



**HAL**  
open science

# Inferring the physical properties of yeast chromatin through Bayesian analysis of whole nucleus simulations

Jean-Michel Arbona, Sébastien Herbert, Emmanuelle Fabre, Christophe Zimmer

► **To cite this version:**

Jean-Michel Arbona, Sébastien Herbert, Emmanuelle Fabre, Christophe Zimmer. Inferring the physical properties of yeast chromatin through Bayesian analysis of whole nucleus simulations. *Genome Biology*, 2016, 18 (1), pp.81. 10.1186/s13059-017-1199-x . inserm-01517883

**HAL Id: inserm-01517883**

**<https://inserm.hal.science/inserm-01517883>**

Submitted on 3 May 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH

Open Access



# Inferring the physical properties of yeast chromatin through Bayesian analysis of whole nucleus simulations

Jean-Michel Arbona<sup>1,2</sup>, Sébastien Herbert<sup>1,2,3</sup>, Emmanuelle Fabre<sup>4</sup> and Christophe Zimmer<sup>1,2\*</sup> 

## Abstract

**Background:** The structure and mechanical properties of chromatin impact DNA functions and nuclear architecture but remain poorly understood. In budding yeast, a simple polymer model with minimal sequence-specific constraints and a small number of structural parameters can explain diverse experimental data on nuclear architecture. However, how assumed chromatin properties affect model predictions was not previously systematically investigated.

**Results:** We used hundreds of dynamic chromosome simulations and Bayesian inference to determine chromatin properties consistent with an extensive dataset that includes hundreds of measurements from imaging in fixed and live cells and two Hi-C studies. We place new constraints on average chromatin fiber properties, narrowing down the chromatin compaction to ~53–65 bp/nm and persistence length to ~52–85 nm. These constraints argue against a 20–30 nm fiber as the exclusive chromatin structure in the genome. Our best model provides a much better match to experimental measurements of nuclear architecture and also recapitulates chromatin dynamics measured on multiple loci over long timescales.

**Conclusion:** This work substantially improves our understanding of yeast chromatin mechanics and chromosome architecture and provides a new analytic framework to infer chromosome properties in other organisms.

**Keywords:** Chromatin, Chromosomes, Nuclear architecture, Polymer models, Yeast

## Background

The mechanical properties of chromatin and the spatial arrangement of chromosomes play an important role in genome functions, but in general remain poorly known [1, 2]. The structure of the chromatin fiber has remained elusive and controversial. The classical 30 nm structure, in which nucleosomes are tightly stacked on top of each other, is now called into question [3–9]. Chromosome architecture has been intensely studied in recent years using genome-wide chromosome conformation capture (Hi-C), which provides precious quantitative information about chromatin folding and has revealed biologically important features such as promoter-enhancer interactions and chromosome partitioning into functional domains

[10–15]. A key challenge is to build mechanistic models that are able not only to explain these observations, but also to predict the 3D organization of chromosomes and its alterations *de novo*. It is increasingly evident that polymer physics provides an adequate framework for this purpose [5, 16–26]. This is particularly clear for the extensively studied budding yeast nucleus [27–34]. We and others have shown that many aspects of yeast nuclear architecture can be reproduced *in silico* by modeling chromosomes as generic, semi-flexible polymers, with only a small number of sequence-specific constraints [35–37]. Such models can also make predictions on functional features, such as differences in DNA repair efficiency by homologous recombination [37, 38].

Nevertheless, two important questions remain unanswered. The first question relates to the mechanical and structural properties of the chromatin fiber, including its compaction and rigidity. The compaction  $C$  can be defined as the number of base pairs per unit length

\* Correspondence: [czimmer@pasteur.fr](mailto:czimmer@pasteur.fr)

<sup>1</sup>Unité Imagerie et Modélisation, Institut Pasteur, 25 rue du Docteur Roux, 75015 Paris, France

<sup>2</sup>UMR 3691, CNRS; C3BI, USR 3756, IP CNRS, Paris, France

Full list of author information is available at the end of the article



along the fiber (bp/nm). The bending rigidity can be measured by the persistence length  $P$ , the curvilinear distance along the fiber over which the direction (tangent vector) of the fiber becomes uncorrelated, such that more rigid fibers have higher  $P$ . Both parameters are key to the mechanical behavior of the chromatin during functional processes such as transcription and replication, but remain poorly characterized, and the very structure of chromatin remains uncertain. Previous estimates of  $C$  in yeast were in the range of 30–150 bp/nm and estimates of  $P$  were in the range of less than 30 nm to 200 nm or more [4, 12, 39–43].

