# CONTENTS OF SUPPLEMENTARY INFORMATION

## **Supplementary Figures** (page 29-80)

Number of intra-chromosomal interactions per *HindIII* fragment as a function of chromosome size

**Supplementary Figure 8**

Intra-chromosomal interactions in each yeast chromosome.

**Supplementary Figure 9**

Inter-chromosomal interactions between each pair of the 16 yeast chromosomes.

**Supplementary Figure 10**

The mean frequency of inter-chromosomal interactions is significantly higher in each of the experimental libraries relative to the controls, particularly for interactions between centromere regions and those between telomere regions.

**Supplementary Figure 11**

Enrichment of inter-chromosomal interactions with respect to 27 groups of gene loci.

**Supplementary Figure 12**

The two tRNA clusters

**Supplementary Figure 13**

The clusters of early replication origins

**Supplementary Figure 14**

Inverse correlation between the ratio of inter-chromosomal interaction to intra-chromosomal interaction of each chromosome and the number of *HindIII* fragments for each of 16 chromosomes.

**Supplementary Figure 15**

Interaction probability between each chromosome

**Supplementary Figure 16**

Interaction probability between each chromosome arm

**Supplementary Figure 17**

Functions describing the relationship between interaction frequency and spatial distance

**Supplementary Figure 18**

Relationship between interaction frequency and spatial distance of each chromosome in the H-Mp library.

## PDB file of 3D model

3d_model_of yeast_genome.pdb

## Summary of Results

### I. Key technical advances of the method (Figure 1).

1. Our method uses massively parallel sequencing to detect long-range chromosomal interactions on a genome-wide scale, de novo.
2. Our method requires no preexisting knowledge of interacting sequences.
3. The frequency of interactions identified by our method can be used to infer physical distances between interacting segments.
4. Using a series of negative controls we assessed signal to noise ratios, and validated a set of more than 4 million interactions at a false discovery rate of 1%.
5. Our method is compatible with next generation sequencing approaches such as Roche 454 pyrosequencing, Illumina GA, ABIO SOLiD and Helicos' Heliscope.
6. Our method can be used to examine the three dimensional architecture of any genome.

### II. Key findings

1. The intra-chromosomal interaction density and the relationship between interaction frequency and spatial distance are quite similar across chromosomes.
2. Chromosome XII has a unique conformation with a lack of interactions between its distal arms, implicating the nucleolus as a major barrier.
3. Inter-chromosomal interactions are dominated by interactions between centromeres.
4. There is a high degree of interaction between the telomeres of similarly sized chromosome arms.
5. Our findings are consistent with a Rabl configuration, in which centromeres and telomeres occupy opposite poles of the nucleus.
6. Our findings revealed colocalizations of tRNAs, early origins of DNA replication, and chromosomal breakpoints.
7. The yeast chromosome arms are more flexible than their mammalian counterparts.
8. Based on comprehensive mapping of the yeast chromosomal interactions, we constructed a three dimensional model of the yeast genome.

### III. Conclusions

Our experimental approach yields a view of the architecture of the yeast genome with unprecedented resolution.

## Supplementary Methods

### Preparation of yeast cells (cross-linked and uncross-linked)

*Saccharomyces cerevisiae* (genotype: *Mat***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 bar1::KanMX)* were cultured at 30° C with shaking overnight in 50 mls of YEP media plus 2% glucose. Cultured cells were diluted the next morning to an OD660 = .2 in 1 liter of YEP plus 2% glucose. Cells were incubated with shaking at 30° C until reaching an OD660 = 1.0. At this time cells were treated with 27.7 mls of 37% formaldehyde (final concentration = 1%) for 10 minutes with constant stirring. Fixation was quenched with 52.6 mls of 2.5 M glycine (final concentration = 0.125 M) for 15 minutes at room temperature with constant stirring. Fixed cells were collected via centrifugation (1500xg - 5 minutes) and resuspended in 50 mls of spheroplast buffer (1.0 M sorbitol, 100 mM potassium phosphate pH 7.5) plus 30 mM dithiothreitol **(**DTT). Fixed cells were recollected via centrifugation (1500xg - 5 minutes) and resuspended in 50 mls of spheroplast buffer plus 1 mM DTT. Fixed cells were converted to spheroplasts with Zymolyase 20T (MP Biomedicals LLC.) (.66g/L) treatment at 30° C with gentle rotation. Conversion to spheroplasts was confirmed via microscopy and spheroplasts collected via centrifugation at 4° C (1500xg- 5 minutes). Spheroplasts were washed twice in 50 mls spheroplast buffer and collected via centrifugation at 4° C (1500xg- 5 minutes) prior to resuspension in 50 mls restriction enzyme buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT pH 7.9). Un-fixed cells were treated identically without formaldehyde treatment.

Uncross-linked DNA was prepared as described (Miele et al (2006))

### RE1 enzyme digestion and first (3C step) ligation

Fixed spheroplasts (approximately 5 or 10 x $10^9$) were resuspended in 15 mls 1x NEBuffer 2 plus 0.3% SDS and split into 30- 1.5 ml micro-centrifuge tubes. Resuspended spehroplasts were incubated at 65$^o$C for 20 minutes followed by 1 hour at 37$^o$C with shaking. Following incubation 6 μl 10x NEBuffer 2 and 50 μl 20% Triton X-100 were added to each tube, mixed carefully and incubated 37$^o$C for one hour with shaking. Spheroplasted cells were digested overnight with 1200 Units of RE1 (HindIII or EcoRI, NEB) at 37$^o$C with shaking. Restriction enzymes were inactivated by adding 112 μl 10% SDS to each tube and incubating at 65$^o$C for 20 minutes.

For purified yeast genomic DNA, approximately 4 μg of DNA was digested overnight at 37°C with 400 Units of RE1 (HindIII or EcoRI, NEB) in NEBuffer 2. RE1 digestion efficiency was checked by DNA electrophoresis and quantitative real time PCR.

Restriction digested DNA was pooled into 10 equal reactions and each reaction was diluted into 24 ml ligation buffer (66 mM Tris-HCl, pH 7.5; 5 mM DTT; 10 mM MgCl2; 10 mM ATP; 1 mg/ml BSA; 1% Triton X-100). After incubation at 37°C for 1 hour, 250 Units T4 DNA ligase (5 Units/μl, Fermentas) were added per tube and ligation was carried out at 16°C for 4 hours and 25°C for one hour.

**DNA purification**

Following ligation, 200 μl of 20 mg/ml proteinase K and 1 ml 10% SDS was added to each ligation mixture and the tubes were incubated overnight at 65°C. The next day an additional 50 μl 20 mg/ml proteinase K was added to each tube and the incubation was continued at 55°C for another 2 hours. DNA was precipitated with 20 μl GlycoBlue (Ambion), 2.4 ml 3M Na-acetate (pH 5.2) and 24 ml iso-propanol (-80°C- 2 hours). Precipitated DNA was pelleted via centrifugation (4°C- 2 hours 4150 rpm Sorvall 75006441 swinging bucket rotor). Each DNA pellet was dissolved in 1.5 ml 1xTE buffer and transferred to two 1.5 ml microcentrifuge tubes. Each sample was treated with 4 μl 1 mg/ml RNAse A at 37°C for 30 minutes. Following RNA digestion, DNA was extracted three times with 750 μl phenol:chloroform (Invitrogen). Following the third extraction, DNA was precipitated with iso-propanol and washed three times with 70% ethanol. The pellets were air-dried for 10 minutes and dissolved in 60 μl water. The resuspended DNA was pooled and its concentration was determined by measurement on a Nanodrop-1000 spectrophotometer (Thermo Scientific).

**RE2 digestion**

Ligated DNA (~30 μg) was digested overnight (37°C) at a concentration of 10 ng/μl with 1000 Units of RE2 (MspI or MseI, NEB). Following overnight incubation, an additional 100 Units of RE2 was added and incubated at 37°C for 2 additional hours. DNA was precipitated with iso-propanol and purified via the Qiaquick PCR purification kit (Qiagen) according to the manufacturer's instruction. Purified DNA concentration was determined by measurement on a Nanodrop-1000 spectrophotometer.

**Second (4C step) ligation, circular DNA purification and RE1 re-linearization**

For each library, approximately 12 µg RE2-digested DNA was treated with 200 Units T4 DNA ligase (Fermentas, 5U/µl) in a 24 ml reaction overnight at 16°C. Ligated DNA was precipitated with iso-propanol, pelleted via centrifugation and washed with 70% ethanol as described previously. Precipitated DNA was air-dried briefly and dissolved in 657.5 µl water. Linear DNA was then degraded with 100 Units ATP-dependent DNase (10 U/µl, Epicentre) in 1x ATP-dependent DNase Buffer containing 1 mM ATP overnight at 37°C. Circular DNA was purified via the Qiaquick PCR kit column and linearized with RE1

(200 Units corresponding RE1 enzyme (FastDigest® HindIII or EcoRI, Fermentas) in 200 µl 1x FastDigest® buffer 37°C- 30 minutes). Linearized DNA was purified via the Qiaquick PCR purification kit (Qiagen) and DNA concentration was determined by measurement on a Nanodrop-1000 spectrophotometer.

**EcoP15I methylation, EcoP15I adaptor ligation, and Biotin labeling**

Approximately 3ug RE1 re-linearized DNA was methylated overnight at 37°C with 30 Units EcoP15I (NEB) in 1x NEBuffer 3 containing 380 µM S-adenosylmethionine and 1 µg/µl BSA. The DNA fragments were then purified with the Qiaquick PCR kit (Qiagen) and DNA concentration was determined by measurement on a Nanodrop-1000 spectrophotometer.

The methylated DNA was ligated to the corresponding Ecop15I adaptor (Table Sx) at 25°C for 30 minutes in 100µl 1x Fast ligation buffer (Fermentas) containing 400 pmol Ecop15I adaptor and 25 Units T4 DNA ligase (5 U/µl, Fermentas). Ligated DNA fragments were isolated from the excess adaptors via agarose gel electrophoresis and recovered with the Qiaquick Gel Extraction Kit (Qiagen). DNA concentration was determined by measurement on a Nanodrop-1000 spectrophotometer. After ligation with the corresponding EcoP15I adaptor, the RE1 sites were disrupted---a feature which was designed to eliminate the chimerical DNA resulted from random ligation during the above EcoP15 adaptor ligation reaction. The purified DNA was therefore digested by RE1 (HindIII or EcoRI) at 37°C for 2 hours and purified with the Qiaquick PCR kit (Qiagen) and DNA concentration was determined with a Nanodroper 1000.

.

For each library, approximately 0.7 µg adaptor-ligated DNA was circularized in a 750 µl reaction (1x T4 DNA ligase Buffer containing 2.7 pM Biotin labeled adaptor (Supplementary Table 11) and 30 Units T4 DNA ligase (Fermentas) overnight- 16°C). The DNA was precipitated with iso-propanol and purified via the Qiaquick PCR Purification Kit (Qiagen). Linear DNA fragments

were degraded overnight at 37°C with 50 Units ATP-dependent DNase (10 U/µl, Epicentre) in 100 µl 1x ATP-dependent DNase Buffer containing 1 mM ATP. Circular DNA was purified via the Qiaquick PCR kit (Qiagen) and DNA concentration was determined by measurement on a Nanodrop-1000 spectrophotometer.

**EcoP15I digestion, library end-repair and sequencing adaptor ligation**

All subsequent steps were carried out in DNA LoBind tubes (Eppendorf). Circular DNA was digested overnight at 37°C with 20 Units EcoP15I (10Units/µl, NEB) in 100 µl 1x NEBuffer 3 containing 1ug/µl BSA, 2mM ATP, and 100uM Sinefungin . An additional 0.5 µl Ecop15I (10 Units/µl), 0.5 µl 100mM ATP, and 1 µl 10mM Sinefungin was added to the reaction the next day and the incubation was continued for another 2 hours. The EcoP15I was then inactivated by incubation at 65°C for 20 minutes. The digestion was incubated on ice for 5 minutes and unpaired 5' overhangs generated by EcoP15I digestion were repaired with in the presence of 1.5 µl 25 mM dNTPs (Invitrogen) and 1 µl Klenow (5 Units/µl, NEB) (25°C for 30 minutes). The reaction was stopped by incubating at 65°C for 20 minutes and incubated on ice for 5 minutes.

For sequencing adaptor ligation, 1µl 1M $MgCl_2$, 2 µl 100mM ATP, 60 µl 25% PEG-8000, 2 µl 40uM Illumina-PE-Adaptor-A (Table Sx), 2 µl 40uM Illumina-PE-Adaptor-B (Supplementary Table 11), and 5 µl Quick Ligase (NEB) were added to each sample. The reaction was incubated at 25°C for 30 minutes and terminated by incubation at 65°C for 20 minutes. After incubation on ice for 5 minutes, 100 µl water was added to the tube to make the total volume to 300 µl.

**Biotin pull-down, Nick repair, library amplification and purification**

Biotin-labeled, paired-end adaptor ligated DNAs were immobilized to Dynabeads M-280 Streptavidin beads (Invitrogen) as follows. Fifteen µl M-280 beads were washed with 200 µl 1x Binding and Washing (B&W) buffer (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl). The beads were isolated from bulk solution via a DynaMag-Spin magnet (Invitrogen). The beads were then washed once with 200 µl 1x BSA (NEB) and once with 200 µl 1x B&W buffer. Isolated beads were resuspended in 300 µl 2x B&W buffer and combined with the 300 µl labeled library DNA. The reaction was incubated at 25°C for 20 minutes and constant rotation. The supernatant was removed, the DNA-bound M-280 beads resuspended in 200 µl 1x B&W buffer and transferred to a new tube. The beads were washed twice with 200 µl 1x B&W buffer and once with 200 µl 1x NEBuffer 2.

The beads were resuspended in 34.5 µl 1x NEBuffer 2. Four µl 25 mM dNTPs mix (Invitrogen) and 1.5 µl DNA polymerase I (10 Units/µl, NEB) was added and the mixture was incubated at 16$^{o}$C for 30 minutes with shaking. The beads were then washed once with 200 µl EB (Qiagen) and resuspended in 40 µl EB.

The libraries were PCR amplified for a variable number of cycles (between 17-23) of with Phusion High-Fidelity DNA polymerase (NEB) and the Illumina-lib-PCR-A (5'-seq- 3') and Illumina-lib-PCR-B (5'-seq- 3') primer pair (Supplementary Table 11). Approximately 12- 50 µl PCR reactions were carried out for each library. PCR products for each library were pooled and resolved in 3% low melting agarose gel. The paired end tags with the desired size range (207-209 bp) were isolated and recovered with the Qiaquick Gel Extract Kit (Qiagen). DNA concentration was determined by measurement on a Nanodrop-1000 spectrophotometer. The libraries were then subjected to paired-end sequencing on an Illumina Genome Analyzer 2.

**3C Confirmation**

3C is carried out as described (Miele, A. et al., 2006) with the following modifications. About 2x $10^{8}$ cross-linked yeast cells were resuspended in 500 µl 1x NEBuffer 2 containing 0.3% SDS and were incubated at 65$^{o}$C for 20 minutes followed by one hour at 37$^{o}$C with shaking. A volume of 6 µl 10x NEBuffer 2 and 50 µl 20% Triton X-100 was added to each tube and mixed gently. After incubation at 37$^{o}$C for one hour with shaking, cells were digested overnight at 37$^{o}$C with shaking and 1200 Units of RE1 (HindIII or EcoRI, NEB). Restriction enzymes were inactivated by adding 112 µl 10% SDS to each tube and incubating at 65$^{o}$C for 20 minutes. Digested nuclei were diluted into 8 ml ligation buffer (66 mM Tris-HCl, pH 7.5; 5 mM DTT; 10 mM MgCl2; 10 mM ATP; 1 mg/ml BSA; 1% Triton X-100). After incubation at 37$^{o}$C for 1 hour, 80 Units T4 DNA ligase (5 Units/µl, Fermentas) were added and ligation was carried out at 16$^{o}$C for 4 hours followed by one hour at 25$^{o}$C. Crosslinks were reversed by incubation at 65$^{o}$C overnight in the presence of 100 µl 20 mg/ml proteinase K . An additional 50 µl 20 mg/ml proteinase K was added the following day and incubated at 55$^{o}$C for 2 hours. DNA was purified by three consecutive phenol-chloroform extractions followed by iso-propanol precipitation. RNA was degraded via incubation with RNase.

A randomized positive control library was generated for each RE1 3C library to normalize the PCR amplification efficiency of each primer pair used in this assay. This control library was

created by restriction digest and inter-molecular ligation of yeast genomic DNA (Miele et al (2006)).

3C interaction frequencies were assessed via quantitative PCR (ABI 7500) with SybrGreen. The identities of qPCR products were confirmed by both DNA gel electrophoresis and melting curve analysis. The data were normalized to an *ARP2* loading control. Primers used in 3C assay were listed in Supplementary Table 11.

## Computational methods

### Mapping of sequence data

Sequence files from the Illumina Genome Analyzer 2 were processed as follows. Paired 20 bp reads were extracted and mapped individually to the *S. cerevisiae* genome with the MAQ tool (Li et al, 2008) with default parameters. Pairs of reads were extracted such that each read mapped to a location that was consistent with expectation relative to the position of a restriction enzyme recognition site in the reference genome (HindIII or EcoRI), but without imposing constraints on the relative locations of the mapped pair. We further allowed for ±1 bp shifts with respect to the precise length of the sequencing tags generated by EcoP15I digestion. We only considered those reads that mapped to the reference genome with an MAQ score of at least 20.

### Statistical confidence estimation

Our assay produces as output a collection of paired reads, corresponding to interacting, distal restriction enzyme (RE) sites. First, we convert the RE sites to RE *fragments*, as follows. The reads that mapped downstream of a site (MAQ orientation '-') belong to the downstream fragment and vice-versa. Throughout this manuscript, when we refer to a fragment we mean the mid-location between the bordering RE sites. Second, we eliminate sequencing tiling duplicates (that can appear when a DNA molecule has extremely high concentration during sequencing). Third, we eliminate all intra-chromosomal interactions in which the two RE fragments are separated by <20 kbp. Fourth, we eliminate all ligations between adjacent fragments, because such interactions are observable only in case of self-ligation (Supplementary Table 12).

For a given interaction involving two RE fragments A and B, there are 72 possible isoforms: 3 (possible number of sizes of one paired tag (25-27 bp)) x 3 (possible number of sizes of the other paired tag) x 2 (number of orientations) x 4 (number of fragment combinations) = 72. We observed a high correlation between sequence frequency and the number of isoforms; i.e., the higher the sequence frequency for an interaction, the more isoforms were observed for that interaction. Therefore, it was not necessary to consider different isoform types independently.

We split all the possible RE (HindIII or EcoRI) fragments in the yeast genome into two categories: mappable and unmappable. A fragment is deemed mappable if it was mapped to the reference genome at MAQ mapping threshold of 20 at least $K$ times (all libraries were considered), where $K$ was determined using the following equation:

$$K = mean - 2 \times stddev,$$

where $mean$ is the average number of mapped reads per RE fragment (excluding those with zero mapped reads), and $stddev$ is the standard deviation. For HindIII we obtained $K = 236$, and for EcoRI we obtained $K = 1$. For HindIII, this filter eliminates 264 of 4457 total fragments (5.9%) in the yeast genome; for EcoRI, the filter eliminates 125 of 4384 fragments (2.9%).

Finally, we eliminated all interactions that contained at least one unmappable fragment. In our heat maps and Circos images throughout this article, we represent the mappable HindIII fragments in green and the unmappable ones in black.

To assign statistical confidence estimates to inter-chromosomal interactions, we use a uniform probability model to convert the observed frequencies into p-values. To do so, we count the total number $M$ of inter-chromosomal pairs of RE fragments (HindIII or EcoRI) in the yeast genome. Note that, in counting these fragments, we only consider pairs in which both fragments are mappable, as defined above. Assuming that the probability of observing any particular interaction is uniform, that probability is $m = 1/M$. We then count the total number $n$ of observed inter-chromosomal interactions. We can compute the probability of observing a given interaction pair exactly $k$ times via the binomial distribution:

$$P(K = k) = \binom{n}{k} m^k (1-m)^{n-k}.$$

The p-value is the probability that each pair is observed *at least* $k$ times, and is:

$$\text{p-value} = \sum_{i=k}^{n} P(K = i).$$

The p-value calculation for intra-chromosomal interactions is slightly more complicated. Because each chromosome acts like a polymer in which the spatial distance between two loci increases with increasing genomic distance, we observe a very strong inverse relationship between the intra-chromosomal distance of two interacting RE fragments and the frequency with which the paired fragments are observed in our assay. To control for this distance effect, we subdivide the observed intra-chromosomal interactions into 5 kbp bins; i.e., all interactions from 20-25 kb, 25-30 kb, etc. At larger distances, we require a minimum of 100 observed interactions per bin, merging adjacent bins as necessary to achieve this minimum value. We then perform the p-value calculation separately for each distance bin. The resulting p-value is conditioned on the distance bin in which the given pair of RE fragments occurs. Hence, the p-

values are directly comparable to one another across bins and between inter-chromosomal and intra-chromosomal interactions.

Finally, we must correct for multiple testing, because we are simultaneously measuring the interactions among many pairs of RE fragments. We do this using the Q-value software (Storey, 2002), which assigns to each p-value a corresponding q-value, defined as the minimal false discovery rate threshold at which a given score is deemed significant. We perform this q-value estimation procedure separately for the intra-chromosomal interactions and the inter-chromosomal interactions. Importantly, we provide as input to the multiple testing correction procedure the p-values associated with all pairs of RE fragments, including pairs for which no interactions were observed.

To combine two libraries $A$ and $B$ (for example H-Mp and H-Me) and obtain one library (for example HindIII) at a given FDR, we proceed as follows. Given a q-value $q$, let $A_q$ and $B_q$ denote the subsets of $A$ and $B$ at $q$. Then, the q-value of the combined library $q_{combined}$ is:

$$q_{combined} = \frac{q \mid A_q \mid + q \mid B_q \mid}{\mid A_q \cup B_q \mid},$$

where $\mid A \mid$ is the number of interactions in library $A$. This approach is conservative, because we assume that false positive interactions in $A_q$ are distinct from the false positive interactions in $B_q$. We compute $q_{combined}$ for all $q$ values from 0.005 to 0.05 and a step of 0.001, and we select that value that gives a $q_{combined}$ value as close as possible to 0.01.

**Heat maps**

We draw heat maps of the binary interaction values at a 1 kb resolution (if there are several fragments within the same 1 kb bin, all interactions are considered). For improved visualization, we apply a Gaussian density smoothing with standard deviation 3 kb (roughly matching the average distance between the HindIII and EcoRI sites in yeast). Using p-values instead of binary values gave very similar results.

**Chromosomal arm proximity**

Following Lieberman-Aiden et al. (2009), the interaction scores in Figure 4c and Supplementary Figures 15 and 16 were computed as ratios between the observed and expected number of interactions for every pair of chromosomal arms (i.e., left and right of the centromere):

$$\text{score of arm } i \text{ vs. arm } j = \frac{N_{i,j}}{Ni/N \times Nj/N \times N},$$

where $N_{i,j}$ is the total number of observed interactions between chromosomal arms $i$ and $j$, $N_i$ is the total number of observed inter-arm interactions involving arm $i$, and $N$ is the total number of observed inter-arm interactions.

**Reproducibility between H-Mp libraries**

We calculate the reproducibility between the two biological replicates H-Mp-A and H-Mp-B (shown in Figure 2c) as follows. First, we construct three sub-libraries (one from H-Mp-A and two from H-Mp-B) such that the number of non-distinct interactions (before FDR) is the same. Because H-Mp-B is more than twice as large as H-Mp-A, the first sub-library includes the entire H-Mp-A, and the remaining two include interactions from H-Mp-B selected uniformly at random. The resulting number of intra-chromosomal interactions is 1,652,946. Then, for each sub-library, we select the interactions that pass a q-value threshold of 1%. We compute the matrices corresponding to the heat maps as described, and we compute the p-value of correlation for all entries of the pairwise matrices.

**Clustering**

To find clusters of strongly interacting tRNAs, we perform the following procedure. First, we find the closest HindIII fragment to the left and right of each tRNA location. Next, we build an interaction matrix by assigning a value of 0 (no interaction) to 4 (every possible interaction between the four HindIII fragments) for every interaction found in our HindIII library (at FDR 1%) between the two HindIII fragments for each pair of tRNAs. We then apply a hierarchical average-link clustering algorithm using a distance function of 4 minus the number of interactions between each pair of tRNAs. We cut the resulting dendrogram to yield 32 clusters. We perform the same clustering procedure for origins of replication, but we use 16 rather than 32 clusters.

**Co-localization analysis**

To compute whether various genomic functional groups (e.g., centromeres, telomeres, tRNAs) are co-localized (Figure 5a and Supplementary Figure 13), we proceed as follows. First, we count the number $N$ of all possible inter-chromosomal interactions, and the number $n$ of all observed distinct inter-chromosomal interactions. We consider a 5 kbp window centered on the area of interest and count $m$ and $k$, the numbers of all possible and distinct observed, respectively, inter-chromosomal interactions that have both ends inside these windows. Assuming that all possible interactions are distributed uniformly at random, the probability that we encounter exactly $k$ interactions inside our windows is given by a hypergeometric distribution:

$$P(K = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}.$$

Then, the p-value, i.e., the probability of observing an event at least as extreme as the one observed, is:

$$\text{p-value} = \sum_{i=k}^{n} P(K = i).$$

Note that in this case we only use distinct interactions, and ignore the interaction frequencies. This is a more conservative approach.

**Co-localization between tRNA clusters, centromeres and rDNA**

We analysed whether the two tRNA clusters we obtained are co-localized with the centromeres or rDNA (which is known to reside in the nucleolus; Supplementary Figure 11). To answer this question, we took an approach very similar to the hypergeometric approach used for co-localization of functional groups, except that when we computed $m$ and $k$ we required one interaction fragment to be within a 10 kb window of a tRNA, and one fragment to be within a 10 kb window of a centromere or rDNA. Note that our final set of HindIII interactions does not include any interactions with the rDNA regions. This is because the reference yeast genome contains two copies of the rDNA region; hence, all sites in this region receive a MAQ score of 0 (signifying non-unique mapping). Therefore, for the purposes of this experiment, we selected all interactions with the rDNA region that had MAQ score 0. We obtained 15,448 inter-chromosomal interactions with frequency 6 or more, which corresponds to the frequency that passed the FDR threshold of 1% for all other inter-chromosomal interactions. Using the hypergeometric calculation described above, we concluded that the bright cluster co-localizes with the centromeres (p-value < 1e-300), and the dim cluster co-localizes with the rDNA (p-value < 0.001).

**Receiver Operating Curve (ROC) analysis**

To generate an ROC curve with respect to, e.g., centromeric loci, the complete set of interactions (q<0.01) are first ranked by p-value. Next, each interaction is labeled "positive" if it links two loci that are close to (<2.5kb) a centromere, and "negative" if the interaction links two loci that are not close to a centromere. Interactions involving a centromeric locus and a non-centromeric locus are removed from the ranked list. Finally, the ROC curve is generated by traversing the ranked list of interactions and plotting, for each p-value threshold, the percent of "positive" and "negative" interactions above the threshold. If a given class of loci is enriched for strong interactions (i.e., interactions with low p-values), then the corresponding ROC curve will be significantly above the line y=x (shown as a gray dashed line). For the "telomere" class, a locus is deemed positive if it is <30kb from a telomere end.

**Inferring 3D structure**

To infer the 3D structure of the genome in the nucleus, we start from our final set of *HindIII* interactions. We developed an optimization procedure to convert the observed interactions into a 3D map. To do so, we represent each chromosome as a series of beads in 3D space, spaced

10 kbp apart. Let $p = (x, y, z)$ denote such a bead, and let $\pi$ denote the collection of all beads. We map each RE fragment to its closest bead. Given two beads $p = (x_1, y_1, z_1)$ and $q = (x_2, y_2, x_2)$, let $dist(p,q)$ denote the Euclidean distance between these two points:

$$dist(p,q) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.$$

The main idea of our algorithm is to place each pair of beads at a distance that is inversely proportional to their interaction frequency.

To convert the interaction frequency into distances, we assume that an interaction observed with frequency $f$ has the same distance between its loci as an intra-chromosomal interaction that has expected frequency $f$ by polymer packing. To obtain these values, we compute the mean observed frequencies for every 5 kbp distance bin in our H-Mp and H-Me libraries (Supplementary Figure 17). Following (Bystricky et al, 2004), every 110-150 bp of packed chromatin has a length of 1 nm, so we use an intermediate value of 130bp/nm. Using this information, we can compute the expected Euclidean distance corresponding to any genomic distance; e.g., 1000 bp will occupy 7.7 nm. Therefore, given an interaction in our data set with frequency *f*, we estimate the corresponding Euclidean distance from the mean of the curves obtained from the H-Mp and H-Me libraries (green curve in Supplementary Figure 17).

To construct a 3D model for the yeast genome, we minimize an objective function that tries, as much as possible, to place two interacting loci at the expected distance $\delta$:

$$\min_{\pi} \sum_i \left(dist(p_i, q_i) - \delta\right)^2,$$

where $dist$ is the distance between the two loci $p_i$ and $q_i$ involved in interaction $i$.

We then introduce the following constraints on the locations of those beads:

1. All beads must lie within a spherical nucleus with radius 1 μm (1000 nm). This is motivated by the fact that the yeast nucleus roughly resembles a sphere with this radius (Berger et al, 2008).

   $$-1 \le p \le 1, \qquad \text{for all points } p \in \pi.$$

   For convenience, we center the sphere at the origin.

2. The distance between every two beads adjacent on the chromosome must be within a given range. 1000 base pairs of chromatin occupy a length of 6.6nm -9.1nm (Bystricky et al, 2004). Therefore, for our resolution of 10kbp, we use the following constraints:

$$0.066^2 \le dist^2(p,q) \le 0.091^2, \quad \text{for all } p,q \in \pi, \quad p,q \text{ adjacent}$$

3. To prevent clashes, no two beads along the same chromosome can lie closer than 30nm apart. This number is motivated by the thickness of the chromatin fiber of packed nucleosomes (Amzallag et al, 2006)

$dist^2(p,q) \ge 0.03^2$, for every two distinct points along the same chromosome $p,q \in \pi$.

4. To prevent inter-chromosomal crossings between segments connecting adjacent beads, we imposed a minimum distance of 0.075 between beads of different chromosomes (this distance is sufficient to guarantee no crossings at resolution of 10 kbp). $dist^2(p,q) \ge 0.075^2$, for every two points on different chromosomes $p,q \in \pi$.

5. We extend the chromosome XII of the reference *S. cerevisiae* genome to include 148 additional ribosomal DNA repeats, for a total of 150 (that area is know to be repeated between 100-200 times). The length of each repeat is 9.1kb. We constrain the rDNA to be inside the nucleolus, within a sphere with the center at $c = (0.7,0,0)$ μm and radius 0.3 μm:

$$dist^2(p,c) \le 0.3^2, \qquad \text{for all rDNA beads } p.$$

6. The centromeres were shown to cluster on the diametrically opposite side from where the nucleolus is located (Berger et al, 2008); therefore, we placed the centromere of chromosome XII within a sphere with the center at $d = (-0.9,0.0)$ μm and radius 0.1 μm:

$$dist^2(p,d) \le 0.1^2, \qquad \text{where } p \text{ is the closest bead to centromere XII.}$$

Other centromeres are not explicitly constrained to be far from the nucleolus. The resulting set of equations defines an optimization problem with non-linear objective and non-convex quadratic constraints (the first part of the second set of constraints and the third and fourth sets of constraints are non-convex, although the remaining ones are). We solve this problem using IPOPT (Wächter and Biegler, 2006), an is open-source software for nonlinear constrained optimization problems that uses interior-point gradient-based methods. Our particular problem (10 kb resolution) had 4,053 variables and 914,746 constraints, and was solved in 30 CPU hours on a single core of an Intel Xeon CPU at 2.33GHz with 4MB cache and 16 GB RAM. Once we obtain the locations of all points, we use a quadratic Bezier curve to smoothly connect

the middle of the segments formed by every two points adjacent along the chromosome. We use Rasmol to display the final 3D structure.

## Supplementary References

Amzallag A, Vaillant C, Jacob M, Unser M, Bednar J, Kahn JD, Dubochet J, Stasiak A, and Maddocks JH, 3D reconstruction and comparison of shapes of DNA minicircles observed by cryo-electron microscopy, Nucleic Acids Res. 2006 October; 34(18): e125.

Berger AB, Cabal GG, Fabre E, Duong T, Buc H, Nehrbass U, Olivo-Marin JC, Gadal O, Zimmer C., High-resolution statistical mapping reveals gene territories in live yeast. Nat Methods. 2008 Dec;5(12):1031-7. Epub 2008 Nov 2.

Bystricky K, Heun P, Gehlen L, Langowski J, Gasser SM., Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. Proc Natl Acad Sci U S A. 2004 Nov 23;101(47):16495-500. Epub 2004 Nov 15.

Li H, Ruan J, Durbin R., Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. 2008 Nov;18(11):1851-8. Epub 2008 Aug 19.

Miele A, Gheldof N, Tabuchi TM, Dostie J, Dekker J. Mapping chromatin interactions by chromosome conformation capture. Curr Protoc Mol Biol. 2006 May;Chapter 21:Unit 21.11.

Storey JD. (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B*, **64**: 479-498.

