another, we calculate waiting times for genes in the same community to leave and then re-enter another common community, and calculate the fraction of all genes that were in the same community at least once during interphase, which we call the community cross-communication fraction.

From these analyses, we discovered a novel dynamic self-organization regime, wherein the rigid, persistent communities at relatively short-lived crosslink timescales ($\mu = .09sec$) transition at slightly longer-lived crosslink timescales ($\mu = 0.19sec$) to more mobile (literally, the clusters diffuse faster) communities that interact far more frequently, each interaction corresponding to merger, subsequent division, and an exchange of genes. We refer to this regime as flexible community structure with enhanced cross-communication. Furthermore, we discovered non-monotonicity in the dynamic self-organization behavior: the community cross-communication fraction is maximized, coincident with a minimum waiting time between genes departing and returning to common communities, with a crosslink timescale of $\mu = 0.19sec$. Both properties fall off, albeit in different ways, for shorter and longer timescales.

We emphasize that these network tools and fast algorithms are amenable to any 4D dataset from polymer models. While we restricted the analysis in this study to the nucleolus during G1 of budding yeast where SMC proteins are allowed to transiently crosslink 5k bp domains, the same analyses can be applied to data with tandem SMC crosslinking and loop generation, for any cell type and for any phase of the cell cycle. Moreover, given the growing interest in network-based analyses for Hi-C data, network modeling is well-positioned to provide a fruitful direction for data assimilation efforts aimed at connecting simulated and empirical 4D chromosome conformation data. An important challenge facing this pursuit is the development of improved data pre-processing and community-detection methodology for temporal and multimodal network datasets [52, 53].

Materials and methods

Model

Chromatin dynamics within confined yeast nuclei has been widely modeled using Rouse polymer bead-spring chains; see earlier citations. We employ the identical model and code in [21], resolving the entire yeast genome into 2803 total beads, each of which represents approximately 5k base pairs (bp). As such, each bead is interpreted as a chromosome domain or a tension blob, as explained in the polymer physics literature. The beads are arranged on 32 chromosome arms having lengths that reflect their experimentally identified lengths. Each

arm is tethered at both ends to the nuclear wall: all emanating from the centromere at one end, with the other end tethered to one of six telomeres. Along each arm, there are entropic, nonlinear springs (the wormlike chain model is used here) that connect neighboring beads. The beads also experience Brownian, entropic, repulsive and hydrodynamic drag forces, and are physically confined to the nucleus. We simulate approximately 20 minutes of G1 during interphase, and each simulation is initialized with 32 chromosome arms tethered at the centromere and one of six telomeres on the nuclear wall, otherwise randomly located within the nucleus (idealized as a spherical domain).

We model the effect of SMC proteins by transiently crosslinking pairs of non-neighboring beads in the nucleolus, represented by a contiguous chain of 361 beads on chromosome XII. (We note in passing that [21] also studied, experimentally and computationally, when the nucleolus is split onto separate chromosome arms, showing this to have a negligible affect on the clustering and interaction results). A transient crosslink is modeled by a pair of stochastic events—a binding and unbinding of two non-adjacent nucleolar beads—and the timescale of these events is tuned by a single parameter μ (measured in seconds). We defer to [21] for the details, but elaborate here on the role of μ. To implement crosslinks, all nucleolar beads are assigned a state of "active" or "inactive," and crosslinks are allowed only between active beads. Each bead's state fluctuates stochastically as follows: an active bead becomes inactive after a duration that is a random number drawn the normal distribution $N(\mu, (\mu/5)^2)$; and an inactive bead becomes active after a duration drawn from $N(\mu/9, (\mu/45)^2)$. If two nucleolar beads are both active and the distance between them is less than $\tilde{d} = 90$ nm, then a crosslink is formed between them (i.e., they bind). Each bead may be involved in at most one crosslink at a time, decided on the basis of pairwise proximity among all active beads at each timestep. If either bead becomes inactive, then the crosslink is broken (i.e., they unbind). Hence, crosslinks are established and broken stochastically in the nucleolus, and the single parameter μ dictates the kinetic timescales for crosslinking. See [21] for additional model details.