Second, significant discrepancies still exist between model predictions and experimental observations [35]. If these discrepancies are due to errors in the model rather than the data, one must determine if they reflect incorrect values of the basic mechanical parameters of the model, including  $P$  and  $C$ , or rather reflect local effects possibly associated to a functional process, such as chromatin decondensation during transcriptional activation of a specific gene [27, 28, 44, 45]. Whereas the first type of discrepancy might be remedied simply by adjusting the global structural parameters, the second type of discrepancy calls for a more complex model, in which chromatin structure can change locally, e.g. depending on gene expression. Thus, correcting for global parameter mismatches will make it easier to identify discrepancies pointing to local modifications of chromatin architecture that may relate to biological functions.

In this paper, we present a computational framework that simultaneously addresses these two questions, using hundreds of polymer simulations with different chromatin parameters, combined with hundreds of experimental data points from multiple independent studies. Our analysis provides new constraints on the mechanical properties of the chromatin fiber, as well as a considerably more accurate predictive model of 3D nuclear architecture and chromatin dynamics in yeast.

## Methods

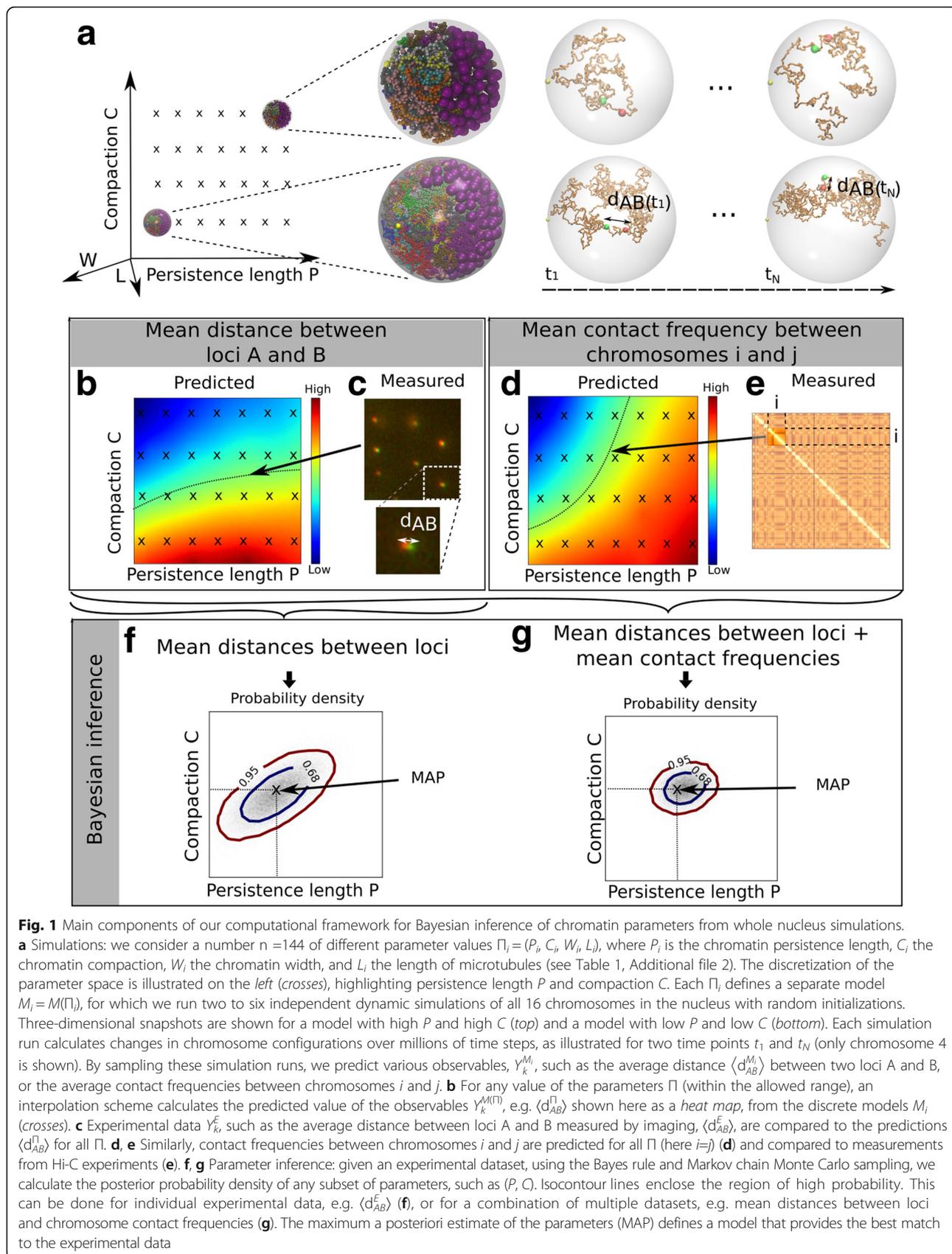
### Framework to infer chromatin parameters from whole-nucleus simulations