Wächter A and Biegler LT, On the Implementation of a Primal-Dual Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming, *Mathematical Programming* 106(1), pp. 25-57, 2006

## Validation of the Method

Because all 3C-based technologies are encumbered by low signal-to-noise ratios[16,19], we included a series of controls to establish the reliability of our method. First, we varied the pairs of restriction enzymes to construct four libraries that differed in the combinations of first (*HindIII* or *EcoRI*) and second (*MseI* or *MspI*) restriction enzymes. Second, because DNA concentration during proximity-based ligation critically influences signal-to-noise ratios, we constructed two independent sets of experimental libraries differing by DNA concentration at the 3C step (Supplementary Table 1 and Supplementary Methods). Third, to assess noise resulting from random inter-molecular ligations, we constructed five control libraries, one using the combination of *HindIII* + *MseI* in non-cross linked cells and four using purified yeast genomic DNA with each of the four different restriction enzyme combinations.

Signal-to-noise ratios were determined to be acceptable by four measures. First, the ratio of long-range (defined as >20 kb, non-adjacent, see Supplementary Methods) intra-chromosomal interactions between restriction fragments flanked by *HindIII* or *EcoRI* sites was significantly higher in each of the experimental libraries than in any of the control libraries (Supplementary Table 2). Second, each experimental library exhibited a strong inverse correlation between the frequency of intra-chromosomal ligations of *HindIII* or *EcoRI* flanked restriction fragments and their genomic distance (Figure 2a). This feature is consistent with previous 3C and 4C data[16] and indicates that cross-linking retains chromosome conformation. In contrast, this polymer-like behavior was not observed in any of the control libraries. Third, because our method, like 4C[11], involves multiple rounds of restriction enzyme digestion and DNA circularization that can introduce a potential source of bias, we examined each of our libraries for differences in restriction enzyme digestion and ligation efficiency. We observed no significant differences in digestion efficiency at different *HindIII* or *EcoRI* sites (Supplementary Figure 1) across the genome. While restriction enzyme site-dependent differences in ligation efficiency were observed, the replicate experimental libraries produced results that were far more similar to each other (Pearson's $r$=0.85, $P$=$10^{-16}$) than to the corresponding control libraries ($r$=0.18, $P$=0.2) (Figure 2b and Supplementary Figure 2a). Similar results were also observed for restriction fragment ligation efficiencies (Supplementary Figure 2b and 2c). Fourth, for any combination of two restriction fragments, we observed high correlations (Pearson's r from 0.97 to 0.99*)* between captures from the 5' ends versus the 3' ends (Supplementary Table 3).

To estimate a false discovery rate (FDR), we eliminated self-ligations, ligations between adjacent restriction fragments, and ligations between restriction fragments separated by less than 20 kb at their midpoints. To account for the strong influence of genomic proximity on ligation frequency, we subdivided the remaining intra-chromosomal interactions into 5kb bins as measured by the genomic distance between the midpoints of the two ligated fragments. Inter-chromosomal interactions were placed into a separate bin. In each bin, we ranked the observed interactions according to their sequence frequency and assigned a p value relative to all other possible interactions in the same bin. Lastly, we converted the p value of each interaction into a q value (defined as the minimal FDR threshold at which the interaction is deemed significant) and used this value to rank interactions library-wide (Supplementary Methods). This approach allowed us to distinguish a propensity for relatively short-range interactions, which arise from the polymer-like behavior of chromosomes, from interactions due to higher order chromosome folding (Figure 2a and Supplementary Methods). To further evaluate the reproducibility of our assay, we constructed interaction matrices (using an FDR threshold of 1%, Supplementary Methods) for the two independent libraries generated using the *HindIII* + *MspI* combination (Figure 2c, Supplementary Table 1). Although these libraries were constructed using slightly different conditions (see above), they displayed a high degree of correlation ($r$=0.70, $P$<$10^{-300}$) relative to that observed between randomly segregated halves of a single library ($r$=0.89, $P$<$10^{-300}$). This observation also highlights the importance of sequence depth. Therefore, to achieve maximum sequence depth we combined, in subsequent analyses, two independent libraries for each of the four restriction enzyme combinations.

We derived sets of interactions between restriction fragments flanked by *HindIII* or *EcoRI* sites across the yeast genome at several FDR thresholds (Supplementary Table 4, Supplementary Figures 3 and 4, Supplementary Methods). A total of 65,683 distinct intra- and 240,629 inter-chromosomal interactions were identified at the FDR threshold of 1% from the *HindIII* libraries, whereas the *EcoRI* libraries had corresponding values of 30,578 and 72,860 (Supplementary Tables 4-8). Since the sequence depth of the *EcoRI* libraries was much lower than that of the *HindIII* libraries, we used the *EcoRI* libraries to confirm the validity of interactions identified in the *HindIII* libraries. We observed that >90% of intra-chromosomal and >70% of inter-chromosomal interactions from the *EcoRI* libraries were supported within a 10 kb radius by a corresponding interaction from the *HindIII* libraries (Supplementary Figure 5). We then confirmed our results using conventional 3C for 24 chromosomal interactions (Figure 2D, Supplementary Figure 6). Taken together, the reproducibility of results from independent

libraries and from libraries constructed using different restriction enzyme combinations, as well as the proximity-dependent efficiency of ligation in experimental versus control libraries, argue that our method is reliable and robust. The yeast genome architecture described below is based on interactions generated from the *HindIII* libraries using an FDR threshold of 1%, and the main architectural features were confirmed by interactions from the *EcoRI* libraries at the same threshold.

Supplementary Table 1 Summary of sequencing information of each library

| Library | | RE1+RE2 | Mapped reads | |
|---|---|---|---|---|
| | | | A | B |
| Cross-linked | H-Mp | HindIII+MspI | 29,737,006 | 22,073,244 |
| | H-Me | HindIII+MseI | 15,343,651 | 13,131,690 |
| | E-Mp | EcoRI+MspI | 18,917,316 | 10,269,318 |
| | E-Me | EcoRI+MseI | 18,761,765 | 8,330,175 |
| Un-cross-linked | U-H-Me | HindIII+MseI | 2,401,077 | 3,502,486 |
| Purified genomic DNA | N-H-Mp | HindIII+MspI | 5,272,095 | - |
| | N-H-Me | HindIII+MseI | 7,249,911 | - |
| | N-E-Mp | EcoRI+MspI | 3,276,861 | - |
| | N-E-Me | EcoRI+MseI | 2,929,860 | - |

A: libraries with the DNA concentration of ~0.5 μg/ml at the 3C step
B: libraries with the DNA concentration of ~0.3 μg/ml at the 3C step

Supplementary Table 2 the ratio of Long-distance (>=20kb), non-adjacent signals in each library

| Library | | Intra- (>=20 Kb) | | Inter- | |
|---|---|---|---|---|---|
| | | A | B | A | B |
| Cross-linked | H-Mp | 0.079 | 0.0326 | 0.164 | 0.080 |
| | H-Me | 0.08 | 0.0472 | 0.176 | 0.110 |
| | E-Mp | 0.0336 | 0.0265 | 0.091 | 0.060 |
| | E-Me | 0.0449 | 0.0286 | 0.115 | 0.074 |
| Un-cross-linked | U-H-Me | 0.0135 | 0.0075 | 0.171 | 0.0613 |
| Purified genomic DNA | N-H-Mp | 0.00267 | - | 0.0338 | - |
| | N-H-Me | 0.00437 | - | 0.055 | - |
| | N-E-Mp | 0.00481 | - | 0.059 | - |
| | N-E-Me | 0.00475 | - | 0.060 | - |

A: libraries with the DNA concentration of ~0.5 μg/ml at the 3C step
B: libraries with the DNA concentration of ~0.3 μg/ml at the 3C step

Supplementary Table 3 Pearson's correlation between captures identified from the 5' ends versus that from 3' ends for any combination of two restriction fragments in each library

| Library | r value |
|---|---|
| H-Mp | 0.97 |
| H-Me | 0.98 |
| E-Mp | 0.99 |
| E-Me | 0.98 |
| U-H-Me | 0.99 |
| N-H-Mp | 0.97 |
| N-H-Me | 0.98 |
| N-E-Mp | 0.97 |
| N-E-Me | 0.97 |

Supplementary Table 4 Statistics of data analysis

| Library | Initial mapped reads | Initial distinct reads | Final distinct fragments for FDR analysis | FDR 1% | Union of libraries with the same RE1, target FDR of union 1% | Union of libraries with the same RE1, target FDR of union 0.5% | Union of libraries with the same RE1, target FDR of union 0.1% | Union of libraries with the same RE1, target FDR of union 0.01% |
|---|---|---|---|---|---|---|---|---|
| **H-Mp** | 51,810,250 intra <20kb = 42,097,847 intra >=20kb = 3,068,835 inter = 6,643,568 | 4,341,215 intra <20kb = 84,522 intra >=20kb = 591,801 inter = 3,664,892 | intra >=20kb = 304,460 inter = 2,664,839 | intra >=20kb = 51,732 inter = 205,674 | intra>=20kb = 65,683 inter = 240,629 | intra>=20kb = 60,136 inter = 167,270 | intra>=20kb = 48,280 inter = 103,251 | intra>=20kb = 37,094 inter = 66,127 |
| **H-Me** | 28,475,341 intra <20kb = 22,489,101 intra >=20kb = 1,846,622 inter = 4,139,618 | 3,505,988 intra <20kb = 86,753 intra >=20kb = 547,684 inter = 2,871,551 | intra >=20kb = 317,109 inter = 2,312,898 | intra >=20kb = 24,596 inter = 44,056 | | | | |
| **E-Mp** | 29,186,634 intra <20kb = 26,273,545 intra >=20kb = 846,617 inter = 2,066,472 | 1,467,570 intra <20kb = 24,941 intra >=20kb = 182,080 inter = 1,260,549 | intra >=20kb = 182,080 inter = 1,260,549 | intra >=20kb = 21,071 inter = 45,411 | intra>=20kb = 30,578 inter = = 72,860 | intra>=20kb = 27,188 inter = 72,860 | intra>=20kb = 20,620 inter = 34,180 | intra>=20kb = 15,405 inter = 27,651 |
| **E-Me** | 27,091,940 intra <20kb = 23,828,694 intra >=20kb = 910,395 inter = 2,352,851 | 1,748,361 intra <20kb = 30,283 intra >=20kb = 228,107 inter = 1,489,971 | intra >=20kb = 228,107 inter = 1,489,971 | intra >=20kb = 13,718 inter = 29,022 | | | | |

Supplementary Table 14 the average spatial distance between each pair of the 16 centromere regions (10 kb region around centromere, 5 kb left and 5 kb right)

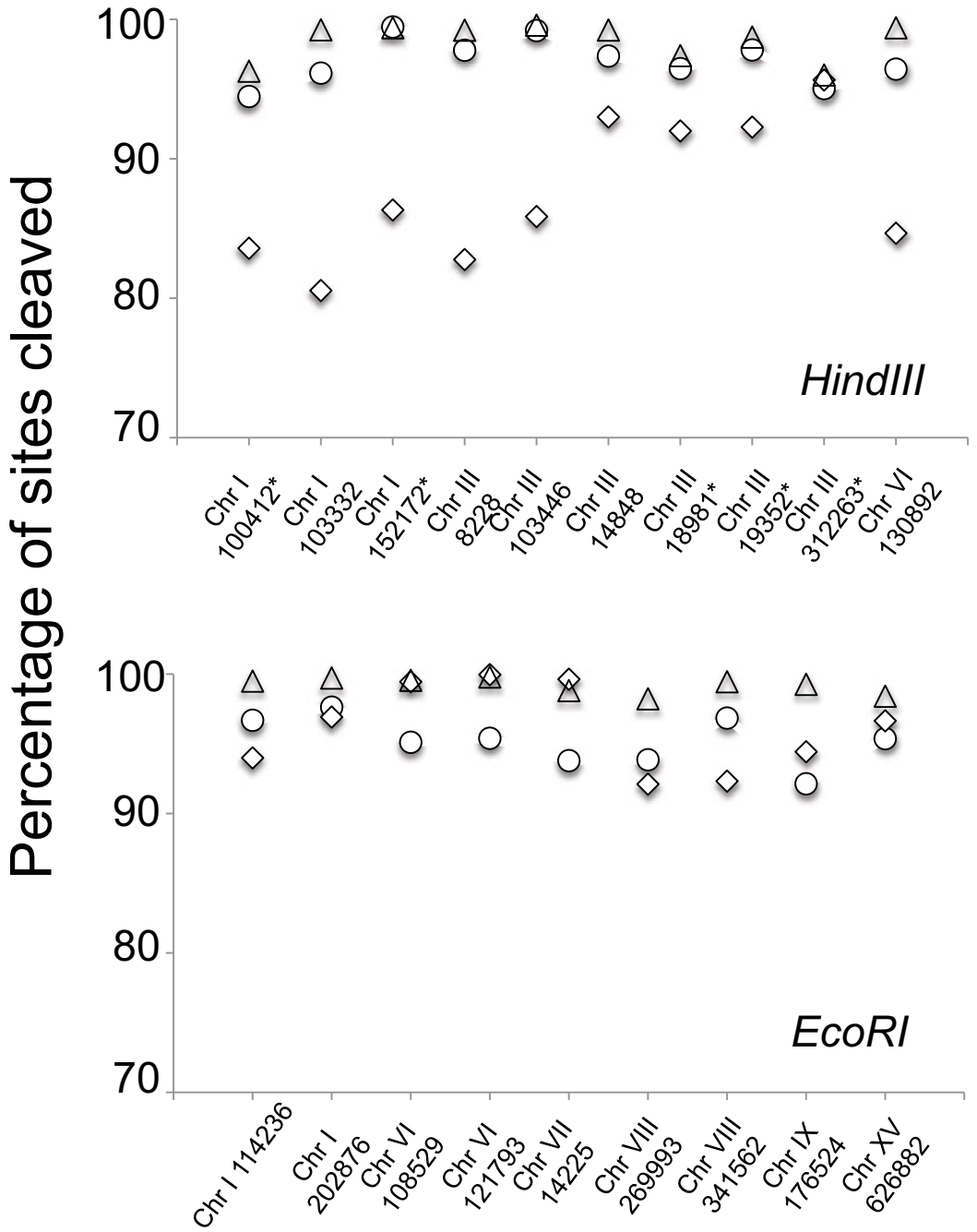
| | I | II | III | IV | V | VI | VII | VIII | IX | X | XI | XII | XIII | XIV | XV | XVI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **I** | 0 | 532 | 532 | 457 | 457 | 551 | 551 | 532 | 419 | 551 | 551 | 532 | 551 | 551 | 627 | 532 |
| **II** | 532 | 0 | 551 | 475 | 457 | 551 | 494 | 627 | 381 | 457 | 551 | 532 | 475 | 475 | 589 | 532 |
| **III** | 532 | 551 | 0 | 532 | 494 | 551 | 627 | 419 | 475 | 741 | 551 | 551 | 551 | 589 | 836 | 627 |
| **IV** | 457 | 475 | 532 | 0 | 419 | 532 | 438 | 532 | 381 | 419 | 457 | 381 | 381 | 381 | 494 | 475 |
| **V** | 457 | 457 | 494 | 419 | 0 | 475 | 494 | 532 | 381 | 381 | 475 | 419 | 343 | 419 | 494 | 475 |
| **VI** | 551 | 551 | 551 | 532 | 475 | 0 | 551 | 551 | 475 | 419 | 551 | 457 | 419 | 494 | 494 | 494 |
| **VII** | 551 | 494 | 627 | 438 | 494 | 551 | 0 | 381 | 381 | 457 | 551 | 475 | 419 | 457 | 494 | 475 |
| **VIII** | 532 | 627 | 419 | 532 | 532 | 551 | 381 | 0 | 627 | 741 | 551 | 589 | 741 | 551 | 741 | 684 |
| **IX** | 419 | 381 | 475 | 381 | 381 | 475 | 381 | 627 | 0 | 419 | 438 | 324 | 551 | 475 | 494 | 419 |
| **X** | 551 | 457 | 741 | 419 | 381 | 419 | 457 | 741 | 419 | 0 | 419 | 381 | 475 | 494 | 475 | 419 |
| **XI** | 551 | 551 | 551 | 457 | 475 | 551 | 551 | 551 | 438 | 419 | 0 | 532 | 684 | 589 | 684 | 494 |
| **XII** | 532 | 532 | 551 | 381 | 419 | 457 | 475 | 589 | 324 | 381 | 532 | 0 | 475 | 475 | 551 | 381 |
| **XIII** | 551 | 475 | 551 | 381 | 343 | 419 | 419 | 741 | 551 | 475 | 684 | 475 | 0 | 475 | 551 | 494 |
| **XIV** | 551 | 475 | 589 | 381 | 419 | 494 | 457 | 551 | 475 | 494 | 589 | 475 | 475 | 0 | 684 | 532 |
| **XV** | 627 | 589 | 836 | 494 | 494 | 494 | 494 | 741 | 494 | 475 | 684 | 551 | 551 | 684 | 0 | 551 |
| **XVI** | 532 | 532 | 627 | 475 | 475 | 494 | 475 | 684 | 419 | 419 | 494 | 381 | 494 | 532 | 551 | 0 |

Supplementary Table 15 a the maximum spatial distance (nm) between each pair of the 16 centromere regions (10 kb region around centromere, 5kb left and 5 kb right)

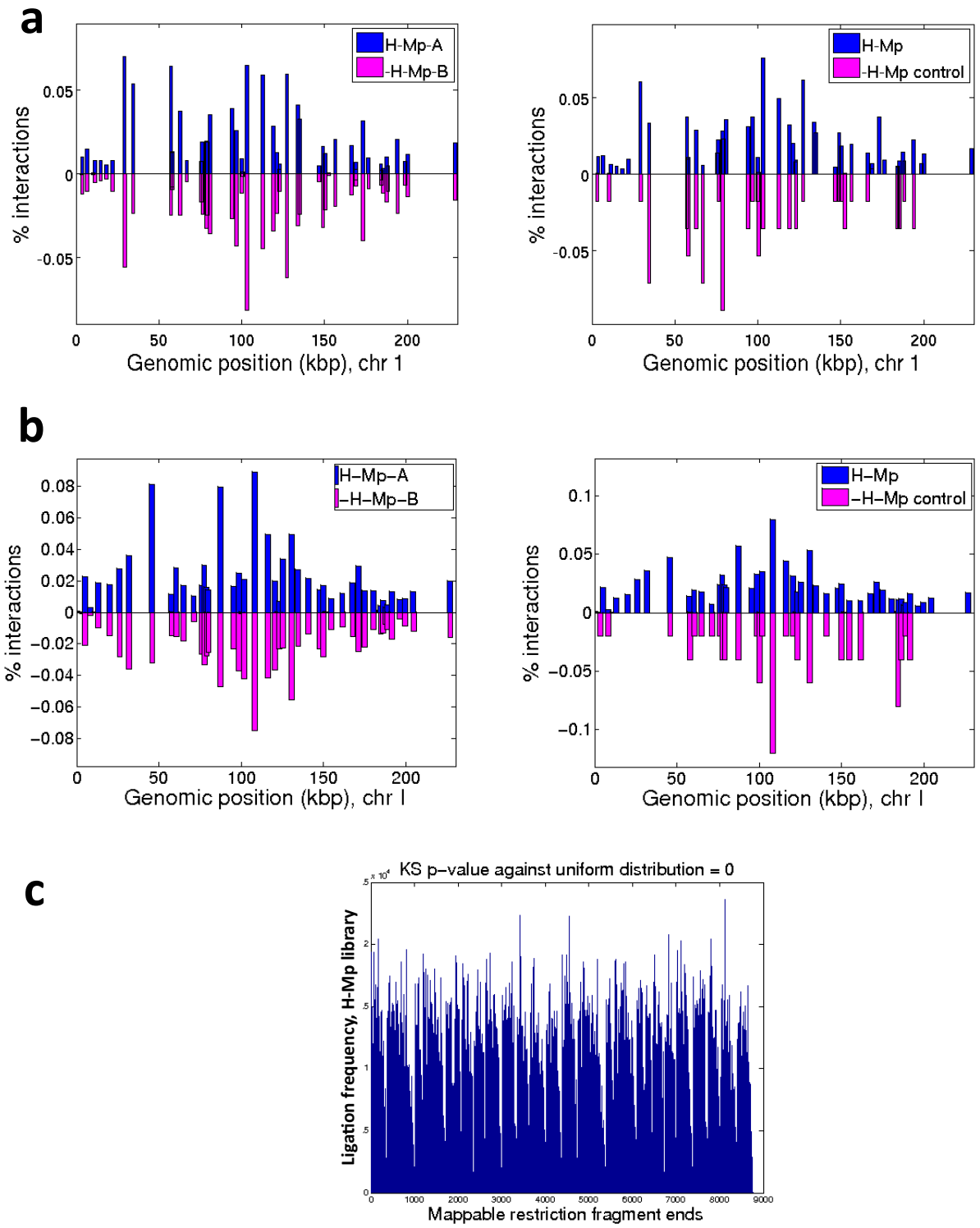
| | I | II | III | IV | V | VI | VII | VIII | IX | X | XI | XII | XIII | XIV | XV | XVI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I | 0 | 950 | 741 | 950 | 741 | 950 | 950 | 950 | 836 | 836 | 950 | 950 | 950 | 836 | 950 | 950 |
| II | 950 | 0 | 836 | 950 | 950 | 950 | 950 | 950 | 950 | 950 | 950 | 950 | 950 | 950 | 950 | 950 |
| III | 741 | 836 | 0 | 950 | 950 | 836 | 741 | 551 | 741 | 950 | 741 | 950 | 551 | 836 | 950 | 950 |
| IV | 950 | 950 | 950 | 0 | 950 | 950 | 836 | 684 | 950 | 950 | 950 | 836 | 741 | 950 | 950 | 950 |
| V | 741 | 950 | 950 | 950 | 0 | 950 | 950 | 950 | 627 | 589 | 950 | 950 | 551 | 684 | 836 | 950 |
| VI | 950 | 950 | 836 | 950 | 950 | 0 | 950 | 741 | 950 | 950 | 950 | 627 | 741 | 950 | 741 | 950 |
| VII | 950 | 950 | 741 | 836 | 950 | 950 | 0 | 475 | 950 | 950 | 836 | 950 | 684 | 836 | 684 | 741 |
| VIII | 950 | 950 | 551 | 684 | 950 | 741 | 475 | 0 | 836 | 836 | 950 | 950 | 741 | 684 | 836 | 950 |
| IX | 836 | 950 | 741 | 950 | 627 | 950 | 950 | 836 | 0 | 950 | 741 | 741 | 741 | 741 | 950 | 741 |
| X | 836 | 950 | 950 | 950 | 589 | 950 | 950 | 836 | 950 | 0 | 684 | 950 | 627 | 627 | 684 | 836 |
| XI | 950 | 950 | 741 | 950 | 950 | 950 | 836 | 950 | 741 | 684 | 0 | 836 | 741 | 836 | 836 | 950 |
| XII | 950 | 950 | 950 | 836 | 950 | 627 | 950 | 950 | 741 | 950 | 836 | 0 | 627 | 950 | 950 | 950 |
| XIII | 950 | 950 | 551 | 741 | 551 | 741 | 684 | 741 | 741 | 627 | 741 | 627 | 0 | 684 | 551 | 950 |
| XIV | 836 | 950 | 836 | 950 | 684 | 950 | 836 | 684 | 741 | 627 | 836 | 950 | 684 | 0 | 741 | 950 |
| XV | 950 | 950 | 950 | 950 | 836 | 741 | 684 | 836 | 950 | 684 | 836 | 950 | 551 | 741 | 0 | 836 |
| XVI | 950 | 950 | 950 | 950 | 950 | 950 | 741 | 950 | 741 | 836 | 950 | 950 | 950 | 950 | 836 | 0 |

Supplementary Table 15 b the minimun spatial distance (nm) between each pair of the 16 centromere regions (10 kb region around centromere, 5kb left and 5 kb right)

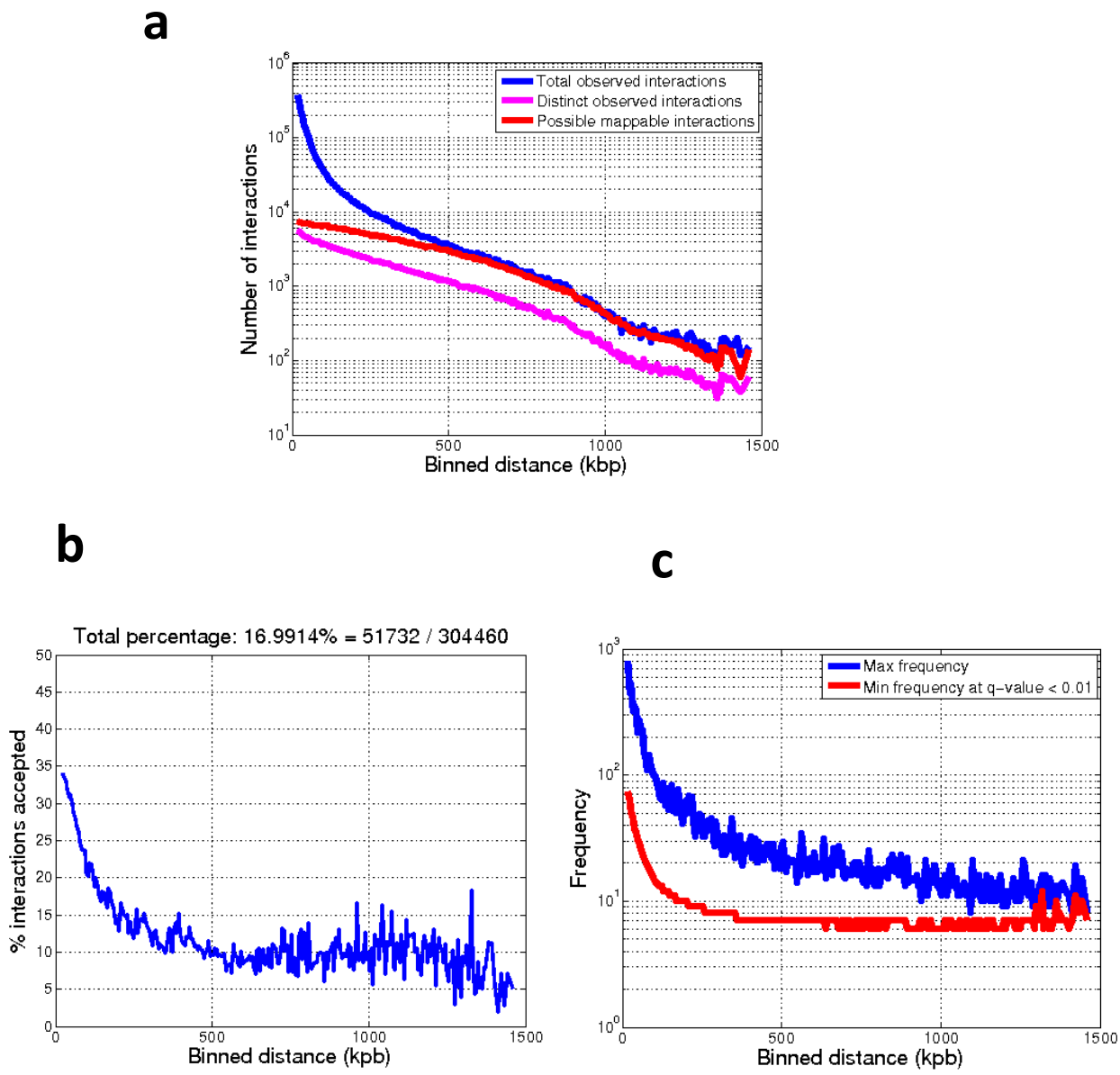
| | I | II | III | IV | V | VI | VII | VIII | IX | X | XI | XII | XIII | XIV | XV | XVI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I | 0 | 229 | 419 | 191 | 267 | 267 | 381 | 324 | 229 | 381 | 343 | 305 | 324 | 381 | 457 | 343 |
| II | 229 | 0 | 457 | 286 | 229 | 229 | 381 | 475 | 267 | 324 | 248 | 381 | 419 | 305 | 438 | 324 |
| III | 419 | 457 | 0 | 438 | 343 | 475 | 475 | 343 | 381 | 627 | 381 | 419 | 551 | 494 | 741 | 419 |
| IV | 191 | 286 | 438 | 0 | 267 | 324 | 248 | 381 | 229 | 172 | 305 | 267 | 248 | 229 | 286 | 229 |
| V | 267 | 229 | 343 | 267 | 0 | 191 | 267 | 343 | 172 | 248 | 248 | 267 | 286 | 267 | 267 | 324 |
| VI | 267 | 229 | 475 | 324 | 191 | 0 | 438 | 419 | 286 | 229 | 305 | 305 | 286 | 324 | 324 | 286 |
| VII | 381 | 381 | 475 | 248 | 267 | 438 | 0 | 324 | 286 | 286 | 419 | 343 | 343 | 324 | 419 | 267 |
| VIII | 324 | 475 | 343 | 381 | 343 | 419 | 324 | 0 | 494 | 741 | 419 | 475 | 741 | 475 | 741 | 589 |
| IX | 229 | 267 | 381 | 229 | 172 | 286 | 286 | 494 | 0 | 267 | 267 | 191 | 419 | 305 | 324 | 229 |
| X | 381 | 324 | 627 | 172 | 248 | 229 | 286 | 741 | 267 | 0 | 267 | 248 | 419 | 381 | 381 | 248 |
| XI | 343 | 248 | 381 | 305 | 248 | 305 | 419 | 419 | 267 | 267 | 0 | 286 | 627 | 419 | 589 | 324 |
| XII | 305 | 381 | 419 | 267 | 267 | 305 | 343 | 475 | 191 | 248 | 286 | 0 | 381 | 267 | 343 | 191 |
| XIII | 324 | 419 | 551 | 248 | 286 | 286 | 343 | 741 | 419 | 419 | 627 | 381 | 0 | 381 | 551 | 324 |
| XIV | 381 | 305 | 494 | 229 | 267 | 324 | 324 | 475 | 305 | 381 | 419 | 267 | 381 | 0 | 684 | 343 |
| XV | 457 | 438 | 741 | 286 | 267 | 324 | 419 | 741 | 324 | 381 | 589 | 343 | 551 | 684 | 0 | 475 |
| XVI | 343 | 324 | 419 | 229 | 324 | 286 | 267 | 589 | 229 | 248 | 324 | 191 | 324 | 343 | 475 | 0 |



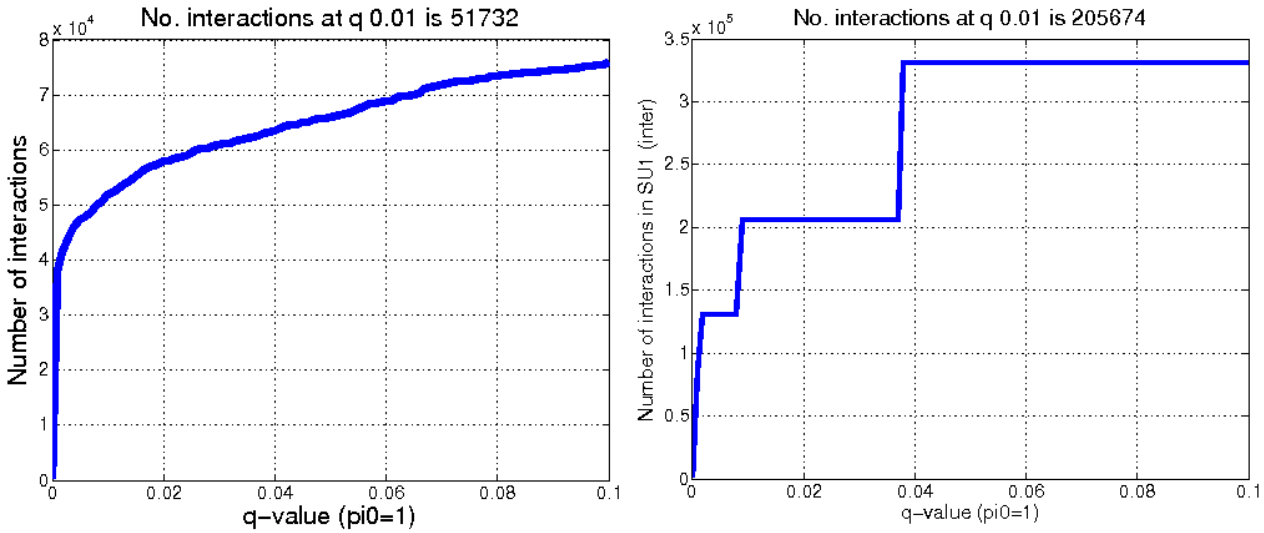

Supplementary Figure 1. Restriction enzyme digestion efficiency (*HindIII,* top; *EcoRI,* bottom) at the 3C step of our method was equally high across a series of sites including those (indicated with *) were underrepresented among interactions identified in our experimental libraries. This suggests that the observed difference in their involvement of interactions between different enzyme sites is not attributable to differences in the efficiency of restriction enzyme digestion. Circles show results for cross-linked libraries, triangles show results for uncross linked libraries, and diamonds show results for libraries generated using purified DNA. The uncross linked *EcoRI* library was prepared but not subjected to high throughput sequencing.