Pairwise-distance maps for high-throughput chromosome conformation capture (Hi-C)

Each simulation of the Rouse-like polymer model yields time-series data $\{\mathbf{x}^i(t)\} \in \mathbb{R}^3$ that defines the 3D location of each bead $i \in \{1, \ldots, N\}$ at each discrete time step $t = 0, 1, \ldots$. We establish a connection between our simulated data and the state-of-the-art in chromosome imaging—namely, high-throughput conformation capture (Hi-C)—by constructing and analyzing *pairwise-distance maps*. Hi-C "images" the conformation of chromosomes using a combination of proximity-based ligation and massively parallel sequencing, which yields a map that is correlated with the pairwise distances between gene segments. While the actual pairwise distances between gene segments cannot be directly measured, Hi-C implements spatially constrained ligation followed by a locus-specific polymerase chain reaction to obtain *pairwise count maps* that are correlated with spatial proximity: the count between two gene segments monotonically decreases as the physical 3-dimensional distance between them increases.

To provide an analogue to Hi-C imaging, we construct pairwise-distance maps for our simulated data $\{\mathbf{x}^i(t)\}$. Let

$$F : \left\{\mathbf{x}^i\right\}_{i=1}^{i=N} \mapsto \mathbb{R}^{N \times N} \tag{1}$$

define a *map* (used here in the mathematical sense) from a set of N points $\{\mathbf{x}^i\} \in \mathbb{R}^3$ to a matrix such that each entry (i, j) in the matrix gives the distance between point i and point j. Whereas Hi-C imaging aims to study the positioning of chromosomes using noisy measurements that are inversely correlated with pairwise distances, for our simulations we have access

to the complete information about the chromosome positioning. We therefore define and study several variations for pairwise distance maps, which will allow us to also study artifacts that can arise under different preprocessing techniques, such as averaging the time series data across time windows and/or averaging across multiple simulations with different initial conditions. We define the following pairwise distance maps:

- An *instantaneous* pairwise distance map $X(t) = F(\{\mathbf{x}^i(t)\})$ encodes pairwise distances between beads (i.e., chromosome domains) at a particular timestep t.

- A *time-averaged* pairwise distance map $Y(\tau) = \frac{1}{|\tau|}\sum_{t\in\tau} X(t)$ encodes the pairwise distance between beads averaged across a set of timesteps τ.

- A *population-averaged* pairwise distance map $Z(t) = \langle X(t)\rangle_p$ encodes the pairwise distance at timestep t between beads, which are averaged across several simulations that have different initial conditions (which are chosen uniformly at random).

These pairwise distance maps represent the data that is sought after, but cannot be directly measured, by Hi-C imaging. Moreover, by defining several distance maps we are able to study "averaging" artifacts that can arise due to various limitations of Hi-C imaging. For example, Hi-C imaging obtains measurements that are typically averaged across a large heterogeneous distribution of cells that are subjected to nonidentical conditions and exist at nonidentical states in their cell cycles.

Microscope image acquisition, processing and analysis

Yeast strains for experiment. The budding yeast strains used in this study were obtained by transforming the yeast strain EMS219 (Mat alpha, his5 leu2-3,212 ura3-50 CAN1 asp5 gal2 (form I1 rDNA::leu2 URA3+)) with CDC14-GFP:KAN^R, to label the nucleolus, and SPC29-RFP:HYG^R, to label the spindle pole body, to generate the yeast strain DCY1021.1. DCY1021.1 was transformed to knock out FOB1 and HMO1 to generates DCY1055.1 and DCY1056.1 respectively.

Image acquisition and baseline processing. Fluorescent image stacks of unbudded yeast cells were acquired using a Eclipse Ti wide-field inverted microscope (Nikon) with a 100× Apo TIRF 1.49 NA objective (Nikon) and Clara charge-coupled device camera (Andor) using Nikon NIS Elements imaging software (Nikon). Each image stack contained 7 Z-planes with 200 nm step-size.

Image stacks of experimental images were cropped to 7 Z-plane image stacks of single cells using ImageJ and saved as TIFF files. The cropped Z-stacks were read into MATLAB 2018b (MathWorks), converted into maximum intensity projections, and the projections of hmo1Δ and fob1Δ were cropped to 55 × 55 pixels, to match the dimensions of WT projections, using MATLAB function padarray with `replicate` option specified to extend outer edge of pixel values to ensure the center of all cropped images was the brightest pixel. The intensity values all projections were normalized by subtracting all intensity values by the minimum value and then dividing the resulting values by the maximum intensity value after subtraction. The normalized intensity values were stored with double point precision, preventing any loss in dynamic range.