We begin with an overview of our analysis framework (Fig. 1). Briefly, we systematically varied the parameters  $\Pi$  of a whole nucleus chromosome simulation, which include the chromatin compaction  $C$  and the rigidity  $P$  (more parameters will be detailed below). We then aimed to determine the parameter values for which the model  $M(\Pi)$  agrees well with a range of experimental data and to quantify the plausibility of these parameters. The workflow comprises the following four components: (1) a set of hundreds of independent simulations of all chromosomes in the yeast nucleus, corresponding to 144 different parameter values  $\Pi_i$ , ( $i = 1..144$ ) (Fig. 1a);

(2) nine sets of experimental data on nuclear architecture compiled from several distinct studies, each comprising many independent measurements  $Y_k^E$  ( $k = 1..N$ , where  $N$  is the number of measurements and “ $E$ ” stands for ‘empirical’), such as the mean distance between two chromatin loci (Fig. 1c) or the mean contact frequency between two chromosomes (Fig. 1e); (3) a module that, for any given parameter value  $\Pi$ , computes the model predictions  $Y_k^{M(\Pi)}$  corresponding to these measurements, by interpolating the predictions from discrete parameter values  $\Pi_i$  (Fig. 1b, d); and (4) an algorithm that samples the parameter space and computes the posterior probability density of the parameters for a given dataset (Fig. 1f) or for multiple datasets taken together (Fig. 1g). We describe each of these components in more detail below.

### Whole-nucleus chromosome simulations and chromatin parameters

We simulated the spatial configurations and dynamics of chromosomes in the yeast nucleus using an approach similar to that described previously [35, 37] (see Additional file 1: Supplementary Methods for details). Briefly, our model simulated 16 randomly moving yeast chromosomes, each of which was represented as a chain of beads of diameter  $W$ , connected by non-linear spring potentials. The number of beads followed from the assumed compaction  $C$  and each chromosome’s genomic length (ranging from 230 Kb for chromosome 1 to 1531 Kb for chromosome 4). Triplets of consecutive beads were linked by a potential that penalizes bending and whose strength followed from the assumed persistence length  $P$ . The nucleus was modeled as a confining sphere of radius  $R_N = 1 \mu\text{m}$ . We included two additional constraints specific to the yeast nucleus. First, budding yeast centromeres are linked by a single microtubule to the spindle pole body (SPB), a macromolecular complex embedded in the nuclear envelope [27, 46]. We therefore introduced a harmonic (spring-like) potential between the centromeric bead of each chromosome and a point on the nuclear sphere representing the SPB, with an equilibrium length  $L$ . Second, telomeres are tethered to the nuclear envelope by two redundant pathways [27, 47]; we therefore applied a purely radial short-range outward force to the 32 telomeric beads to bring them in close vicinity to the nuclear envelope. Third, for the ~1–2 Mb chromosomal region encoding the ribosomal DNA (rDNA), we used beads of a diameter  $W_{\text{rDNA}} > W$ . This was done to account for the steric constraints exerted by the rDNA, which undergoes particularly intense transcription and gives rise to the nucleolus [27, 48]. A short-range repulsive potential prevented all beads and chromosome chains from traversing each other. Starting from an arbitrary initial configuration of all chromosomes within the



nucleus (Additional file 1: Figure S1a, b), we used Langevin dynamics to simulate their movements, i.e. at each time step, each bead was subjected to a randomly oriented force, and its displacement was computed from the equations of motion resulting from this force and those derived from the abovementioned potentials. We let the simulation run for  $\sim 10^9$ – $10^{10}$  iterations to reach equilibrium and sampled the trajectories to predict observables  $Y_k^{M(\Pi)}$  such as the mean distance between two loci or the average contact frequency between two chromosomes (Fig. 1b, d; Additional file 1: Figure S1c, d; Figure S2a, b).