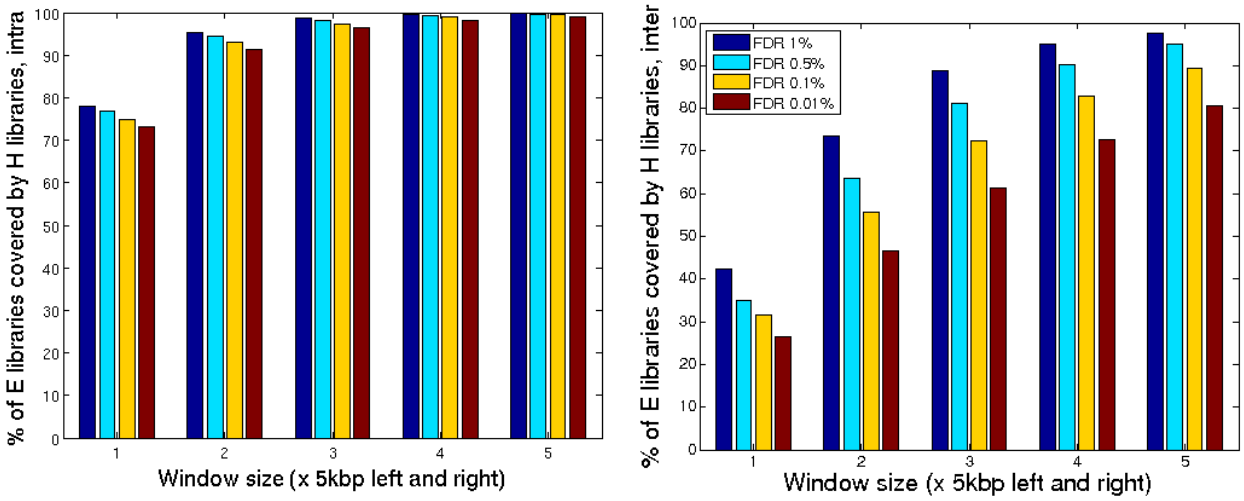

Supplementary Figure 2. The fraction of instances that each *HindIII* site (a) or fragment (b) along chromosome I was engaged in an intra-chromosomal interaction was very similar for two independently derived experimental H/Mp libraries designated A and B (left panel) but differed significantly between experimental and non-cross linked control H-Mp libraries (right panel). c. distribution plot with number of reads versus mappable *HindIII* fragment ends in H-Mp library.



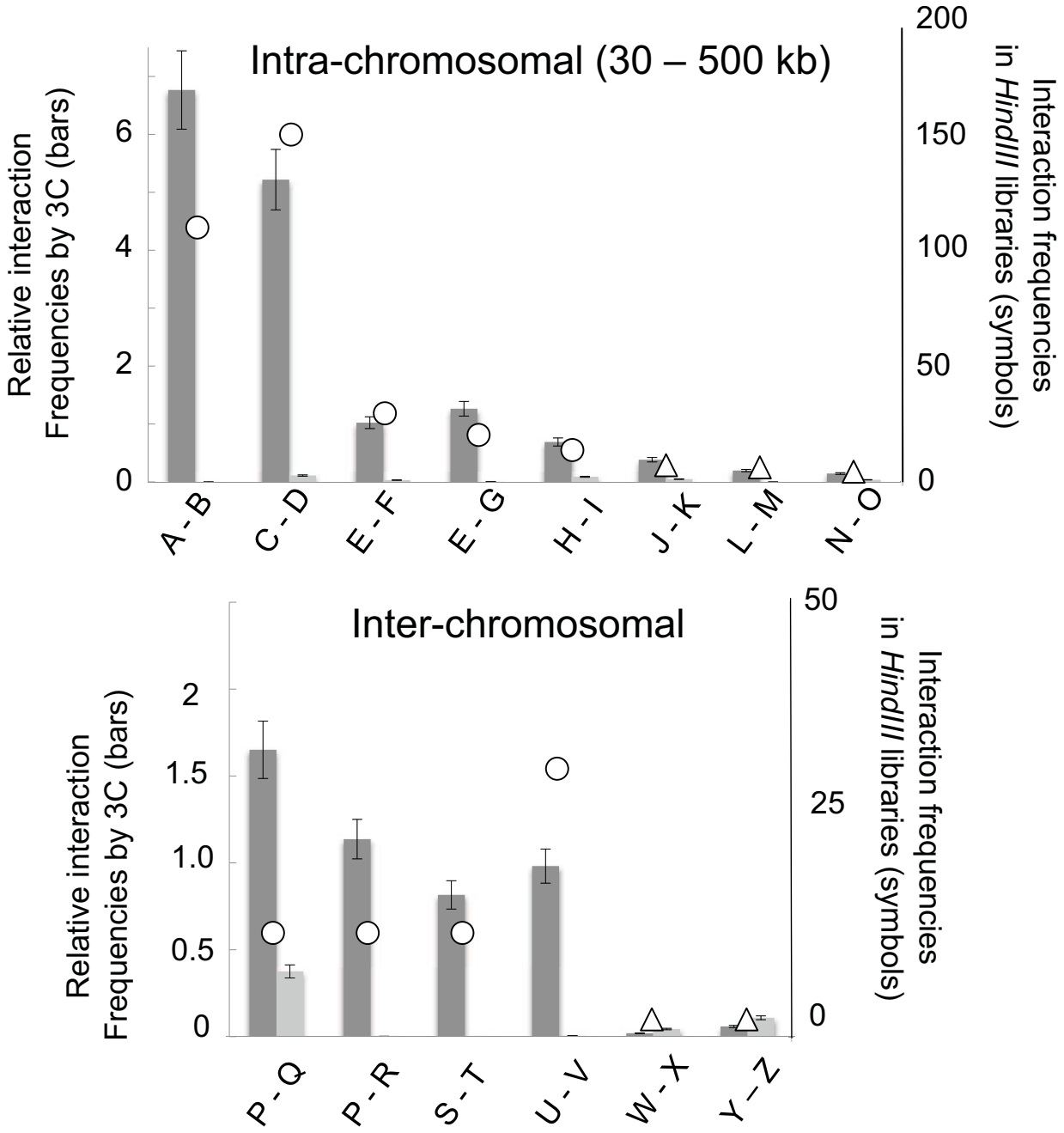

Supplementary Figure 3. Examples showing the details of our FDR analysis for intra-chromosomal interactions in the *HindIII-MspI* experimental library. a. Sequence depth at each distance bin. The numbers of possible unique mappable (red), observed distinct (pink) and observed total interactions (blue) in each bin are indicated. b. The percentage of distinct interactions in each distance bin that passed the FDR 1% threshold. The overall percentage of distinct interactions that passed the FDR 1% threshold is also indicated. c. The range of sequence frequencies (minimum, red; maximum, blue) for distinct interactions in each distance bin that passed the FDR threshold of 1%.



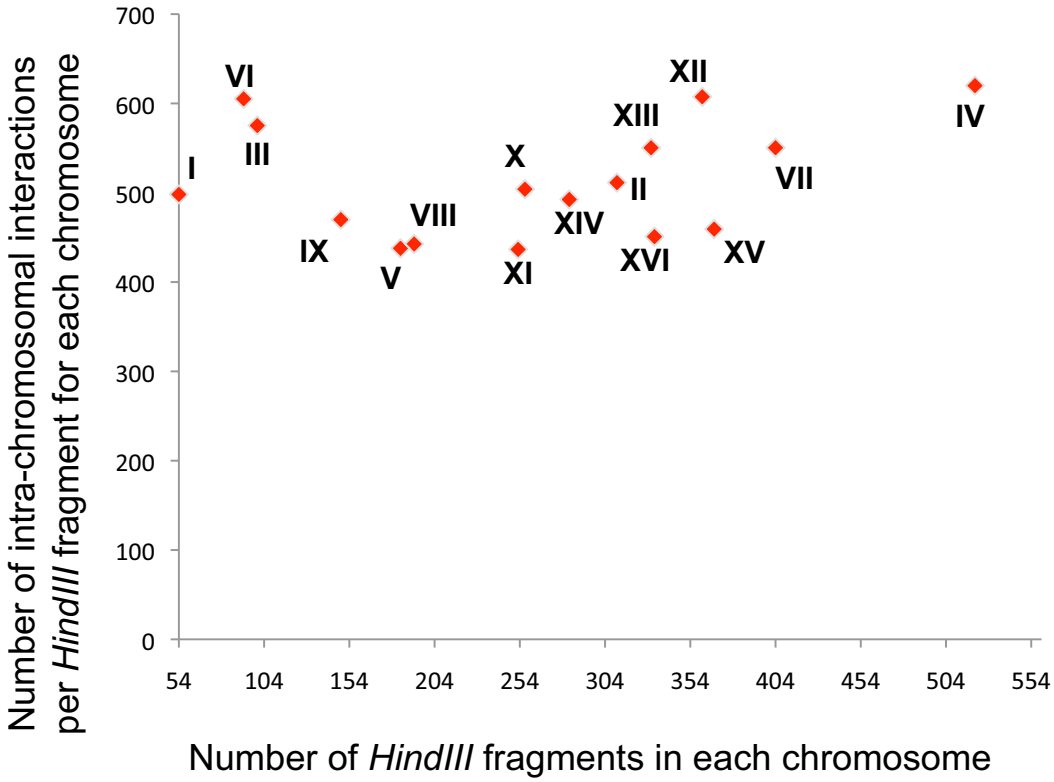

Supplementary Figure 4. Examples showing details of our FDR analysis for intra-chromosomal and inter-chromosomal interactions in the *HindIII-MspI* experimental library. Left panel: The distribution of intra-chromosomal interactions relative to q-value. The total number of intra-chromosomal interactions in the library with a q-value less than 0.01 is indicated. Right panel: The distribution of inter-chromosomal interactions relative to q-value. The total number of inter-chromosomal interactions with q-values of less than 0.01 is indicated.




Supplementary Figure 5. Fraction of intra-chromosomal interactions (left panel) and inter-chromosomal interactions (right panel) in the *EcoRI* libraries that are encompassed by interactions in the *HindIII* libraries, as a function of different window sizes. Y axis indicates the percentage of interactions in the *EcoRI* libraries that are also detected in the *HindIII* libraries. X axis reflects five window sizes ranging from 10 to 50 kb. Colored bars reflect results at increasingly stringent false discovery rates (FDR), as indicated. The large majority of interactions observed in the *EcoRI* libraries at the threshold of FDR 1% fall within 20kb on both ends of a corresponding interaction identified in the *HindIII* libraries at the same threshold.



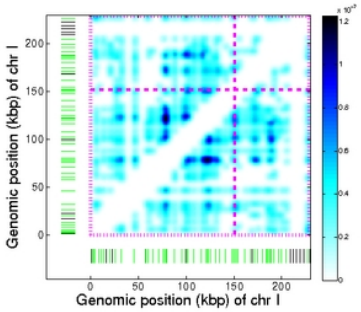

Supplementary Figure 6. High degree of correlation between absolute interaction frequencies as determined by our method (symbols) versus relative interaction frequencies as determined by conventional 3C using cross-linked (dark bars) and uncross-linked (light bars) libraries. Top graph: Results for 8 potential intermediate distance (30 – 500 kb) intra-chromosomal interactions, of which 5 passed (circles) and 3 did not pass (triangles) an FDR threshold of 1%. Bottom graph: Corresponding results for inter-chromosomal interactions of which 4 passed (circles) and 2 did not pass (triangles) an FDR threshold of 1%. Chromosomal positions (midpoints of *HindIII* restriction fragments) are as follows: A: Chr XII 43985; B: Chr XII 87022; C: Chr III 60064; D: Chr III 96173; E: Chr VI 17411; F: Chr VI: 249743; G: Chr VI: 255566; H: Chr XII 661926; I: Chr XII 923206; J: Chr II 474127; K: Chr II 797920; L: Chr XV 463436; M: Chr XV 944437; N: Chr XV 319599; O: Chr XV 736358; P: Chr I 101871; Q: Chr VI 123588; R: Chr XIII 282471; S: Chr I 171164; T: Chr V 166959; U: Chr XI 481783; V: Chr XV 317227; W: Chr XII 347076; X: Chr XIV 110551; Y: Chr X 380259; Z: Chr XV 472430. Error bars denote standard deviations of 3 determinations.

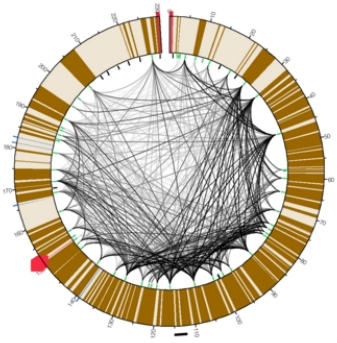



Supplementary Figure 7. Number of intra-chromosomal interactions per *HindIII* fragments as a function of chromosome size. Results indicate relatively similar levels of overall chromosome compaction, irrespective of chromosome size.

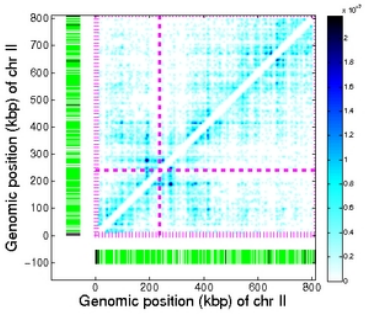
Supplementary Figure 8. Intra-chromosomal interactions in each yeast chromosome. 2D heat maps (left), Circos diagrams (middle), and 3 dimensional pictures (right) were generated using the interactions identified from the HindIII libraries at an FDR threshold of 1%.

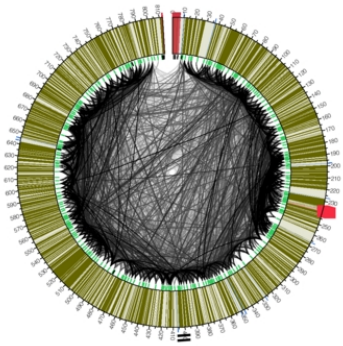
I



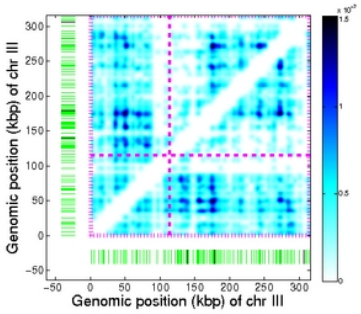



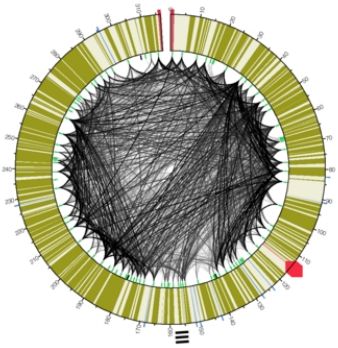
II



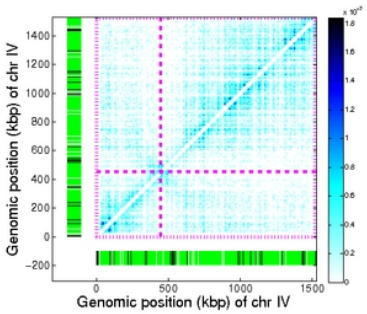
III

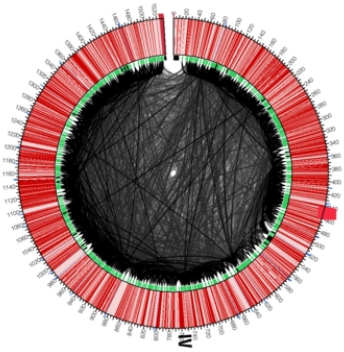


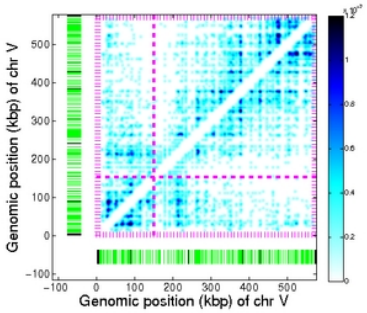
Supplementary Figure 8, continued.

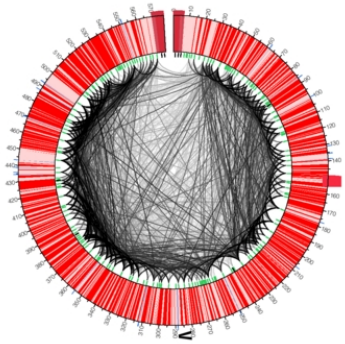
IV



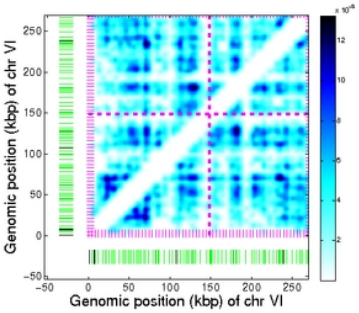



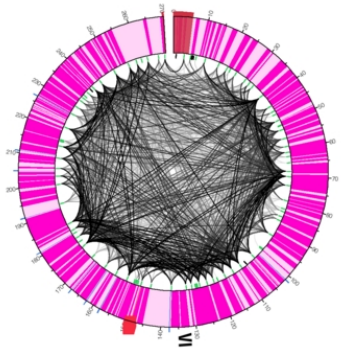

V



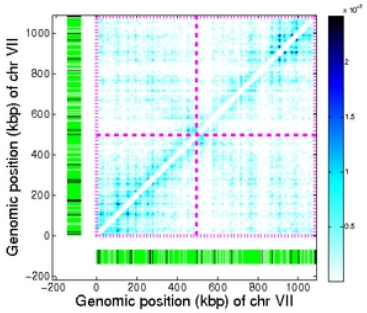



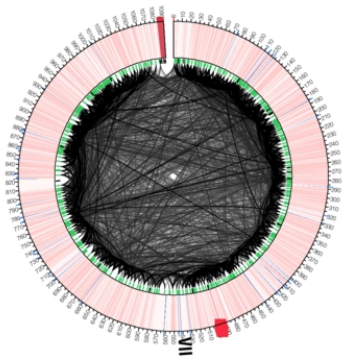

VI



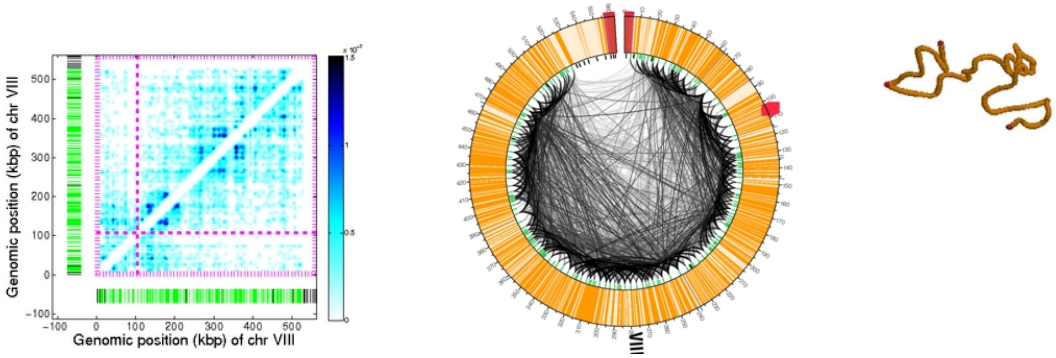

VII



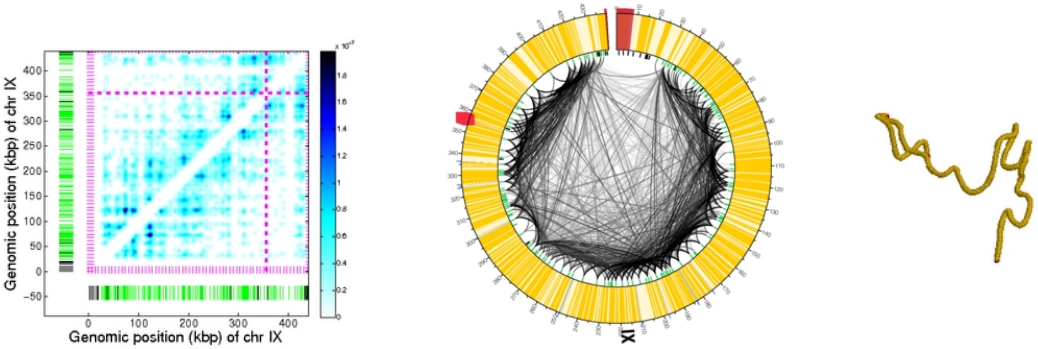




Supplementary Figure 8, continued.

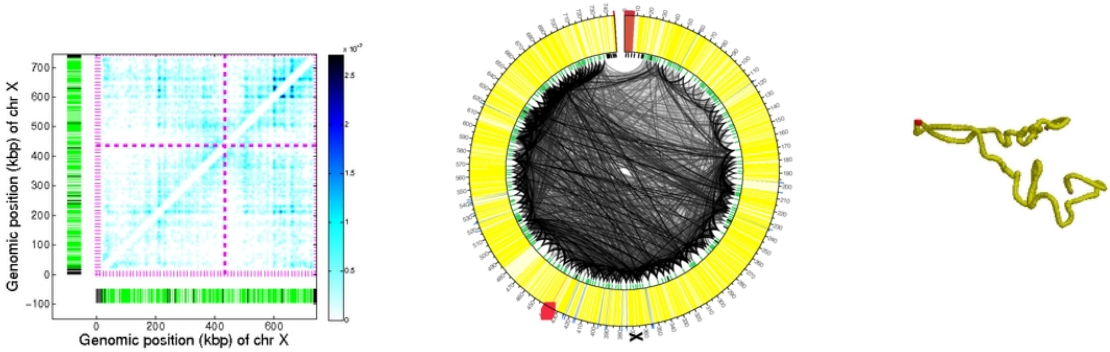
VIII



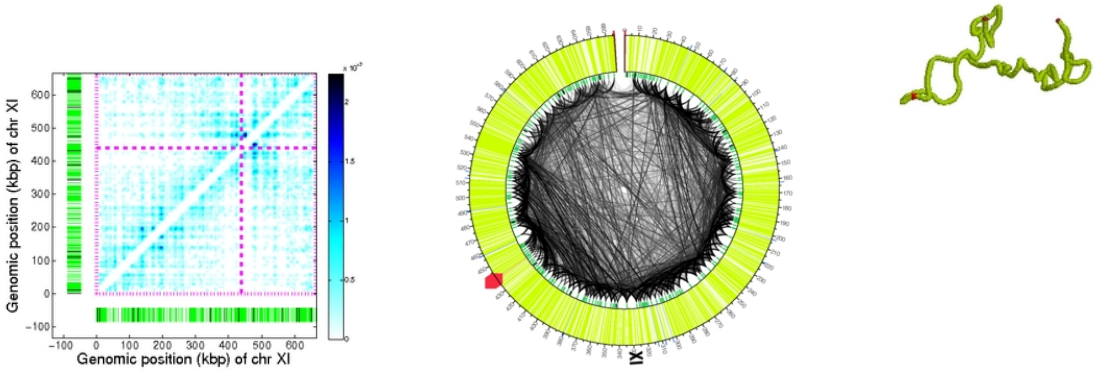

IX




X

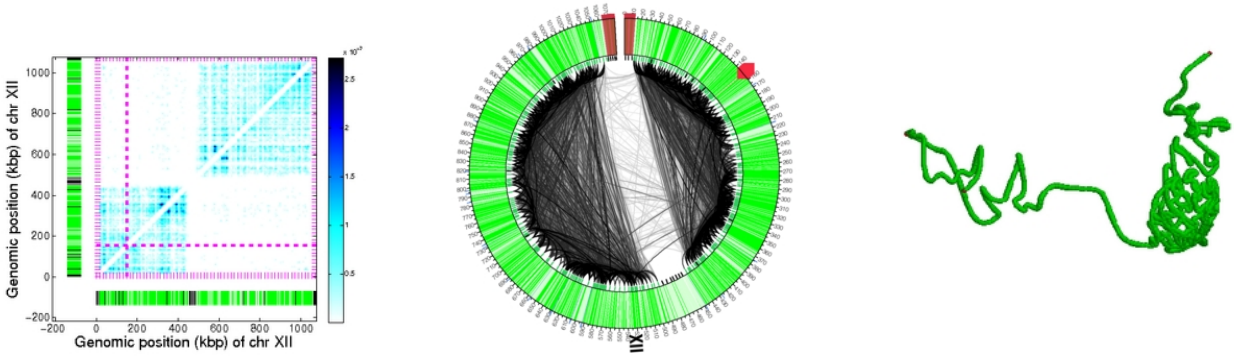



XI

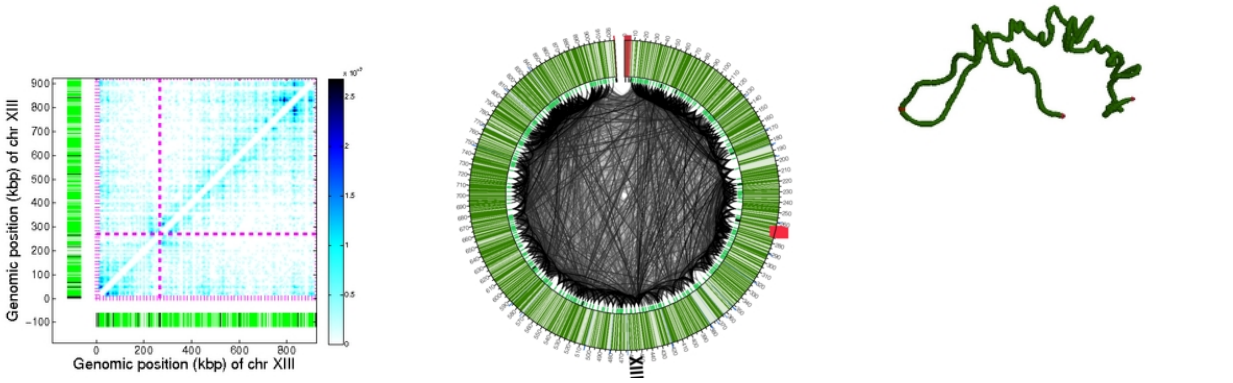


Supplementary Figure 8, continued.

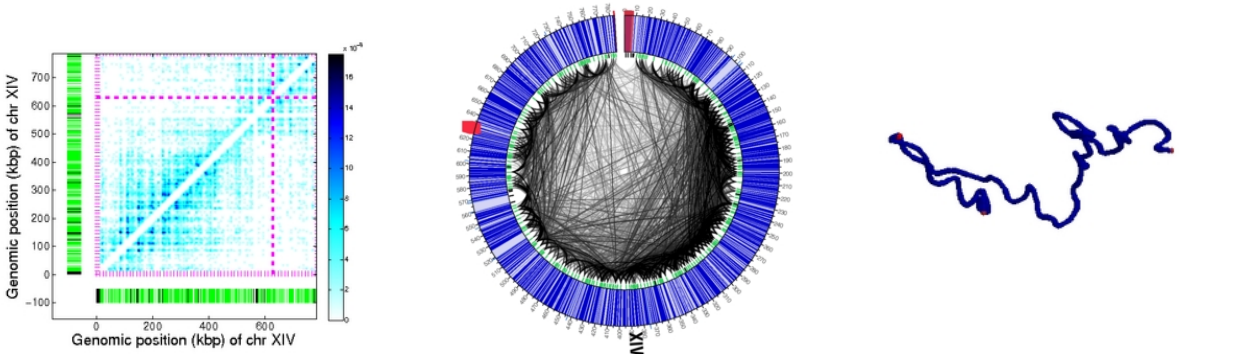
XII



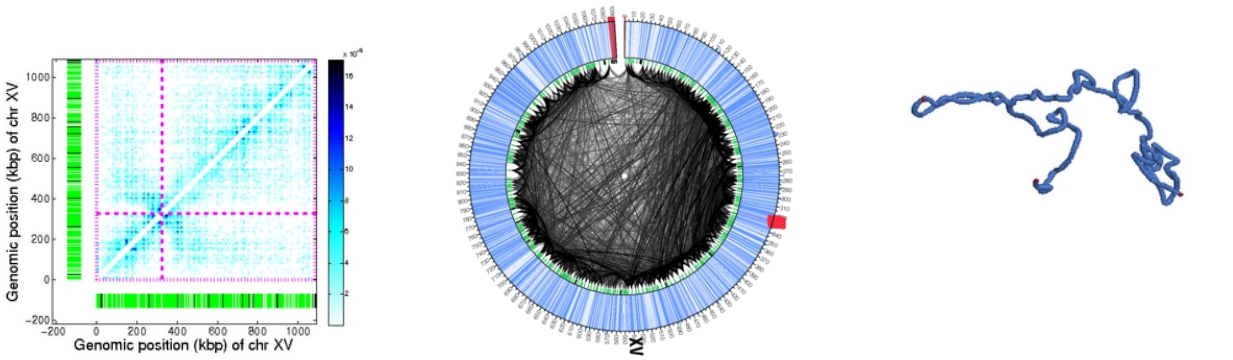

XIII




XIV

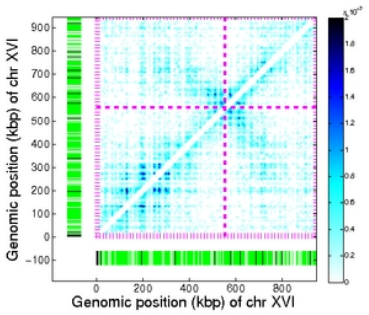


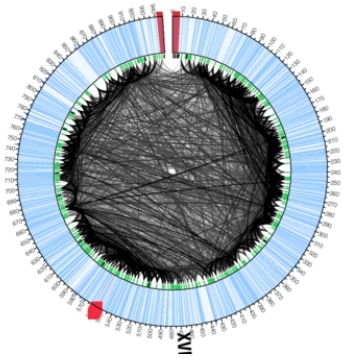

XV



Supplementary Figure 8, continued.

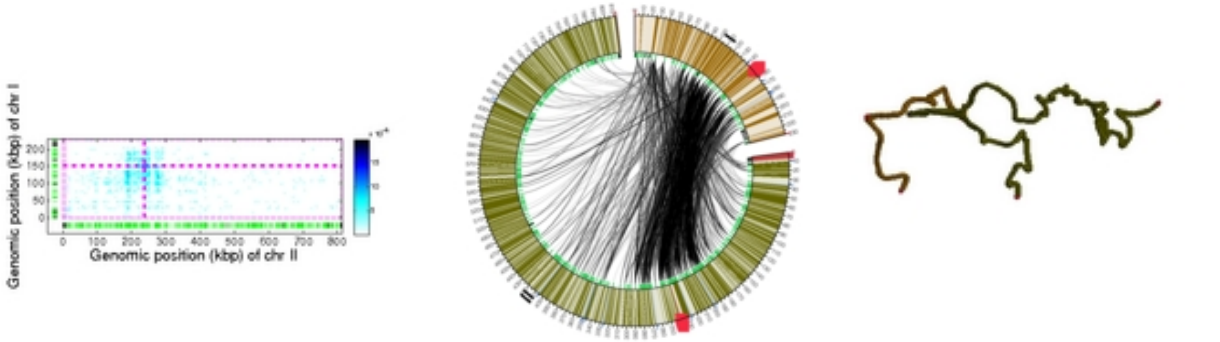
XVI



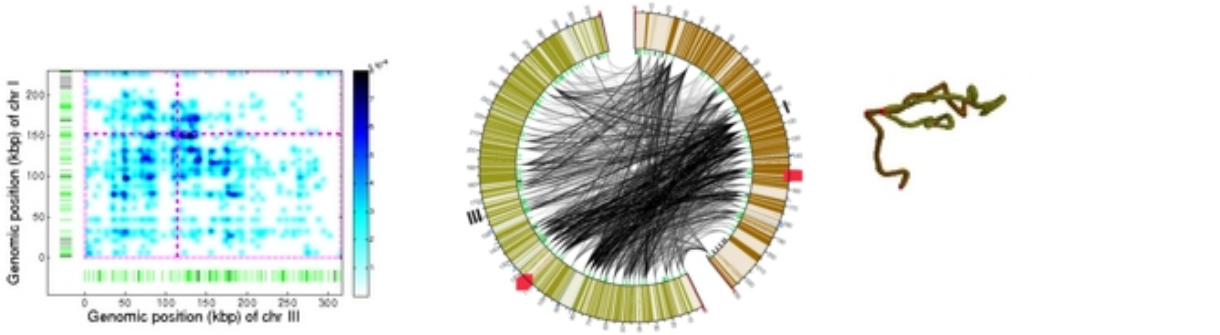




Supplementary Figure 9. Inter-chromosomal interactions between each pair of the 16 yeast chromosomes. 2D heat maps (left), Circos diagrams (middle), and 3 dimensional pictures (right) were generated using the interactions identified from the HindIII libraries at an FDR threshold of 1%.

