Physical tethering and volume exclusion determine higher-order genome organization in budding yeast

Harianto Tjong[1], Ke Gong[1], Lin Chen[1,2,*], Frank Alber[1,*]

[1]Molecular and Computational Biology, Department of Biological Sciences, University of Southern California, 1050 Childs Way, Los Angeles, CA 90089, USA

[2]Department of Chemistry and Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA

*Correspondence should be addressed to L.C. (linchen@usc.edu) and F.A. (alber@usc.edu)

Supplemental Text

1. Contact Analyses

1.1. Definition of a contact

A bead contact is defined when the center-to-center distance between the two beads is not larger than 1.5 times the sum of both bead radii (i.e. 45 nm).

1.2. Contact frequency matrices

The contact frequency matrix was defined as a K x K matrix, $\mathbf{C}_K = (c_{ij})_{K\times K}$, where K is the number of chromosomal regions and the entry c_{ij} is equal to the total number of contacts between regions i and j in the structure population. Several different contact frequency matrices were generated at variable resolution with different number of K.

Normalization of contact frequency maps

For comparison between experiment and structure population, contact frequency maps from the structure populations and experiment were normalized in a manner previously described in other articles (Duan et al. 2010; Lieberman-Aiden et al. 2009). The normalization is performed as:

$$\overline{\mathbf{C}}_K = (\overline{c}_{ij})_{K\times K}$$

$$\overline{c}_{ij} = c_{ij}\frac{\left(\sum_{k=1}^{K\text{-}1}\sum_{l=k+1}^{K} c_{kl}\right)}{\left(\sum_{k=1}^{K} c_{ik}\right)\left(\sum_{k=1}^{K} c_{kj}\right)}$$

where $\overline{c}_{ij}$ is the normalized contact frequency between segments i and j. The terms in the denominator are the total sums of all contacts in the corresponding rows and columns of the contact matrix. The term in the numerator is the total sum of all contacts in the matrix. For the experimental maps, c_{ij} is equal to the total number of reads observed in sequencing that were mapped to the two corresponding regions i and j in the genome.

Normalized contact frequency maps were generated at different levels of resolution (K) from both the experiment and the structure population.

1.2.1. Genome-wide contact frequency map at 32 kb resolution $\mathbf{C}_{369}^{32\text{kb}} = (c_{ij})_{369\times369}$

The contact frequency map $\mathbf{C}_{369}^{32\text{kb}} = (c_{ij})_{369\times369}$ was defined as a 369 x 369 matrix, where each region was represented by 32 kb of consecutive genome sequence. c_{ij} is equal to the total sum of observed contacts between any one of the beads in one region (i) to any one of the beads of the second region (j).

In the conformation capture experiment contacts between chromatin regions separated by less than 20 kb are not considered (Duan et al. 2010). For comparison, contacts in the structure population were also not considered if the corresponding beads are separated by less than 6 beads in the chain sequence. These contacts were excluded in the following analysis, if not stated otherwise.

The contact frequency heat map for the structure population is shown in **Figure 3E** and the heat map from experiment is shown in **Figure 3F**; both heat maps have been normalized using the method described above.

To assess the reproducibility of our structure population, we divided the population into two groups each with 100,000 structures. For each group contact frequency matrices are constructed and the Pearson's correlation between the two matrices calculated. The Pearson's r = 0.999 (two tailed t-test H_0=the two matrices are uncorrelated, p-value < 10^{-15}) indicating that the contact frequency maps are highly reproducible.

1.2.2. Chromosome arm contact frequency map $\mathbf{C}_{32}^{\text{arm}} = (c_{ij})_{32\times32}$

The chromosome arm contact frequency map $\mathbf{C}_{32}^{\text{arm}} = (c_{ij})_{32\times32}$ describes the number of contacts observed between two chromosome arms and was defined as a 32 x 32 matrix, where c_{ij} is equal to the total sum of all observed bead contacts between chromosome arms i and j.

Experimental data on contacts of regions located within 10 kb from centromere are not available. Correspondingly, for consistency all contacts between beads

located within 3 beads from the centromere were excluded. Finally, the chromosome arm contact frequency map is normalized following the approach described above (Duan et al. 2010). The heat map of the structure population is shown in **Figure 3C** and the heat map from experiment is shown in **Figure 3D**.

1.2.3. Chromosome contact frequency map $\mathbf{C}_{16}^{\text{chrom}} = (c_{ij})_{16\times16}$

The chromosome contact frequency map $\mathbf{C}_{16}^{\text{chrom}} = (c_{ij})_{16\times16}$ contains the number of contacts observed between two chromosomes and was defined as a 16 x 16 matrix, where c_{ij} is equal to the total sum of all observed bead contacts between chromosomes i and j. In order to compare chromosome contact frequencies between the structure population, the experiment, and the random control population, the chromosome contact frequency map is normalized by following procedure: first, each element c_{ij} is normalized with respect to the chromosome lengths of the two corresponding chromosomes i and j leading to the length normalized element w_{ij} defined as

$$w_{ij} = \frac{c_{ij}}{L_i L_j},$$

where L_i is the sequence length of chromosome i. Then w_{ij} is further normalized with respect to the total sum of all normalized contact frequencies,

$$\overline{w}_{ij} = \frac{1}{\sum_{i=1}^{16}\sum_{j=1}^{16} w_{ij}} w_{ij}.$$

The corresponding chromosome contact frequencies for structure population, experiment and random control population are plotted in **Figures 3A and Supplemental Fig. 2.**

1.3. Row-based Pearson's correlation between two contact frequency maps

Let $\overline{\mathbf{C}}_K^{\text{Exp}} = (\overline{c}_{ij}^{\text{Exp}})_{K\times K}$ and $\overline{\mathbf{C}}_K^{\text{P}} = (\overline{c}_{ij}^{\text{P}})_{K\times K}$ be the contact frequency maps from experiment and structure population, respectively. To compare the similarity between the row vectors i, in experiment and structure population, we calculate the Pearson's correlation coefficient of row i between the two matrices as

$$r_i = \frac{n\sum_{j\neq i}^{n} \overline{c}_{ij}^{\mathrm{P}} \overline{c}_{ij}^{\mathrm{Exp}} - \sum_{j\neq i}^{n} \overline{c}_{ij}^{\mathrm{P}} \sum_{j\neq i}^{n} \overline{c}_{ij}^{\mathrm{Exp}}}{\sqrt{n\sum_{j\neq i}^{n} \left(\overline{c}_{ij}^{\mathrm{P}}\right)^2 - \left(\sum_{j\neq i}^{n} \overline{c}_{ij}^{\mathrm{P}}\right)^2} \sqrt{n\sum_{j\neq i}^{n} \left(\overline{c}_{ij}^{\mathrm{Exp}}\right)^2 - \left(\sum_{j\neq i}^{n} \overline{c}_{ij}^{\mathrm{Exp}}\right)^2}},$$

where the summation j runs either over all indices, or only the indices for intra- or inter-chromosomal contacts, depending the type of comparison. Then the average of the correlation coefficients across all regions can indicate how similar the two matrices are.

1.4. Maximal correlation

Several factors may affect the strength of the correlation between experiment and simulation. First of all, the sampling of inter-chromosomal contacts is not complete, both in simulation and experiment. In the following section we analyze what effect a limited fragment sampling would have on the expected maximal possible accuracy for contact detection.

