

# SCIENTIFIC REPORTS



OPEN

## The Rabl configuration limits topological entanglement of chromosomes in budding yeast

Maxime Pouokam<sup>6</sup>, Brian Cruz<sup>2</sup>, Sean Burgess<sup>1</sup>, Mark R. Segal<sup>3</sup>, Mariel Vazquez<sup>4,5</sup> & Javier Arsuaga<sup>1,4</sup>

The three dimensional organization of genomes remains mostly unknown due to their high degree of condensation. Biophysical studies predict that condensation promotes the topological entanglement of chromatin fibers and the inhibition of function. How organisms balance between functionally active genomes and a high degree of condensation remains to be determined. Here we hypothesize that the Rabl configuration, characterized by the attachment of centromeres and telomeres to the nuclear envelope, helps to reduce the topological entanglement of chromosomes. To test this hypothesis we developed a novel method to quantify chromosome entanglement complexity in 3D reconstructions obtained from Chromosome Conformation Capture (CCC) data. Applying this method to published data of the yeast genome, we show that computational models implementing the attachment of telomeres or centromeres alone are not sufficient to obtain the reduced entanglement complexity observed in 3D reconstructions. It is only when the centromeres and telomeres are attached to the nuclear envelope (i.e. the Rabl configuration) that the complexity of entanglement of the genome is comparable to that of the 3D reconstructions. We therefore suggest that the Rabl configuration is an essential player in the simplification of the entanglement of chromatin fibers.

All organisms pack their genomes in small volumes in a manner that allows for basic DNA transactions to take place<sup>1–4</sup>. It is, however, well known that spatial confinement promotes the topological entanglement of DNA molecules<sup>5–16</sup> and that topological entanglement inhibits basic DNA transactions including gene expression and chromosome segregation<sup>17–20</sup>. How the eukaryotic cell balances function and topological entanglement is poorly understood. At large scales, eukaryotic chromosomes are organized into territories<sup>21,22</sup> and compartments<sup>23,24</sup>. Both structures are present across cell types and experimental conditions<sup>25</sup>, with compartments being partially determined by the epigenetic state of the genome<sup>23,26</sup>. At the megabase-pair resolution, in an apparently self-similar fashion<sup>24,27–29</sup>, chromosomes are partitioned into domains and loops<sup>30–36</sup> in an evolutionary preserved manner<sup>37,38</sup>. Several biophysical mechanisms, such as active loop extrusion<sup>39–41</sup> and protein bridging<sup>42</sup>, and their corresponding models which include block co-polymers and random branched trees<sup>43,44</sup> have been proposed to explain the possibly diverse origin of chromosomal domains and loops. Interestingly, these levels of organization are not independent of each other suggesting a sophisticated crosstalk of chromosome organization across scales<sup>25,45</sup>.

First described at the end of the nineteenth century<sup>46</sup>, the Rabl configuration is an evolutionary conserved feature of the 3D nuclear organization<sup>47,48</sup>. The Rabl configuration is characterized by the clustering of centromeres on one side of the nuclear envelope<sup>49,50</sup> and is believed to reflect the orientation of chromosomes from the preceding mitosis<sup>51</sup>. Analysis of fluorescently tagged chromosomal loci in yeast revealed that, in the Rabl configuration, subtelomeric regions are positioned non-randomly near the nuclear periphery with their relative position and distance from subtelomeric regions in other chromosome arms influenced by chromosome arm length<sup>51,52</sup>. This analysis also revealed an axis that passes through the center of the nucleus, with centromeres clustered opposite to the nucleolus at one end of the nuclear periphery<sup>53</sup>. Functionally, the preservation or disruption of the Rabl configuration has been associated with chromosome segregation<sup>54</sup> and DNA repair processes<sup>55</sup>.

<sup>1</sup>Department of Molecular and Cellular Biology, University of California, Davis, CA, 95616, USA. <sup>2</sup>Department of Mathematics, University of California, Berkeley, CA, 94720, USA. <sup>3</sup>Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, 94143, USA. <sup>4</sup>Department of Mathematics, University of California, Davis, CA, 95616, USA. <sup>5</sup>Department of Microbiology and Molecular Genetics, University of California, Davis, CA, 95616, USA. <sup>6</sup>Department of Statistics, University of California, Davis, CA, 95616, USA. Correspondence and requests for materials should be addressed to J.A. (email: [jarsuaga@ucdavis.edu](mailto:jarsuaga@ucdavis.edu))

Here we hypothesized that the Rabl configuration significantly reduces the incidence of chromosome entanglements during interphase. To test this hypothesis we turned to budding yeast. The three dimensional organization of the budding yeast genome is much simpler than that of the human genome, its chromosome territory structure is much less pronounced, and its overall organization is believed to be closer to equilibrium than that of the mammalian genome<sup>56,57</sup>. On the other hand, the budding yeast genome preserves some of the basic packaging features of higher organisms<sup>21,53,58</sup> and presents very strong evidence for the Rabl configuration<sup>49,50,59,60</sup>. We used previously published 3D reconstructions obtained from chromosome conformation capture (CCC) data which have successfully illuminated many aspects of 3D chromosomal architecture in yeast including chromatin structure, functional compartmentalization and DNA repair<sup>61–63</sup>.

Quantification of the entanglement complexity between two chains is a well-defined mathematical problem when the chains are circular; in this case entanglement complexity is commonly quantified by topological invariants, including the linking number<sup>64</sup>. Since chromosomes are open chains (i.e. non-circular), new mathematical tools are required for quantifying their entanglement<sup>65–67</sup>. Here we propose a method inspired by advances in knot identification in proteins<sup>68</sup>. There, the knot type of the protein backbone is identified by producing ensembles of closed circular chains associated to a single protein backbone conformation. The topology of the protein is probabilistic in nature and determined by the proportion of different knots observed in the associated ensembles. We extend this approach by examining the entanglement of pairs of open chains, and by introducing the linking proportion for two open chains as a measure of their entanglement complexity. The distribution of the linking proportions associated to each pair of chromosomes in a CCC reconstruction quantifies the entanglement complexity of the reconstruction and provides a measure to compare experimentally derived reconstructions against each other and against theoretical models.

To quantify the entanglement complexity of the yeast genome, we compared the distributions of linking proportions associated with CCC reconstructions to the distributions of linking proportions associated to randomly embedded semiflexible (wormlike) chains in confinement. We found that the entanglement complexity of all reconstructions was lower than the entanglement complexity predicted by the wormlike chain model. This finding validates our approach and provides further evidence that the yeast genome is not randomly organized in the interphase nucleus<sup>52,69–71</sup>. We also show that the entanglement complexity of a random organization of simulated yeast chromosomes can be significantly reduced by attachments of centromeres or telomeres to the nuclear envelope; although these cannot fully account for the low entanglement complexity observed in the 3D reconstructions of the yeast genome. The absence of entanglement observed in the reconstructions is achieved only with the implementation of the Rabl configuration. We therefore suggest that the Rabl configuration is a key organizational feature that prevents the yeast chromosomes from becoming entangled.