Table 1 lists the main parameters used in our simulations, including those that had the same value for all simulations and those that we systematically varied, namely  $P$ ,  $C$ ,  $W$ , and  $L$ . We hereafter refer to the latter four parameters as “structural parameters” and note their combination  $\Pi = (P, C, W, L)$ . In our previous study, we used  $\Pi_{\text{Wong}} = (P_{\text{Wong}}, C_{\text{Wong}}, W_{\text{Wong}}, L_{\text{Wong}}) = (30 \text{ nm}, 83 \text{ bp/nm}, 20 \text{ nm}, 380 \text{ nm})$  based on the assumption of a compact fiber structure and early biophysical studies of chromatin reconstituted in vitro [35, 43, 49]. However, as already mentioned, the structural properties of the chromatin fiber remain largely undetermined. We therefore varied these parameters over the following ranges, which encompass most previous estimates [5, 39–43]:  $P$ : 27–252 nm,  $C$ : 25–110 bp/nm,  $W$ : 30–60 nm,  $L$ : 200–400 nm. We discretized the parameter space into 144 distinct values of  $\Pi_i = (P_i, C_i, W_i, L_i)$  (Additional file 2). For each  $\Pi_i$ , we ran two to six independent (replica) simulations with random initialization, to assess the effect of purely statistical differences and check for equilibration. In total, we run 473 independent simulations, which we sampled every  $10^4$  iterations, resulting in hundreds of millions of different chromosome configurations. We hereafter refer to these simulations as “core simulations.”

### Compilation of experimental data on yeast chromatin architecture

We gathered a large amount of quantitative measurements on yeast nuclear architecture based on light microscopy in live or fixed cells and on Hi-C. The microscopy data contained mostly distances between pairs of loci on the same chromosome (*cis*) or on different chromosomes (*trans*), which were typically measured in hundreds or thousands of individual cells [35, 39, 50, 51]. We also added newly measured distances between 12 pairs of loci along the right arm of chromosome 4, spanning pericentromeric, internal, and subtelomeric regions (Additional file 1: Supplementary Methods). The Hi-C data provided genome-wide maps of contact frequencies between DNA segments averaged over millions of cells [29, 30]. Most of the imaging and Hi-C data were acquired and quantified by independent groups [29, 30, 35, 39, 50, 51]. While these data contain many thousands of individual measurements, we chose to condense them into a more manageable set of summary statistics. For example, instead of the entire distribution of distances between two loci, we only considered the mean (or median) distance over all cells. Similarly, we reduced the Hi-C data to the following quantities: (1) average contact frequencies of each chromosome with itself; (2) average contact frequencies of each chromosome with all other chromosomes; (3) intrachromosomal contact frequencies for different genomic distances, averaged over the entire genome; (4) intrachromosomal contact frequencies relative to the centromere, averaged over all chromosomes. Table 2 summarizes the experimental datasets and summary statistics. Additional file 3 provides a detailed list of all image-based measurements we have used. In the following, we separately pooled image-based measurements from fixed cells (28 data points  $Y_k^f$ ) and live cells (126 data points) because of the potentially important effect of fixation on nuclear architecture. We also considered the two Hi-C studies

**Table 1** Simulation

Parameter	Notation	Unit	Range	Wong 2012	MAP	Best model
Chromatin persistence length	$P$	nm	27–252	30	88	69
Chromatin compaction	$C$	bp/nm	25–110	83	61	50
Chromatin fiber width	$W$	nm	30–60	20	37	30
Microtubule length	$L$	nm	200–400	380	390	400
rDNA diameter	$W_{\text{rDNA}}$	nm	194 <sup>a</sup>	200	NA	194
Radius of nucleus	$R_N$	nm	1000	1000	NA	1000