In genome-wide conformation capture experiments a contact is determined by the sequencing of ligated DNA fragments comprised from two different regions in the genome. However, only a relatively small sample of all ligated DNA fragments is used, which is then amplified through PCR and sequenced. Here, we estimate the effect of limited fragment sampling on the expected accuracy for contact detection. The accuracy of a prediction is calculated as the cross-correlation between two contact frequency heat maps: first, the ground truth, which is defined as the contact frequency map at 3.2 kb resolution determined from the complete structure population; and second, maps that are generated from randomly sampling only a fraction of all contacts in the structure population. The exact number of ligated DNA fragments sampled in experiment is unknown. Instead, we approximate the number and types of DNA fragments with the contacts observed in the structure population. The structure population contains a total 826 million contacts in the 200,000 structures. We then randomly pick a certain fraction (i.e. 0.0001, 0.001, 0.01, or 0.1) of contacts from the pool of all contacts in the structure population (i.e. as the hypothetical “reads” in a 3C-based experiment) and build the corresponding ‘sampled’ contact frequency matrices at 32-kb resolution. The Pearson

correlations between the "sampled matrices" and the "true matrix" provide a first estimate of the expected error due to limited sampling (**Supplemental Fig. 3**). This estimate is only a lower bound of the error and the generated cross-correlation value is therefore referred to as the "maximal" expected cross-correlation, because in this ideal simulated case effects due to experimental errors and limitations, such as those sequenced regions that cannot be mapped unambiguously to the genome, false positive reads and also the bias due to PCR amplifications are not taken into account.

1.5. Contact similarity between structures in the population

To assess the contact similarities between individual structures in the population we measure the overlap of the contacts between all structures. A binary vector v^i is defined for structure i containing the list of all contacts (except direct neighbors in the chain). If a contact is present in a structure the corresponding vector element is set to 1, and 0 otherwise. An overlap index O_{ij} between a structures i and j is then calculated as

$$O_{ij} = \frac{\sum_{k=1}^{N} v_k^i v_k^j}{\sum_{k=1}^{N} \min\left(1, v_k^i + v_k^j\right)}$$

where v^i_k is the contact vector for structure i for the k-th bead pair; and the total number of inter- and intra-chromosomal bead pair is N=7,134,769. O_{ij} is the fraction of the total number of vector elements where both structures carry 1, divided by the total number of vector positions that at least one structure carries 1.

We calculated the contact overlaps for a sample of 5,000 structures from the population yielding 12.497 million pairs of structures. As expected from a highly heterogeneous population we found that, on average, structures in our population had 0.22% and 0.01% intra-and inter-chromosomal contact overlaps, respectively.

2. Localization Probability Density (LPD) of Gene and Chromosome

To generate the gene 2D localization probability density (LPD) map from the structure population, we collected all the 3D coordinates, (x',y',z') for the beads of interest (e.g. a

particular gene or a group of beads representing a chromosome) and then projected them into a 2D coordinate reference frame (Berger et al. 2008),

$$(z, \rho)=(x', \sqrt{y'^2 + z'^2})$$

Density grid projection: The grid size was chosen as Δ = 10 nm resulting in a 2D grid of 200×200 pixels representing a 2×2 μm map. Once a point (z_c,ρ_c) is mapped onto the grid, a Gaussian blur is applied centered at the pixel following the expression

$$G_{ij} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z_c - z_i)^2 + (\rho_c - \rho_j)^2}{2\sigma} \right)$$

where z_c and ρ_c denote the center pixel in z- and ρ-axis, respectively, (i,j) denote the neighboring pixels and σ = 30nm.

Normalization: The density is normalized with respect to the radial volume around the z-axis. The normalized density value G_{ij} of voxel ij is defined as

$$\bar{G}_{ij} = \frac{G_{ij}}{\pi\Delta\left(\left(\rho_j + \Delta \right)^2 - \rho_j^2 \right)}$$

where Δ is the grid size. Finally, all $\bar{G}_{ij}$ are divided by a constant so that the maximum value is 1.

In the contour plots shown in **Figures 2, 5, and Supplemental Fig. 1** the bottom half of the 2D density plot is mirrored from the top half for visual convenience.

3. Additional Structure Populations

Besides the random control population, we have generated the following populations.

3.1. Structure populations of single chromosomes

In order to estimate the influence of excluded volume effects on chromosome location and chromosome contacts**,** we have performed a simulation of a nucleus that contains only a single chromosome (i.e., the target chromosome) (**Figure 2C and Supplemental Fig. 1**). Simulations of a single chromosome were performed

with identical setup to the complete landmark model except without the presence of all other chromosomes.

3.2. Structure population with modified chromosome 12 nucleolus constraints

We have also generated a structure population with a modified rDNA position constraint for chromosome 12. In this constraint the first 3 rDNA beads are allowed as far away as 850 nm from the surface of nucleolus instead of being constraint to the surface of the nucleolus. A structure population is calculated with 100,000 structures. The inter-chromosomal contact frequency correlation is improved from 0.54 to 0.58 in comparison to the initial structure population (see **Supplemental Fig. 4**)

4. Change Point Analysis

To detect the change of slope in the linear regression analyses on the telomere-telomere distances, we employ a change point analysis (Zeileis et al. 2003) following the approach previously reported (Therizols et al. 2010). The optimal number of change points is determined according to the derivatives of the Bayesian information criterion (Zeileis et al. 2003), then a linear regression is performed to fit each segment raised by the change points. We set the minimal length of a segment as 30% of the total data points and only 1 change point is detected for every reference telomere (**Figure 6 and Supplemental Fig. 5**)

5. Spatial Clustering of Functionally Related Genes or Loci

We analyzed the spatial distributions of three groups of functional related loci, namely all early replication sites, all late replication sites, and all tRNA gene loci in the genome (**Figure 7**). We asked whether the pair distances between the sites in each group are significantly different to the distances between randomly chosen sites in the genome. More specifically, for each structure in the population we calculated the mean of the pair distances between all sites in a group. This mean pair distance is normalized by the mean pair distance of all beads in each structure. We then determined the distribution of

all mean values from the 100,000 structures sampled in the population (shown with each histogram in **Figure 7**).

5.1. Statistical assessment

We also generated a mean pair distance distribution using the same number of randomly selected sites from the same structure.

For testing if the shift in the mean values between both distributions is statistically significant we used the paired t-test (e.g. H_0: the means are the same; H_1: mean of early/late replication sites is less/greater than that of random background) and the p-value is collected. The randomization is repeated k = 1000 times and the corresponding 1000 p-values are combined using Stouffer's Z-transform test (Stouffer et al. 1949; Whitlock 2005). Each p-value (p_i) is transformed to a standard normal z-score (z_i) and the cumulative evidence of the common null hypothesis can be inferred from Z_S which is determined using the following formula,

$$Z_S = \frac{\sum_{i=1}^{k} z_i}{\sqrt{k}}$$

where k is the number of independent tests.

The groups include early replication origins, late replication origins, two categories of replication origins, and tRNAs genes. The number of loci in each group are as follows: 1) CDR-based replication sites (McCune et al. 2008): 77 early start sites and 123 late start sites; 2) Rad53-based replication sites: 101 early replication sites and 99 late replication sites (Feng et al. 2006); 3) GINS-based replication sites: 169 category-1 sites, and 135 category-2 sites (Sekedat et al. 2010); 4) 275 tRNA sites.

5.2. Centromere bias test

The analysis was repeated by excluding early replication origins that are located within 100 kb distance from the centromeres. The results and conclusion did not change, i.e. the early replication origins have a statistically significant lower mean pair distance. This analysis is also applied to the late replication origins where only sites located within 300 kb from centromere are included and the same conclusion