## Data, Models and Methods

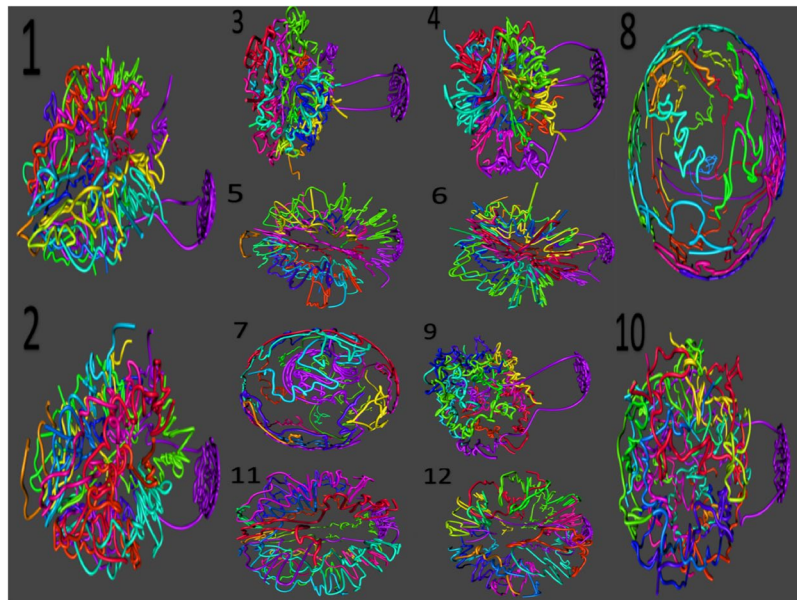
**Data.** In previous works we generated a set of eleven 3D reconstructions of the yeast genome<sup>72</sup> using data published by Duan and colleagues<sup>62</sup>. Reconstructions in Figure 1 were obtained using those published data. These reconstructions were the product of two consecutive restriction reactions with different enzymes, different false discovery rate (FDR) threshold<sup>73</sup> at the time of significance assessment of contacts, and/or different physical distances. Reconstruction 1 was the reconstruction reported in the publication by Duan and colleagues; It was obtained from contact maps using HindIII followed by the combination of MseI and MspI (denoted by  $MseI \cup MspI$ ) restriction assays and a FDR threshold of 0.01. Reconstructions 2–4 were obtained by changing the FDR threshold (0.01, 0.1 and 1.0). Reconstructions 5, 6, 7 and 8 used restriction enzyme HindIII followed by MseI (Reconstruction 5), MspI (Reconstruction 6) and  $MseI \cup MspI$  (reconstructions 7 and 8). Additionally, reconstructions 7 and 8 used recomputed physical distances. Reconstruction 9 used the common fragments obtained from EcoRI and HindIII (denoted by  $EcoRI \cap HindIII$ ) followed by  $MseI \cup MspI$ ; reconstruction 10 used EcoRI followed by  $MseI \cap MspI$  and reconstructions 11 and 12 used EcoRI followed by MseI and MspI respectively<sup>72</sup>. Distance units in the reconstructions reported here coincide with those used in<sup>72</sup>, and are proportional to, the experimental measurements<sup>62</sup>.

## Quantification of the entanglement complexity of the yeast genome and classification of 3D reconstructions.

To determine the topological complexity of the yeast genome we defined a new geometrical invariant that extends the concept of linking number for open curves. We use the single stochastic closure method<sup>68,74</sup> to associate an ensemble of closed trajectories to each chromosome reconstruction. For this purpose, we define a sphere  $S$  with its center coinciding with the center of mass of the original reconstruction (all 16 chromosomes), and radius  $R > r$ , where  $r$  is the radius of the smallest sphere containing the reconstruction (e.g.  $r \approx 110$  and  $R = 150$ ). We then define a circular trajectory for each chromosome by tracing two rays connecting the telomeres of each chromosome reconstruction with a point  $P$  chosen at random on the surface of the sphere  $S$  (Fig. 2a). We obtained an ensemble of closed trajectories associated to the chromosome by repeating this process for  $P_i, i = 1, \dots, N$ . In the results reported below we used  $N = 10^3$ . To validate our method, we recomputed the linking proportion for different values of  $R, P$  and  $N$  and obtained consistent results. We define the linking proportion for the reconstruction of two chromosomes  $l_1$  and  $l_2$  in a given reconstruction or simulation as

$$PLk(l_1, l_2) = \frac{1}{N} \sum_{k=1}^N X_k(i, j)$$

Here  $X_k(i, j) = 1$  if  $I(C_1^i, C_2^j) \neq 0$  and  $I(C_1^i, C_2^j) = 0$  otherwise, where  $I$  is a topological invariant, which in our study is the linking number, of two circular trajectories  $C_1^i$  and  $C_2^j$  associated, through the single stochastic



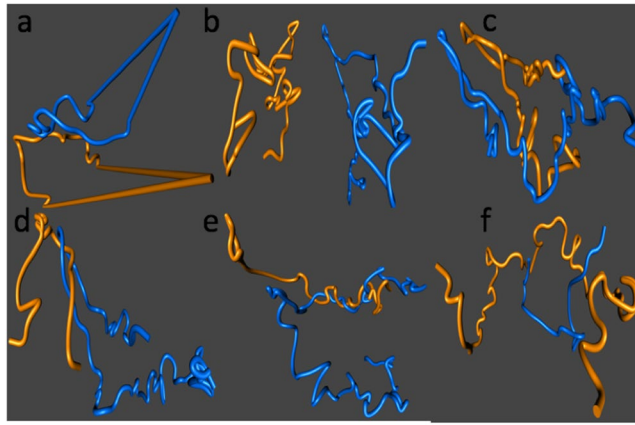
**Figure 1.** Reconstructions of the yeast genome derived from CCC data. Panel (1) shows the reconstruction using data from<sup>62</sup> and Panel (2) shows a replicate of Panel (1) obtained using the same input parameters. Panels (2–12) Reconstructions obtained in<sup>72</sup> using different experimental conditions and reconstruction parameters. All reconstructions are consistent with CCC data.