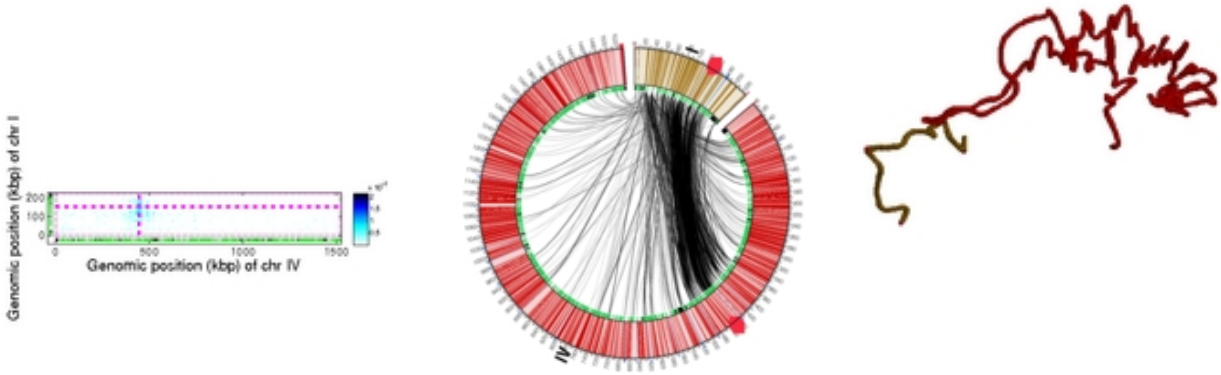
I/II




I/III

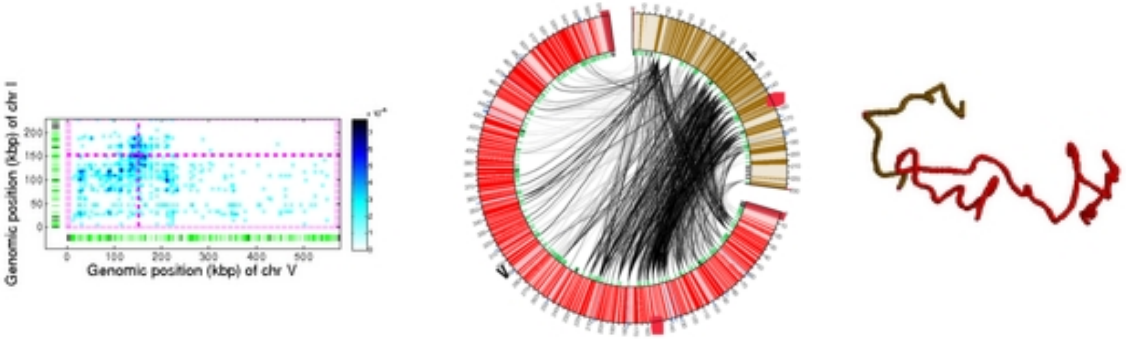



I/IV

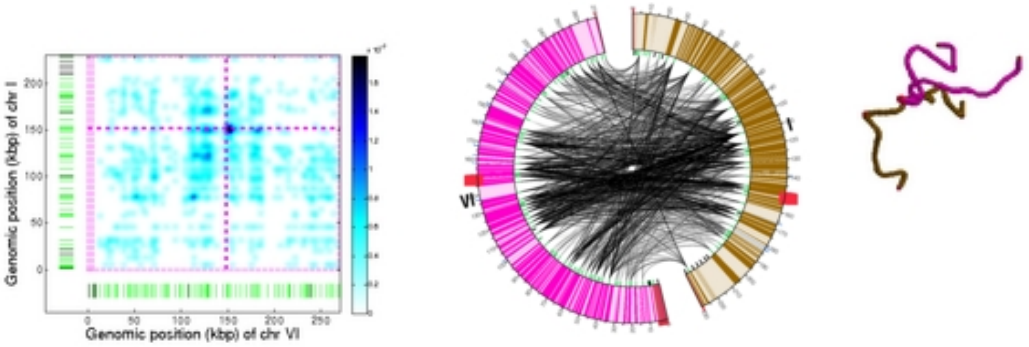


Supplementary Figure 9, continued.

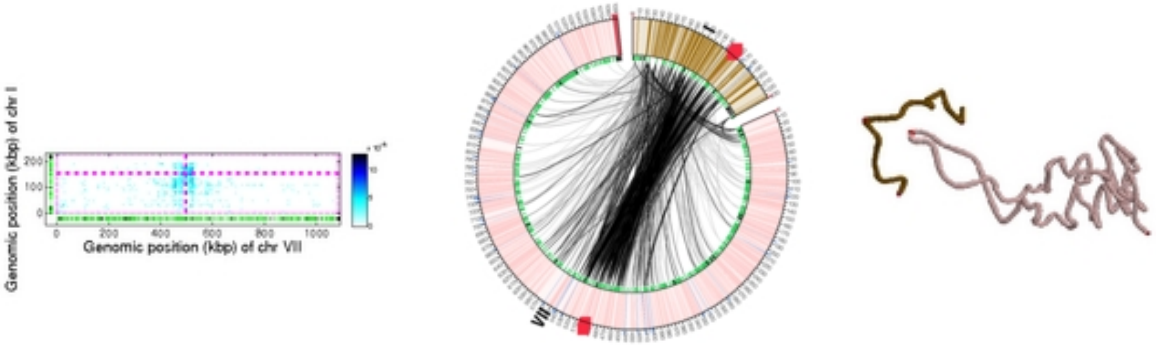
I/V



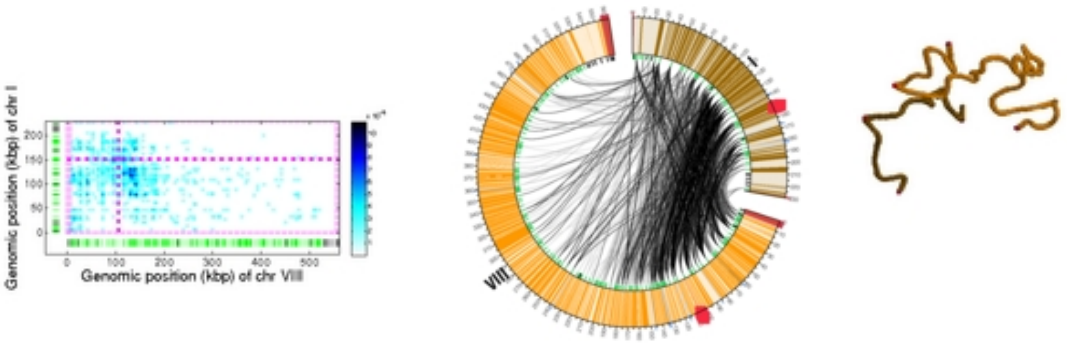

I/VI




I/VII

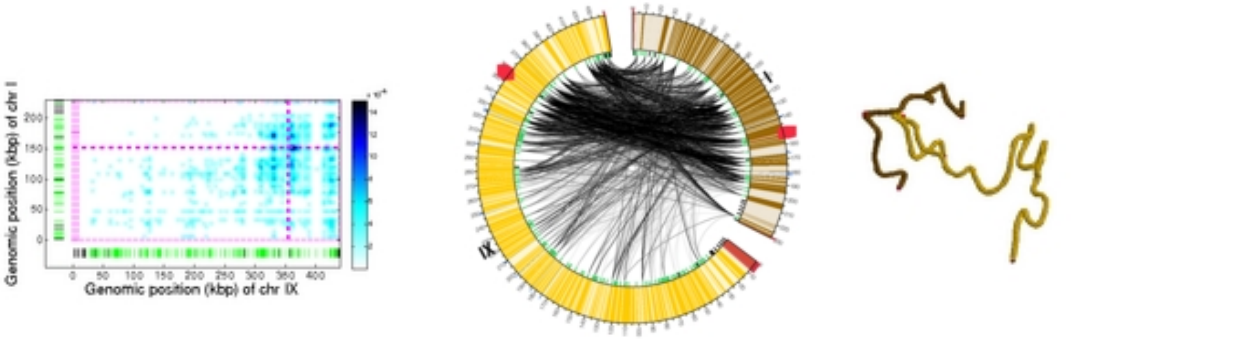



I/VIII

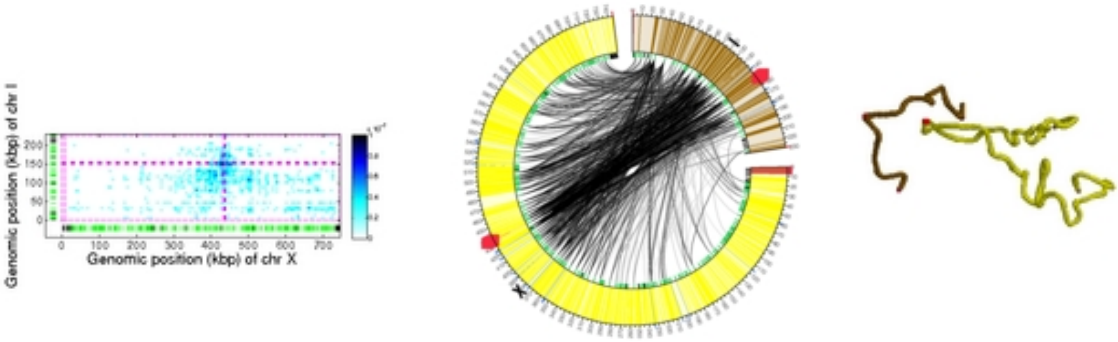


Supplementary Figure 9, continued.

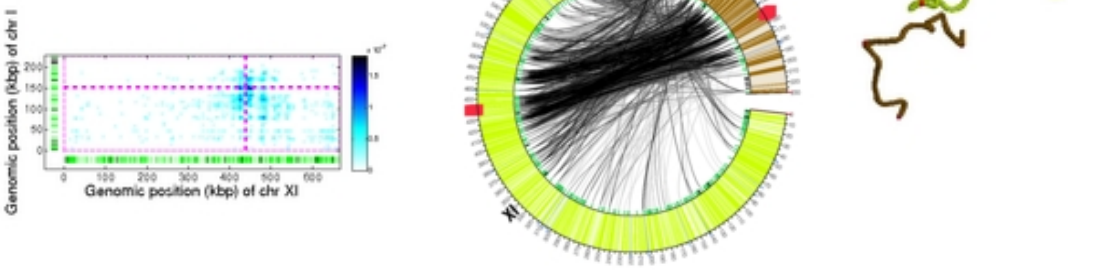
I/IX



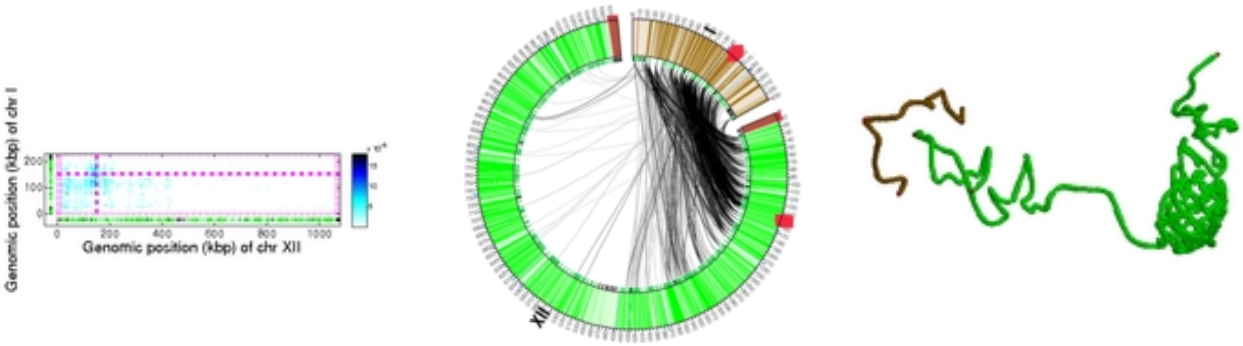

I/X




I/XI

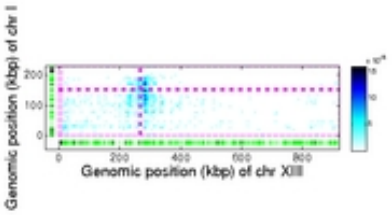


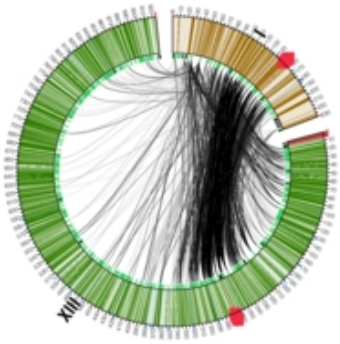

I/XII



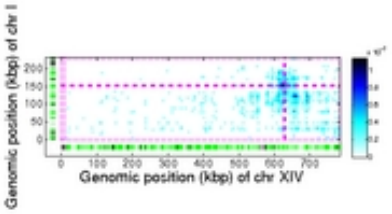
Supplementary Figure 9, continued.

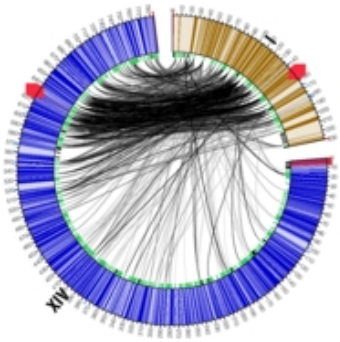
I/XIII



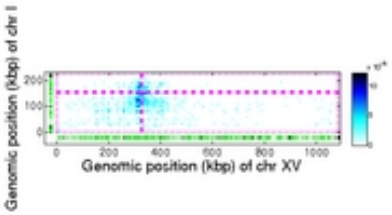



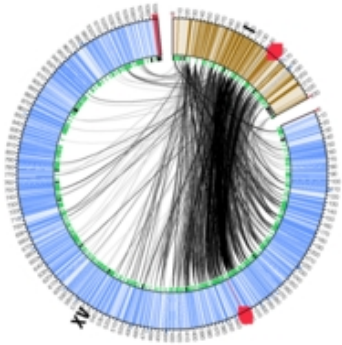

I/XIV



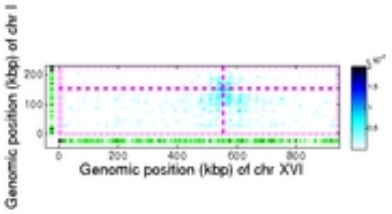



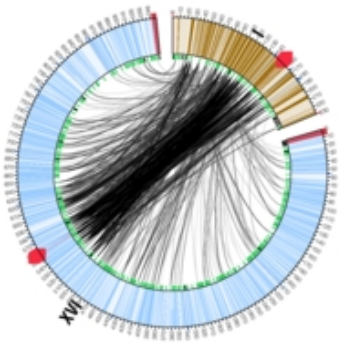

I/XV



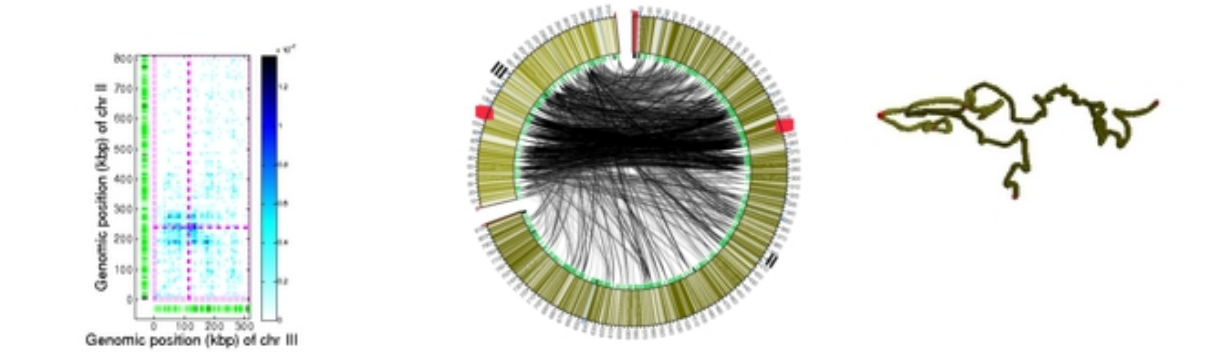

I/XVI



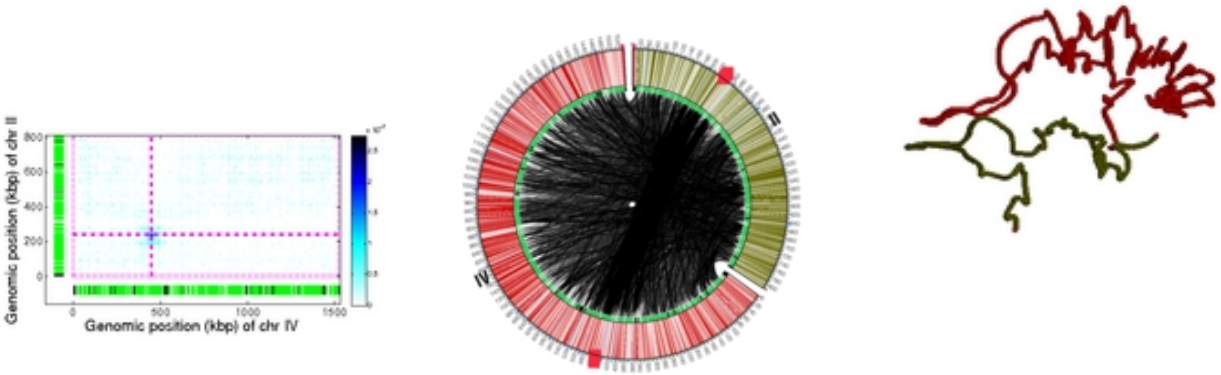




Supplementary Figure 9, continued.

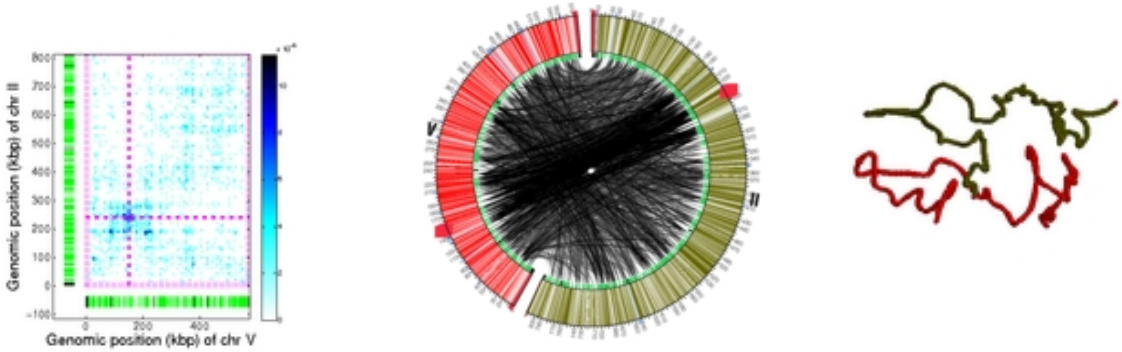
II/III



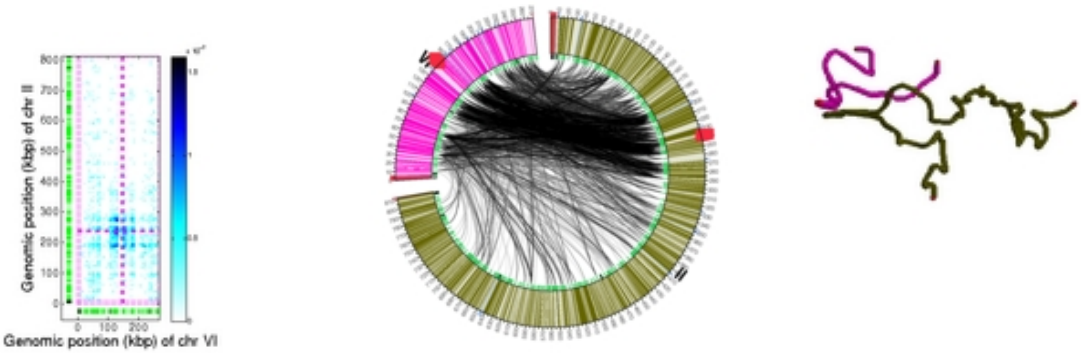

II/IV




II/V

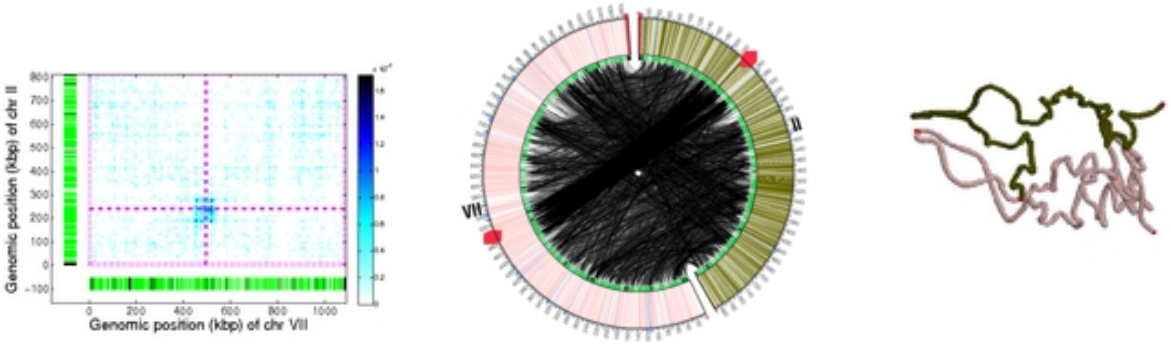



II/VI

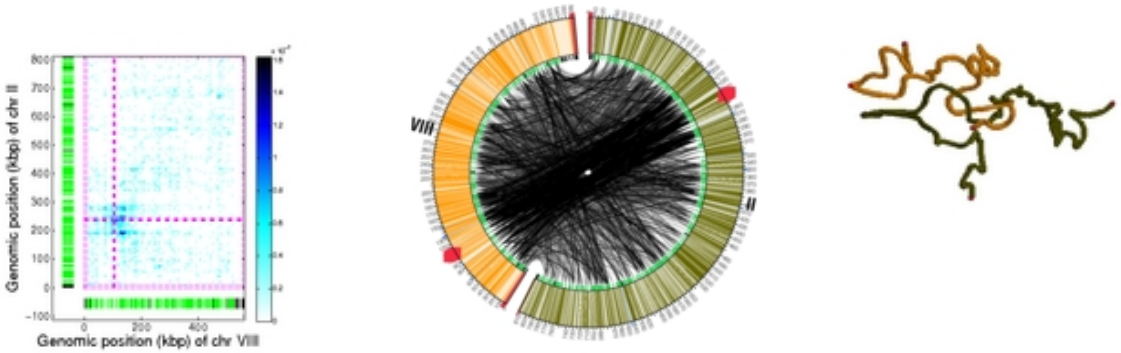


Supplementary Figure 9, continued.

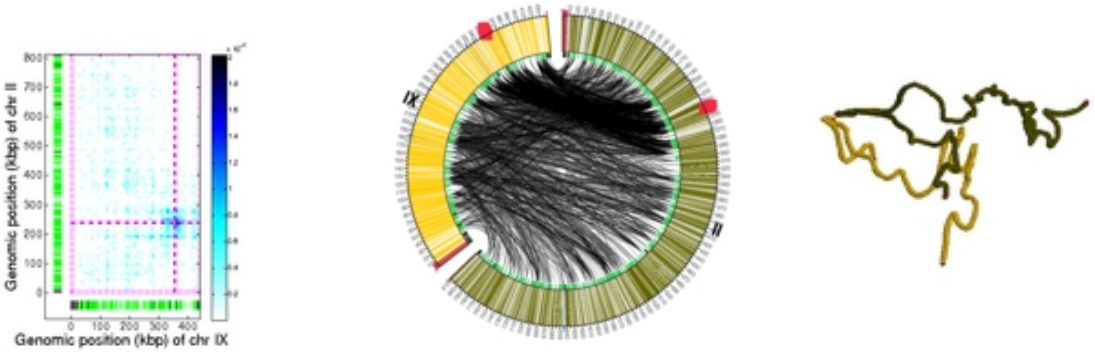
II/VII



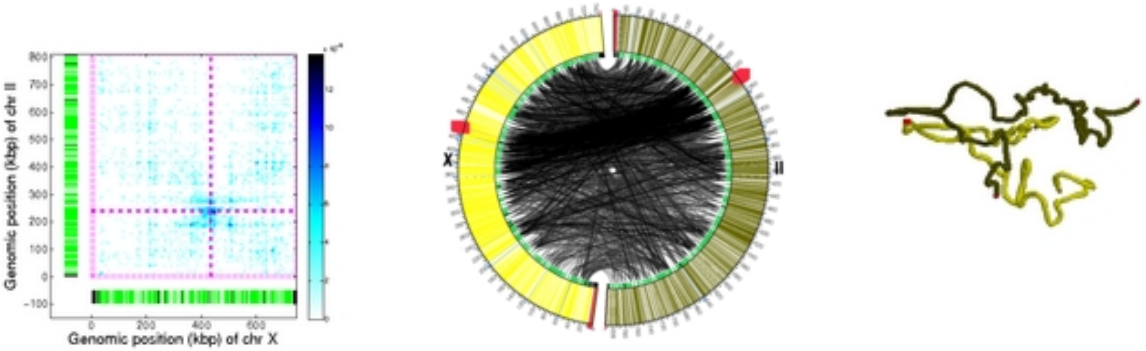

II/VIII




II/IX

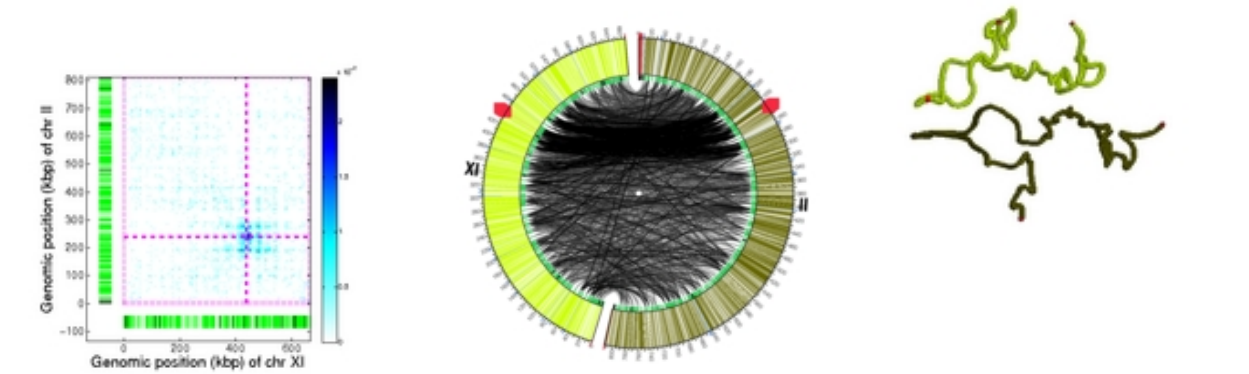



II/X

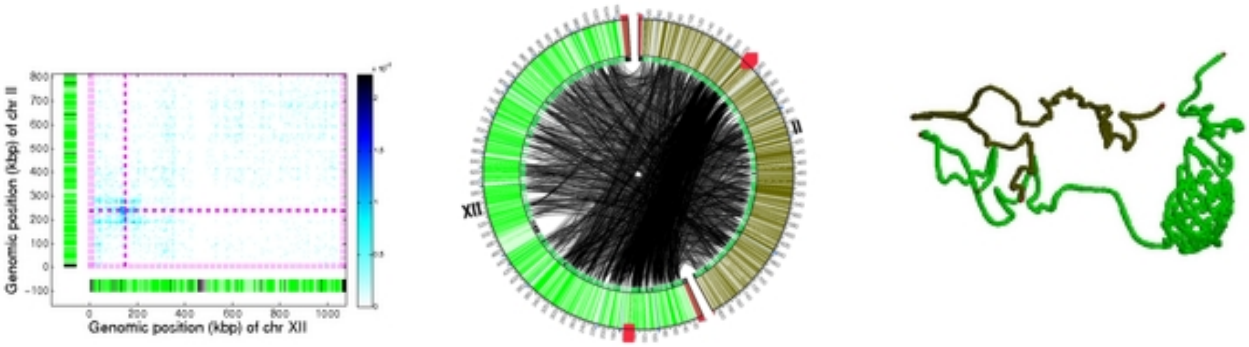


Supplementary Figure 9, continued.

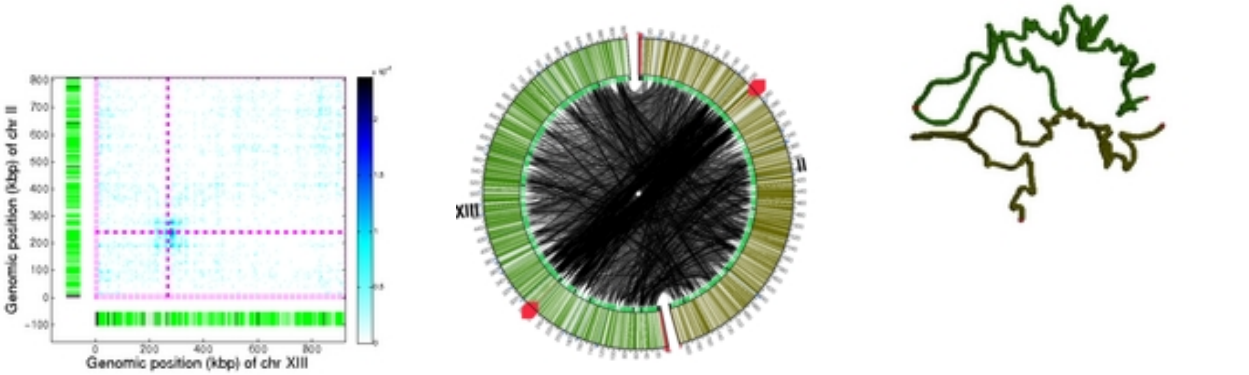
II/XI



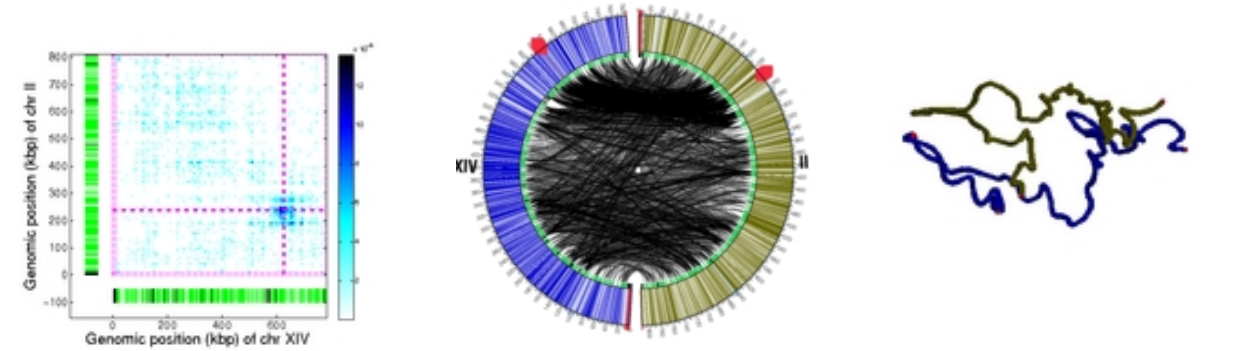

II/XII




II/XIII

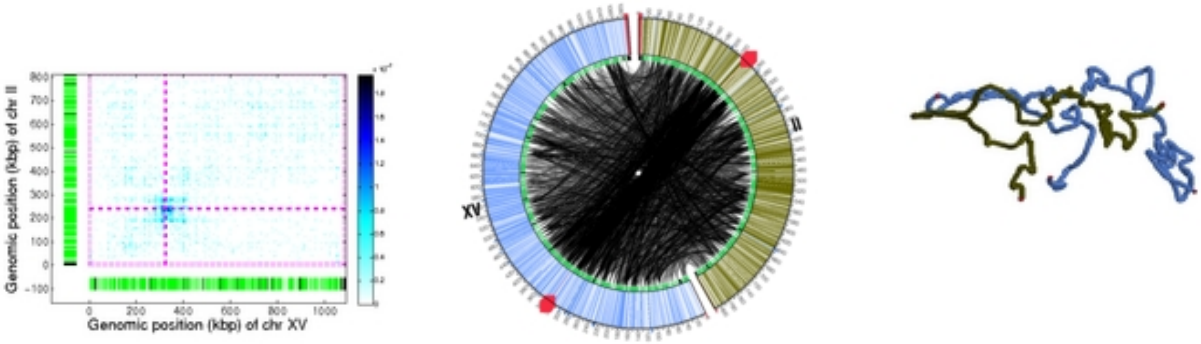



II/XIV

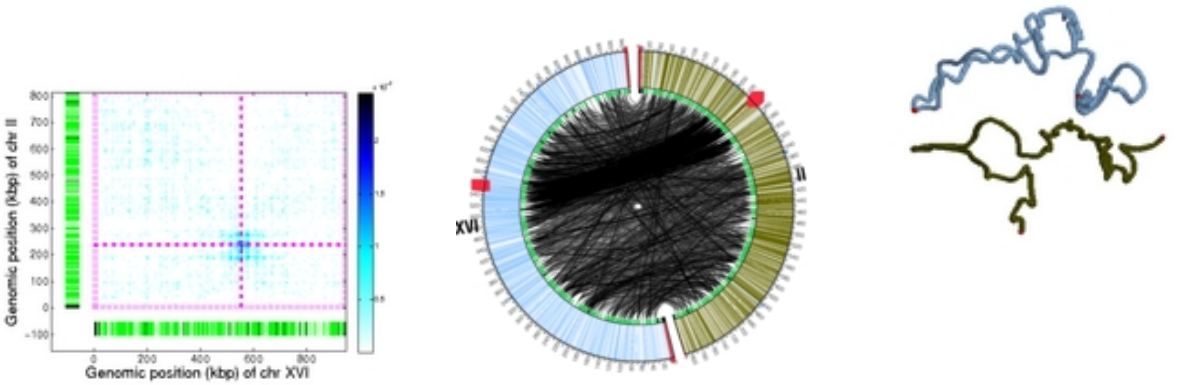


Supplementary Figure 9, continued.