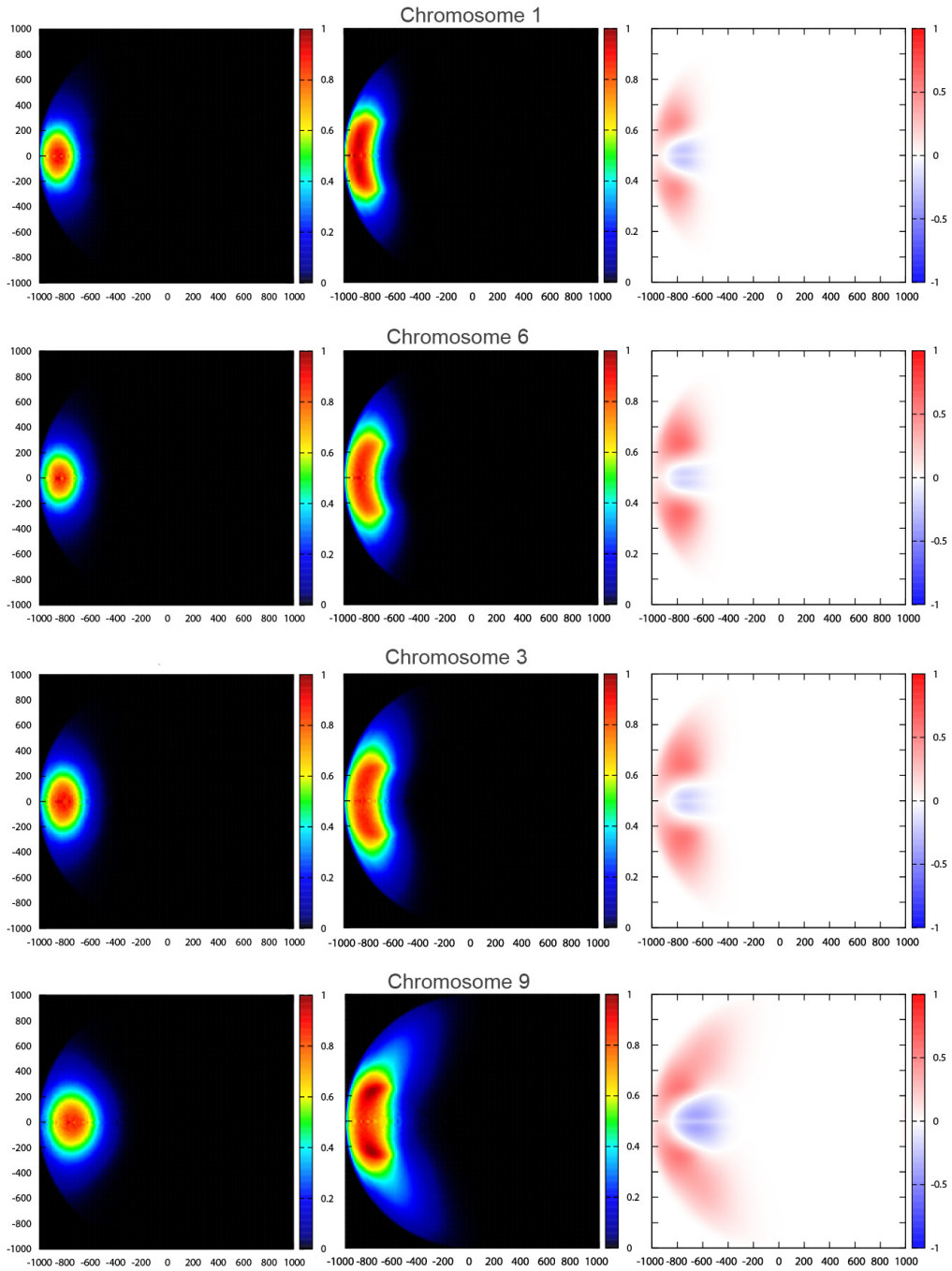
still holds, i.e. late replication origins have a statistically significant larger mean pair distance than randomly selected loci.

5.3. Early versus late replication origins

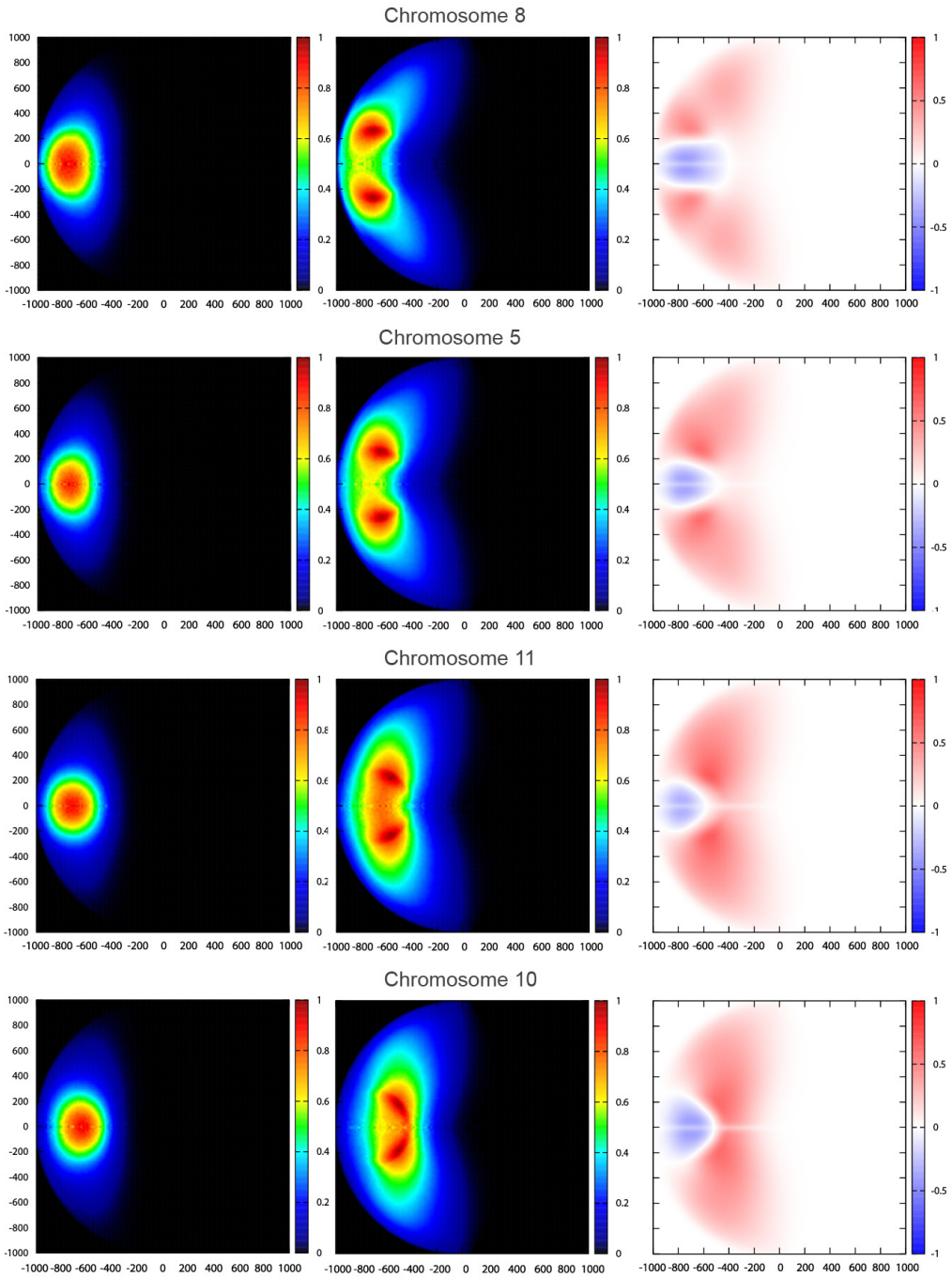
We compared the sequence distance to the centromere between early and late replication origins. Each start site is measured by the sequence distance to centromere normalized by the length of the chromosome arm it is located; 0 and 1 mark the locus at centromere and telomere, respectively. The distributions of early and late replication origins were collected for all 3 data source aforementioned, and Wilcoxon signed-rank test was performed for each data set. The null hypothesis, namely that the sets of early and late replication sites are equally distributed with respect to the centromere, was rejected with all p-values $< 10^{-5}$ for the 3 data set, and the alternative hypothesis that early replication origins on average are closer to centromere than the late origins was accepted.