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI
I	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
II	8.3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
III	5.1	79.3	—	—	—	—	—	—	—	—	—	—	—	—	—	—
IV	3.9	13.6	5.6	—	—	—	—	—	—	—	—	—	—	—	—	—
V	13.7	15.7	9.2	8.6	—	—	—	—	—	—	—	—	—	—	—	—
VI	14.8	8.1	8.0	17.6	9.5	—	—	—	—	—	—	—	—	—	—	—
VII	9.8	25.9	15.3	11.8	35.6	7.4	—	—	—	—	—	—	—	—	—	—
VIII	7.7	29.4	83.9	9.4	10.7	12.9	15.0	—	—	—	—	—	—	—	—	—
IX	12.4	14.9	9.1	5.2	48.9	10.9	24.0	12.4	—	—	—	—	—	—	—	—
X	24.2	8.2	4.0	11.1	16.6	12.9	10.6	7.3	13.8	—	—	—	—	—	—	—
XI	9.2	8.5	5.7	30.2	9.0	23.5	7.5	11.9	7.5	8.9	—	—	—	—	—	—
XII	12.1	29.6	11.8	15.7	39.5	9.8	48.9	21.6	66.5	11.6	11.1	—	—	—	—	—
XIII	5.1	16.9	10.0	16.8	11.8	85.1	14.5	17.1	9.1	5.9	20.6	13.6	—	—	—	—
XIV	18.2	16.7	6.7	24.3	31.9	11.6	20.1	9.9	23.4	80.0	10.2	22.8	11.5	—	—	—
XV	10.3	13.1	4.9	25.9	14.7	17.8	12.5	9.4	14.2	70.9	11.5	20.4	14.0	45.6	—	—
XVI	8.1	23.2	7.7	26.1	11.7	13.9	15.7	21.1	11.5	7.7	20.5	20.8	42.3	12.3	11.4	—

**Table 1.** The linking proportions of the 16 chromosomes in Reconstruction 1 (Fig. 1). The entries reported are given in %. The entry in row  $i$  and column  $j$  corresponds to  $PLk(l_i, l_j)$ , the linking proportions of the linear chromosomes  $l_i$  and  $l_j$ . Entries highlighted in red indicate a linking proportion greater than 50%.

closure method, to the linear trajectories  $l_1$  and  $l_2$ . We calculated the linking number using the double Gaussian integral form<sup>75</sup> and compiled the linking proportion results in a lower triangular matrix, as illustrated in Table 1. Using the Kolmogorov-Smirnov (KS) test<sup>76</sup> we determined differences between reconstructions by testing whether the distribution of entries in the tables in their vectorized form<sup>77</sup> were derived from the same distribution (Section 2.3). Differences between reconstructions and simulated configurations were estimated using the Wilcoxon test (Section 2.5). In both statistical tests we assumed samples were independent and identically distributed. All p-values were corrected using FDR.

**Simulation of Yeast Genomes.** We modeled yeast chromosomes as a discrete approximation of a semiflexible chain with no torsional constraint, a configuration known as the wormlike chain<sup>78</sup>. Each chromosome  $C$  consisted of  $n_C$  segments  $\{e_1, \dots, e_{n_C}\}$  of equal length  $l$  and an energy given by



**Figure 2.** Circularization algorithm and examples of linking proportions for chromosome reconstructions from reconstructions 1, 9 and 10. **(a)** The circularization algorithm applied to two chromosome reconstructions. **(b)** Reconstructions for Chromosomes X (yellow) and XVI (blue) with  $PLk = 7.7\%$  **(c)** Chromosomes X (yellow) and XIV (blue) with  $PLk = 80.0\%$  **(d)** Chromosomes VI (yellow) and XIII (blue) with  $PLk = 85\%$  **(e)** Chromosomes IX (yellow) and XVI (blue) from Reconstruction 9, with  $PLk = 51.3\%$  **(f)** Chromosomes II (yellow) and VI (blue) from Reconstruction 10, with  $PLk = 53.8\%$ .

$$E(C) = \frac{1}{2}k \sum_{i=1}^{n_c-1} \theta_i^2$$

where  $k$  is the bending rigidity constant of the  $10\text{ nm}$  fiber, and  $\theta_i$  is the exterior angle between edges  $\mathbf{e}_i$  and  $\mathbf{e}_{i+1}$ . Given the experimentally estimated value for the persistence length of the yeast genome  $L_p = 197 \pm 62\text{ nm}$ <sup>52</sup>, we calculated the value of the bending rigidity constant  $k$  and the corresponding number of discrete segments necessary to represent each Kuhn length<sup>78</sup>. Each chromosome realization was obtained by applying simulated annealing to a freely jointed chain that was gradually confined inside a sphere of fixed radius (*i.e.* the cell nucleus) and simultaneously minimized with respect to the energy of the wormlike chain. To generate Rabl configurations the same annealing algorithm was used with the additional conditions that centromeres were clustered and telomeres were tethered to variable locations near their experimentally measured locations<sup>79,80</sup>. The energy potential that binds telomeres and centromeres to a specific location uses the  $L^1$  norm, which allows for more movement at higher temperatures yet more restricted movement at lower temperatures, and a smoother transition during the cooling schedule (See SI for more details).

**Generation of ensembles of open chains from closed chains.** To test whether our algorithm distinguishes between conformations that are in close proximity and entangled from those that are in close proximity but untangled (Section 2.2) we associated ensembles of open chains to pairs of closed chains. We generated statistically independent ensembles of closed freely jointed chains of fixed length with their center of masses separated by a fixed distance. For a pair of closed chains  $C_1$  and  $C_2$ , we randomly selected a segment in each of the closed chains ( $s_1$  and  $s_2$ ) removed four segments consecutive to each of the selected segments  $s_i$ ,  $i = 1, 2$  and applied the stochastic closure algorithm to compute their linking proportion. This process was repeated for different  $s_i$ . Proportions were averaged over all sets of open chains and used as a measure of entanglement of  $C_1$  and  $C_2$ .

## Results

### The circularization algorithm quantifies the entanglement complexity of 3D reconstructions.

We used the linking proportion between pairs of chromosomes (*i.e.* open chains  $l_1$  and  $l_2$ ) as a measure of entanglement complexity. Although this geometrical invariant changes with chromosome length and with the distance between the center of masses of the chromosomes, it detects entanglement between two open curves and it is robust with respect to noise inherent to CCC data, a problem that has previously obscured the geometrical interpretation of CCC data<sup>12,23,24,27,33</sup>. We illustrate the properties of this algorithm with some examples. Figure 2 shows 3D reconstructions corresponding to some of the entries in Table 1 (Fig. 2a–c), Supplementary Table 9 (Fig. 2e) and Supplementary Table 10 (Fig. 2f). Panel *b* shows the relative positions of chromosomes X (yellow) and XVI (blue). These chromosomes are large (745 kb and 948 kb respectively); their centers of masses are far from each other (83 reconstruction units) and their linking proportion was just 7.7%. Panel *c* shows chromosomes X (yellow) and XIV (blue), with the latter of size 784 kb. They are in close proximity (with the distance between their center of masses equal to 14 reconstruction units), and are highly intermingled producing a linking proportion of 80%. Panel *d* shows chromosomes VI (yellow) and XIII (blue). Chromosome XIII is also one of the large chromosomes in the genome (924 kb) while chromosome VI is only 270 kb. The distance between their centers of mass is large (70 reconstruction units) but they are highly entangled, with a linking proportion of 85%. Panel *e* shows chromosomes IX (yellow) and XVI (blue), chromosome IX is 440 kb, their centers of mass are 37 units apart and have a linking proportion of 51.3%. Panel *f* shows chromosomes II (yellow) and VI (blue), chromosome

	1	2	3	4	5	6	7	8	9	10	11	12
7	0.008	0.085	0.069	0.039	0.035	0.039	—	—	0.035	0.004	0.794	0.465
8	0.105	0.131	0.043	0.057	0.084	0.105	—	—	0.003	0.003	0.159	0.105
7	0.003	0.015	0.012	0.005	0.015	0.018	—	—	0.00	0.00	0.072	0.036
8	0.00	0.00	0.00	0.00	0.015	0.014	—	—	0.00	0.00	0.17	0.036

**Table 2.** Table of p-values comparing reconstructions 7 and 8 (obtained using the recomputed physical distances) with other reconstructions. The top two rows show the p-values obtained using the K-S test for the linking number defined for open (non-circular) chains. The bottom two rows show the p-values obtained using the linking proportion. The p-values were adjusted using False Discovery Rate (FDR) for multiple testing for the 12 reconstructions.