Image analysis. The areas of nucleolar signals were determined by setting all values below threshold, calculated using `multithresh` function, to NaN and then summing number of values that were not NaNs. That pixel count was converted to μm^2 by multiplying the sums by 0.0648^2, the area of each pixel in μm^2.

To calculate the standard deviation of the intensities of the nucleolar signal, the non-NaN values remaining after thresholding were re-normalized using the same method described above, and the standard deviation of those values was measured.

To count clusters within the nucleolar signal, the normalized images were deconvolved with 5 × 5 Gaussian structural element, using deconvblind function, and underwent two rounds of background subtraction by setting all intensity values below threshold value, calculated using default `multithresh` function, to NaN and then all NaN values to 0. Clusters were identified using the `imregionalmax` function and counted using `bwconncomp` function.

The simulated images generated from our simulations were analyzed as described above with the additional step of measuring the standard deviation of the each simulated maximum intensity projection. All WT images were analyzed using the script wtExpIm.m. All hmo1Δ and fob1Δ images were analyzed using the script cropExpIm.m. All simulated images were analyzed using the script clusterCountLoop.m.

All MATLAB scripts have been made available at [49].

Cluster identification via network community detection

Gene-interaction networks from pairwise-distance data. Given a pairwise distance map $X \in \mathbb{R}^{N \times N}$ in which each entry X_{ij} gives the (possibly averaged) Euclidean distance between beads i and j as described in Pairwise-distance maps for high-throughput chromosome conformation capture (Hi-C), we construct a network model in which there are weighted edges (i.e., interactions) only between beads that are in close proximity to each other and for which each edge weight $A_{ij} \geq 0$ decreases monotonically with distance X_{ij}. We propose a model with two parameters, d^* and s, which represent a distance threshold and a decay rate, respectively. In particular, we define a network adjacency matrix A having entries

$$A_{ij} = \begin{cases} e^{-sX_{ij}}, & X_{ij} < d^* \\ 0, & X_{ij} \geq d^*. \end{cases} \tag{2}$$

Note that there exists an undirected edge between i and j (i.e., $A_{ij} > 0$) only when $X_{ij} < d^*$, and s controls the rate in which the edge weight A_{ij} decreases with increasing distance X_{ij}. Because the edge weights exponentially decrease with distance, the community detection algorithms we study are insensitive to the choice for d^*, provided that d^* is sufficiently large so that the network is connected. Our choice $d^* = 325$ in Section: Temporal stability of clusters ensures there is an edge between all beads in the same cluster (recall Fig 6 in the main text) and was found to yield qualitatively similar results for other choices of d^*.

Eq (2) defines a map between a distance matrix and an affinity matrix that encodes a network. Note that for any such adjacency matrix A, we can equivalently define the network using the graph-theoretic formulation $G(\mathcal{V}, \mathcal{E})$. Here $\mathcal{V} = \{1, \ldots, N\}$ denotes the set of nodes (i.e., the beads in the chromosome model) and $\mathcal{E} = \{(i, j, A_{ij}) : A_{ij} > 0\}$ denotes the set of weighted edges (i.e., a set encoding which beads are interacting as well as their interaction strengths).

In Pairwise-distance maps for high-throughput chromosome conformation capture (Hi-C) we defined several versions of pairwise distance maps—instantaneous, time-averaged, and population-averaged maps—and a network model can be constructed for any of these maps:

- An *instantaneous interaction network* refers to a network associated with an instantaneous pairwise distance map $X(t)$.

- A *time-averaged interaction network* refers to a network associated with a time-averaged pairwise distance map $Y(\tau)$. We point out that due to the nonlinearity of Eq (2), a network

associated with a time-averaged distance map can in general differ from a temporal average of instantaneous interaction networks, averaged across the same time interval.

- A *population-averaged interaction network* refers to a network associated with an population-averaged distance map $\langle Z(t) \rangle_p$.