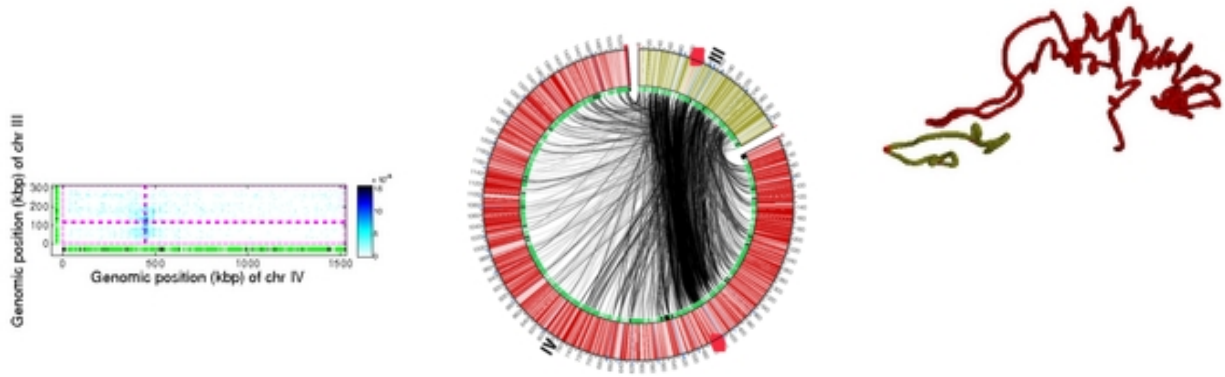
II/XV



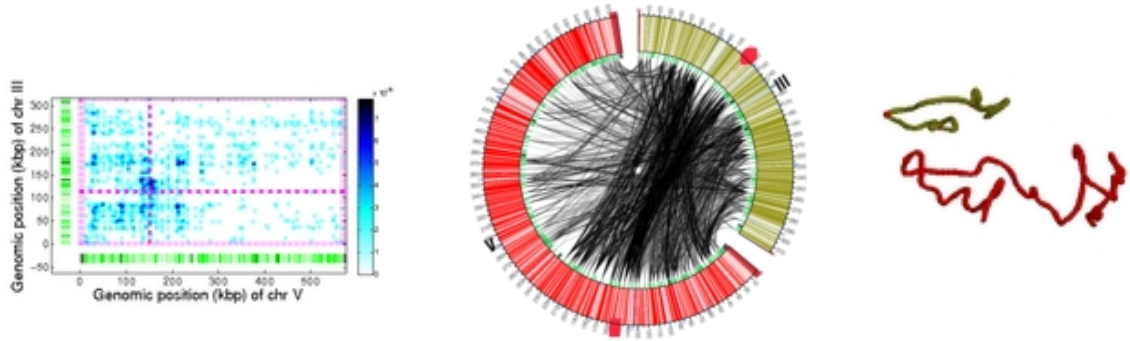

II/XVI




III/IV

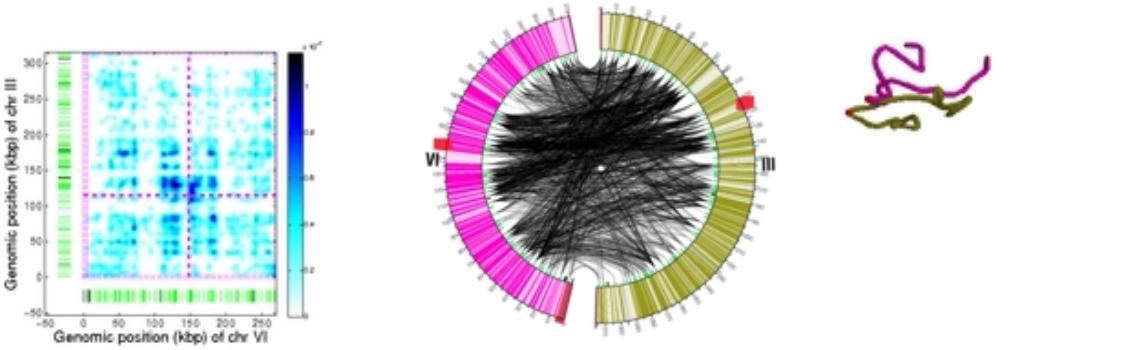



III/V

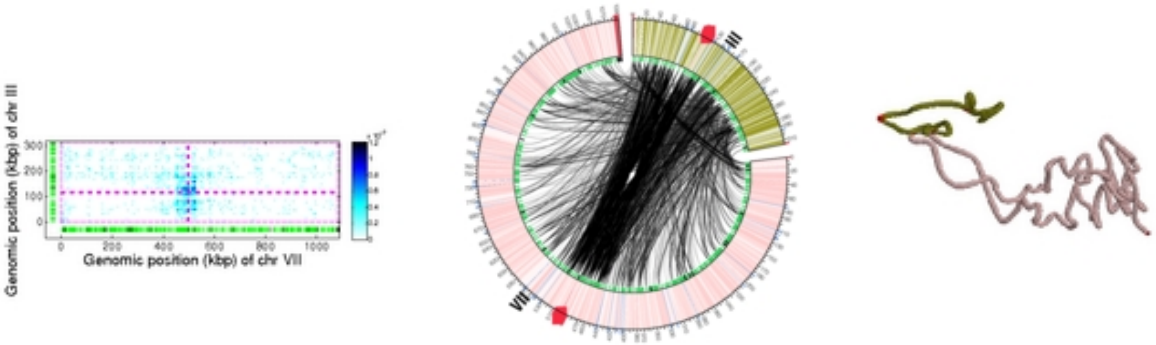


Supplementary Figure 9, continued.

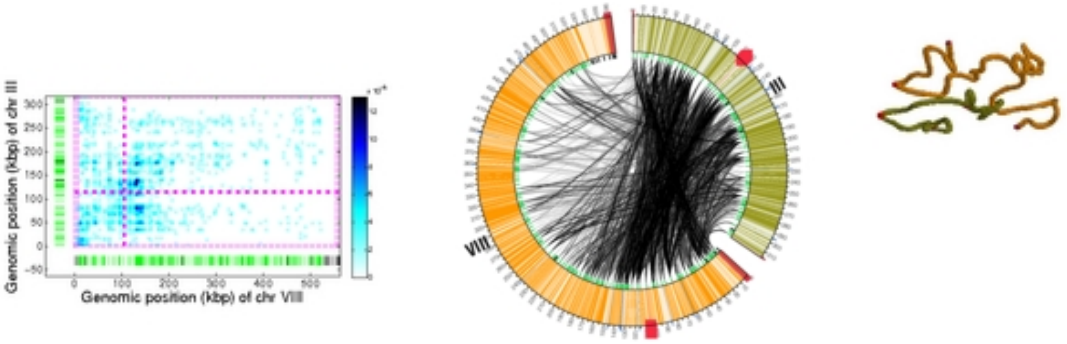
III/VI



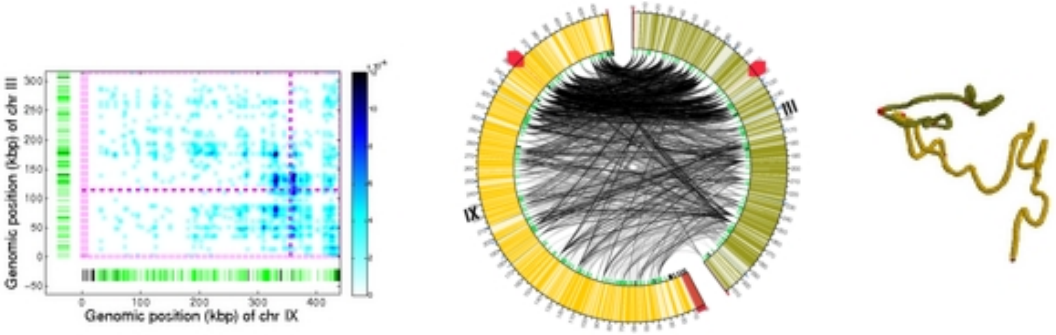

III/VII




III/VIII

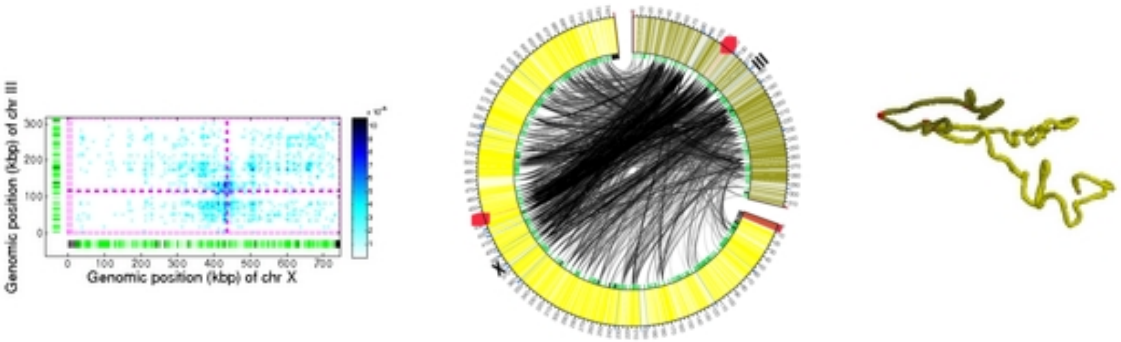



III/IX

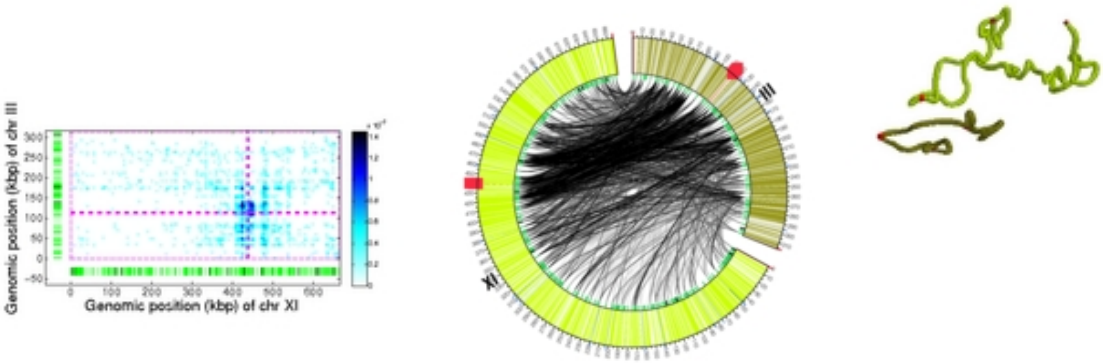


Supplementary Figure 9, continued.

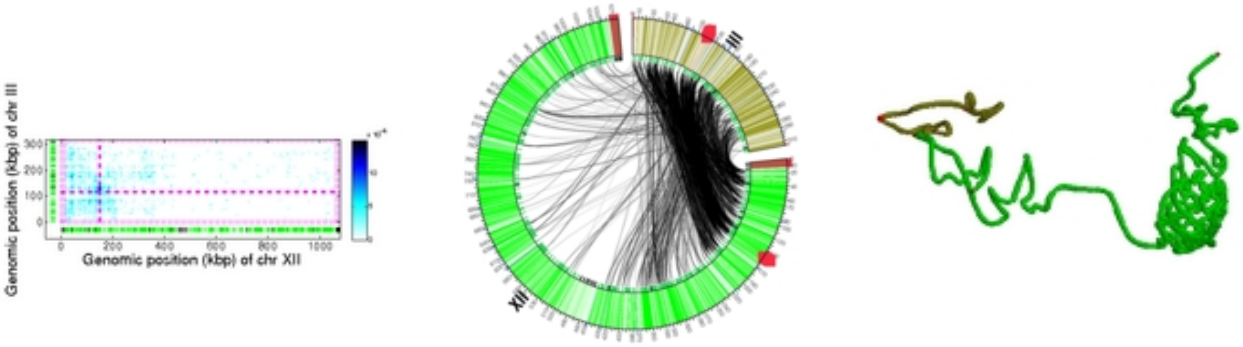
III/X



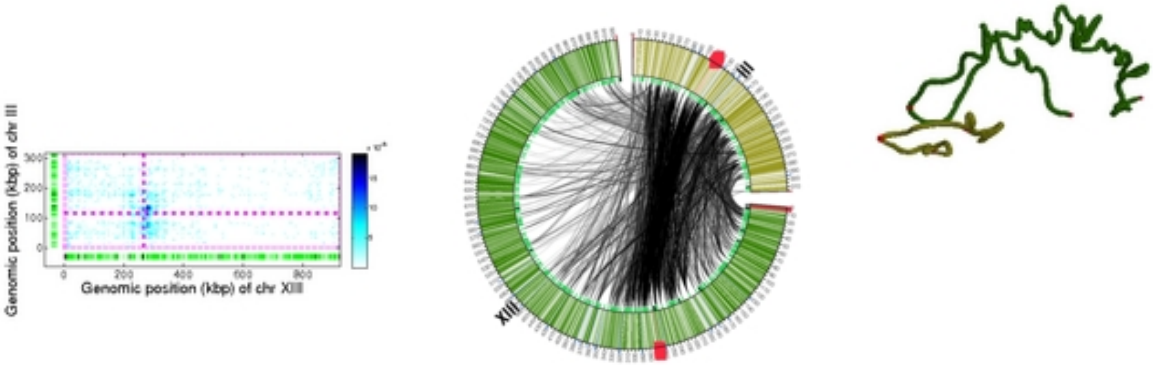

III/XI




III/XII

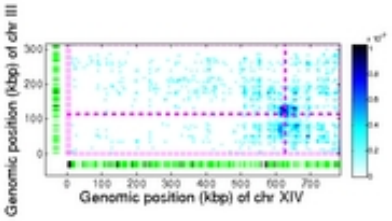


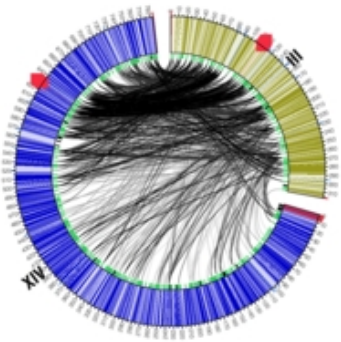

III/XIII



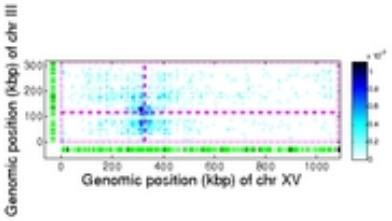
Supplementary Figure 9, continued.

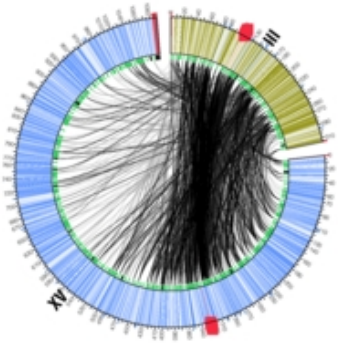
III/XIV



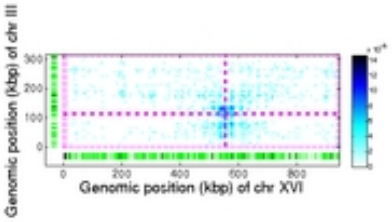



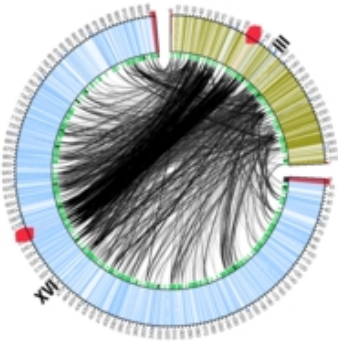

III/XV



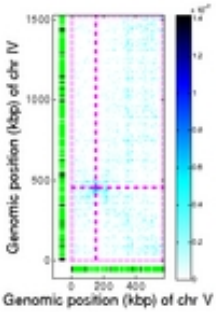



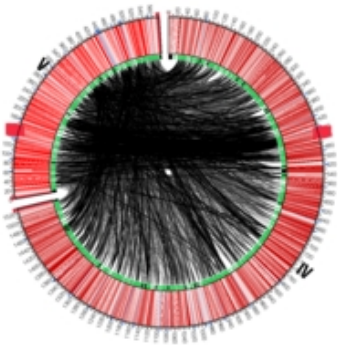

III/XVI



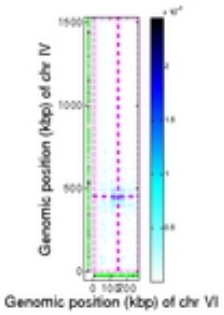

IV/V

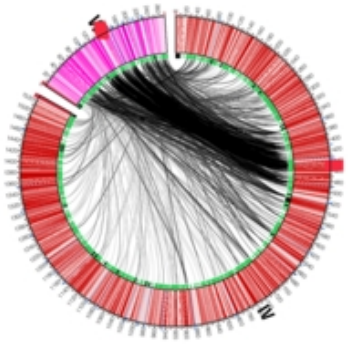


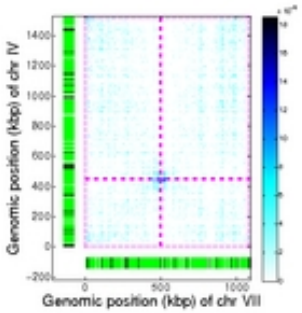

Supplementary Figure 9, continued.

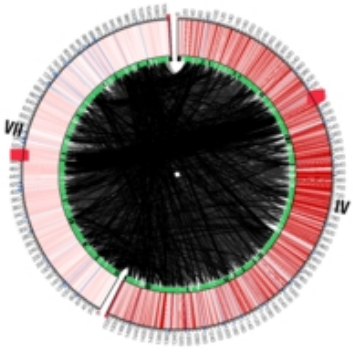
IV/VI



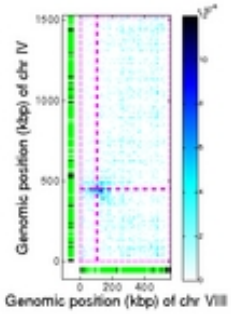



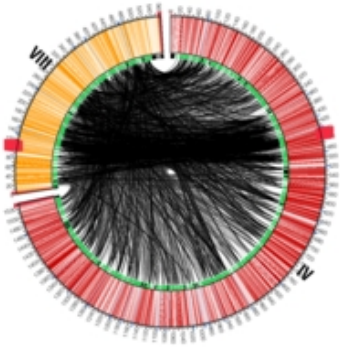

IV/VII



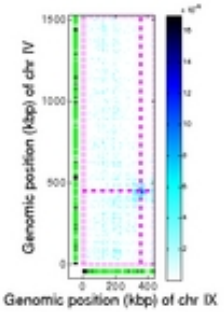



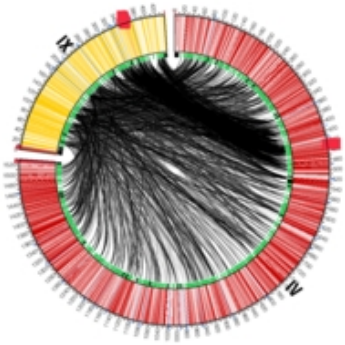

IV/VIII



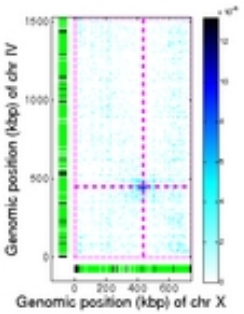

IV/IX

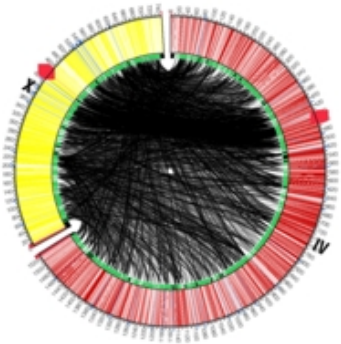


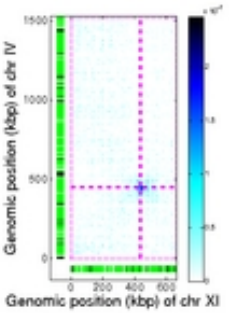

Supplementary Figure 9, continued.

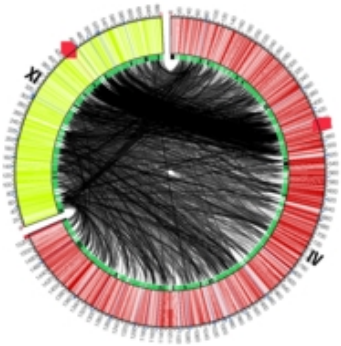
IV/X



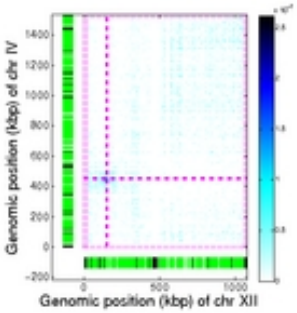



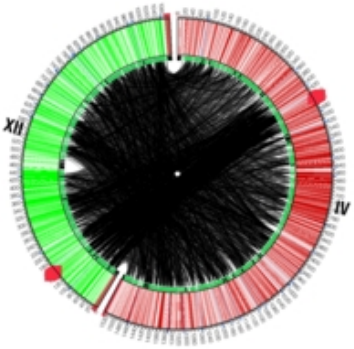

IV/XI



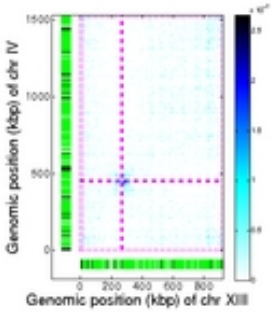



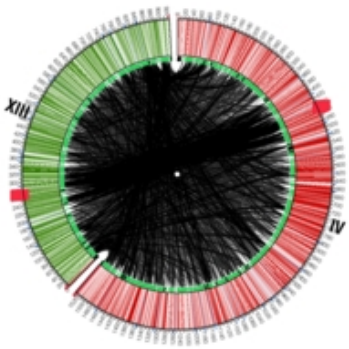

IV/XII

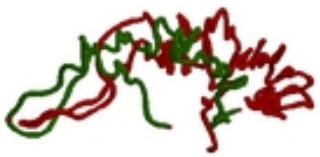



IV/XIII

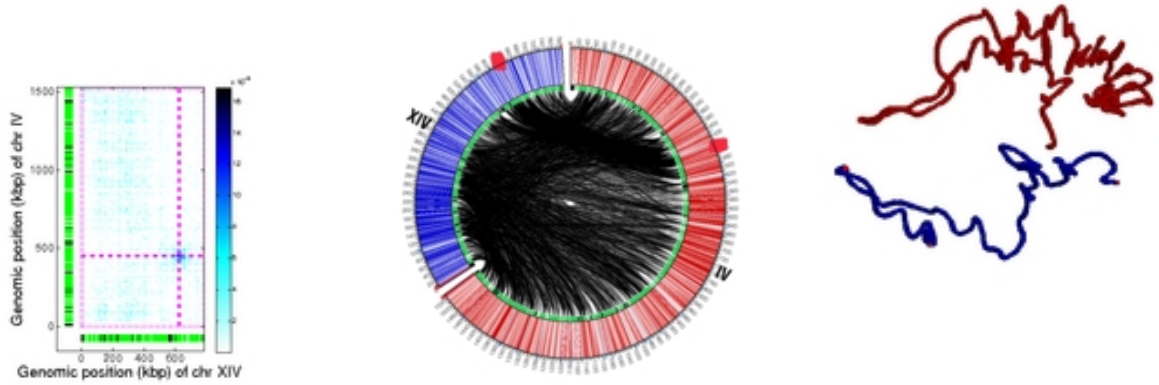


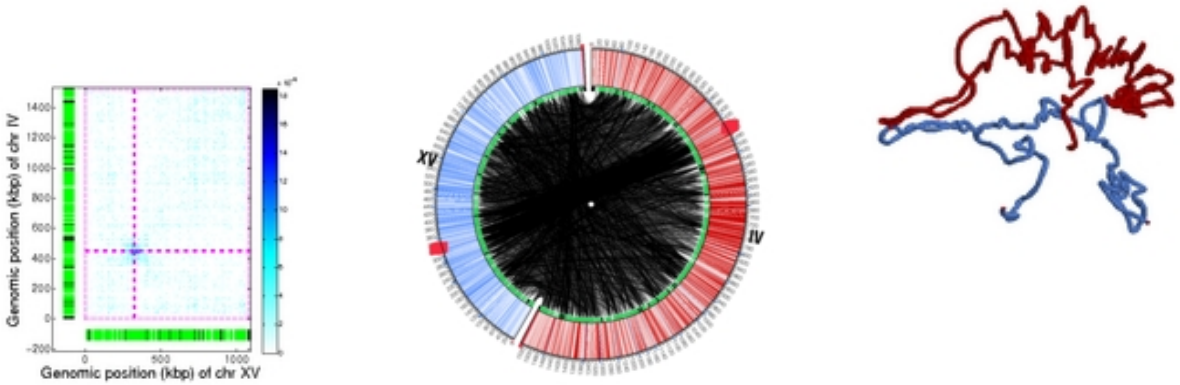



Supplementary Figure 9, continued.

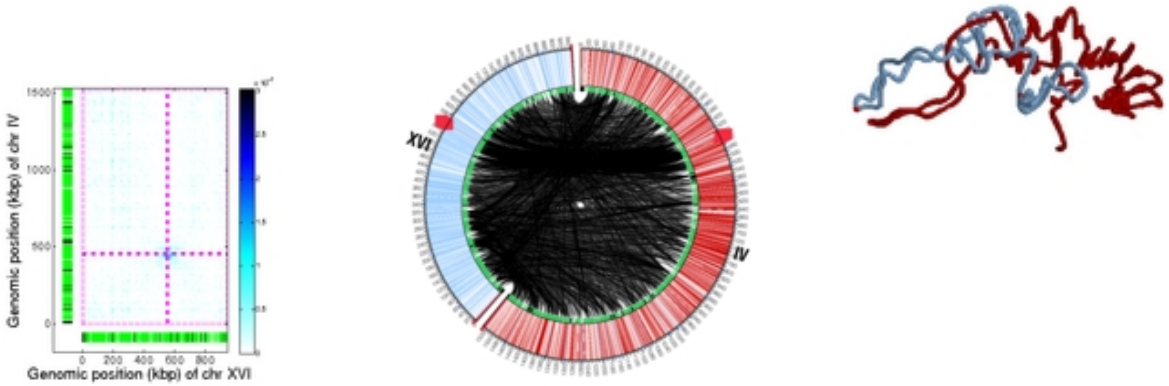
IV/XIV



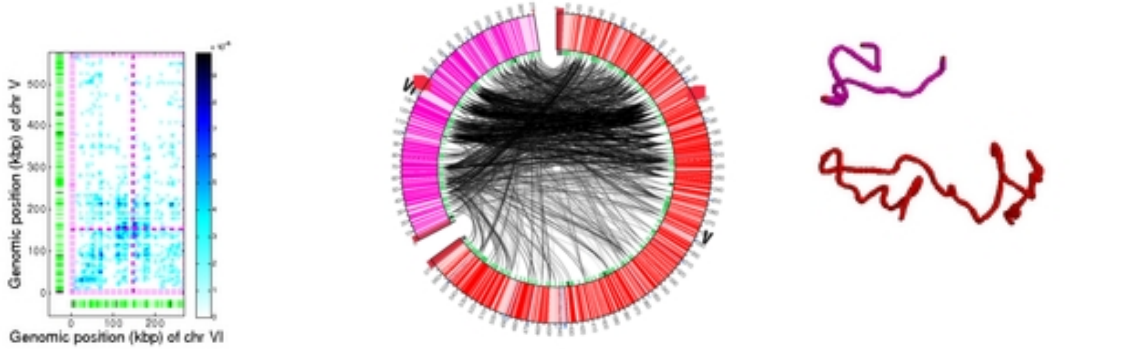

IV/XV




IV/XVI

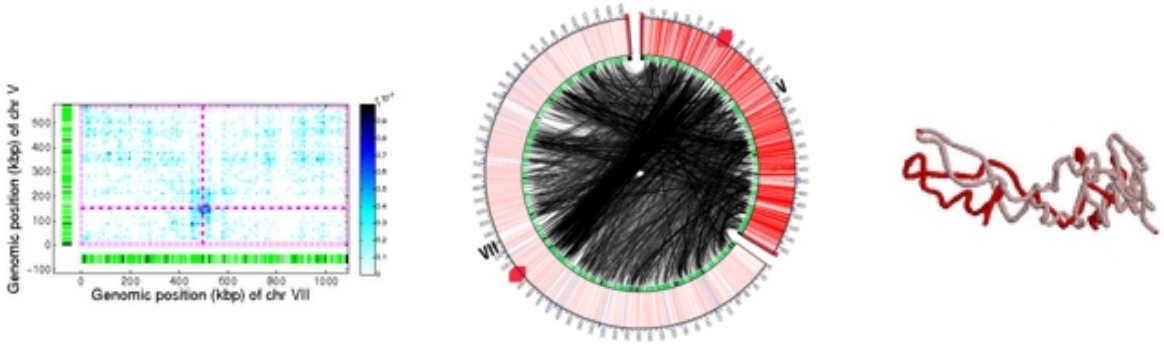



V/VI

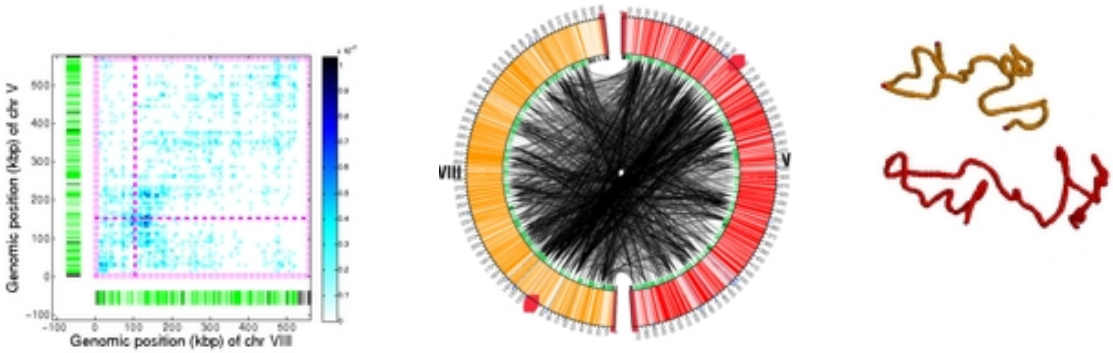


Supplementary Figure 9, continued.

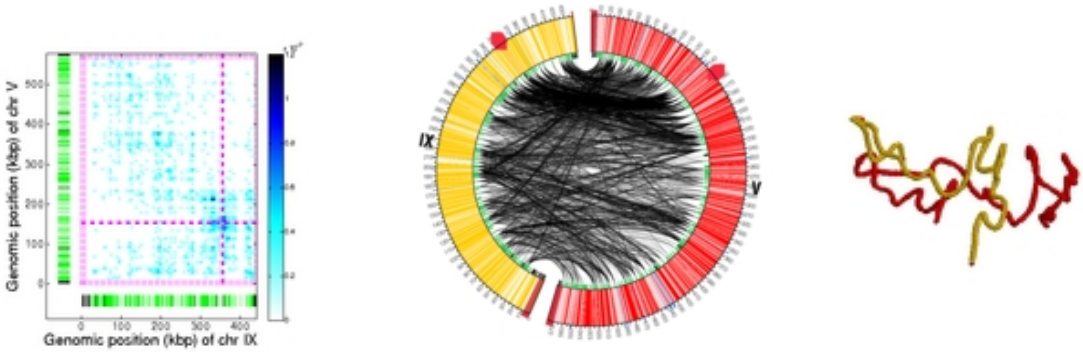
V/VII



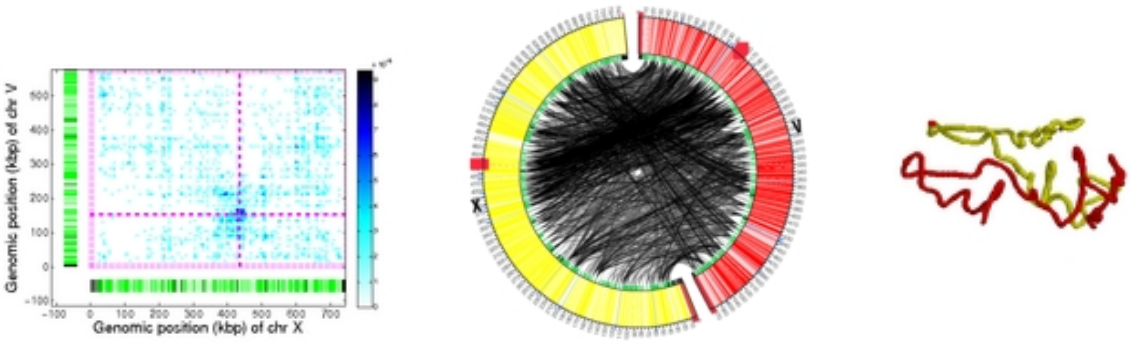

V/VIII




V/IX

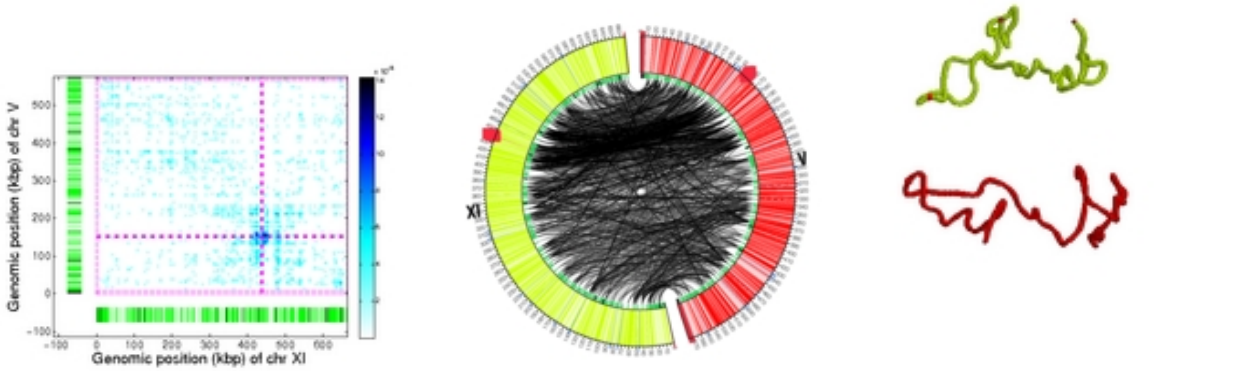



V/X

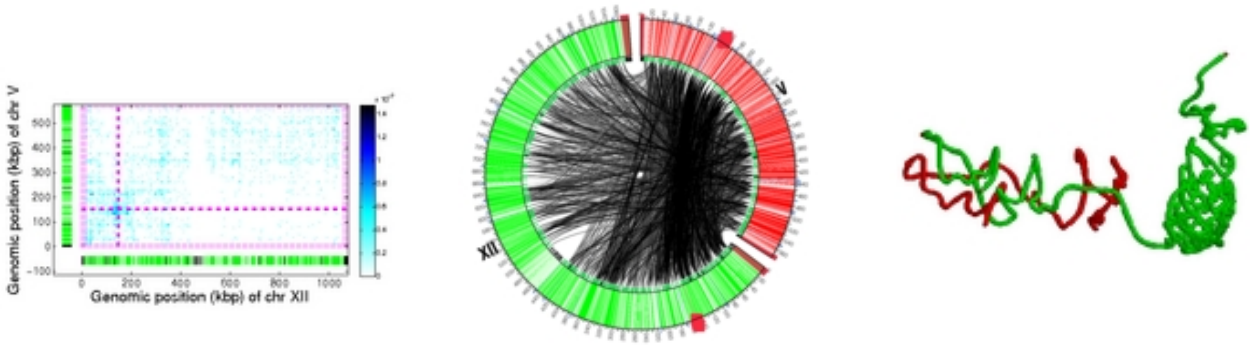


Supplementary Figure 9, continued.