Reconstructions	1	2	3	4	5	6	7	8	9	10	12	Rabl
mean	18.2	18.4	18.4	19.0	14.0	16.1	14.4	12.7	21.6	25.6	14.0	24.1
median	12.7	13.1	14.35	14.2	11.5	12.1	10.8	9.5	17.4	20.2	9.9	15.6
min	3.9	3.1	3.6	2.1	3.30	2.8	1.4	2.2	3.2	4.7	3.2	2.2
max	85.1	93.8	91.5	83.5	73.7	91.0	99.3	71.3	93.5	92.8	90.5	94.6

**Table 3.** Linking proportion mean, median and range values for all reconstructions reflect the simulated Rabl configuration. Each column corresponds to one reconstruction.

It is 813 in length and the distance between the center of mass is 24 reconstruction units. Their linking proportion is 53.8%. Results for all linking proportions for Reconstruction 1 are shown in Table 1. Tables for all other reconstructions can be found in the supplementary material. On the basis of this work, we conclude that the circularization algorithm captures the entanglement complexity for 3D reconstructions of chromosomes.

### The single stochastic closure algorithm can distinguish between entangled and unentangled open reconstructions.

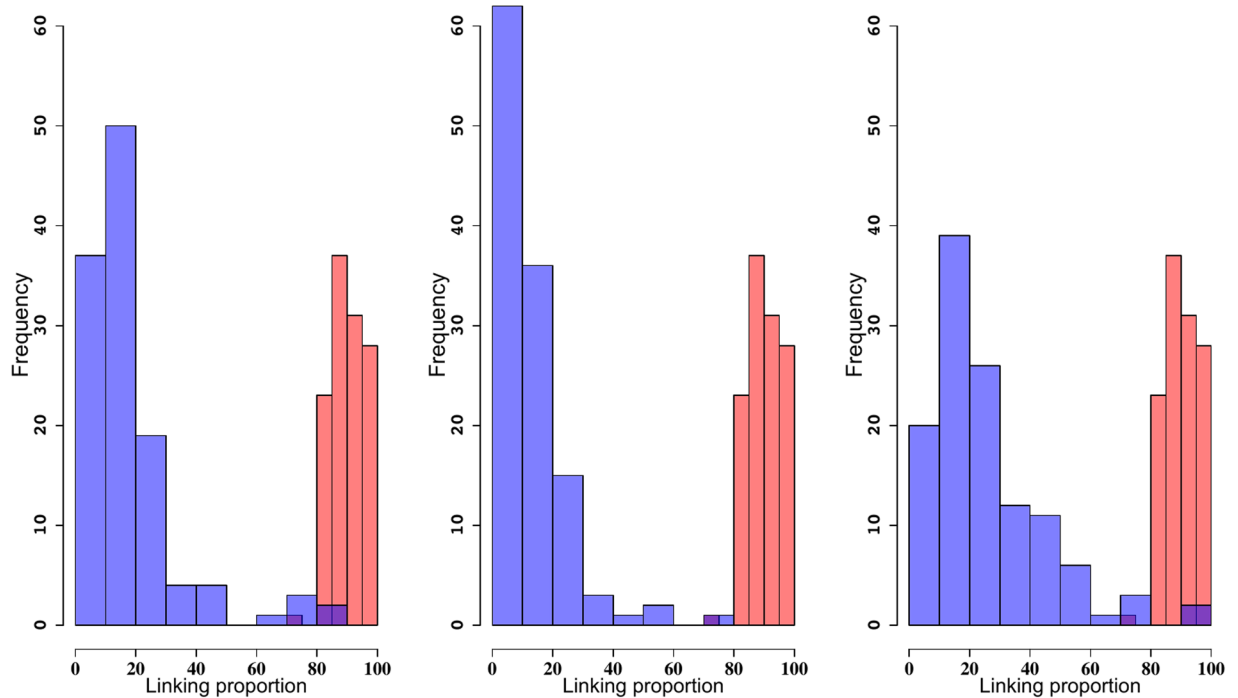
To test whether our method can statistically discriminate between pairs of chains that are in close proximity and entangled from those that are in close proximity but unentangled, we implemented the following statistical test. First, we generated a random sample of 1,000 pairs of closed circular freely jointed chains of equal length and with centers of mass separated by a fixed distance  $d^{81}$ . Second, we split the population into two subpopulations: those with linking number equal to zero and those with linking number different from zero. We computed their associated ensembles of open chains and linking proportions (See Section 1.4). Third, we computed and compared the average value of the linking proportions in each of the subpopulations (with or without linking number equal to zero). The sample mean of the linking proportion was 72% for entangled chains, and 57% for unentangled chains. To test whether these mean values were statistically different we generated the null distribution by permuting the linking proportion values of the entangled and unentangled chains. The p-value for the permutation test was  $\sim 10^{-3}$  (significant for  $\alpha = 0.05$ ). These results were further corroborated by performing the large sample independent t-test. Hence we concluded that the linking proportions can distinguish between entangled and unentangled configurations that are in very close proximity.

### The single stochastic closure algorithm outperforms the linking number for open chains when comparing reconstructions.

To further validate our approach we tested whether the linking proportion obtained by our method could distinguish reconstructions (obtained with different initial conditions) better than the known linking number for open chains. Based on the work in<sup>72</sup>, we would expect that a statistical algorithm should distinguish between reconstructions 7 and 8 from other reconstructions since they were generated using different physical distances. Tables were vectorized and compared using the K-S test (Section 1.2). Results of this comparison are shown in Table 2. The rows and columns in the table correspond to different reconstructions and the entries show the p-value associated to the linking number for open chains (first two rows) and the stochastic closure method proposed here (second two rows). Our results clearly show that the stochastic closure method can distinguish all of the reconstructions from reconstructions 7 and 8 (except for reconstruction 11). The linking number for open chains, on the other hand, fail to distinguish eleven of them. Based on these results we conclude that the stochastic closure algorithm outperforms other standard methods to measure entanglement of open curves obtained through the CCC data analyzed here.