In addition to the above network models, we are particularly interested in constructing and studying *temporal interaction networks*, which we define as a sequence of time-averaged interaction networks encoded by a sequence of adjacency matrices $\{A_{ij}^s\}$. In particular, given a sequence of timesteps $s \in \{1, 2, \ldots, T\}$, we partition time into a sequence of time windows $\tau_s = \{(s-1)\Delta + 1, \ldots, s\Delta\}$ for $s = 1, 2, \ldots$ of width Δ. We then define a sequence of time-averaged networks $\{G(s)\}$ for $s = 1, 2, \ldots$ associated with these distance maps, which are time-averaged across the non-overlapping time windows $\{\tau_s\}$. The result is a sequence of adjacency matrices so that each entry A_{ij}^s indicates the absence ($A_{ij}^s = 0$) or presence ($A_{ij}^s > 0$) of an edge between i and j during time window τ_s.

In practice, we choose the time window width Δ to be similar to—but slightly larger than— μ. Matching the time-scales of μ and Δ allows the temporal interaction networks to efficiently capture the dynamics of interactions. Specifically, if Δ is too short then the temporal network will be identical across many time steps, which is not an efficient use of computer memory. Moreover, if Δ is too large, then the temporal network data will be too coarse to identify interaction dynamics occurring at a faster time scale. We chose $\Delta = 10$ to aggregate the time-varying bead-location data (which was saved every 0.1 second) into 1-second intervals. We studied 1000 such time windows to produce Fig 8.

Spatiotemporal gene clusters revealed by community detection in temporal networks. We analyze spatiotemporal clustering of chromosomes using community detection methodology for temporal interaction networks, particularly an approach based on multilayer-modularity optimization. Given a sequence of adjacency matrices $\{A^s\}$ for $s \in \{1, 2, \ldots, T\}$, we study the *multilayer modularity measure* [37]

$$Q = \frac{1}{2\mu} \sum_{i,j,s,r} \left[\left(A_{ij}^s - \gamma \frac{k_i^s k_j^s}{2m_s} \right) \delta(s,r) + \omega \delta(i,j) C_{sr} \right] \delta(c_{is}, c_{jr}), \qquad (3)$$

where A_{ij}^s denotes an entry in the adjacency matrix for network layer s (i.e., that associated with time window τ_s), γ is again a tunable "resolution parameter," $k_i^s = \sum_j A_{ij}^s$ is the weighted node degree for node i in layer s, $2m_s = \sum_{ij} A_{ij}^s$ is twice the total number of endirected edges in layer s, $\delta(m, n)$ is again a Dirac delta function, $C_{sr} = \delta(s, r-1) + \delta(s, r+1)$ defines the coupling between consecutive (time) layers and $C_{sr} = 1$ if only if $r = s \pm 1$ (otherwise $C_{sr} = 0$), and $\{c_{is}\}$ are the integer indices that indicate the community for each node i in each layer s. If one wished to analyze just a single network (e.g., a time-averaged or population-averaged network), then one can simply set $C_{sr} = 0$ so that the second term in the square brackets is discarded.

Letting $i \in \{1, \ldots, N\}$ enumerate the nodes and $s \in \{1, \ldots, T\}$ enumerate the network layers, the goal is to assign a community label c_{is} to each node-layer pair (i, s) to maximize Q [37]. Here, $c_{is} = c$ indicates that node i is in community c during time window τ_s. There are many techniques to solve such an optimization problem, and we identify partitions that optimize Q using a variational approach commonly referred to as the Louvain algorithm [38]. To provide some intuition into this optimization problem, we briefly comment on how the different terms in Eq (3) contribute to this optimization problem.

Consider the first term, $\left(A_{ij}^{s} - \gamma \frac{k_i^s k_j^s}{2m_s}\right)$, which is a slight generalization of Newman's original definition of the modularity measure [36] (which assumed $\gamma = 1$). The part $\frac{k_i^s k_j^s}{2m_s}$ is the expected probability of an edge between i and j in layer s according to the configuration null-model for networks [36], and the effect on Q of this null-model comparison is scaled by γ. Thus, as whole, the first term is largest when there exists an edge between i and j in time window τ_s and when the expected probability of such an edge is smallest. Effectively, this term influences optimal partitions to give (i, s) and (j, s) the same community label if there is an edge between nodes i and j in time window τ_s. We next consider the second term, $\omega \delta(i, j) C_{sr}$, which is non-negative since $\omega > 0$ and $C_{sr} \in \{0, 1\}$. Since $C_{sr} = 1$ only when τ_s and τ_r are consecutive time windows (i.e., $|s - r| = 1$), this term influences the community labels c_{is} and c_{ir} to be the same from one time window to the next.