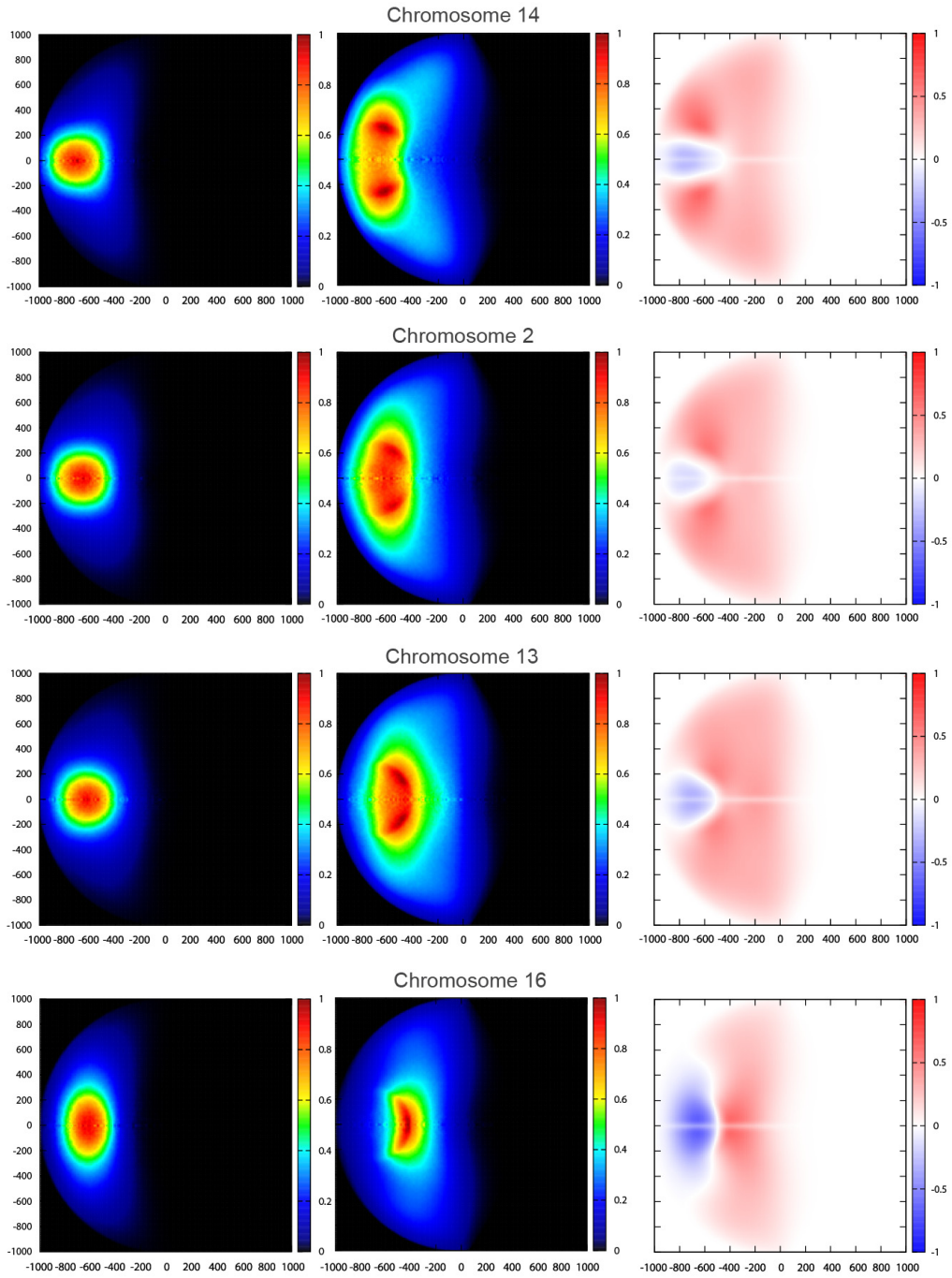
Supplemental Figures

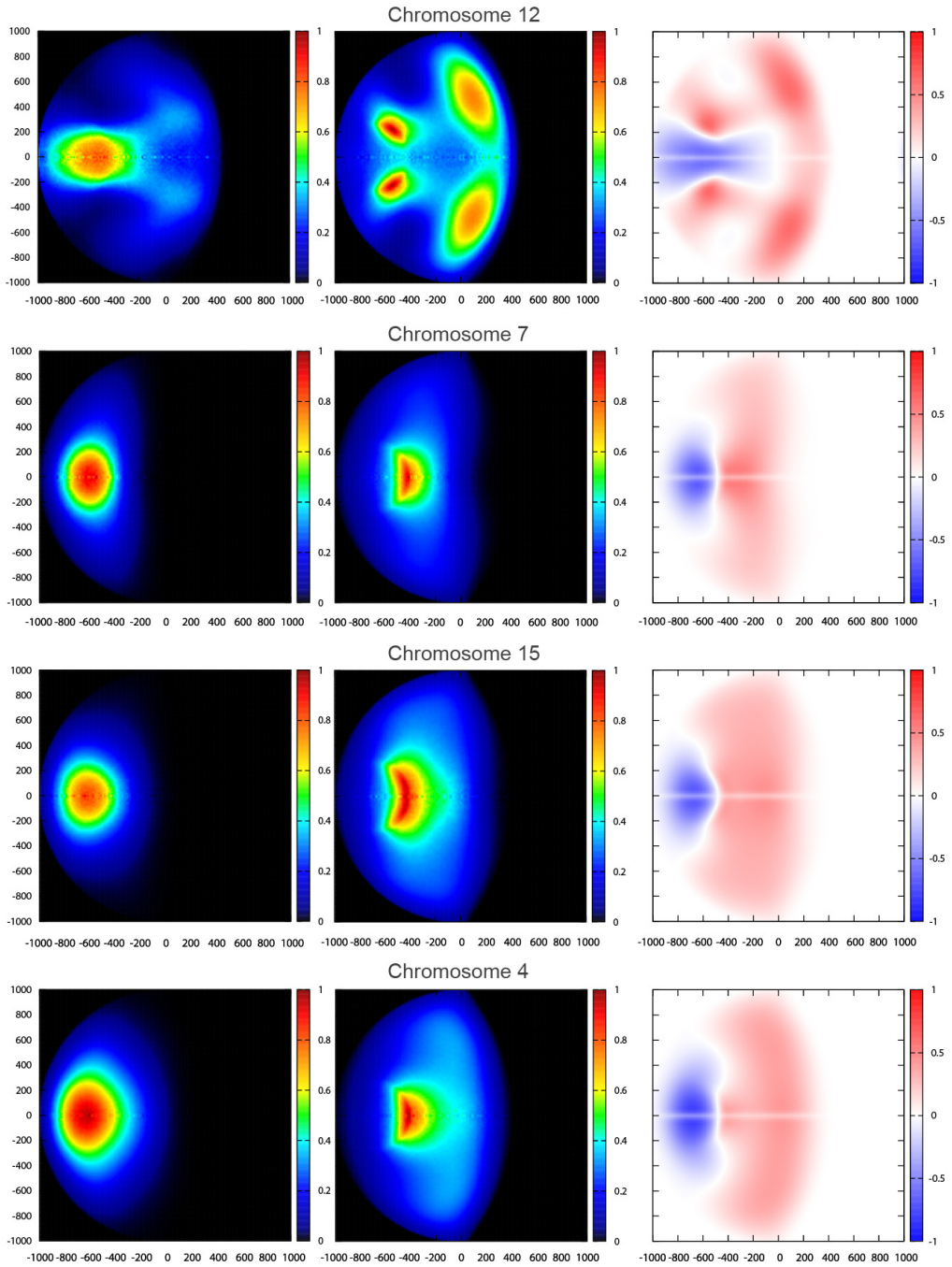
Supplemental Figure 1



Supplemental Figure 1 (continued)