### The entanglement complexity of the yeast genome is lower than predicted by the wormlike chain model.

We compared the observed linking proportion distribution in the reconstructions with those obtained using Monte-Carlo simulations of random embeddings of wormlike chains confined to a spherical volume. The mean, median and range for the distribution of linking proportions are shown in Table 3. Figure 3 shows the histograms representing the frequency of the linking proportions of the CCC reconstructions (blue) and of wormlike chains obtained by computer simulations (red). All the means were significantly smaller than that estimated for the wormlike chain ( $\mu_{wlc} = 81\%$ ). These results clearly show that the entanglement complexity predicted by the wormlike chains is much larger than that of the reconstructions. This finding is consistent with biophysical results of the linking of randomly embedded chains in confined volumes<sup>82</sup> and it provides further evidence that the yeast genome is not randomly organized. We also illustrate this result in Fig. 3, where the



**Figure 3.** Wormlike chains have a larger entanglement complexity than reconstructions. Frequency for the linking proportions of reconstructions 1, 8 and 10 and for randomly embedded wormlike chains. The distribution colored in blue corresponds to the reconstructions and the distribution colored in red corresponds to the wormlike chains. Purple indicates overlapping values.

distribution of linking proportions for three representative reconstructions. Reconstruction 1 (published in<sup>62</sup>), Reconstruction 8 and 10 with the lowest and largest mean/median of the linking proportion values respectively.

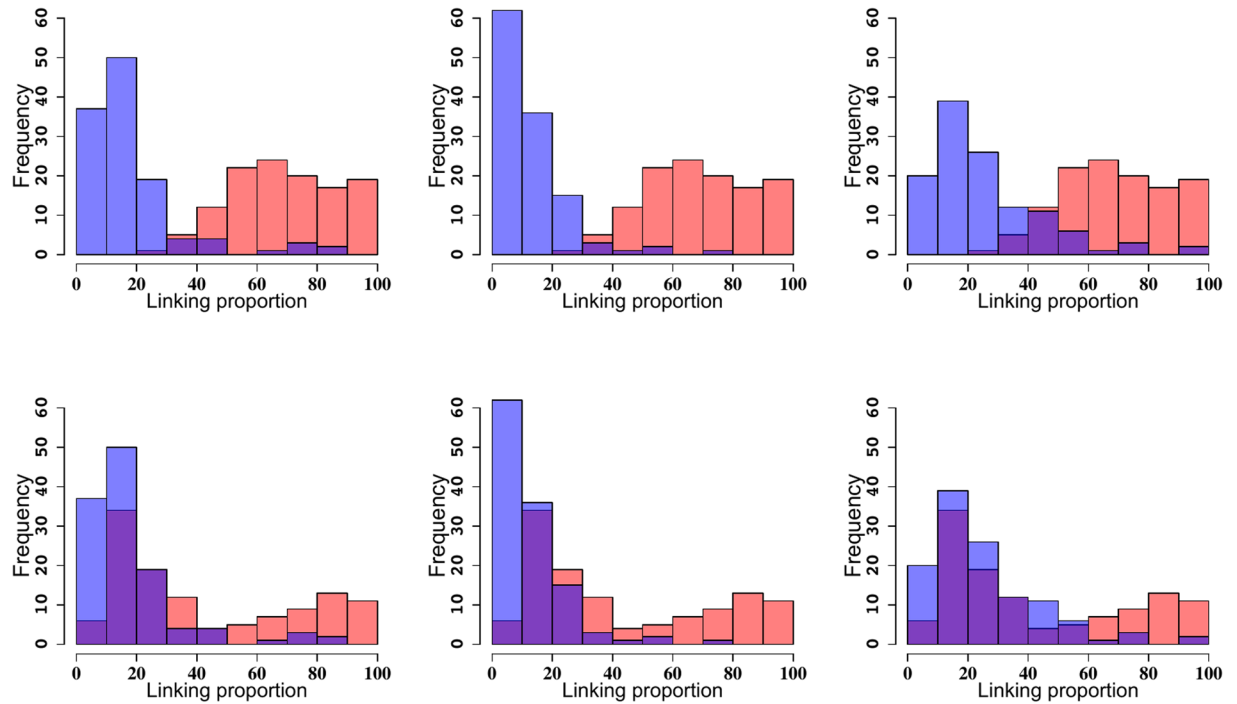
**The Rab1 configuration of the yeast genome helps explain the observed reduced chromosome entanglement.** The main genomic organizational features observed at the level of resolution of chromosome arms and territories are the clustering of centromeres and telomeres in the Rab1 configuration. We therefore posit that the distinct Rab1 configuration plays a key role in preventing the entanglement of chromosomes. To test this hypothesis we performed three different simulation studies: (i) with centromeres clustered near the nuclear envelope and free telomeres; (ii) with only telomeres tethered near the nuclear envelope; and (iii) with both centromeres and telomeres fixed, resembling the Rab1 configuration.

The top three panels in Fig. 4 show the distribution of linking proportions corresponding to simulation (i). The distributions of linking proportions corresponding to the selected reconstructions in (blue) still show much lower linking proportion than when only centromeres are clustered at the nuclear envelope (red) ( $\mu_{cent} = 70 \pm 20\%$ ). The lower three panels show the results for simulation (ii), in which only telomeres are attached near the nuclear envelope<sup>52,70</sup>. Interestingly, in case (ii), the distribution of linking number frequencies ( $\mu_{tel} = 60 \pm 30\%$ ) was closer to that of the reconstructions and much simpler than those of the randomly embedded wormlike chain and those in which only centromeres were attached (Fig. 4, top three panels). These results clearly show that these mechanisms are not sufficient to reduce the entanglement complexity to levels similar to those observed in the reconstruction data.