Note that multilayer modularity involves two parameters γ and ω, which are "tuning knobs" to identify clusters in which their size and temporal coherence are appropriate. In practice, we explore a wide range of parameter values $\gamma \in [\gamma_{\min}, \gamma_{\max}]$ and $\omega \in [\omega_{\min}, \omega_{\max}]$ to study the multiscale organization of clusters. This approach efficiently explores clustering phenomena at multiple spatial and temporal scales, identifying at which scales clustering is most prevalent and at which scales clustering is nonexistent. To identify appropriate values for γ and ω, we used a variety of techniques including the CHAMP algorithm [50] (which utilizes fast algorithms that detect convex hulls in the (γ, ω) parameter space) and comparisons to other community-detection algorithms including the study of connected-components.

Supporting information

S1 Video. One minute of real-time simulation for $\mu = 0.09$, demonstrating association of beads into large, stable clusters. Resource available at https://github.com/bwalker1/chromosome-videos/blob/master/Dataset0_allRed_finer_realtime.mp4.
(MP4)

S2 Video. One minute of real-time simulation for $\mu = 0.19$, showing association of beads into clusters that exhibit frequent changes in membership. Resource available at https://github.com/bwalker1/chromosome-videos/blob/master/Dataset6_allRed_finer_realtime.mp4.
(MP4)

S3 Video. One minute of real-time simulation for $\mu = 1.6$, showing no apparent structure in the bead positions beyond simple pairing. Resource available at https://github.com/bwalker1/chromosome-videos/blob/master/Dataset12_allRed_finer_realtime.mp4.
(MP4)

S4 Video. One minute of real-time simulation for the rigid clustering regime with $\mu = 0.09$. We observe that the stable and well-separated clusters have been distinctly labeled by the community detection algorithm. Resource available at https://github.com/bwalker1/chromosome-videos/blob/master/Dataset0_color_finer_realtime_altView.mp4.
(MP4)

S5 Video. One minute of real-time simulation for the flexible clustering regime with $\mu = 0.19$. Here the clusters are not so clearly separated, but the colored labels still appear consistent with what one would expect. Resource available at https://github.com/bwalker1/chromosome-videos/blob/master/Dataset6_color_finer_realtime_altView.mp4.
(MP4)

S6 Video. One minute of real-time simulation for the non-clustering regime with μ = 1.6. Here there are no clusters present in the data, but the community detection algorithm still tries to give the same label to nearby beads. Due to the lack of stable community structure, beads change label more frequently than for smaller values of μ. Resource available at https://github.com/bwalker1/chromosome-videos/blob/master/Dataset12_color_finer_realtime_altView.mp4.
(MP4)

Author Contributions

Conceptualization: Benjamin Walker, Dane Taylor, Kerry Bloom, M. Gregory Forest.

Data curation: Benjamin Walker, Dane Taylor, Josh Lawrimore, Caitlin Hult, David Adalsteinsson, Kerry Bloom.

Formal analysis: Benjamin Walker, Dane Taylor, Josh Lawrimore, Kerry Bloom.

Investigation: Benjamin Walker, Dane Taylor, Josh Lawrimore, Kerry Bloom, M. Gregory Forest.

Methodology: Benjamin Walker, Dane Taylor, Josh Lawrimore, Caitlin Hult, Kerry Bloom, M. Gregory Forest.

Project administration: Dane Taylor.

Resources: Benjamin Walker, Dane Taylor, Josh Lawrimore, Caitlin Hult, David Adalsteinsson, M. Gregory Forest.

Software: Benjamin Walker, Dane Taylor, Josh Lawrimore, Caitlin Hult, David Adalsteinsson, M. Gregory Forest.

Supervision: Dane Taylor, M. Gregory Forest.

Validation: Benjamin Walker, Dane Taylor, Josh Lawrimore.

Visualization: Benjamin Walker, Dane Taylor, Josh Lawrimore, Caitlin Hult, David Adalsteinsson.