This table lists the main parameters used in our model (first column), together with their notation (second column), and units (third column). The fourth column indicates the range over which the parameters are varied (for parameters that are not varied but held fixed in all simulations, a single value is given). The fifth column indicates values used in our previous model by Wong et al. [35]. The sixth column indicates the maximum a posteriori (MAP) estimate of each inferred parameter as obtained by our method using the whole set of experimental data. The seventh column lists the parameters of the “best model” used for Figs. 4 and 5. NA: not applicable. Additional file 2 lists all 144 models and their parameter values. More details on model parameters are provided in Additional file 1: Supplementary Methods

<sup>a</sup>This diameter corresponds to a net volume of the 150 rDNA beads that equals  $V = 14\%$  of the nuclear volume, i.e. an effective volume (considering free space between rDNA beads) of  $\sim 2V \sim 30\%$  of the nucleus. For the parameter mismatch analysis,  $W_{\text{rDNA}}$  was varied over a range of 138–251 nm, corresponding to a net volume fraction  $V = 5$ – $30\%$ , and an effective fraction of  $\sim 2V = 10$ – $60\%$

**Table 2** Experimental data used for parameter inference

Dataset	Observable	Experimental technique	Data points (n)	Reference
O1	Median 3D distance between pairs of subtelomeric loci	Imaging, live cells	62	[50]
O2	Median angle between locus, nuclear and nucleolar centroids	Imaging, live cells	37	[31, 50]
O3	Median distance between chromosome 12 locus and nucleolus	Imaging, live cells	15	[64]
O4	Median 2D distance between pairs of loci on chromosome 4	Imaging, live cells	12	this study
O5	Mean 3D distances between intrachromosomal pairs of loci on several chromosomes	Imaging, fixed cells (FISH)	13	[51]
O6	Mode of 3D distances between intrachromosomal pairs of loci on chromosome 14	Imaging, fixed cells (FISH)	8	[80]
O7	Mode of 3D distances between SPB and telomeres	Imaging, fixed cell (Immunofluorescence)	7	[80]
O8	Mean contact frequencies within chromosomes (16) and of each chromosome with the other chromosomes (16) + Intrachromosomal contact frequencies for genomic distances 25–85 Kb at 5 Kb intervals, averaged over the genome (12) + relative to the centromere (12)	Hi-C	56	[29]
O9	Same	Hi-C	56	[30]
	Total		266	

This table summarizes the nine experimental datasets used for parameter inference in this study. Additional data used for model validation only are: Hi-C data from [62], Micro-C XL data from [63], and chromatin dynamics data from [64–66]. For more details, see Additional file 3 and Additional file 1: Supplementary Methods

[29, 30] separately because of potentially important differences in protocols and data analysis (56 data points each). In total, our compilation thus consisted of 266 independent data points  $Y_k^E$ , hereafter also called “observables.” See Additional file 1: Supplementary Methods for more details.

### Determining the probability density of chromatin parameters by Bayesian inference

We aimed to explore the space of chromatin parameters  $\Pi$  and to assign a probabilistic score to all possible values of  $\Pi$  based on how accurately the corresponding model  $M(\Pi)$  predicted a given experimental dataset  $D$ . This approach is more informative than an optimization method, which would only determine a single parameter value  $\Pi^*$ , irrespective of how well the model explains the data for  $\Pi \neq \Pi^*$ . Our method employed two main ingredients: (1) a Bayesian formulation that computes the (posterior) probability of an assumed parameter value  $\Pi$  given the data  $D$ ; and (2) a Monte Carlo Markov Chain algorithm that samples the space of parameters  $\Pi$ .