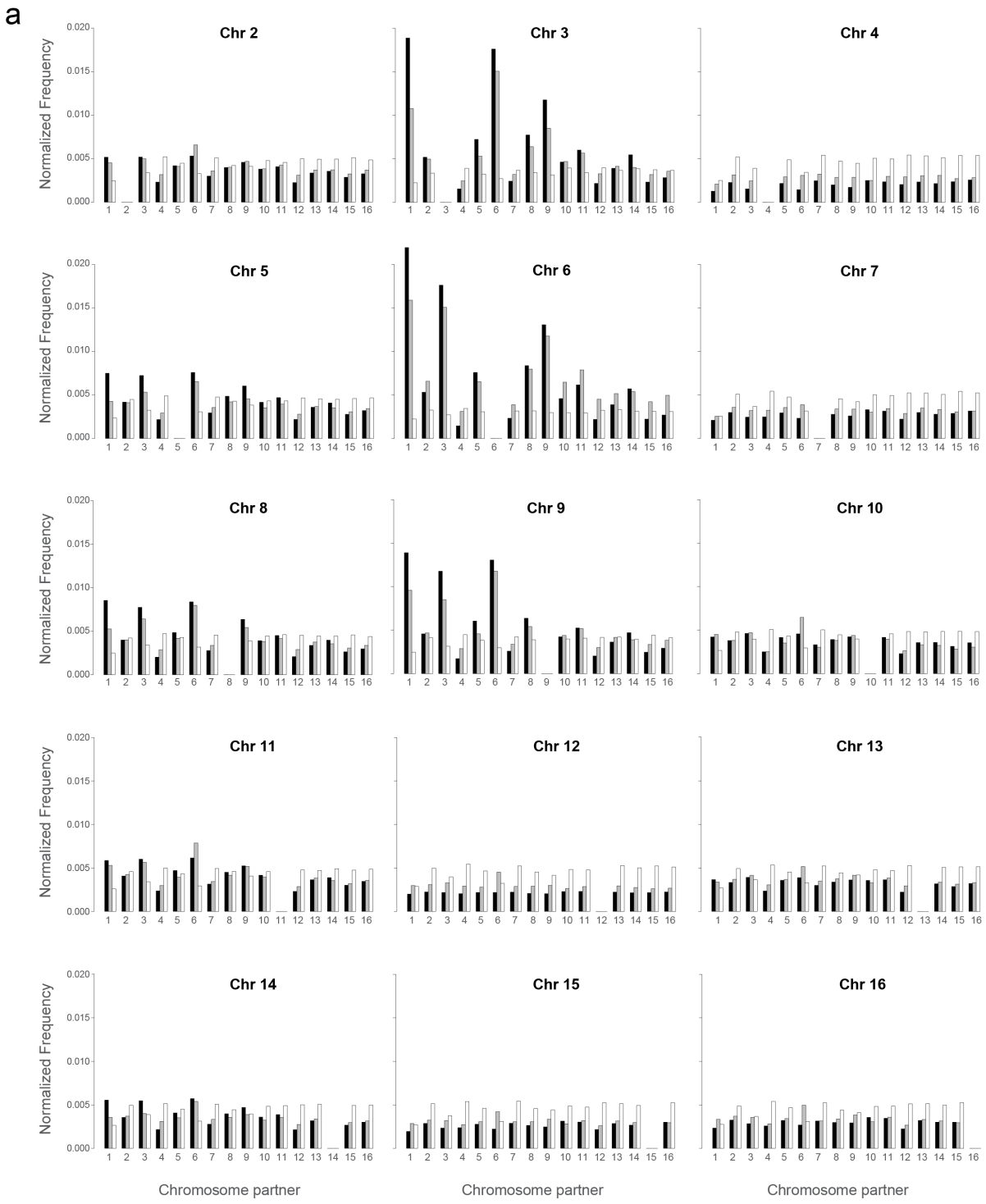


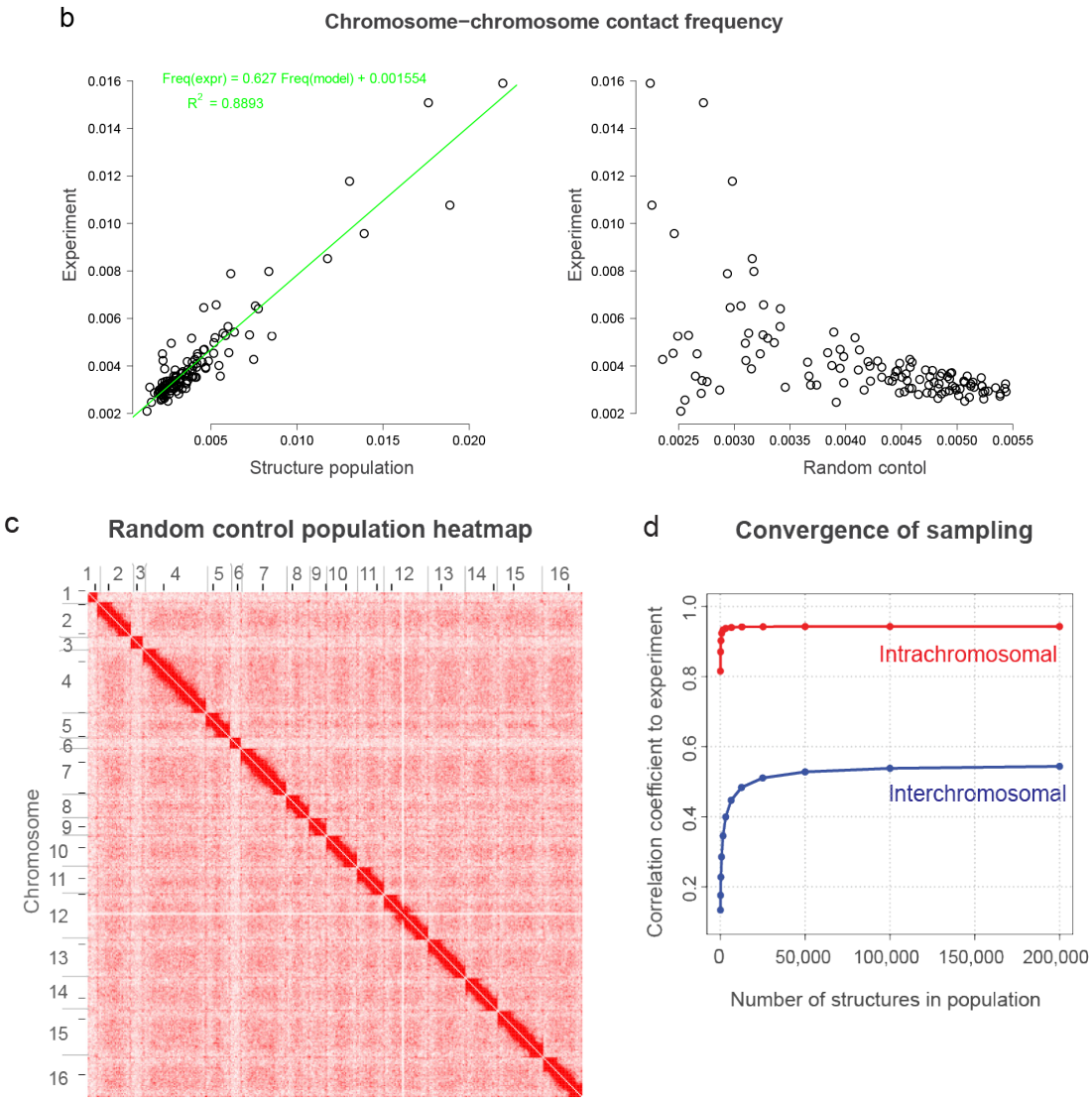

Chromosome 14
Chromosome 2
Chromosome 13
Chromosome 16



Supplemental Fig. 1 Exclusion volume effect for constrained chromosomes. (Left panels) Localization probabilities from the single chromosome structure populations. The chromosome is subject to all landmark constraints but without the presence of the other 15 chromosomes in the nucleus. The plot conventions are the same as in **Figure 2B**. (Middle panels) The localization probability density of the same chromosome from the structure population when all other 15 chromosomes are present in the nucleus. The two density distributions are significantly different. (Right panels) The difference maps between the maps in the left and middle panel. The figure panels, from top to bottom, are ordered by the chromosome size.