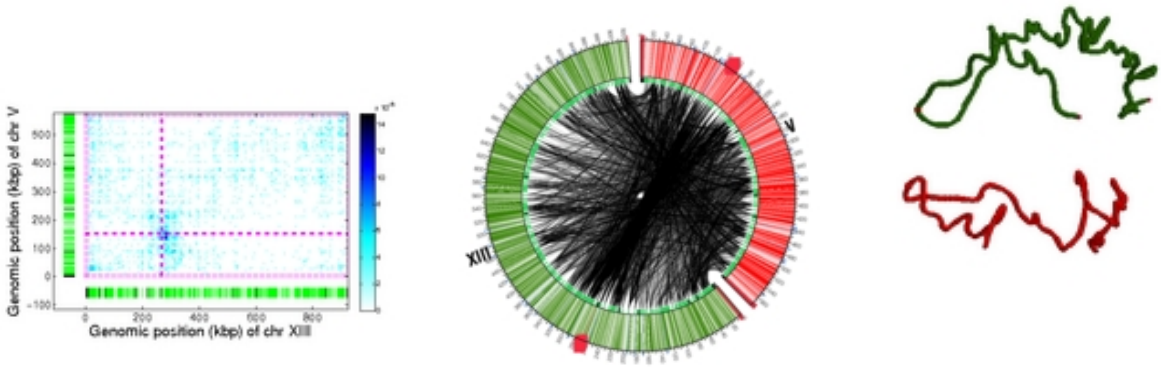
V/XI



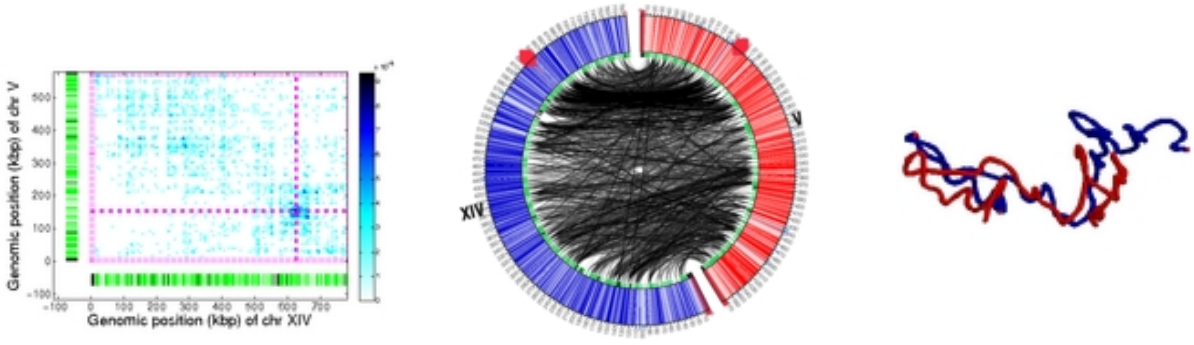

V/XII




V/XIII

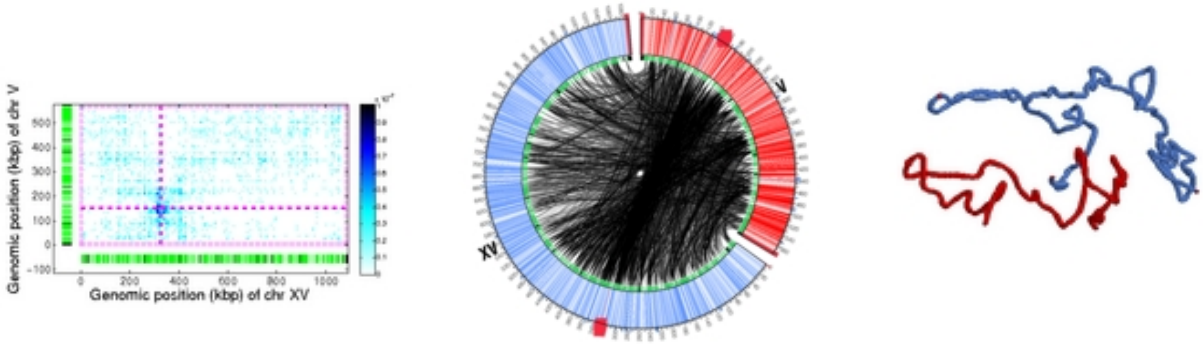



V/XIV

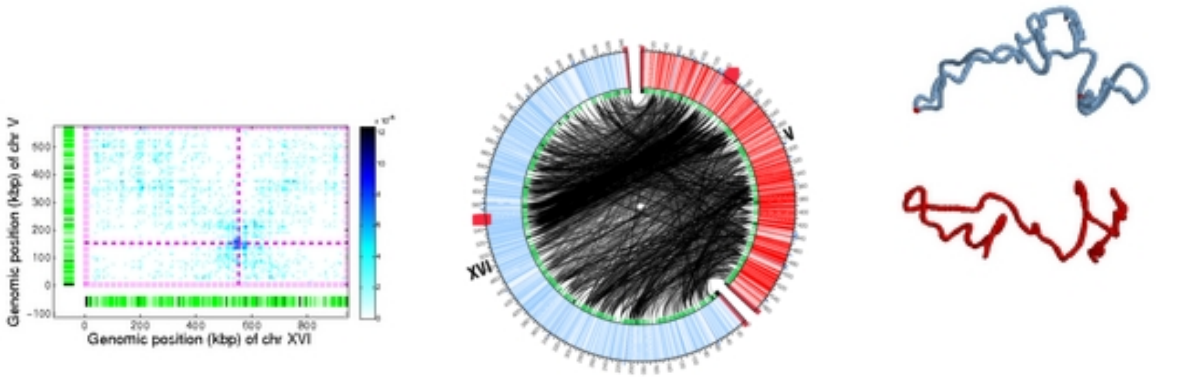


Supplementary Figure 9, continued.

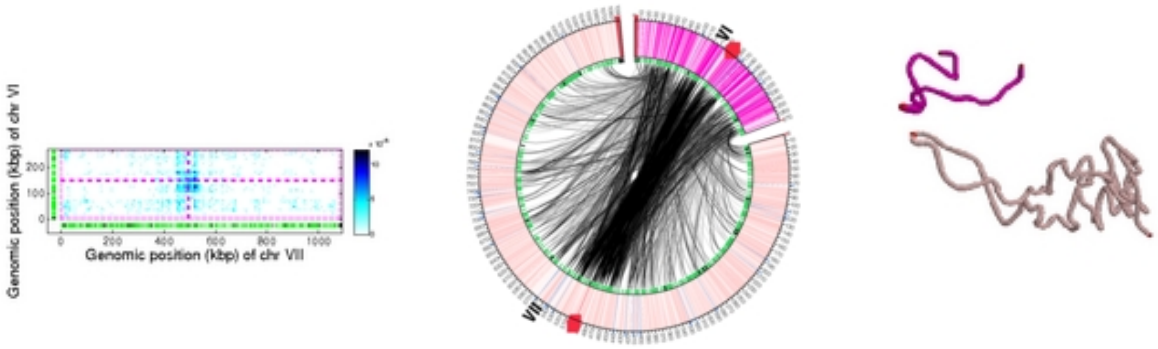
V/XV



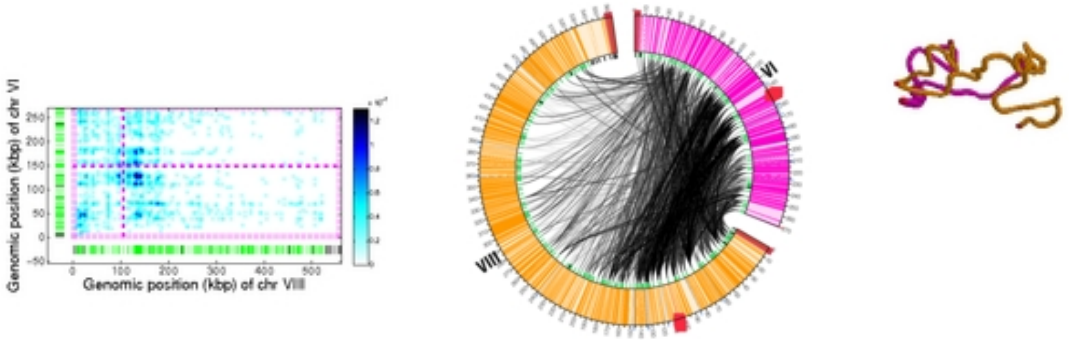

V/XVI




VI/VII

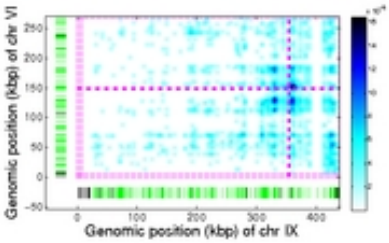


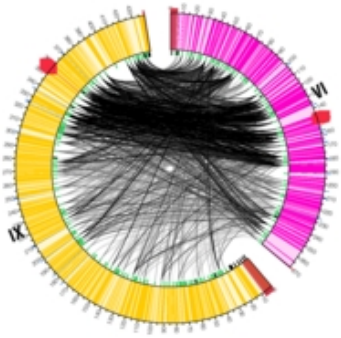

VI/VIII



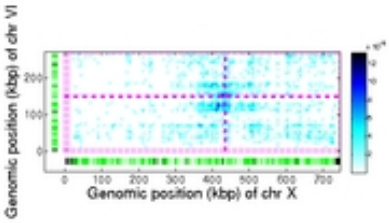
Supplementary Figure 9, continued.

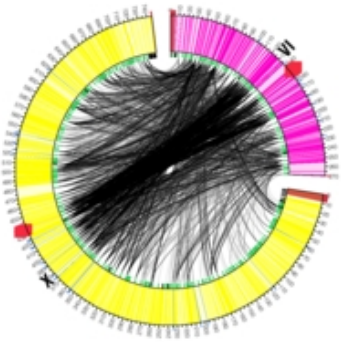
VI/IX



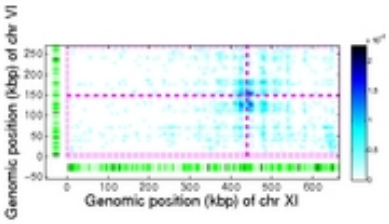



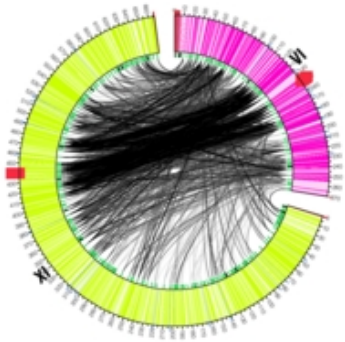

VI/X



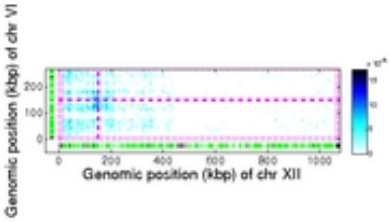



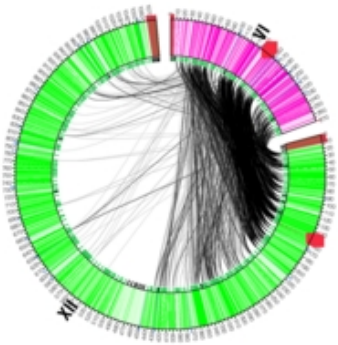

VI/XI



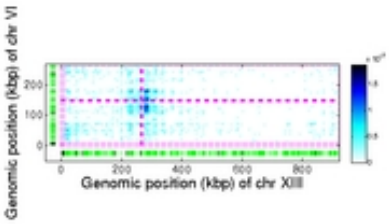

VI/XII

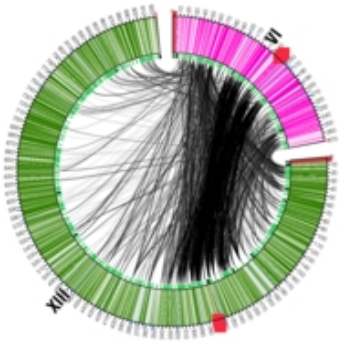


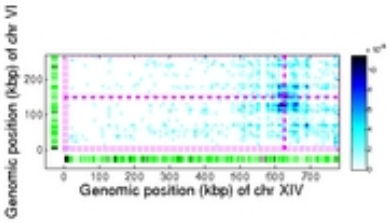

Supplementary Figure 9, continued.

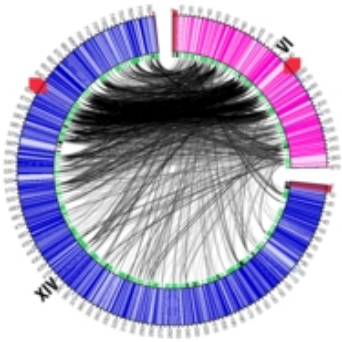
VI/XIII



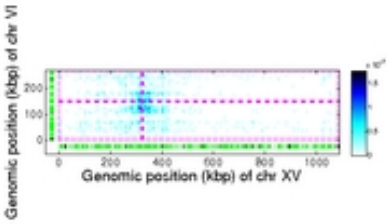



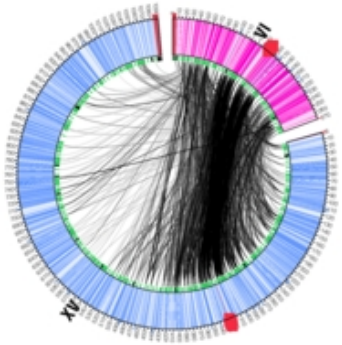

VI/XIV



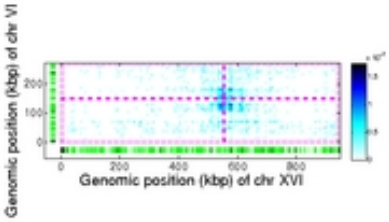



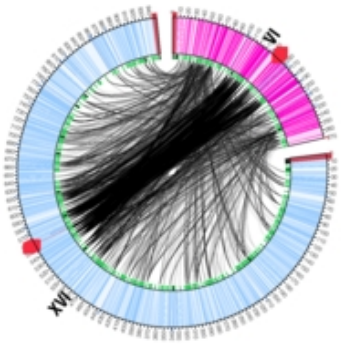

VI/XV



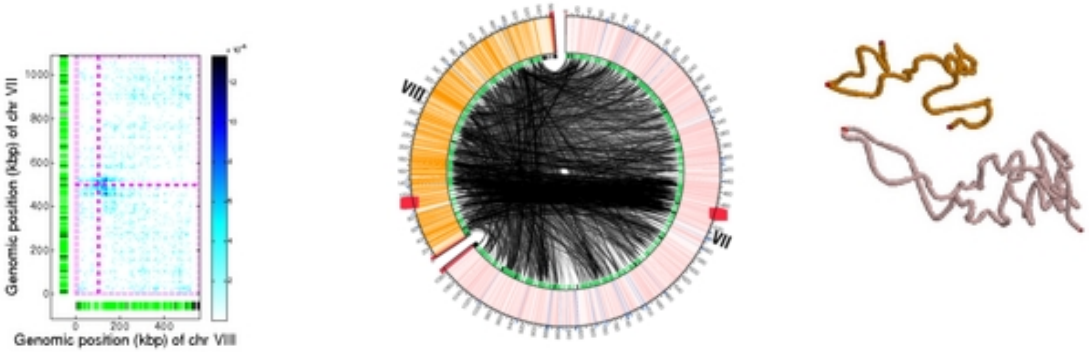

VI/XVI



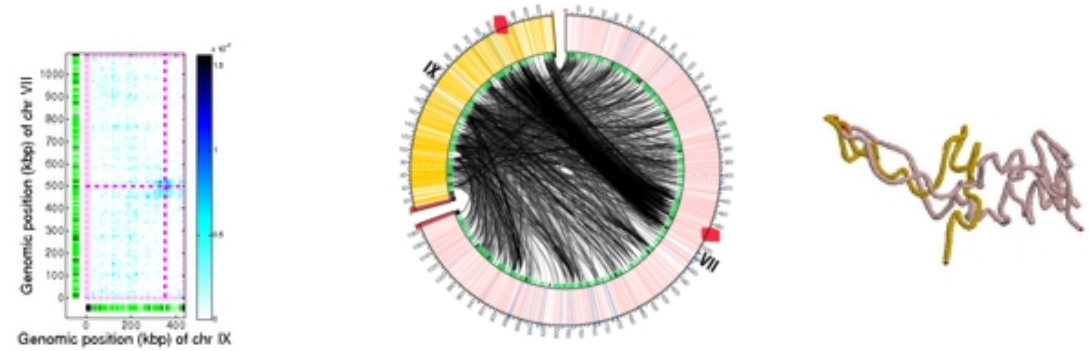




Supplementary Figure 9, continued.

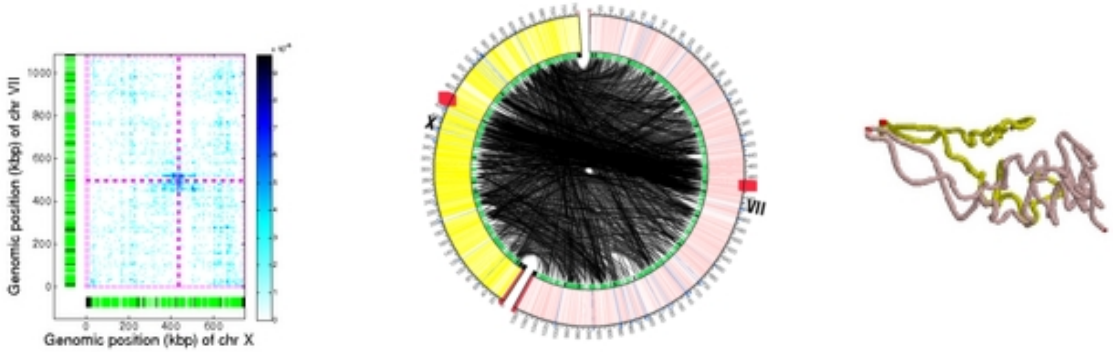
VII/VIII



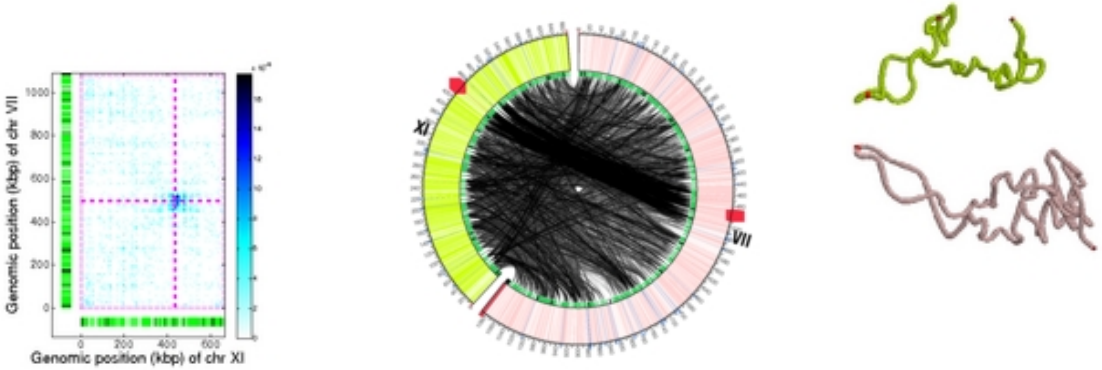

VII/IX




VII/X

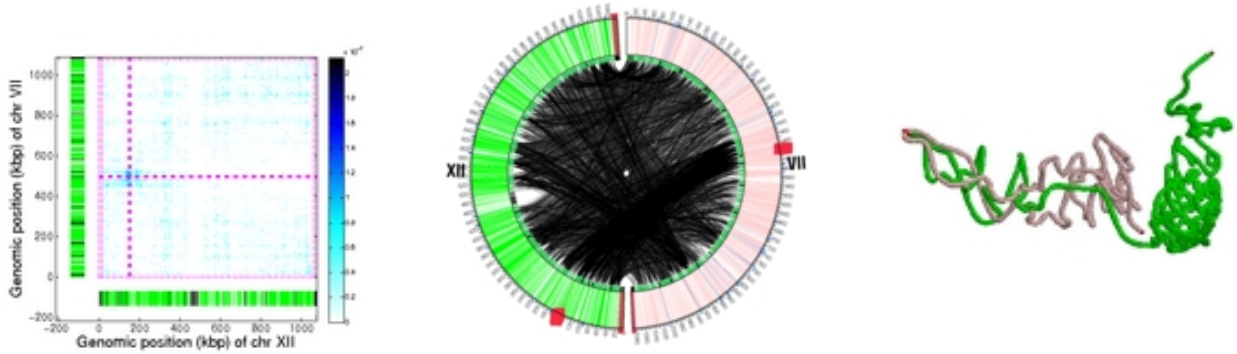



VII/XI

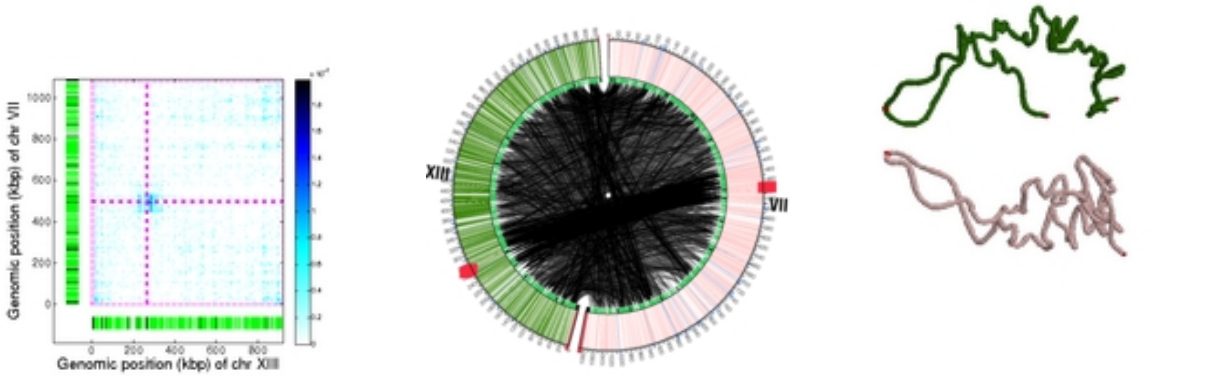


Supplementary Figure 9, continued.

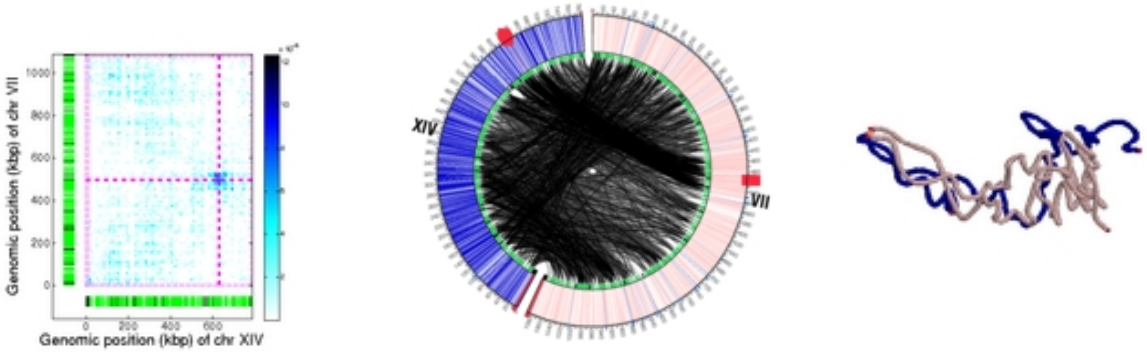
VII/XII



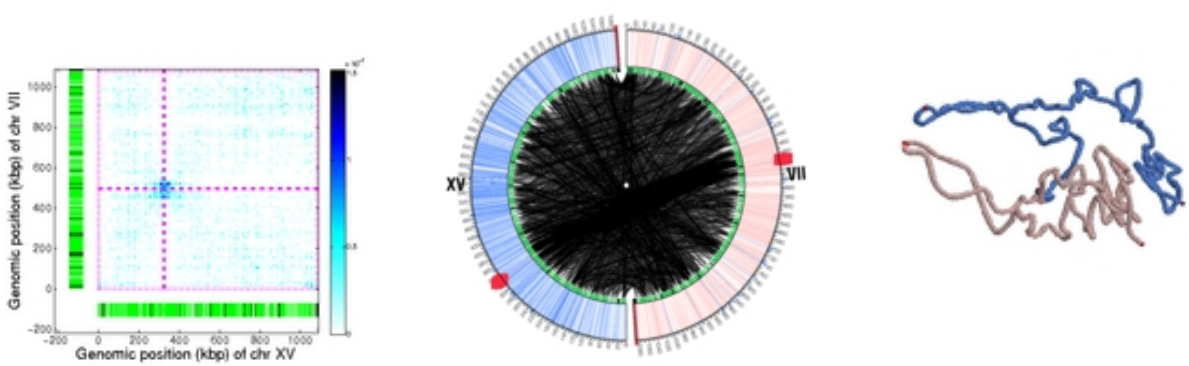

VII/XIII




VII/XIV

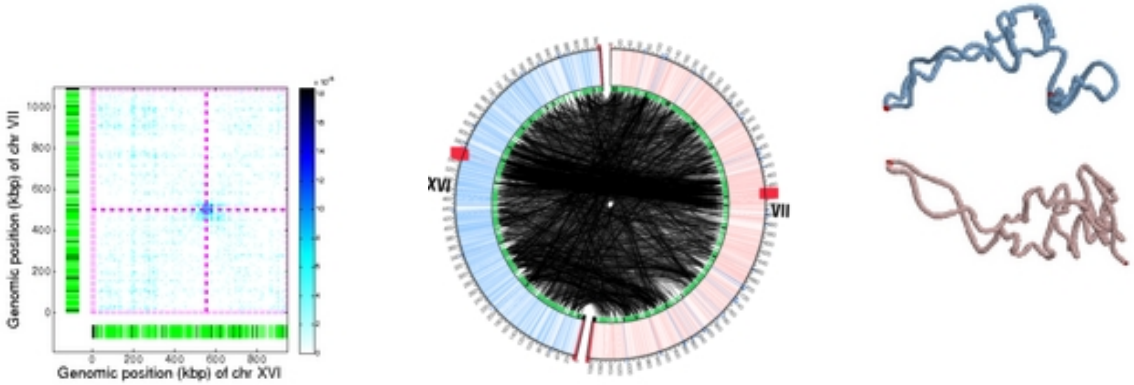



VII/XV

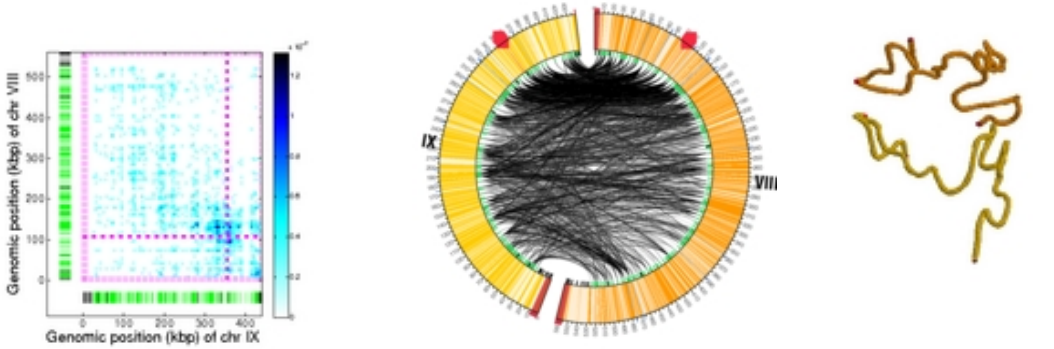


Supplementary Figure 9, continued.

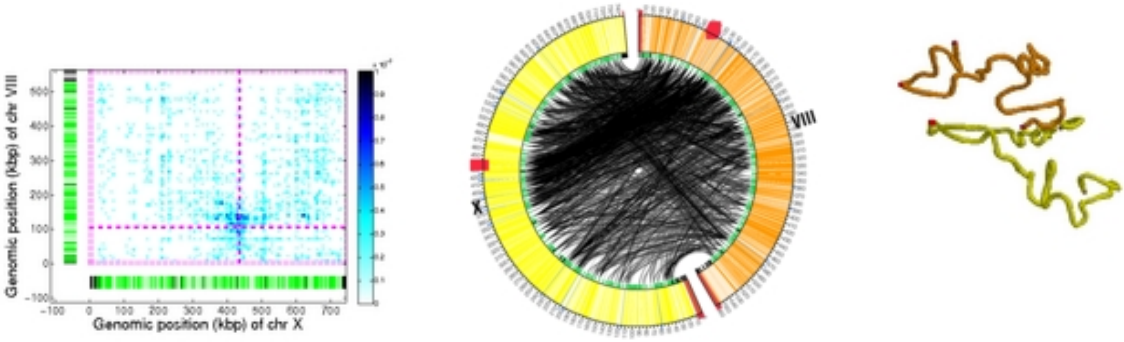
VII/XVI



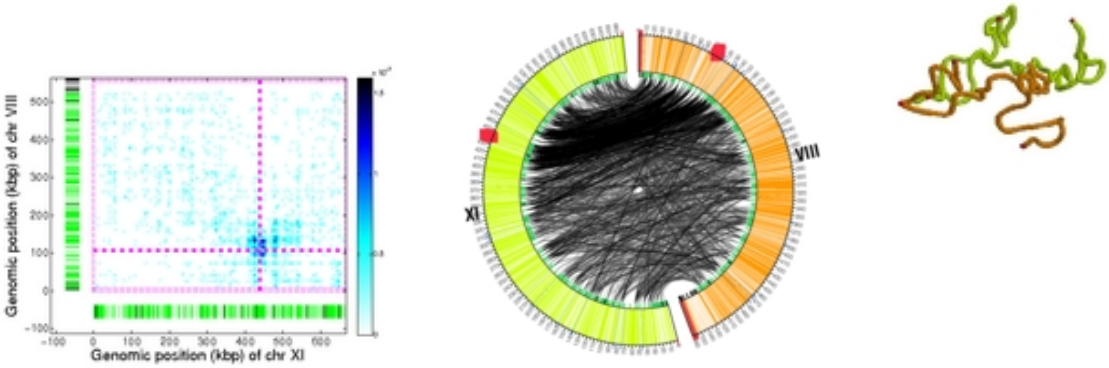

VIII/IX




VIII/X

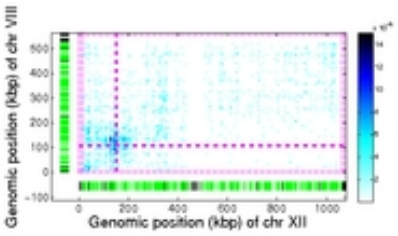


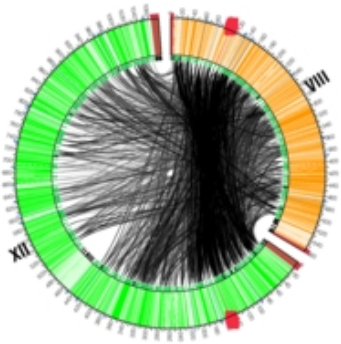

VIII/XI



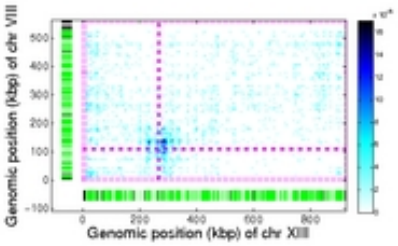
Supplementary Figure 9, continued.