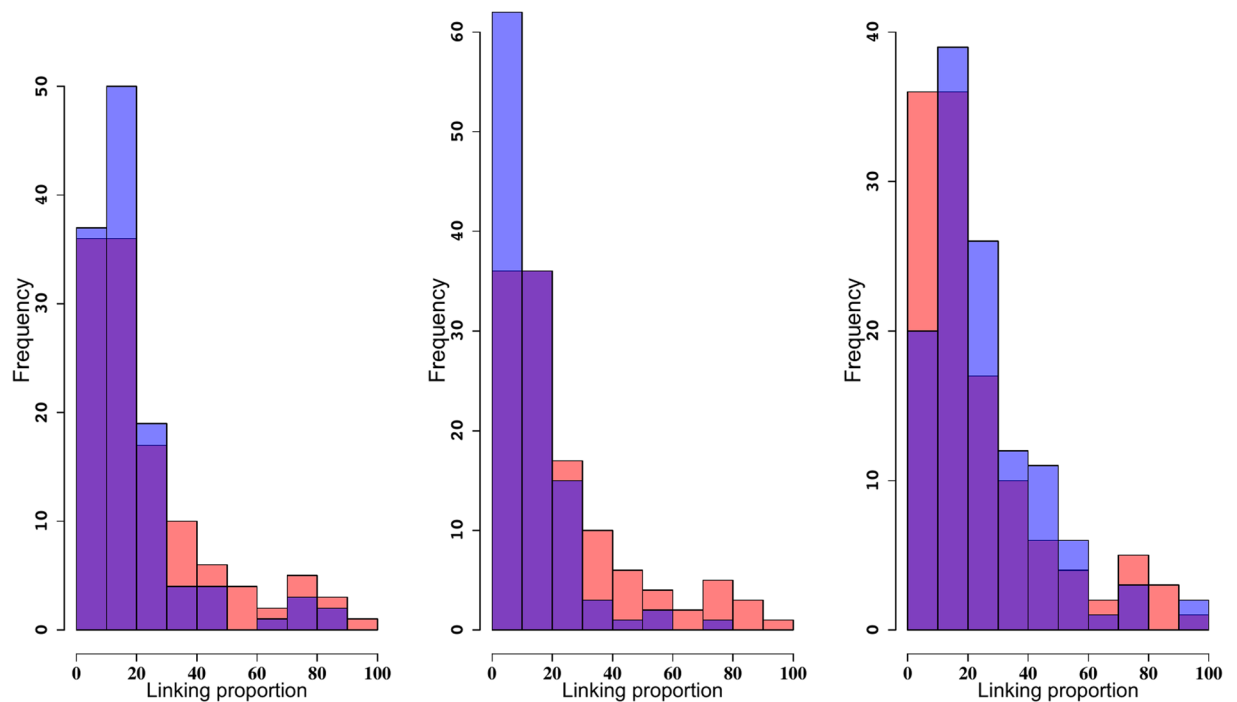
The situation was different when we implemented the Rab1 configuration in simulation (iii). In this case the topological complexity was reduced to levels comparable to those observed in the reconstructions and the mean of the distribution of linking numbers was  $\mu_{Rab1} = 24 \pm 21\%$  a value much closer to those observed in the reconstructions (Fig. 5). In fact there were no significant differences between the simulated configurations and Reconstructions 1 and 10 ( $p = 0.062$  and  $p = 0.062$ ). Interestingly, we still found significant differences between reconstruction 8 and the simulated models. Inspection of this reconstruction revealed that chromatin fibers have more interactions with the nuclear envelope than our proposed Rab1 configuration. This reconstruction suggests a mechanism of entanglement simplification driven by the frequent attachment of chromatin fibers to the nuclear envelope<sup>83</sup>. Based on these results we suggest that the Rab1 configuration is a regulator of the three dimensional organization of genomes that prevents entanglement of chromosome fibers.

## Conclusion and Discussion

The three dimensional organization of genomes is essential for the correct functioning of the cell. Confinement of DNA fibers, however, promotes entanglement as evidenced experimentally by DNA knots and links observed in some viruses<sup>5,84</sup> and in the mitochondrial DNA of kinetoplastids (reviewed in<sup>85</sup>), and by multiple theoretical studies<sup>6,7,9–11,13,86,87</sup>. Organisms have evolved mechanisms to regulate DNA entanglement. Most notable is the presence of topoisomerases and site specific recombinases, enzymes that regulate the topology of genomes and that are



**Figure 4.** Histograms of linking proportions for for reconstructions 1, 8 and 10 (left to right) and for randomly embedded wormlike chains with only centromeres (top three histograms) or only telomeres (bottom three histograms) attached to the nuclear envelope. The blue histogram corresponds to the reconstructions and the red to the wormlike chains. Purple indicate overlapping values. Both models are characterized by entanglement complexity values larger than the values observed in reconstructions.



**Figure 5.** Histograms of the linking proportions for reconstructions 1, 8 and 10 (left to right) and for simulated Rabl configurations (both centromeres and telomeres attached to the nuclear envelope). Blue corresponds to reconstructions and red to simulated data. The histograms overlap showing that the Rabl configurations decreases the entanglement of the genome to levels comparable to those observed in the experimental data.

known to unknot and unlink DNA<sup>88–90</sup>. There is evidence however that the cell has evolved other mechanisms to regulate the entanglement of genomes. For instance, at large scales, the eukaryotic chromosome is confined into territory (reviewed in<sup>22</sup>) and below the megabase scale, genomes are partitioned into domains and loops, an organizational feature that is preserved from bacteria<sup>37,38</sup> to humans<sup>30–34,36,42,91,92</sup>.

On the basis of the work reported here, we propose that the Rabl configuration, an organizational feature that is also preserved through multiple species in evolution, provides another mechanism for topology simplification. We tested this hypothesis using published CCC data and developed a new method in statistical topology to estimate the topological entanglement of genomes. Our method for estimating entanglement has some advantages over the standard linking number for open chains. First, it detects entanglement of curves better than the standard linking number for open curves when analyzing 3D reconstructions and second it can be extended by using topological invariants finer than the linking number. Whether the former is a general property or not remains to be determined. Our results showed that the entanglement complexity of CCC reconstructions of the yeast genome is lower than the entanglement complexity of free randomly embedded chains and chains in which centromeres and telomeres had been tethered to the nuclear envelope; only the implementation of the Rabl configuration yielded an entanglement complexity comparable to that of the CCC reconstructions. These results suggest that the Rabl configuration is a regulator of the entanglement complexity of the genome. Note however that this finding does not exclude the possibility of other mechanisms such as the tethering of other regions of the genome to the nuclear envelope<sup>20,83</sup>. Our method and conclusions are limited by the fact that chromosomes are highly dynamic, specially as the cell goes through the cell cycle. This limitation opens the door to new inquiries on topology regulation during the cell cycle.