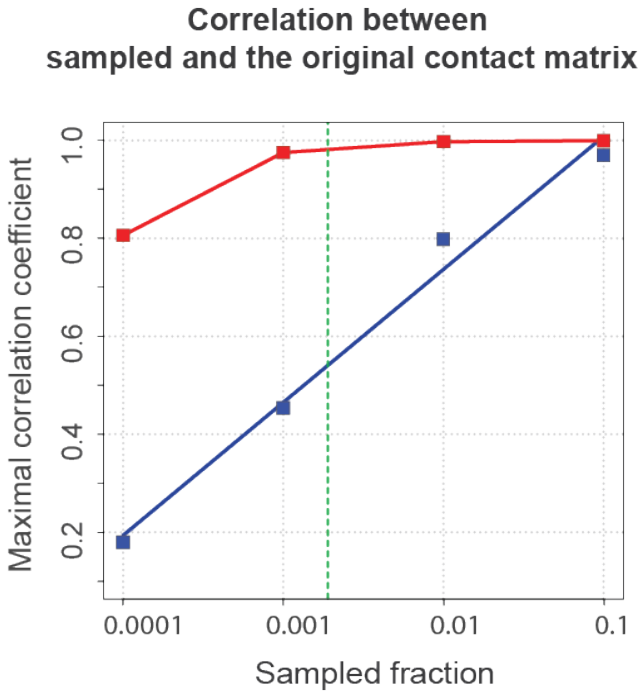
Supplemental Figure 2





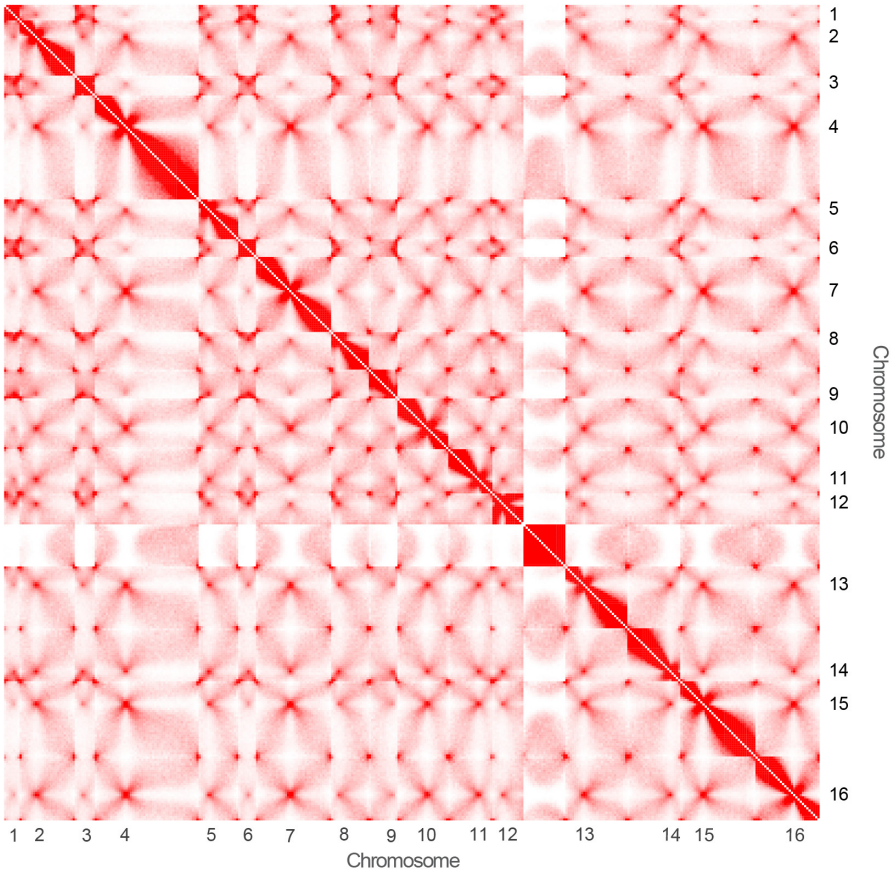
Supplemental Fig. 2 (a) Chromosome-chromosome interaction profiles for chromosomes 2 to 16 from the structure population (black), experiment (light grey), and random control population (white). (b) Comparison of chromosome-chromosome contact frequencies from the genome-wide conformation capture experiments (Duan et al. 2010) with those from the structure population (left panel) and from random control population (right panel). (See **Subsection 1.2.3** in the **Supplemental Text** for details.) (c) The contact frequency heat map of the random control population at 32 kb resolution. (d) The correlation of contact frequency heat maps between the structure population and experiment with increasing number of sampled structures in population. The red curve shows the average row-based Pearson's correlations for only intra-chromosomal contacts while the blue curve shows the average row-based Pearson's correlations between the heat maps for only inter-chromosomal contacts. The standard deviations are smaller than the size of dots. The corresponding samples are generated by randomly selecting structures from the pool of 200,000 structures in the complete structure population. The experiment data in panels a, b, and d are from (Duan et al. 2010).

Supplemental Figure 3



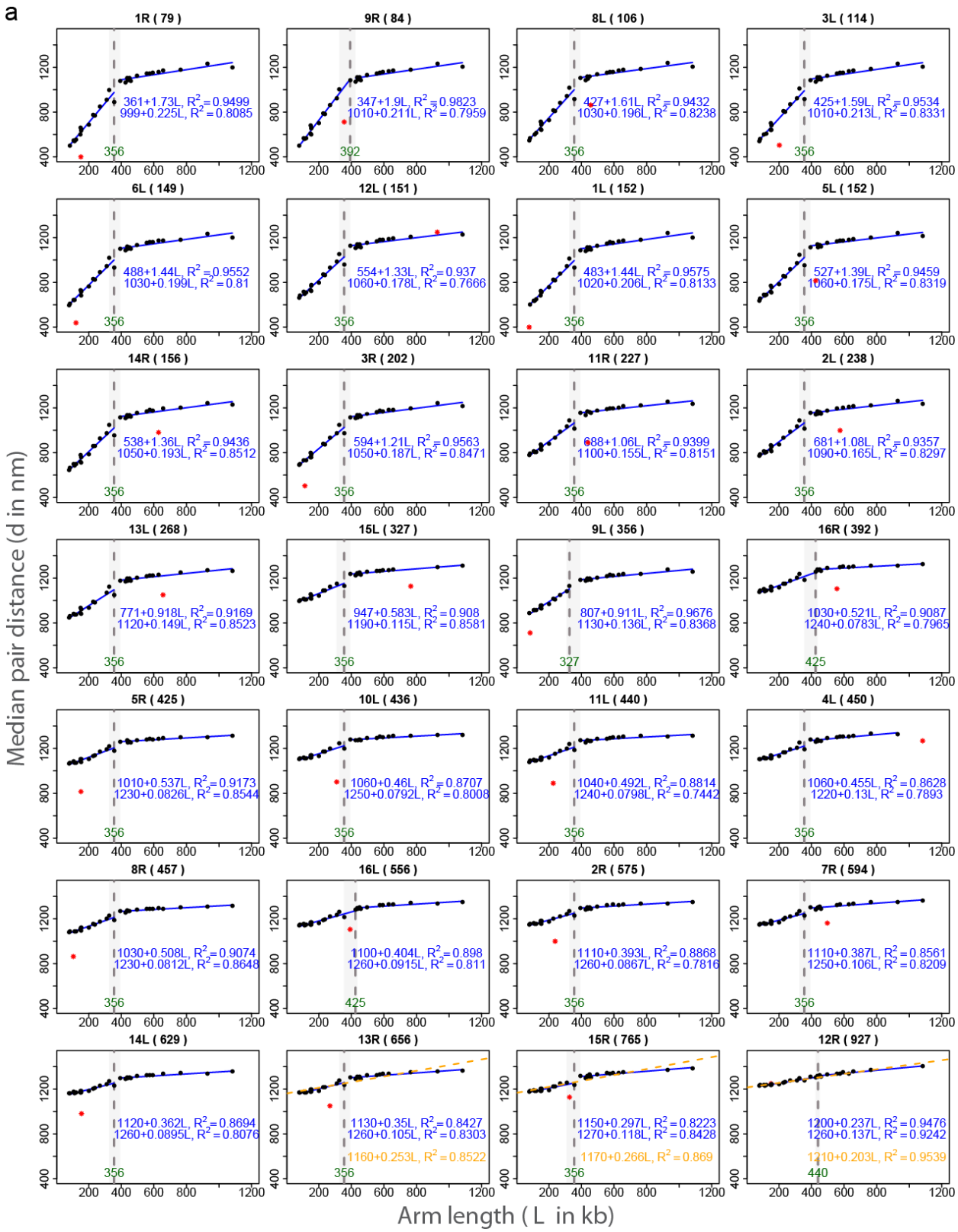
Supplemental Fig. 3 Expected maximal Pearson's correlation between a contact frequency heat map (generated from the structure population) and heat maps that are generated from randomly sampling different portions of contacts from the structure population (see text and **Supplemental material**). The resulting average correlations for only intra-chromosomal contacts are shown in red, and correlation values for only inter-chromosomal contacts are shown in blue. The correlations between intra-chromosomal contact patterns are strong (above 0.95), even at the relatively small sampling rate of 0.1% (**Supplemental Fig. 2d**). In contrast, the correlations between inter-chromosomal contact patterns are greatly affected by sampling error. This is to be expected as there are many more potential inter-chromosomal contacts and each one of them occurs more rarely in the sampling than intra-chromosomal contacts. At a sampling rate of 0.1%, we find that Pearson's correlation between the two inter-chromosomal contact maps (even when assuming an ideal physical model) cannot exceed 0.5. This value is in a similar range as the correlation observed between our structure population and experiment. In our analysis the observed correlation value of 0.54 corresponds to a sampling rate of ~0.2% (green dashed line), which is also the order or magnitude that is expected for the experiment. For example, an experimental sample contains typically tens of millions of cells, which each can be expected to contain hundreds to thousands of chromatin contacts, however, the sequenced non-ambiguous reads, which are typically also enriched by PCR are roughly in the ~5 million range (Duan et al. 2010). Thus, the observed correlation coefficient of 0.54 represents a remarkably good agreement between the inter-chromosomal contact patterns, given that the experimental and computational samplings are finite and cannot be exhaustive.