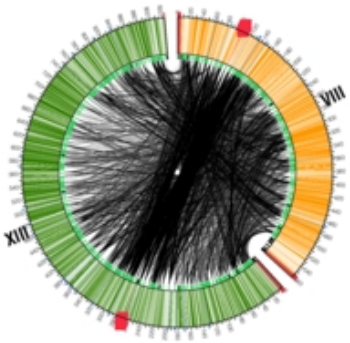
VIII/XII



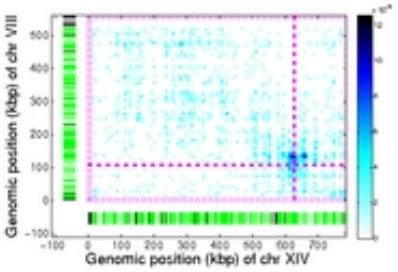



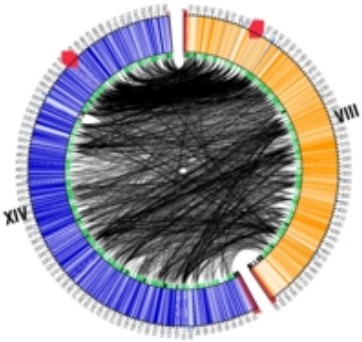

VIII/XIII



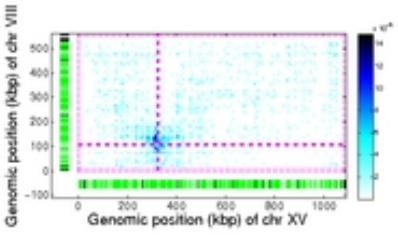



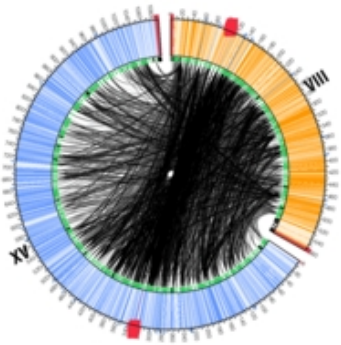

VIII/XIV

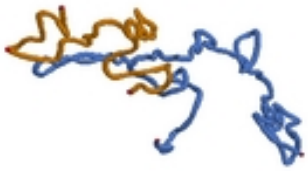



VIII/XV

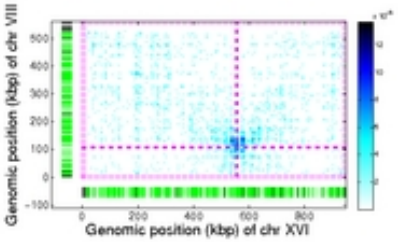


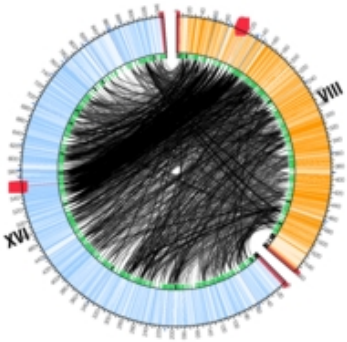



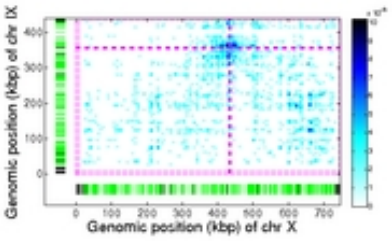
Supplementary Figure 9, continued.

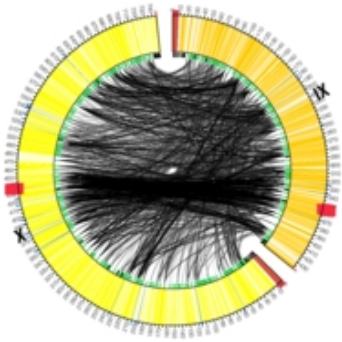
VIII/XVI



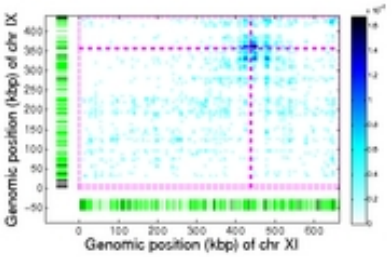



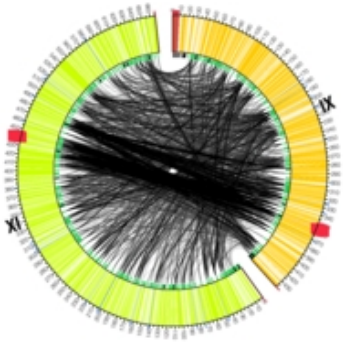

IX/X



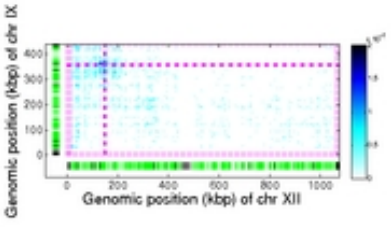



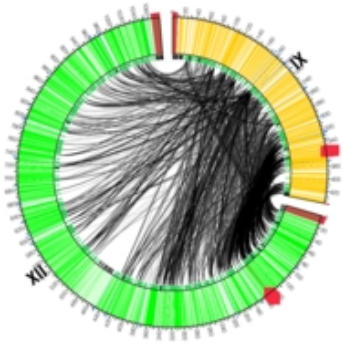

IX/XI



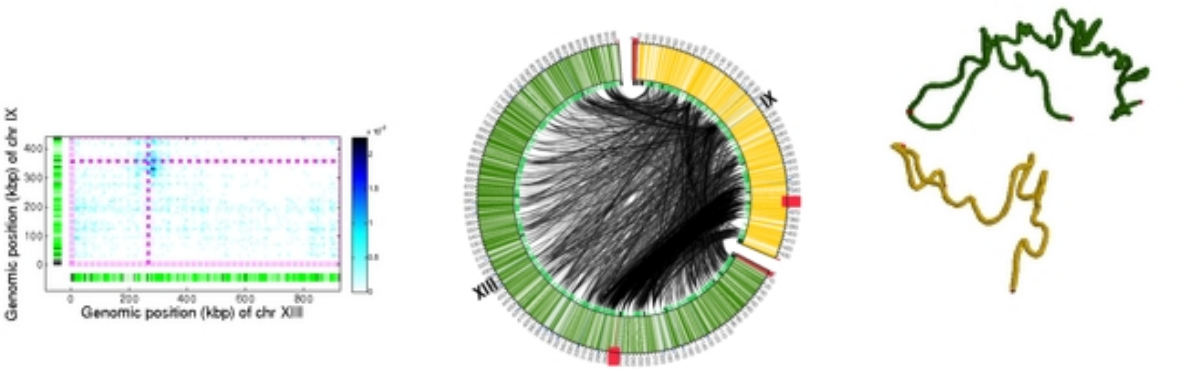

IX/XII



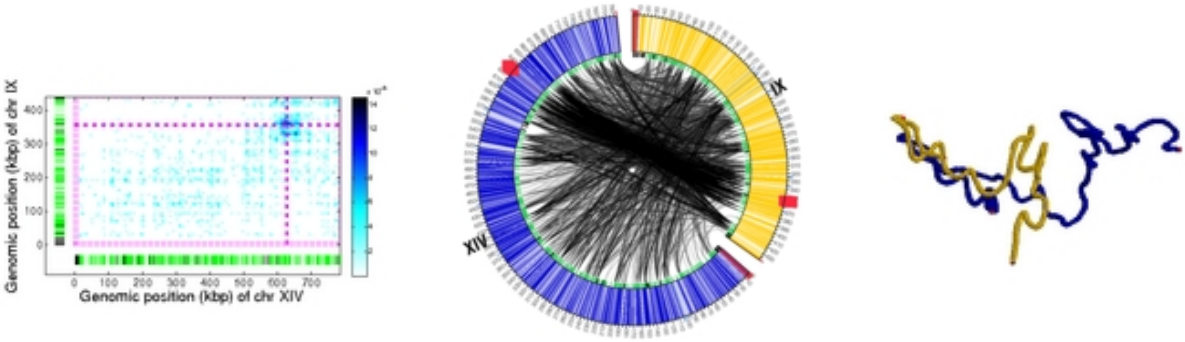




Supplementary Figure 9, continued.

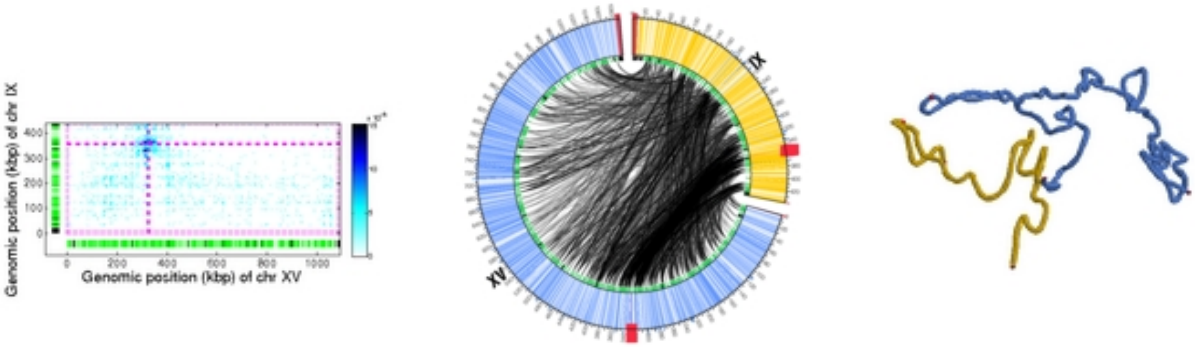
IX/XIII



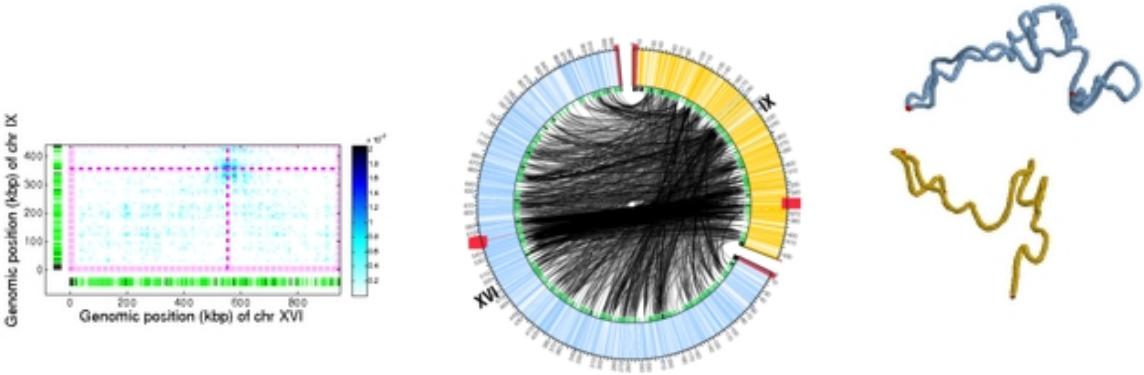

IX/XIV




IX/XV

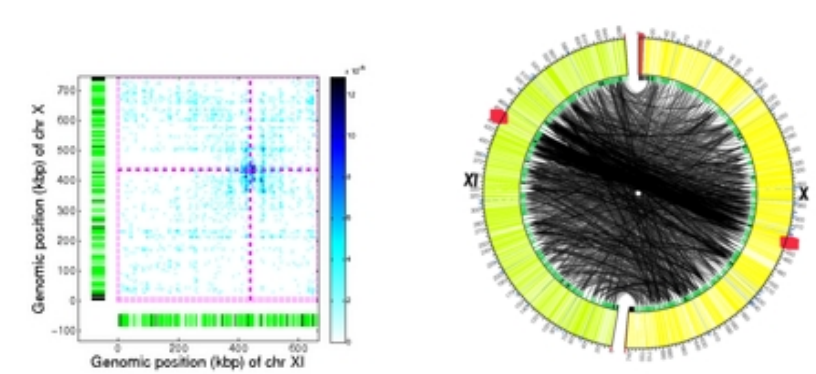



IX/XVI

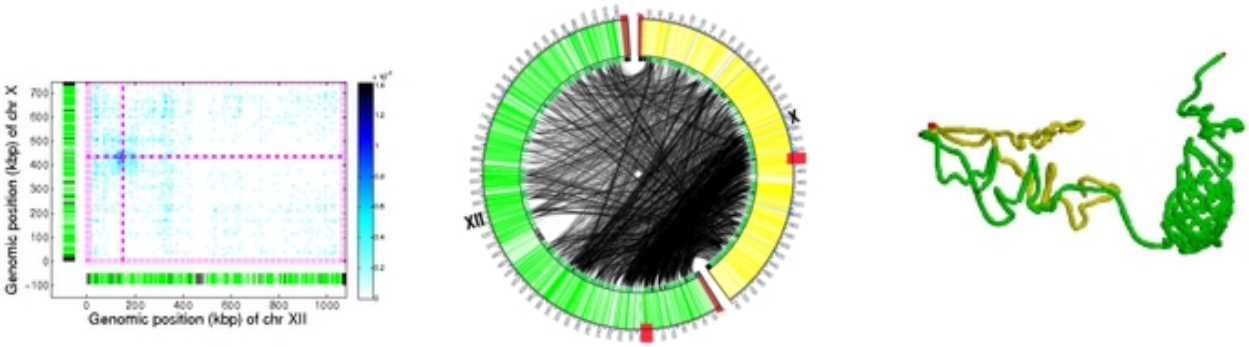


Supplementary Figure 9, continued.

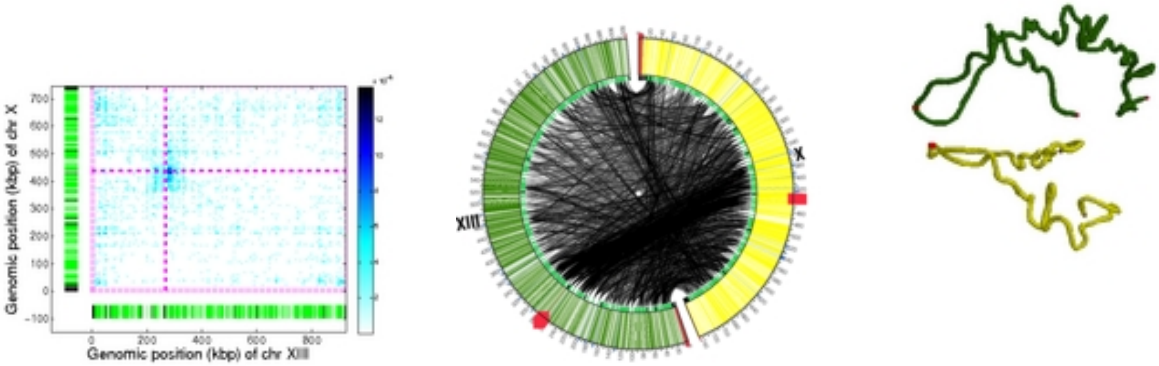
X/XI



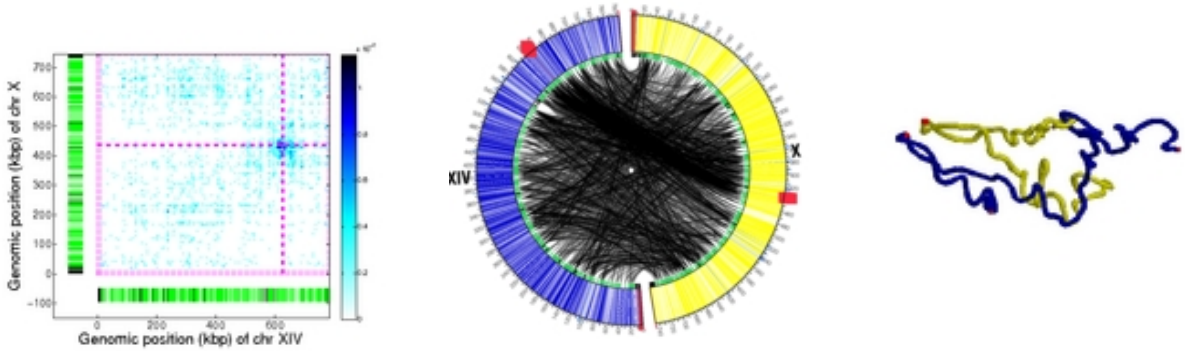

X/XII




X/XIII

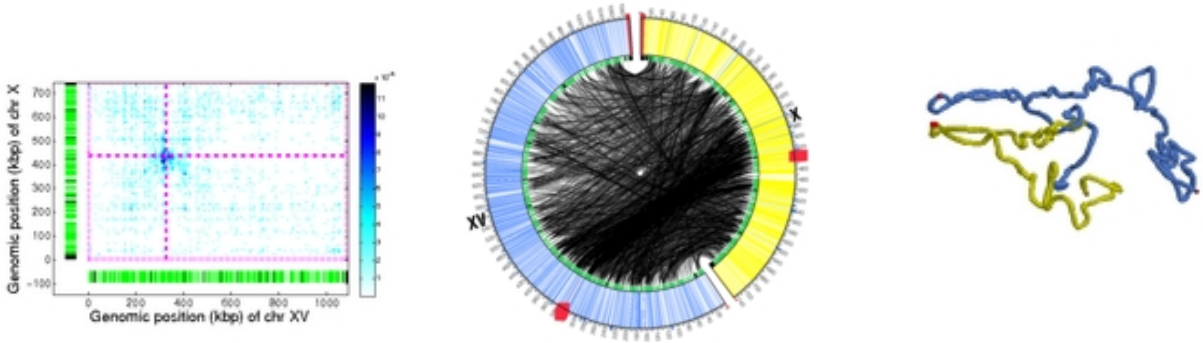



X/XIV

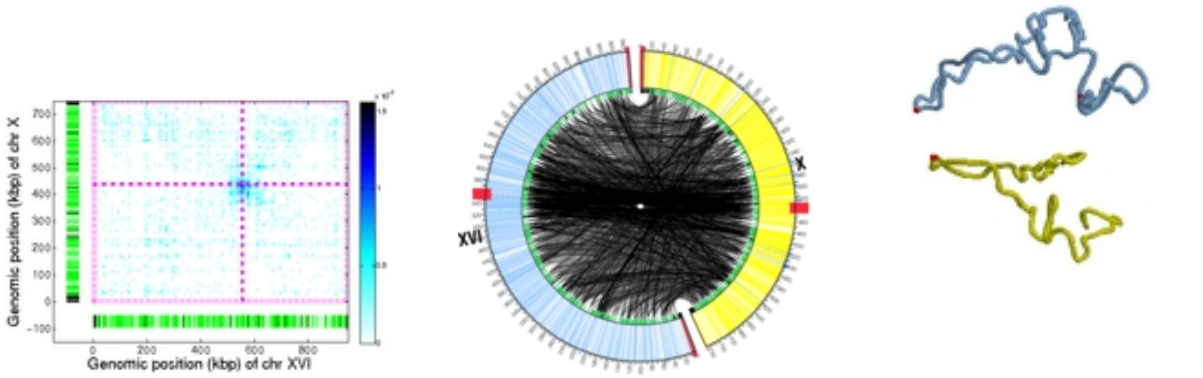


Supplementary Figure 9, continued.

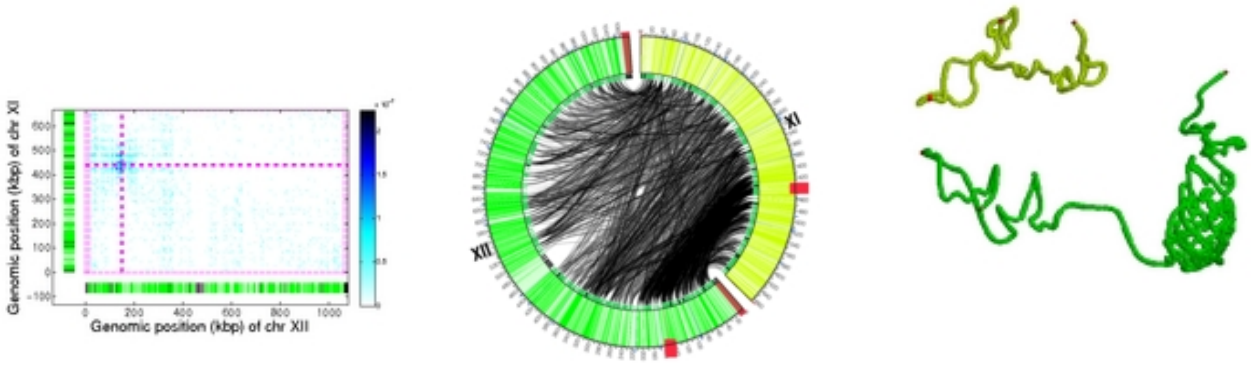
X/XV



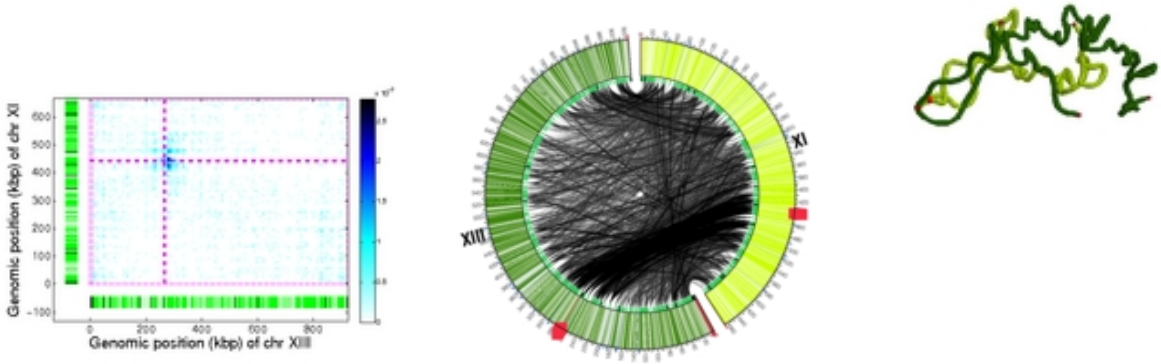

X/XVI




XI/XII

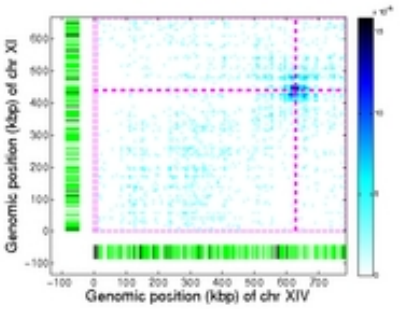


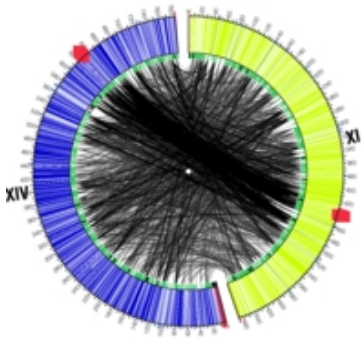

XI/XIII



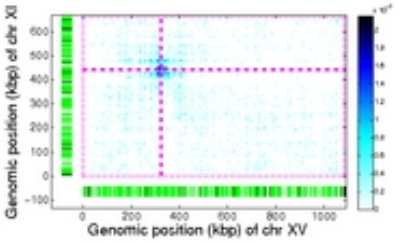
Supplementary Figure 9, continued.

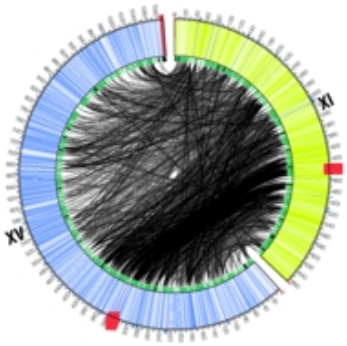
XI/XIV



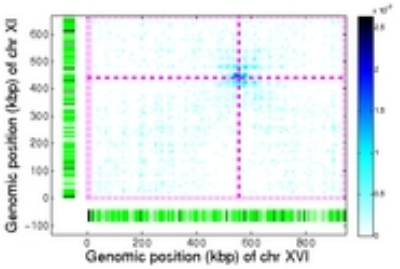



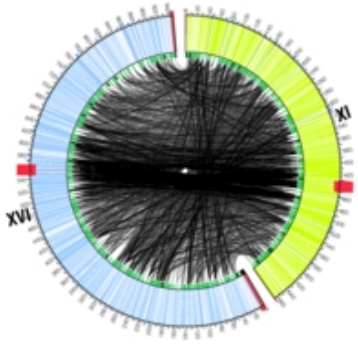

XI/XV

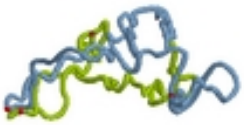


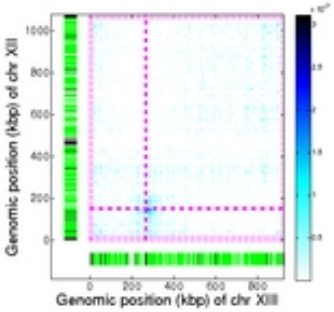

XI/XVI

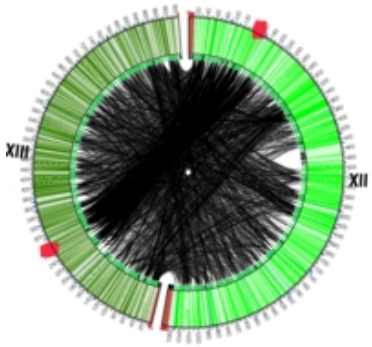


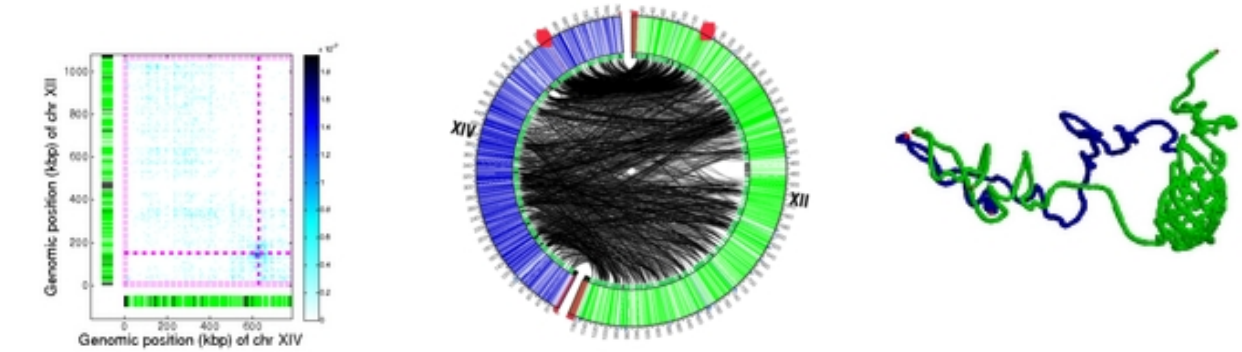
XII/XIII



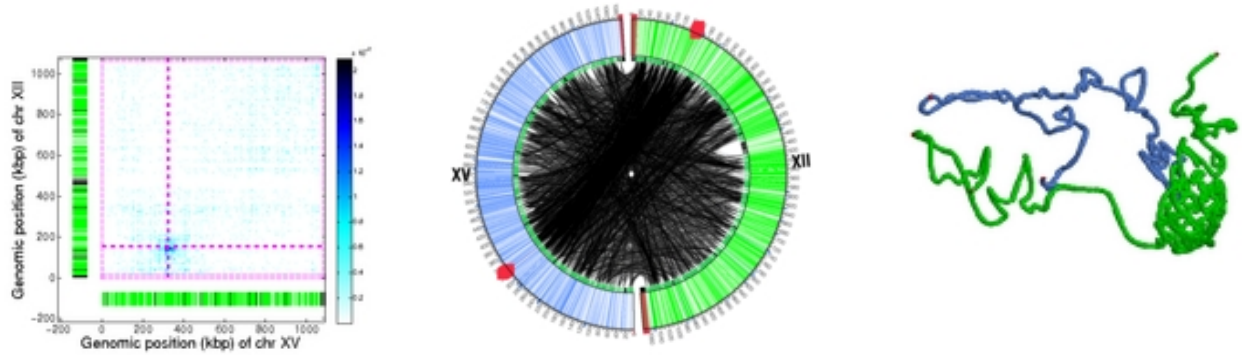




Supplementary Figure 9, continued.

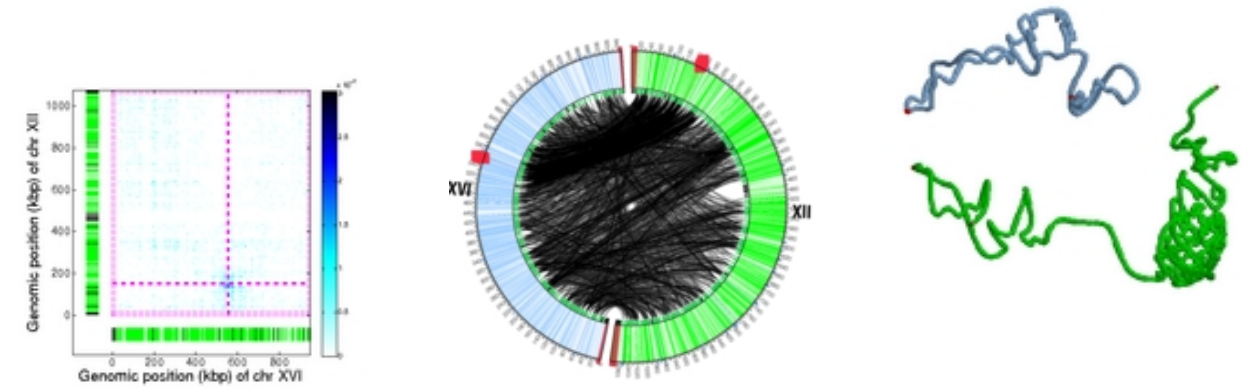
XII/XIV



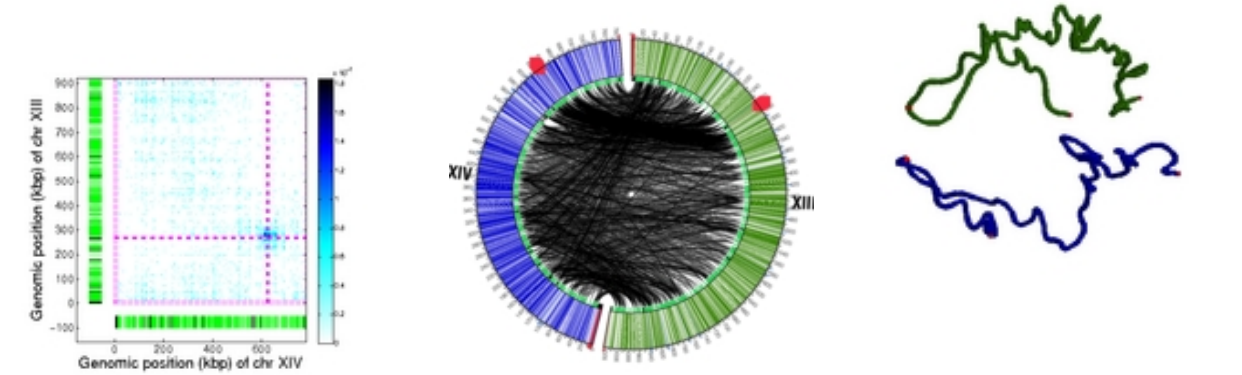

XII/XV




XII/XVI

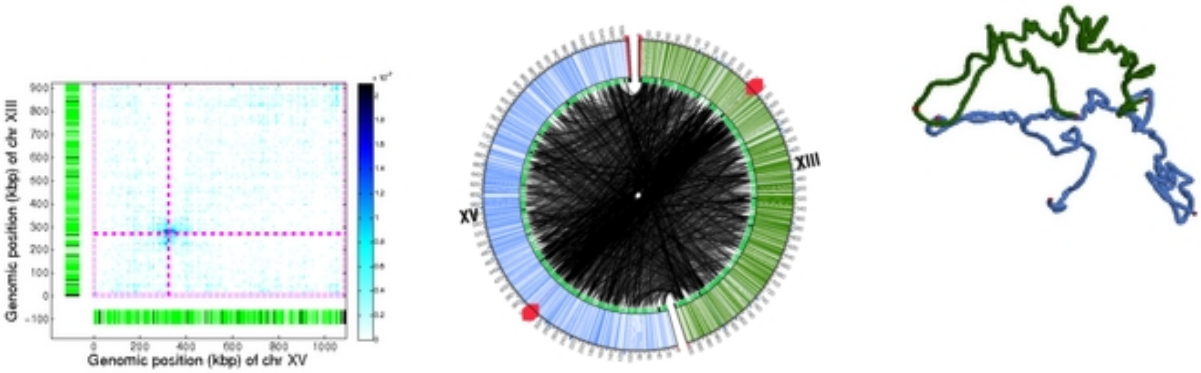



XIII/XIV

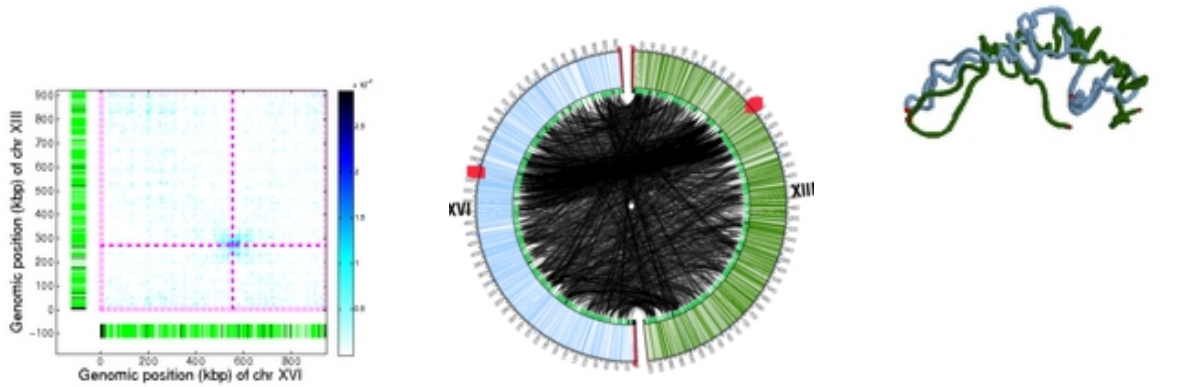


Supplementary Figure 9, continued.