Writing – original draft: Benjamin Walker, Dane Taylor, Josh Lawrimore, Caitlin Hult, Kerry Bloom, M. Gregory Forest.

Writing – review & editing: Benjamin Walker, Dane Taylor, Josh Lawrimore, Caitlin Hult, Kerry Bloom, M. Gregory Forest.

References

1. Dekker J, Belmont AS, Guttman M, Leshyk VO, Lis JT, Lomvardas S, et al. The 4D nucleome project. Nature. 2017; 549(7671):219. https://doi.org/10.1038/nature23884 PMID: 28905911

2. Marko JF, Siggia ED. Fluctuations and supercoiling of DNA. Science. 1994; 265(5171):506–508.

3. Marko J, Siggia E. Statistical mechanics of supercoiled DNA. Physical Review E. 1995; 52(3):2912. https://doi.org/10.1103/PhysRevE.52.2912

4. Wong H, Marie-Nelly H, Herbert S, Carrivain P, Blanc H, Koszul R, et al. A predictive computational model of the dynamic 3D interphase yeast nucleus. Current Biology. 2012; 22(20):1881–1890. https://doi.org/10.1016/j.cub.2012.07.069 PMID: 22940469

5. Wong H, Arbona JM, Zimmer C. How to build a yeast nucleus. Nucleus. 2013; 4(5):361–366. https://doi.org/10.4161/nucl.26226 PMID: 23974728

6. Vasquez PA, Bloom K. Polymer models of interphase chromosomes. Nucleus. 2014; 5(5):376–390. https://doi.org/10.4161/nucl.36275 PMID: 25482191

7. Wang R, Mozziconacci J, Bancaud A, Gadal O. Principles of chromatin organization in yeast: relevance of polymer models to describe nuclear organization and dynamics. Current opinion in cell biology. 2015; 34:54–60. https://doi.org/10.1016/j.ceb.2015.04.004 PMID: 25956973

8. Tjong H, Gong K, Chen L, Alber F. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. Genome research. 2012; 22(7):1295–1305. https://doi.org/10.1101/gr.129437.111 PMID: 22619363

9. Verdaasdonk JS, Vasquez PA, Barry RM, Barry T, Goodwin S, Forest MG, et al. Centromere tethering confines chromosome domains. Molecular Cell. 2013; 52(6):819–831. https://doi.org/10.1016/j.molcel.2013.10.021 PMID: 24268574

10. Ea V, Baudement MO, Lesne A, Forné T. Contribution of topological domains and loop formation to 3D chromatin organization. Genes. 2015; 6(3):734–750. https://doi.org/10.3390/genes6030734 PMID: 26226004

11. Vasquez PA, Hult C, Adalsteinsson D, Lawrimore J, Forest MG, Bloom K. Entropy gives rise to topologically associating domains. Nucleic Acids Research. 2016; 44(12):5540–5549. https://doi.org/10.1093/nar/gkw510 PMID: 27257057

12. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002; 295 (5558):1306–1311. https://doi.org/10.1126/science.1067799 PMID: 11847345

13. Le TB, Imakaev MV, Mirny LA, Laub MT. High-resolution mapping of the spatial organization of a bacterial chromosome. Science. 2013; 342(6159):731–734. https://doi.org/10.1126/science.1242059 PMID: 24158908

14. 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

15. Dekker J. Mapping in Vivo Chromatin Interactions in Yeast Suggests an Extended Chromatin Fiber with Regional Variation in Compaction. The Journal of Biological Chemistry. 2008; 283(50):34532–34540. https://doi.org/10.1074/jbc.M806479200 PMID: 18930918

16. 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

17. Duan Z, Andronescu M, Schutz K, McIlwain S, Kim YJ, Lee C, et al. A three-dimensional model of the yeast genome. Nature. 2010; 465(7296):363. https://doi.org/10.1038/nature08973 PMID: 20436457

18. Tanizawa H, Iwasaki O, Tanaka A, Capizzi JR, Wickramasinghe P, Lee M, et al. Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. Nucleic Acids Research. 2010; 38(22):8164–8177. https://doi.org/10.1093/nar/gkq955 PMID: 21030438

19. Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O, et al. Topologically associating domains are stable units of replication-timing regulation. Nature. 2014; 515(7527):402. https://doi.org/10.1038/nature13986 PMID: 25409831