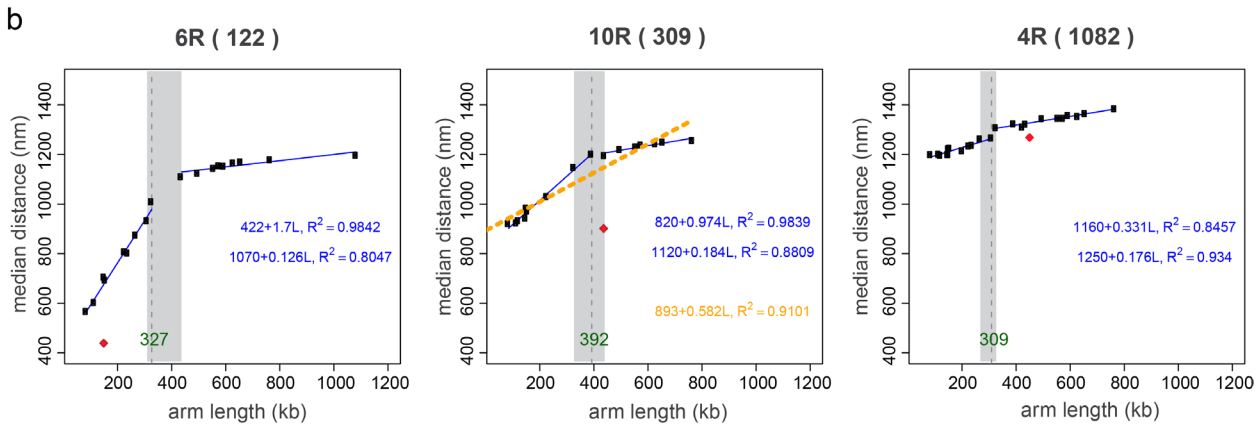
Supplemental Figure 4



Supplemental Fig. 4 Contact frequency heat map from a structure population generated with corrected nucleolar localization constraints (see **Subsection 3.2** of the **Supplemental Text** for description of the constraints). In particular contacts of the small arm of chromosome 12 are more similar to experiment (Duan et al. 2010) (**Figure 3F**). The average region based Pearson's correlation for intra-chromosomal contacts is 0.94, and for inter-chromosomal contacts is 0.58. This value is increased in comparison to the correlation generated from the initial structure population.