## Data Availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

## References

1. Brickner, D. G. & Brickner, J. H. Interchromosomal clustering of active genes at the nuclear pore complex. *Nucleus* **3**, 487–492 (2012).
2. Burgess, S. M. & Kleckner, N. Collisions between yeast chromosomal loci *in vivo* are governed by three layers of organization. *Genes & development* **13**, 1871–1883 (1999).
3. Tang, Z. *et al.* Ctfc-mediated human 3d genome architecture reveals chromatin topology for transcription. *Cell* **163**, 1611–1627 (2015).
4. Wendt, K. S. & Grosveld, F. G. Transcription in the context of the 3d nucleus. *Current opinion in genetics & development* **25**, 62–67 (2014).
5. Arsuaga, J. *et al.* Knotting probability of dna molecules confined in restricted volumes: Dna knotting in phage capsids. *Proc. Natl. Acad. Sci.* **99**, 5373–5377 (2002).
6. Blackstone, T. *et al.* Modeling of chromosome intermingling by partially overlapping uniform random polygons. *J. Math. Biol.* **62**, 371–389 (2011).
7. D'Adamo, G. & Micheletti, C. Molecular crowding increases knots abundance in linear polymers. *Macromolecules* **48**, 6337–6346 (2015).
8. Kawamura, R. *et al.* Mitotic chromosomes are constrained by topoisomerase ii-sensitive dna entanglements. *J. Cell. Biol.* **188**, 653–663 (2010).
9. Marko, J. F. Scaling of linking and writhing numbers for spherically confined and topologically equilibrated flexible polymers. *Journal of statistical physics* **142**, 1353–1370 (2011).
10. Micheletti, C., Marenduzzo, D., Orlandini, E. & Sumners, D. Simulations of knotting in confined circular dna. *Biophys. J.* **95**, 3591–3599 (2008).
11. Poier, P., Likos, C. N. & Matthews, R. Influence of rigidity and knot complexity on the knotting of confined polymers. *Macromolecules* **47**, 3394–3400 (2014).
12. Siebert, J. T. *et al.* Are there knots in chromosomes? *Polymers* **9**, 317 (2017).
13. Sikorav, J. & Jannink, G. Kinetics of chromosome condensation in the presence of topoisomerases: a phantom chain model. *Biophys. J.* **66**, 827–837 (1994).
14. Tang, J., Du, N. & Doyle, P. S. Compression and self-entanglement of single dna molecules under uniform electric field. *Proc. Natl. Acad. Sci.* **108**, 16153–16158 (2011).
15. Trigueros, S. & Roca, J. Production of highly knotted dna by means of cosmid circularization inside phage capsids. *BMC Biotechnol.* **7**, 94 (2007).
16. Valdés, A., Segura, J., Dyson, S., Martnez-Garca, B. & Roca, J. Dna knots occur in intracellular chromatin. *Nucleic acids research* **46**, 650–660 (2017).
17. Deibler, R. W., Mann, J. K., De Witt, L. S. & Zechiedrich, L. Hin-mediated dna knotting and recombining promote replicon dysfunction and mutation. *BMC. Mol. Biol.* **8**, 44 (2007).
18. Portugal, J. & Rodríguez-Campos, A. T7 rna polymerase cannot transcribe through a highly knotted dna template. *Nucleic Acids Res.* **24**, 4890–4894 (1996).
19. Rodríguez-Campos, A. Dna knotting abolishes *in vitro* chromatin assembly. *Journal of Biological Chemistry* **271**, 14150–14155 (1996).
20. van Steensel, B. & Belmont, A. S. Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell* **169**, 780–791 (2017).
21. Cremer, T. *et al.* Rabl's model of the interphase chromosome arrangement tested in chinise hamster cells by premature chromosome condensation and laser-uv-microbeam experiments. *Hum. Genet.* **60**, 46–56 (1982).
22. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292 (2001).
23. Di Pierro, M., Zhang, B., Aiden, E. L., Wolynes, P. G. & Onuchic, J. N. Transferable model for chromosome architecture. *Proc. Natl. Acad. Sci.* **113**, 12168–12173 (2016).
24. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
25. Stevens, T. J. *et al.* 3d structures of individual mammalian genomes studied by single-cell hi-c. *Nature* **544**, 59 (2017).
26. Nazarov, L. I., Tamm, M. V., Avetisov, V. A. & Nechaev, S. K. A statistical model of intra-chromosome contact maps. *Soft matter* **11**, 1019–1025 (2015).



27. Arsuaga, J. *et al.* Current theoretical models fail to predict the topological complexity of the human genome. *Front. Mol. Biosci.* **2**, 48 (2015).
28. Grosberg, A., Rabin, Y., Havlin, S. & Neer, A. Crumpled globule model of the three-dimensional structure of dna. *EPL (Europhysics Letters)* **23**, 373 (1993).
29. Mirny, L. A. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.* **19**, 37–51 (2011).
30. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376 (2012).
31. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin domains: the unit of chromosome organization. *Mol. Cell* **62**, 668–680 (2016).
32. Münkler, C. *et al.* Compartmentalization of interphase chromosomes observed in simulation and experiment. *J. Mol. Biol.* **285**, 1053–1065 (1999).
33. Rao, S. S. *et al.* A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
34. Sachs, R., Van Den Engh, G., Trask, B., Yokota, H. & Hearst, J. A random-walk/giant-loop model for interphase chromosomes. *Proc. Natl. Acad. Sci.* **92**, 2710–2714 (1995).
35. Wang, S. *et al.* Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016).
36. Yokota, H., Van Den Engh, G., Hearst, J. E., Sachs, R. K. & Trask, B. J. Evidence for the organization of chromatin in megabase pair-sized loops arranged along a random walk path in the human g0/g1 interphase nucleus. *J. Cell. Biol.* **130**, 1239–1249 (1995).
37. Le, T. B., Imakaev, M. V., Mirny, L. A. & Laub, M. T. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* **342**, 731–734 (2013).
38. Postow, L., Hardy, C. D., Arsuaga, J. & Cozzarelli, N. R. Topological domain structure of the escherichia coli chromosome. *Genes & development* **18**, 1766–1779 (2004).
39. Fudenberg, G. *et al.* Formation of chromosomal domains by loop extrusion. *Cell reports* **15**, 2038–2049 (2016).
40. Goloborodko, A., Marko, J. F. & Mirny, L. A. Chromosome compaction by active loop extrusion. *Biophys. J.* **110**, 2162–2168 (2016).
41. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci.* **112**, E6456–E6465 (2015).
42. Barbieri, M. *et al.* Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl. Acad. Sci.* **109**, 16173–16178 (2012).
43. Everaers, R., Grosberg, A. Y., Rubinstein, M. & Rosa, A. Flory theory of randomly branched polymers. *Soft matter* **13**, 1223–1234 (2017).
44. Jost, D., Carrivain, P., Cavalli, G. & Vaillant, C. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res.* **42**, 9553–9561 (2014).
45. Schwarzer, W. *et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51 (2017).
46. Rabl, C. Über zelltheilung. *Morphologisches Jahrbuch* **10**, 214–330 (1885).
47. Fang, Y. & Spector, D. L. Centromere positioning and dynamics in living arabidopsis plants. *Mol. Biol. Cell.* **16**, 5710–5718 (2005).
48. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the drosophila genome. *Cell* **148**, 458–472 (2012).
49. Jin, Q.-W., Fuchs, J. & Loidl, J. Centromere clustering is a major determinant of yeast interphase nuclear organization. *J. Cell. Sci.* **113**, 1903–1912 (2000).
50. Jin, Q.-W., Trelles-Sticken, E., Scherthan, H. & Loidl, J. Yeast nuclei display prominent centromere clustering that is reduced in nondividing cells and in meiotic prophase. *J. Cell. Biol.* **141**, 21–29 (1998).
51. Therizols, P., Duong, T., Dujon, B., Zimmer, C. & Fabre, E. Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proc. Natl. Acad. Sci.* **107**, 2025–2030 (2010).
52. Bystricky, K., Heun, P., Gehlen, L., Langowski, J. & Gasser, S. M. Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc. Natl. Acad. Sci. USA* **101**, 16495–16500 (2004).
53. Berger, A. B. *et al.* High-resolution statistical mapping reveals gene territories in live yeast. *Nat. Methods* **5**, 1031 (2008).
54. Buttrick, G. J. *et al.* Nsk1 ensures accurate chromosome segregation by promoting association of kinetochores to spindle poles during anaphase b. *Mol. Biol. Cell.* **22**, 4486–4502 (2011).
55. Strecker, J. *et al.* Dna damage signalling targets the kinetochore to promote chromatin mobility. *Nat. Cell Biol.* **18**, 281 (2016).
56. Halverson, J. D., Smrek, J., Kremer, K. & Grosberg, A. Y. From a melt of rings to chromosome territories: the role of topological constraints in genome folding. *Reports on Progress in Physics* **77**, 022601 (2014).
57. Polovnikov, K., Nechaev, S. & Tamm, M. Effective hamiltonian of topologically stabilized polymer states. *Soft matter* **14**, 6561–6570 (2018).
58. Zimmer, C. & Fabre, E. Principles of chromosomal organization: lessons from yeast. *J. Cell. Biol.* **192**, 723–733 (2011).
59. Shao, Y. *et al.* Creating a functional single-chromosome yeast. *Nature* **560**, 331 (2018).
60. Tjong, H., Gong, K., Chen, L. & Alber, F. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Res.* **22**, 1295–1305 (2012).
61. Capurso, D., Bengtsson, H. & Segal, M. R. Discovering hotspots in functional genomic data superposed on 3d chromatin configuration reconstructions. *Nucleic Acids Res.* **44**, 2028–2035 (2016).
62. Duan, Z. *et al.* A three-dimensional model of the yeast genome. *Nature* **465**, 363 (2010).
63. Lee, C.-S. *et al.* Chromosome position determines the success of double-strand break repair. *Proc. Natl. Acad. Sci.* **113**, E146–E154 (2016).
64. Fuller, F. B. Decomposition of the linking number of a closed ribbon: a problem from molecular biology. *Proc. Natl. Acad. Sci.* **75**, 3557–3561 (1978).
65. Orlandini, E., Tesi, M. & Whittington, S. Polymer entanglement in melts. *J. Phys. A. Math. Gen.* **33**, L181 (2000).
66. Panagiotou, E. The linking number in systems with periodic boundary conditions. *J. Comput. Phys.* **300**, 533–573 (2015).
67. Turaev, V. *et al.* Knotoids. *Osaka Journal of Mathematics* **49**, 195–223 (2012).
68. Millett, K. C., Rawdon, E. J., Stasiak, A. & Sułkowska, J. I. Identifying knots in proteins. *Biochemical Society Transactions.* **41**, 533–537 (2013).
69. Akhtar, A. & Gasser, S. M. The nuclear envelope and transcriptional control. *Nat. Rev. Genet.* **8**, 507 (2007).
70. Funabiki, H., Hagan, I., Uzawa, S. & Yanagida, M. Cell cycle-dependent specific positioning and clustering of centromeres and telomeres in fission yeast. *J. Cell. Biol.* **121**, 961–976 (1993).
71. Taddei, A. & Gasser, S. M. Structure and function in the budding yeast nucleus. *Genetics* **192**, 107–129 (2012).
72. Segal, M. R., Xiong, H., Capurso, D., Vazquez, M. & Arsuaga, J. Reproducibility of 3d chromatin configuration reconstructions. *Biostatistics* **15**, 442–456 (2014).
73. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)* 289–300 (1995).
74. Imakaev, M. V., Tchourine, K. M., Nechaev, S. K. & Mirny, L. A. Effects of topological constraints on globular polymers. *Soft matter* **11**, 665–671 (2015).
75. Klenin, K. & Langowski, J. Computation of writhe in modeling of supercoiled dna. *Biopolymers* **54**, 307–317 (2000).
76. Kirkman, T. Statistics to use: Kolmogorov-smirnov test. *College of Saint Benedict and Saint Johns University. Retrieved October 7, 2008* (1996).