20. Vian L, Pękowska A, Rao SS, Kieffer-Kwon KR, Jung S, Baranello L, et al. The Energetics and Physiological Impact of Cohesin Extrusion. Cell. 2018; 173(5):1165–1178. https://doi.org/10.1016/j.cell.2018.03.072 PMID: 29706548

21. Hult C, Adalsteinsson D, Vasquez PA, Lawrimore J, Bennett M, York A, et al. Enrichment of dynamic chromosomal crosslinks drive phase separation of the nucleolus. Nucleic Acids Research. 2017; 45 (19):11159–11173. https://doi.org/10.1093/nar/gkx741 PMID: 28977453

22. Belmont AS, Straight AF. In vivo visualization of chromosomes using lac operator-repressor binding. Trends in Cell Biology. 1998; 8(3):121–124. https://doi.org/10.1016/S0962-8924(97)01211-7 PMID: 9695822

23. Pearson CG, Maddox PS, Salmon E, Bloom K. Budding yeast chromosome structure and dynamics during mitosis. The Journal of Cell Biology. 2001; 152(6):1255–1266. https://doi.org/10.1083/jcb.152.6.1255 PMID: 11257125

24. Fisher J, Ballenger M, O'Brien E, Haase J, Superfine R, Bloom K. DNA relaxation dynamics as a probe for the intracellular environment. Proceedings of the National Academy of Sciences. 2009; 106 (23):9250–9255. https://doi.org/10.1073/pnas.0812723106

25. Belmont AS, Li G, Sudlow G, Robinett C. Visualization of large-scale chromatin structure and dynamics using the lac operator/lac repressor reporter system. In: Methods in cell biology. vol. 58. Elsevier; 1999. p. 203–222.

26. Rajapakse I, Scalzo D, Tapscott SJ, Kosak ST, Groudine M. Networking the nucleus. Molecular Systems Biology. 2010; 6(1):395. https://doi.org/10.1038/msb.2010.48 PMID: 20664641

27. Rajapakse I, Groudine M. On emerging nuclear order. The Journal of Cell Biology. 2011; 192(5):711–721. https://doi.org/10.1083/jcb.201010129 PMID: 21383074

28. Rajapakse I, Groudine M, Mesbahi M. Dynamics and control of state-dependent networks for probing genomic organization. Proceedings of the National Academy of Sciences. 2011;. https://doi.org/10.1073/pnas.1113249108

29. Liu S, Chen H, Ronquist S, Seaman L, Ceglia N, Meixner W, et al. Genome architecture leads a bifurcation in cell identity. bioRxiv. 2017; p. 151555.

30. Liu S, Chen PY, Hero A, Rajapakse I. Dynamic Network Analysis of the 4D Nucleome. bioRxiv. 2018; p. 268318.

31. Chen J, Hero AO III, Rajapakse I. Spectral identification of topological domains. Bioinformatics. 2016; 32(14):2151–2158. https://doi.org/10.1093/bioinformatics/btw221 PMID: 27153657

32. Cabreros I, Abbe E, Tsirigos A. Detecting community structures in hi-c genomic data. In: Information Science and Systems (CISS), 2016 Annual Conference on. IEEE; 2016. p. 584–589.

33. Chen H, Chen J, Muir LA, Ronquist S, Meixner W, Ljungman M, et al. Functional organization of the human 4D Nucleome. Proceedings of the National Academy of Sciences. 2015; 112(26):8002–8007. https://doi.org/10.1073/pnas.1505822112

34. Delaneau O, Zazhytska M, Borel C, Giannuzzi G, Rey G, Howald C, et al. Chromatin three-dimensional interactions mediate genetic effects on gene expression. Science. 2019; 364(6439):eaat8266. https://doi.org/10.1126/science.aat8266 PMID: 31048460

35. Ronquist S, Patterson G, Muir LA, Lindsly S, Chen H, Brown M, et al. Algorithm for cellular reprogramming. Proceedings of the National Academy of Sciences. 2017; 114(45):11832–11837. https://doi.org/10.1073/pnas.1712350114

36. Newman ME. Modularity and community structure in networks. Proceedings of the National Academy of Sciences. 2006; 103(23):8577–8582. https://doi.org/10.1073/pnas.0601602103