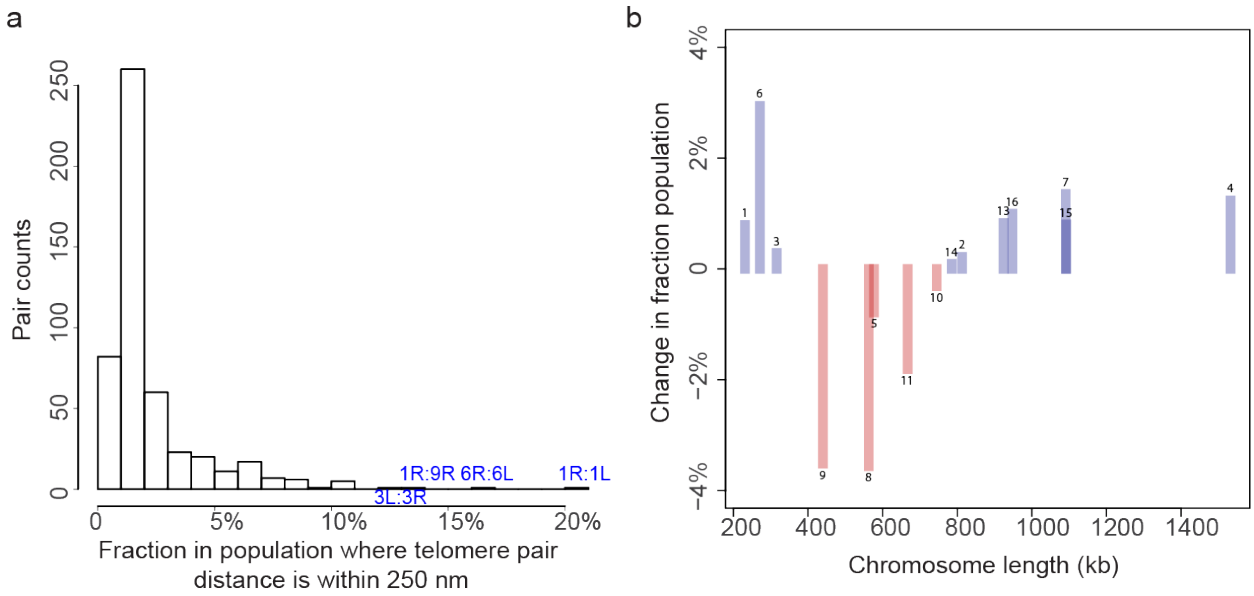
Supplemental Figure 5





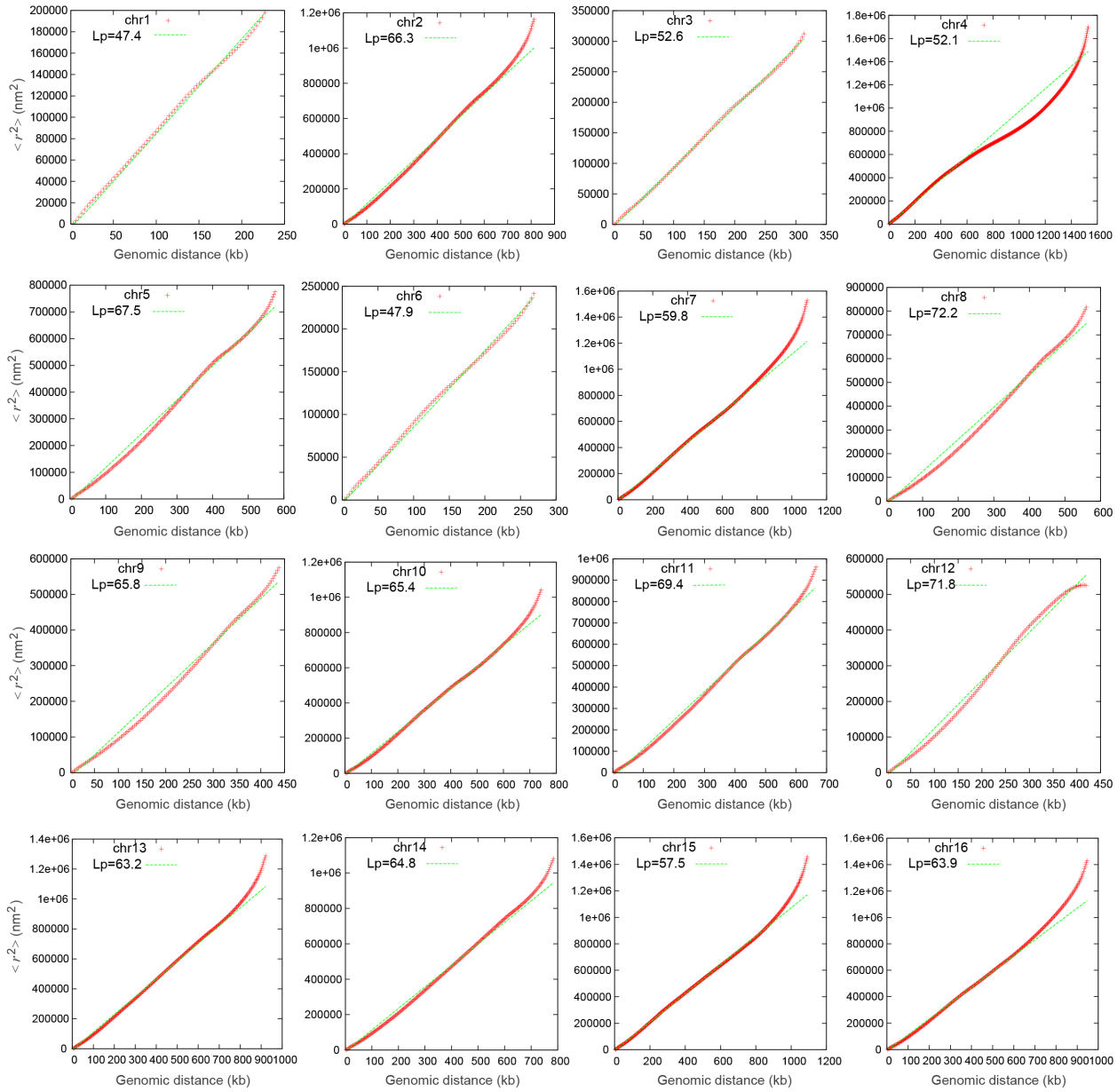
Supplemental Fig. 5 Median telomere-telomere distances in the structure population. (a) The median distances between a telomere of a reference chromosome arm and all other telomeres. The reference chromosome arm and its length (in kb) are indicated at the top of each panel. The vertical dashed line marks the slope change point at 95% confidence interval shown by the shaded area. Red point represents the other telomere on the same chromosome. (b) For comparison with experiment the median telomere-telomere distances are also analyzed using only the subset of chromosomes for which experimental evidence is available (Therizols et al. 2010). The analysis is done for three reference chromosomes: a small, a medium sized and a large chromosome (see comparison to **Figure 6** in the main manuscript). The orange line in the middle panel is a fitted linear regression of all points (without a break point) with the equation shown.

Supplemental Figure 6



Supplemental Fig. 6 Telomere pair co-location frequency and the influence of the excluded volume effect on the telomere pair co-location frequency. (a) Histogram of the pair counts for the fraction of telomere pairs separated by less than 250 nm. Small distances between telomere pairs are infrequent, most of them occur in less than 3% of structures in the population. The most frequent telomere pair association is 1R:1L in about 20% structures in the population. (b) Influence of the excluded volume effect on the telomere co-localization probability. For each chromosome the fraction of models in the population is measured, which have the telomeres on the same chromosome being co-located (telomere co-location defined as in (a)). For each chromosome this fraction is calculated from the structure population containing all chromosomes and the “single chromosome population”, without the presence of all other chromosomes. The spatial competition of all other chromosomes (i.e., the excluded volume effect) dramatically alters the probability of observing co-location of the two telomeres. The histogram shows the relative changes in the fraction of structures with co-located telomeres. Interestingly, the volume exclusion effect causes opposing effects depending on the chromosome length. For the small chromosomes (1, 3, 6) and also the large chromosomes (4, 7, 13, 15, and 16) the volume exclusion effect increases significantly the co-location frequency, while for medium sized chromosomes (5, 8, 9, 10, and 11) the opposite is observed: the volume exclusion effect leads to a dramatic decrease in the telomere co-location probability. For instance, the fraction of co-located telomeres increases by almost 20% for chromosome 6 upon the presence of all other chromosomes in the nucleus, while the co-location probability decreases by 60% for chromosome 8. This result emphasizes the important role of the volume exclusion effect and demonstrates how the total number of chromosomes and relative chromosome lengths of all chromosomes influences interactions and locations of individual gene loci.

Supplemental Figure 7



Supplemental Fig. 7 Mean of squared distance as a function of genomic separation, *s*. The worm-like chain fit is shown as the green line with fitted persistence length indicated as Lp. The Kradky-Porod equation (Krakty and Porod 1949), $\langle r^2 \rangle = 2Lp(Lc/Lp-1+e^{-Lc/Lp})$, is used to fit the points where Lc =*s*/*c* is the contour length of the chain using chromosome density *c* = 3.2 kb/30 nm, and r is the distance between two segments of the chain in the structure population.

Supplemental References

Berger, A.B., G.G. Cabal, E. Fabre, T. Duong, H. Buc, U. Nehrbass, J.C. Olivo-Marin, O. Gadal, and C. Zimmer. 2008. High-resolution statistical mapping reveals gene territories in live yeast. *Nat. Methods* **5:** 1031-1037.

Duan, Z., M. Andronescu, K. Schutz, S. McIlwain, Y.J. Kim, C. Lee, J. Shendure, S. Fields, C.A. Blau, and W.S. Noble. 2010. A three-dimensional model of the yeast genome. *Nature* **465:** 363-367.

Feng, W., D. Collingwood, M.E. Boeck, L.A. Fox, G.M. Alvino, W.L. Fangman, M.K. Raghuraman, and B.J. Brewer. 2006. Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. *Nat. Cell Biol.* **8:** 148-155.

Krakty, O. and G. Porod. 1949. Röntgenuntersuchung gelöster Fadenmoleküle. *Rec. Trav. Chim. Pays-Bas* **68:** 1106-1123.

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

McCune, H.J., L.S. Danielson, G.M. Alvino, D. Collingwood, J.J. Delrow, W.L. Fangman, B.J. Brewer, and M.K. Raghuraman. 2008. The temporal program of chromosome replication: genomewide replication in clb5Δ Saccharomyces cerevisiae. *Genetics* **180:** 1833-1847.

Sekedat, M.D., D. Fenyo, R.S. Rogers, A.J. Tackett, J.D. Aitchison, and B.T. Chait. 2010. GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Mol. Syst. Biol.* **6:** 353.

Stouffer, S.A., E.A. Suchman, L.C. DeVinney, S.A. Star, and R.M.J. Williams. 1949. *The American soldier: adjustment during army life*. Princeton University Press, Princeton.

Therizols, P., T. Duong, B. Dujon, C. Zimmer, and E. Fabre. 2010. Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proc. Natl. Acad. Sci. U.S.A.* **107:** 2025-2030.

Whitlock, M.C. 2005. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J. Evol. Biol.* **18:** 1368-1373.

Zeileis, A., C. Kleiber, W. Krämer, and K. Hornik. 2003. Testing and dating of structural changes in practice. *Comput. Stat. Data Anal.* **44:** 109-123.