77. Cornforth, M. N. *et al.* Chromosomes are predominantly located randomly with respect to each other in interphase human cells. *J. Cell. Biol.* **159**, 237–244 (2002).
78. Frank-Kamenetskii, M., Lukashin, A., Anshelevich, V. & Vologodskii, A. Torsional and bending rigidity of the double helix from data on small dna rings. *Journal of Biomolecular Structure and Dynamics* **2**, 1005–1012 (1985).
79. Jayaram, M. *et al.* Topological similarity between the 2  $\mu$ m plasmid partitioning locus and the budding yeast centromere: evidence for a common evolutionary origin? *Biochemical Society Transactions* **41**, 501–507 (2013).
80. Mortimer, R. K. & Schild, D. Genetic map of *Saccharomyces cerevisiae*. *Microbiol. Rev.* **44**, 519 (1980).
81. Varela, R., Hinson, K., Arsuaga, J. & Diao, Y. A fast ergodic algorithm for generating ensembles of equilateral random polygons. *Journal of Physics A: Mathematical and Theoretical* **42**, 095204 (2009).
82. Arsuaga, J., Blackstone, T., Diao, Y., Karadayi, E. & Saito, M. Linking of uniform random polygons in confined spaces. *Journal of Physics A: Mathematical and Theoretical* **40**, 1925 (2007).
83. Kinney, N. A., Sharakhov, I. V. & Onufriev, A. V. Chromosome–nuclear envelope attachments affect interphase chromosome territories and entanglement. *Epigenetics & Chromatin* **11**, 3 (2018).
84. Llu, L., Perkoča, L., Calendar, R. & Wang, J. Knotted dna from bacteriophage capsids. *Proc. Acad. Sci. USA* **78**, 5498–5502 (1981).
85. Jensen, R. E. & Englund, P. T. Network news: the replication of kinetoplast dna. *Annu. Rev. Microbiol.* **66**, 473–491 (2012).
86. Micheletti, C., Marenduzzo, D., Orlandini, E. & Summers, D. Knotting of random ring polymers in confined spaces. *J. Chem. Phys.* **124**, 064903 (2006).
87. Diao, Y., Hinson, K., Kaplan, R., Vazquez, M. & Arsuaga, J. The effects of density on the topological structure of the mitochondrial dna from trypanosomes. *J. Math. Biol.* **64**, 1087–1108 (2012).
88. Grainge, I. & Sherratt, D. J. Site-specific recombination. In *Molecular Genetics of Recombination*, 443–467 (Springer, 2006).
89. Pommier, Y., Sun, Y., Shar-yin, N. H. & Nitiss, J. L. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev. Mol. Cell Biol.* **17**, 703 (2016).
90. Shimokawa, K., Ishihara, K., Grainge, I., Sherratt, D. J. & Vazquez, M. Ftsk-dependent xercd-dif recombination unlinks replication catenanes in a stepwise manner. *Proc. Natl. Acad. Sci.* **110**, 20906–20911 (2013).
91. Cook, P. R. Predicting three-dimensional genome structure from transcriptional activity. *Nat. Genet.* **32**, 347 (2002).
92. Rosa, A., Becker, N. B. & Everaers, R. Looping probabilities in model interphase chromosomes. *Biophys. J.* **98**, 2410–2419 (2010).

## Acknowledgements

This study was supported by NIH grants R01GM109457 and R01GM075119 and NSF grants DMS-1817156 and DMS-1716987. Figures were done using the program Knotplot.

## Author Contributions

M.P. implemented statistical analysis, B.C. and M.V. implemented biophysical simulations of wormlike chains, M.S. generated CCC reconstructions, S.B. analyzed results and wrote manuscript, J.A. conceived study, designed algorithms, analyzed the results and wrote manuscript. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-42967-4>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019