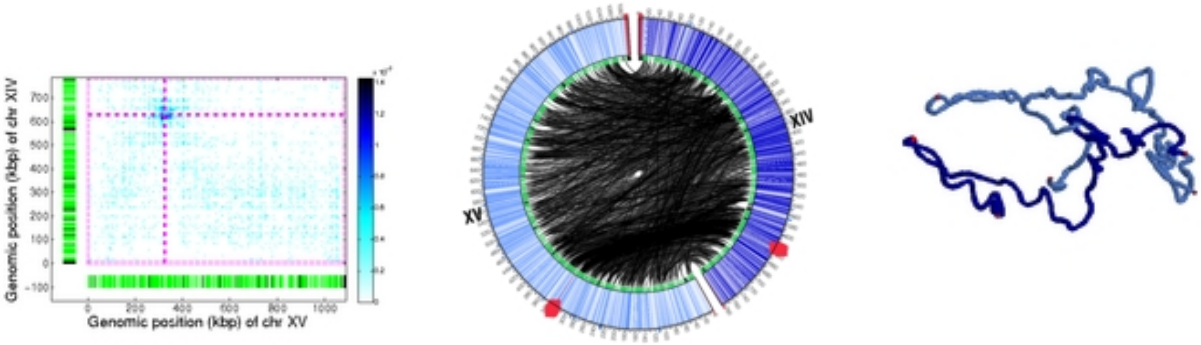
XIII/XV



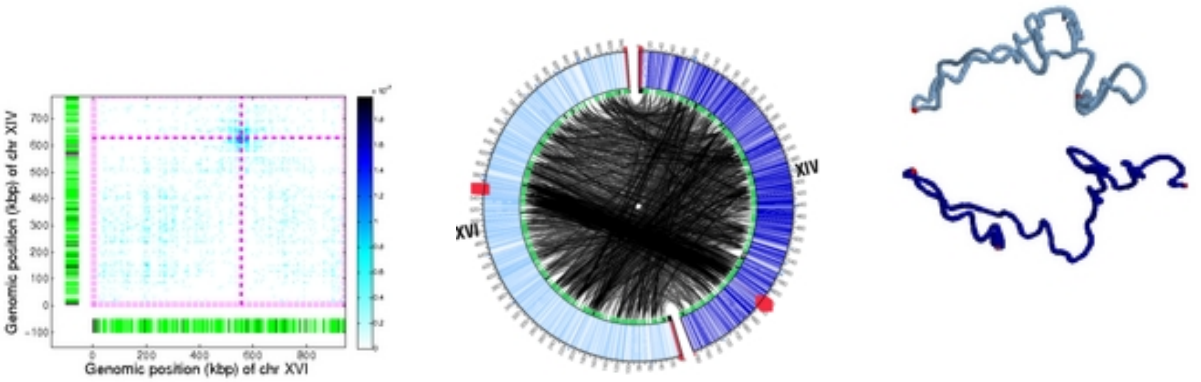

XIII/XVI




XIV/XV

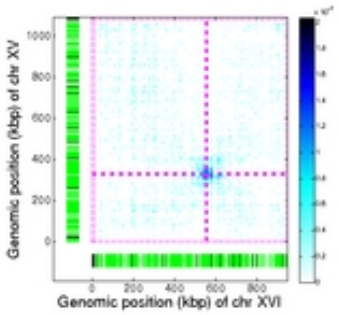


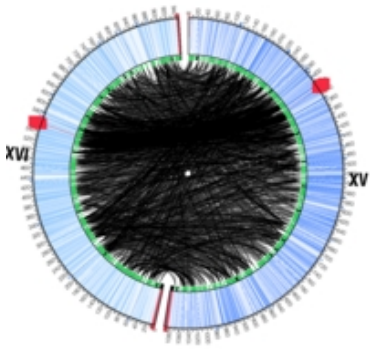

XIV/XVI



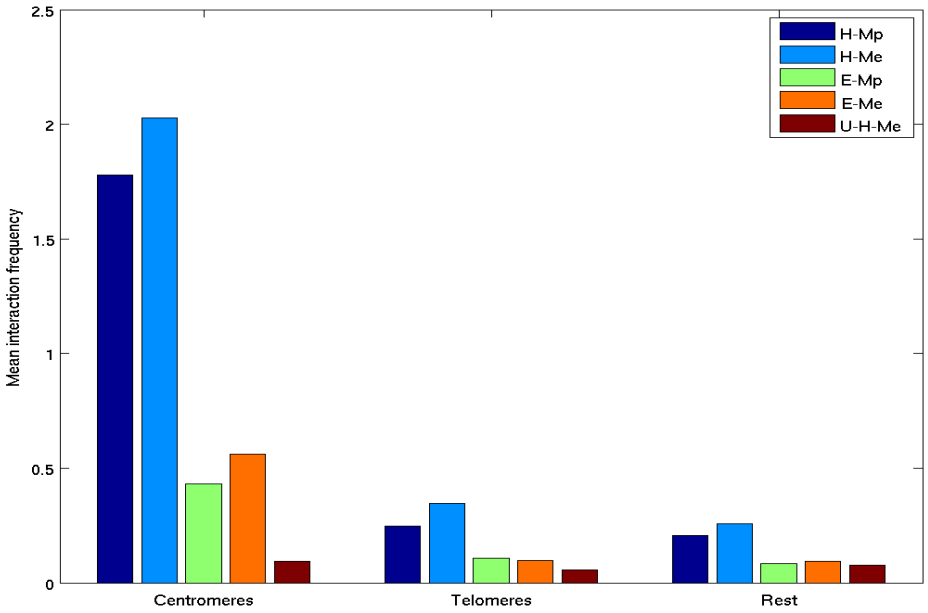
Supplementary Figure 9, continued.

XV/XVI



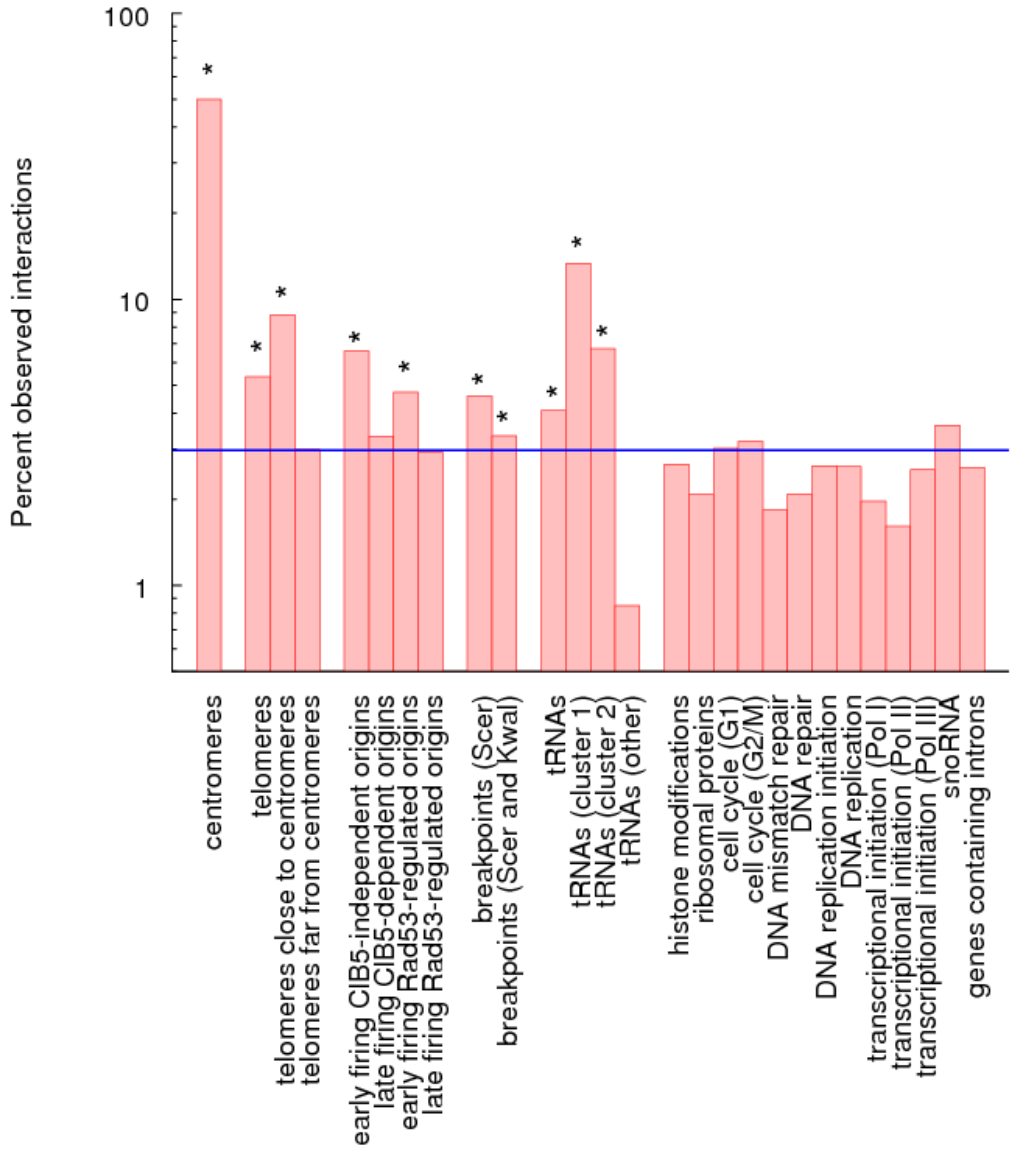

a



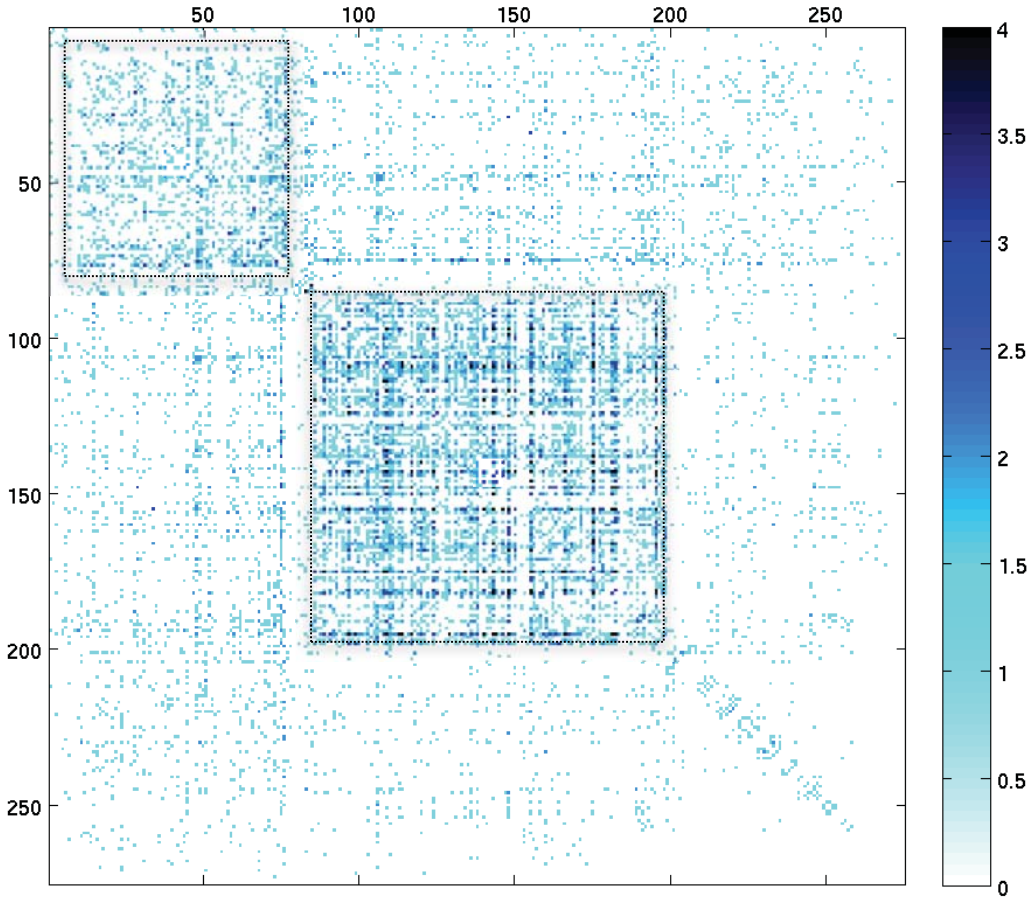

b

| GenomicType | Libraries | Mean inter-chromosomal interaction frequency | 95% confidence Intervals (t-Distribution) |
|---|---|---|---|
| **Centromere** | H-Mp | 1.78035 | 0.0603951 |
| **Centromere** | H-Me | 2.02807 | 0.0531325 |
| **Centromere** | E-Mp | 0.430463 | 0.023094 |
| **Centromere** | E-Me | 0.561537 | 0.025266 |
| **Centromere** | U-H-Me | 0.0939513 | 0.00786968 |
| **Telomere** | H-Mp | 0.248675 | 0.00716433 |
| **Telomere** | H-Me | 0.344907 | 0.00958182 |
| **Telomere** | E-Mp | 0.107585 | 0.00546637 |
| **Telomere** | E-Me | 0.0984609 | 0.00431156 |
| **Telomere** | U-H-Me | 0.0568795 | 0.00347656 |
| **Rest** | H-Mp | 0.205639 | 0.000309491 |
| **Rest** | H-Me | 0.256034 | 0.000386418 |
| **Rest** | E-Mp | 0.0833893 | 0.000180605 |
| **Rest** | E-Me | 0.0950279 | 0.000178 |
| **Rest** | U-H-Me | 0.0780769 | 0.000227029 |

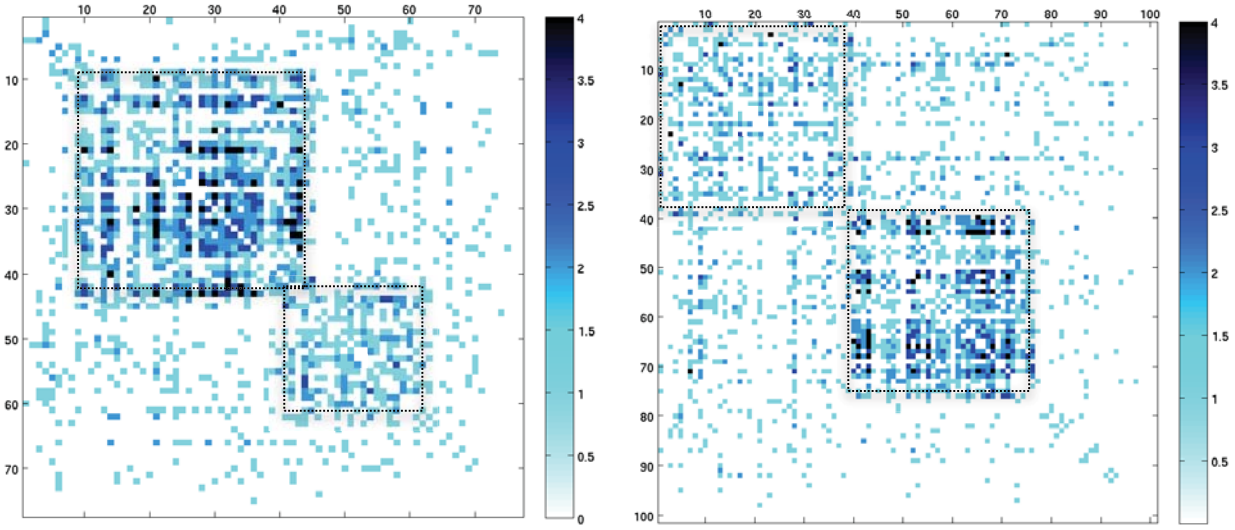
Supplementary Figure 10. a. The mean frequency of inter-chromosomal interactions is significantly higher in each of the experimental libraries relative to the un-crosslinked control library (U-H-Me), particularly for interactions between centromeric regions and between telomere regions. The values used in (a) and the corresponding values of 95% confidence intervals (t-distribution) are provided in (b).



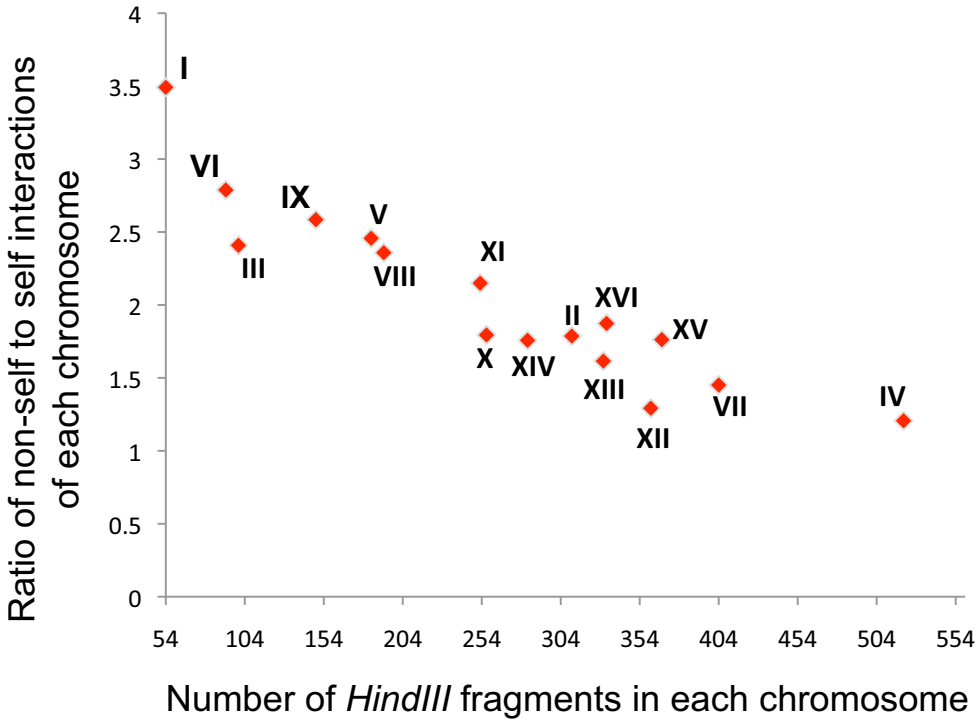

Supplementary Figure 11. Enrichment of inter-chromosomal interactions with respect to 27 groups of gene loci. Each bar depicts the enrichment or depletion relative to the percent (blue line) of all possible interactions that are observed with q<0.01. Bars with an asterisk are significantly enriched (p<0.01, binomial test with Bonferroni correction). In addition to the 11 groups of loci shown in Figure 5, this figure shows that (1) the telomere enrichment remains significant even among telomeres far from centromeres, (2) the difference between early and late origins is significant regardless of how the origins are identified (using CIB5 or Rad53), and (3) none of 13 Gene Ontology categories show significant enrichment.



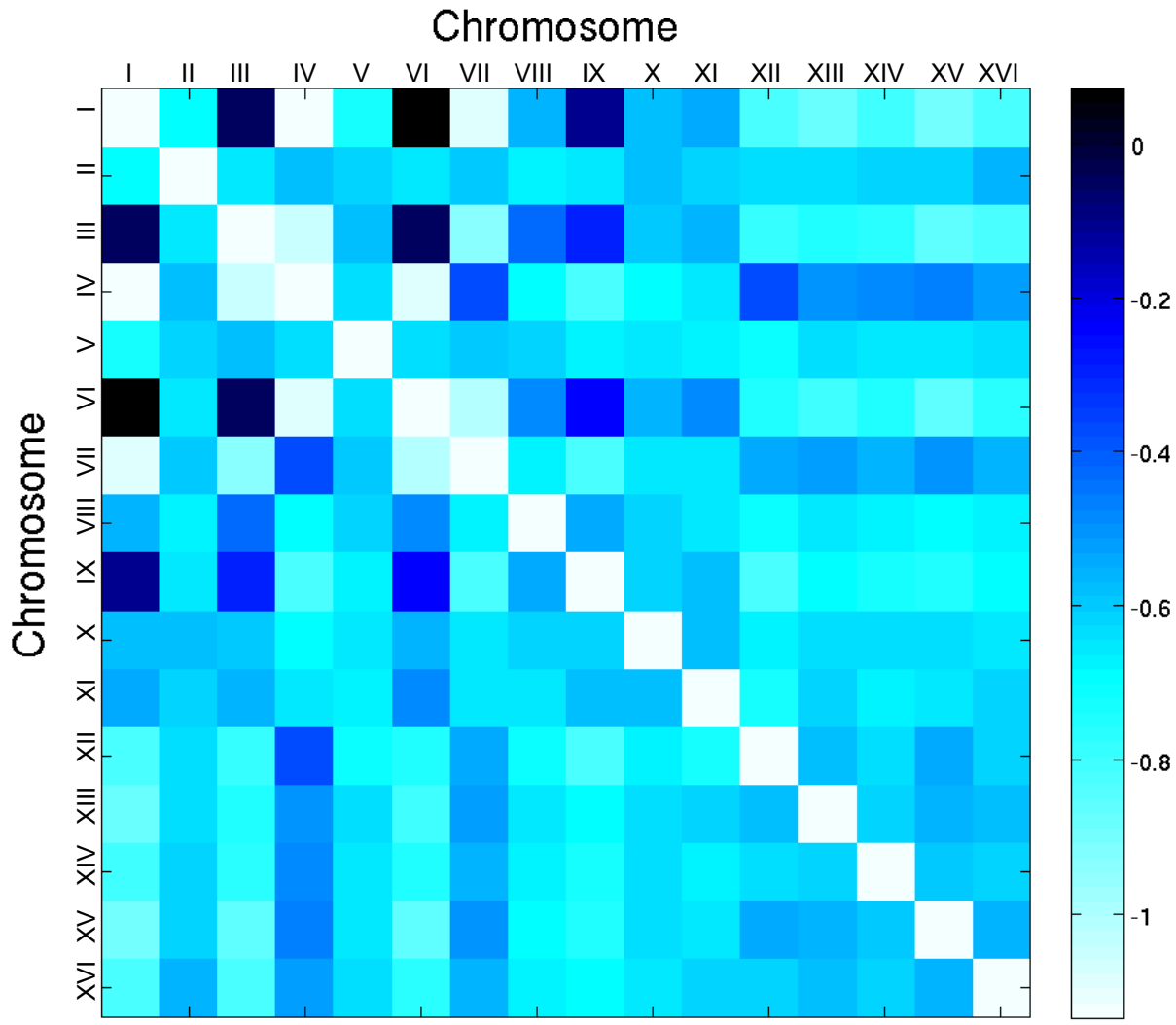

Supplementary Figure 12. Two large clusters of co-localized tRNA genes revealed using a hierarchical average-link clustering algorithm. The dim cluster (upper left) co-localizes with rDNA, implicating their co-localization in the nucleolus, while the bright cluster in the middle co-localizes with centromeres. To improve visualization, dashed lines were drawn to define the borders of the clusters.



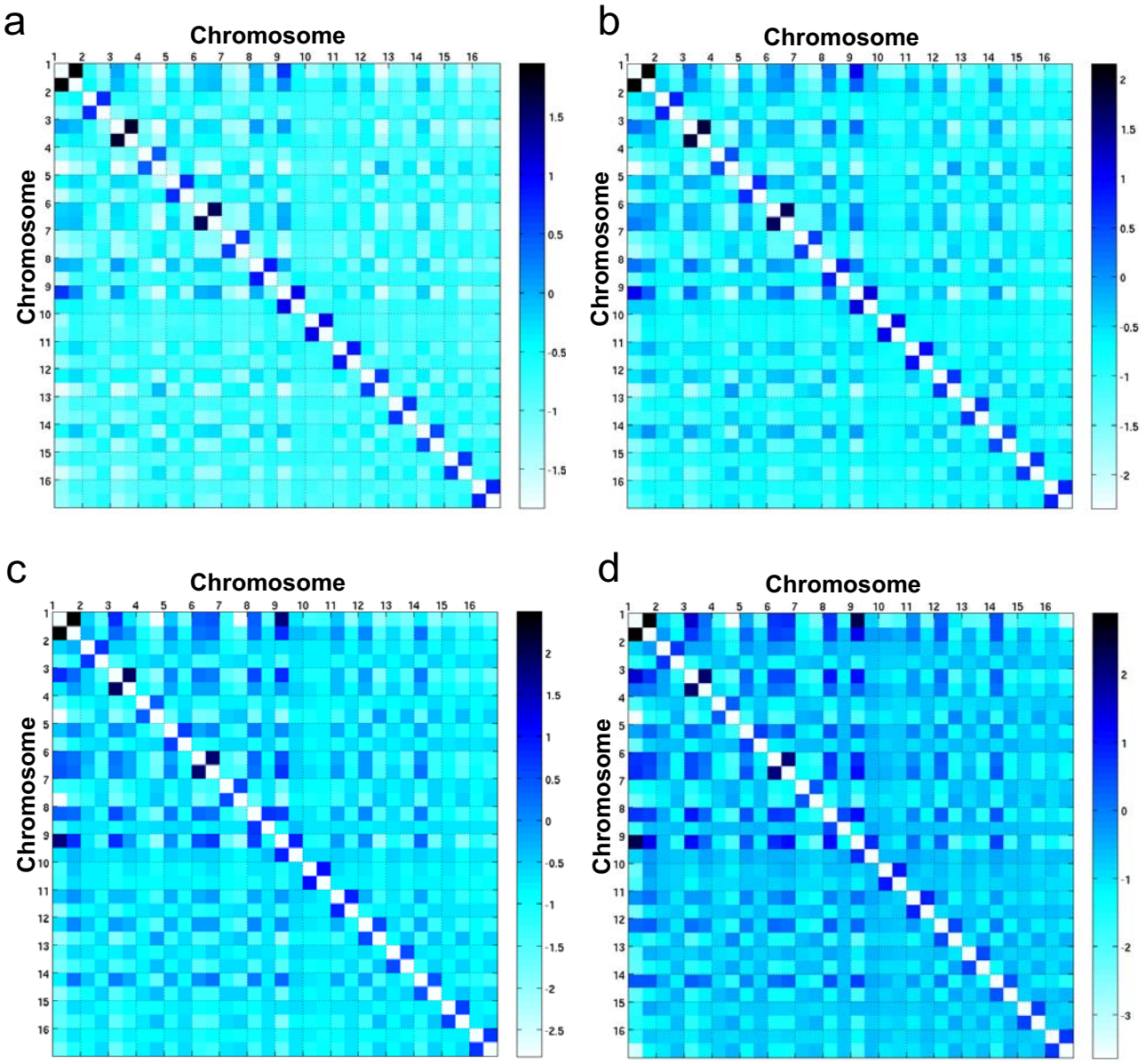

Supplementary Figure 13. Clusters of co-localized early firing replication origins revealed using a hierarchical average-link clustering algorithm. To achieve better visualization, dashed lines were drawn to define the borders of the clusters. Left, early firing Clb5-independent origins (non-CDR), right, early firing Rad53-unregulated origins (unchecked).



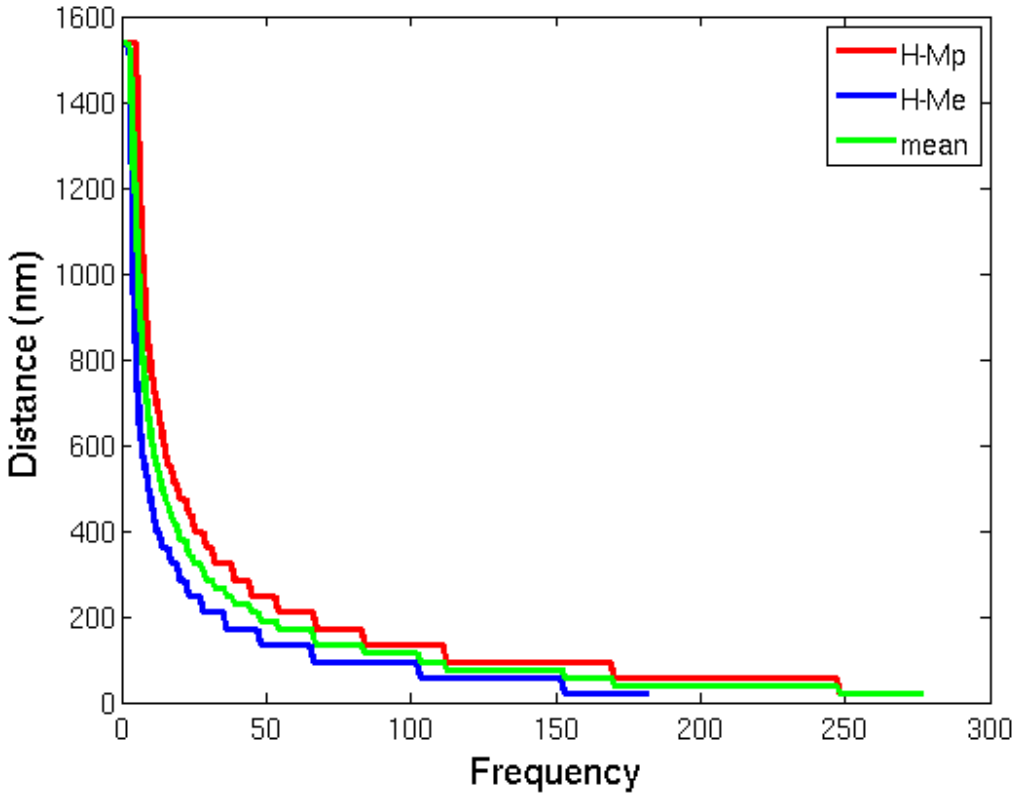

Supplementary Figure 14. Inverse correlation between the ratio of inter-chromosomal to intra-chromosomal interaction versus the number of *HindIII* fragments for each of 16 chromosomes.



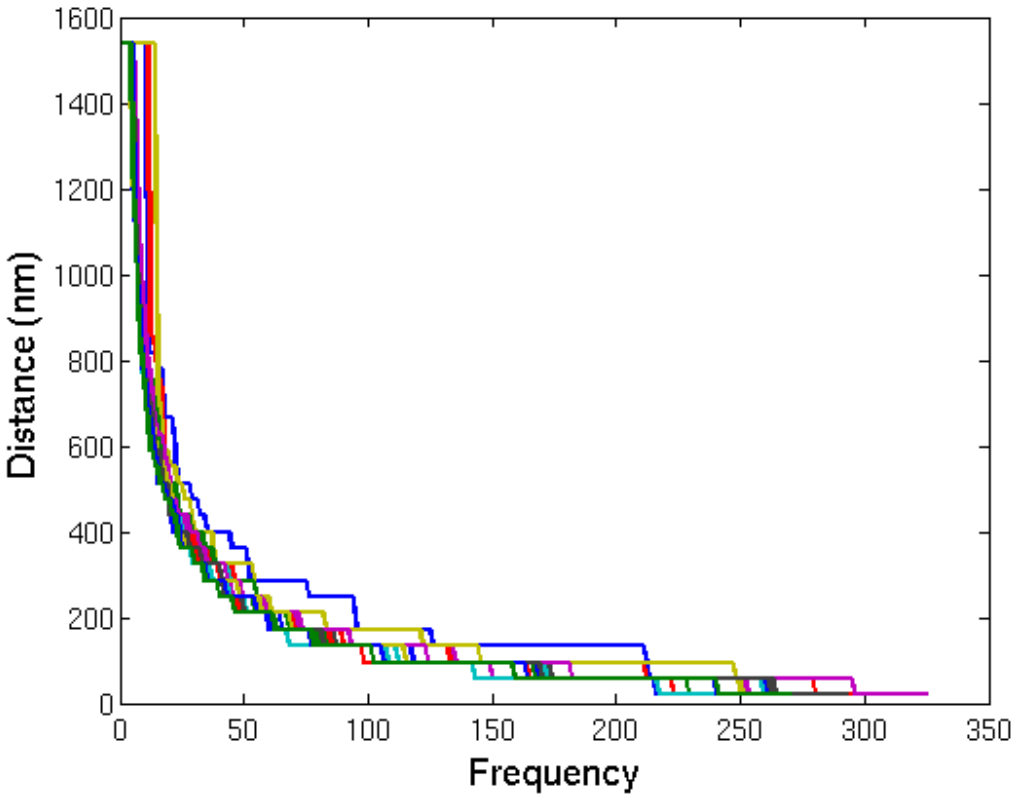

Supplementary Figure 15. Heat map depicting interaction enrichment log ratios among all 16 chromosomes in yeast. Note the strong interactions among the short chromosomes (I, III, VI, IX), as well as the interaction between chromosomes XII and IV.




Supplementary Figure 16. Heat map depicting inter- and intra- chromosomal interactions among the 32 yeast chromosome arms in relationship to increasing distance from centromeric sequences (a: the intact arm; b, c, d: excluding 10kb, 30kb or 50 kb regions starting from the midpoint of centromere, respectively. For each chromosome, the shorter arm is placed before the longer arm. Note, when centromeres are included, the interactions among the 16 intra-chromosomal pairings were stronger than for all inter-chromosomal pairings, except that the two smallest arms (the short arms of chromosomes I and IX) also showed a relatively high interaction enrichment (a). However, with increasing distance from the centromere, there is an increasing propensity for inter-chromosomal pairings (b-d).




Supplementary Figure 17. Function that converts interaction frequencies (x axis) into spatial distances (y axis). Red, H-Mp library, Blue, H-Me library, Green, the average of the two functions. See the main text and Supplementary Methods for details.




Supplementary Figure 18. Relationship between interaction frequency (x axis) and spatial distance (y axis) for each chromosome in the H-Mp library. Each colored curve represents a different chromosome.