37. Mucha PJ, Richardson T, Macon K, Porter MA, Onnela JP. Community structure in time-dependent, multiscale, and multiplex networks. Science. 2010; 328(5980):876–878. https://doi.org/10.1126/science.1184819 PMID: 20466926

38. 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

39. Eser U, Chandler-Brown D, Ay F, Straight AF, Duan Z, Noble WS, et al. Topologically associating genomic domains in yeast. Proceedings of the National Academy of Sciences. 2017; 114(15):3061–3070. https://doi.org/10.1073/pnas.1612256114

40. Schalbetter S, Goloborodko A, Fudenberg G, Belton J, Miles C, Yu M, et al. Structural maintenance of chromosome complexes differentially compact mitotic chromosomes according to genomic context. Nature Cell Biology. 2017; 19(9):1071–1080. https://doi.org/10.1038/ncb3594 PMID: 28825700

41. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of chromosomal domains by loop extrusion. Cell Reports. 2016; 15(9):2038–2049. https://doi.org/10.1016/j.celrep.2016.04.085 PMID: 27210764

42. Dekker J, Marti-Renom MA, Mirny L. Exploring the Three-Dimensional Organization of Genomes: Interpreting Chromatin Interaction Data. Nature Reviews Genetics. 2013; 14(6):390–403. https://doi.org/10.1038/nrg3454 PMID: 23657480

43. Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, et al. Iterative Correction of Hi-C Data Reveals Hallmarks of Chromosome Organization. Nature Methods. 2012; 9(10):999–1003. https://doi.org/10.1038/nmeth.2148 PMID: 22941365

44. Johzuka K, Horiuchi T. The cis element and factors required for condensin recruitment to chromosomes. Molecular Cell. 2009; 34(1):26–35. https://doi.org/10.1016/j.molcel.2009.02.021 PMID: 19362534

45. Johzuka K, Terasawa M, Ogawa H, Ogawa T, Horiuchi T. Condensin loaded onto the replication fork barrier site in the rRNA gene repeats during S phase in a FOB1-dependent fashion to prevent contraction of a long repetitive array in Saccharomyces cerevisiae. Molecular and Cellular Biology. 2006; 26 (6):2226–2236. https://doi.org/10.1128/MCB.26.6.2226-2236.2006 PMID: 16507999

46. Prieto JL, McStay B. Recruitment of factors linking transcription and processing of pre-rRNA to NOR chromatin is UBF-dependent and occurs independent of transcription in human cells. Genes & Development. 2007; 21(16):2041–2054. https://doi.org/10.1101/gad.436707

47. Albert B, Colleran C, Leger-Silvestre I, Berger AB, Dez C, Normand C, et al. Structure-function analysis of Hmo1 unveils an ancestral organization of HMG-Box factors involved in ribosomal DNA transcription from yeast to human. Nucleic Acids Research. 2013; 41(22):10135–10149. https://doi.org/10.1093/nar/gkt770 PMID: 24021628

48. Otsu N. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics. 1979; 9(1):62–66. https://doi.org/10.1109/TSMC.1979.4310076

49. Code Release: Bloom Lab Yeast—Nucleolar Analysis; available at https://github.com/BloomLabYeast/nucleolarAnalysis;.

50. Weir WH, Emmons S, Gibson R, Taylor D, Mucha PJ. Post-processing partitions to identify domains of modularity optimization. Algorithms. 2017; 10(3):93. https://doi.org/10.3390/a10030093 PMID: 29046743

51. Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu AY. An efficient k-means clustering algorithm: Analysis and implementation. IEEE Transactions on Pattern Analysis & Machine Intelligence. 2002;(7):881–892. https://doi.org/10.1109/TPAMI.2002.1017616

52. Taylor D, Shai S, Stanley N, Mucha PJ. Enhanced detectability of community structure in multilayer networks through layer aggregation. Physical Review Letters. 2016; 116(22):228301. https://doi.org/10.1103/PhysRevLett.116.228301 PMID: 27314740

53. Taylor D, Caceres RS, Mucha PJ. Super-resolution community detection for layer-aggregated multilayer networks. Physical Review X. 2017; 7(3):031056. https://doi.org/10.1103/PhysRevX.7.031056 PMID: 29445565