The Bayes rule provides the posterior probability of  $\Pi$  given  $D$  by:  $p(\Pi|D) \propto p(D|\Pi)p(\Pi)$ , where  $p$  indicates probability density and  $p(A|B)$  the probability density of  $A$  conditioned on  $B$ . Let us consider an experimental dataset  $D = (Y_1^E, \dots, Y_N^E)$  consisting of  $N$  independent measurements  $Y_k^E$ . For example,  $Y_k^E$  might represent the mean distance between two loci A and B (Fig. 1c). For each data point  $Y_k^E$ , and for a given parameter value  $\Pi$ , the corresponding model  $M(\Pi)$  provides a single predicted counterpart  $Y_k^{M(\Pi)}$  (Fig. 1b, d). Both the experimental measurements  $Y_k^E$  and the model predictions  $Y_k^{M(\Pi)}$  are, however, affected by noise: distances measured

from images are necessarily corrupted by random localization errors [50, 52, 53], while Hi-C data suffer from counting noise due to limited sequencing depth, random ligations, and other factors [54, 55]. Our model predictions are also subject to random errors because they are computed from finite samples of stochastic simulations (Fig. 1a). For simplicity, we assume that the difference between each measurement  $Y_k^E$  and the corresponding predictions  $Y_k^M$  in the ideal case (i.e. assuming that the model faithfully describes reality) obeys a Gaussian probability density with variance  $\sigma_k^2$ . The variances  $\sigma_k^2$  are usually not known. Depending on the type of data  $Y_k^E$ , they can be expressed as functions of one or more additional parameters  $\Xi$  (nuisance parameters), which must be estimated along with the structural parameters  $\Pi$  [56, 57]:  $p(\Pi, \Xi|D) \propto p(D|\Pi, \Xi)p(\Pi)p(\Xi)$  (where we assumed statistical independence of  $\Xi$  and  $\Pi$ ). For the structural parameters  $\Pi$ , we assumed flat priors, i.e. constant  $p(\Pi)$  over the parameter range mentioned above; the priors  $p(\Xi)$  for the nuisance parameters were different for each observable. See Additional file 1: Supplementary Methods for details. Thus, for any assumed value of the structural and nuisance parameters  $\Pi$  and  $\Xi$ , and for a given dataset  $D$ , the above formulation yielded a posterior probability  $p(\Pi, \Xi|D)$  (to within a normalization constant).

In order to estimate  $p(\Pi, \Xi|D)$  over the entire parameter space, we used a Monte Carlo Markov Chain ensemble sampler that efficiently concentrated the exploration of the high-dimensional parameter space ( $\Pi, \Xi$ ) to the region of high posterior probability [58]. The posterior probability over the structural parameters,  $p(\Pi|D)$  was then readily obtained by integrating

$p(\Pi, \Xi|D)$  over the nuisance parameters  $\Xi$ . Similarly, the posterior probability for a single parameter, e.g. chromatin compaction,  $p(C|D)$ , or the joint probability for a subset of parameters, e.g.  $p(P, C|D)$ , were obtained by integrating over the remaining parameters (Fig. 1f, g). From the joint posterior probability density, we also determined a single maximum a posteriori (MAP) estimate of the parameters as:  $(\hat{P}, \hat{C}) = \arg \max p(P, C|D)$  (Fig. 1f, g).

## Results

### Inference method recovers true chromatin parameters from noisy simulated data

To assess our method's ability to recover the correct parameter values  $\Pi$ , and to better determine how different observables depend on these parameters and can be used to infer them, we first tested our method on synthetic data. To create these synthetic data, we picked a parameter value  $\Pi_0 = \Pi_{i_0}$  among those of our core simulations ( $i_0 \in [1, 144]$ ) (Table 1 and Additional file 2); from the corresponding simulation  $M_0 = M_{\Pi_0}$ , we then computed predictions  $Y_k^{M_0}$  ( $k = 1..266$ ) corresponding to all 266 observables mentioned above (Table 2). We then added random noise to these predictions to simulate experimental errors:  $Y_k^S = Y_k^{M_0} + \varepsilon_k$  ( $k = 1..266$ ), where  $\varepsilon_k$  is a normally distributed random number with mean 0 and variance  $(\sigma_k^S)^2$  (the superscript "S" denotes "simulation"). We tested three different levels of noise, with the highest level chosen such as to be consistent with or exceed the noise in the experimental data (Additional file 1: Supplementary Methods). We then used the noisy simulated data  $D^S = Y_k^S$  ( $k = 1..266$ ) instead of the real data as input to our inference algorithm (Fig. 1c, e) and compared the inferred posterior probabilities (Fig. 1f, g) to the true parameter value  $\Pi_0$ . We performed this comparison for five different values of  $\Pi_0$ , chosen near the boundaries of the explored parameter space (see Additional file 1: Figure S3).

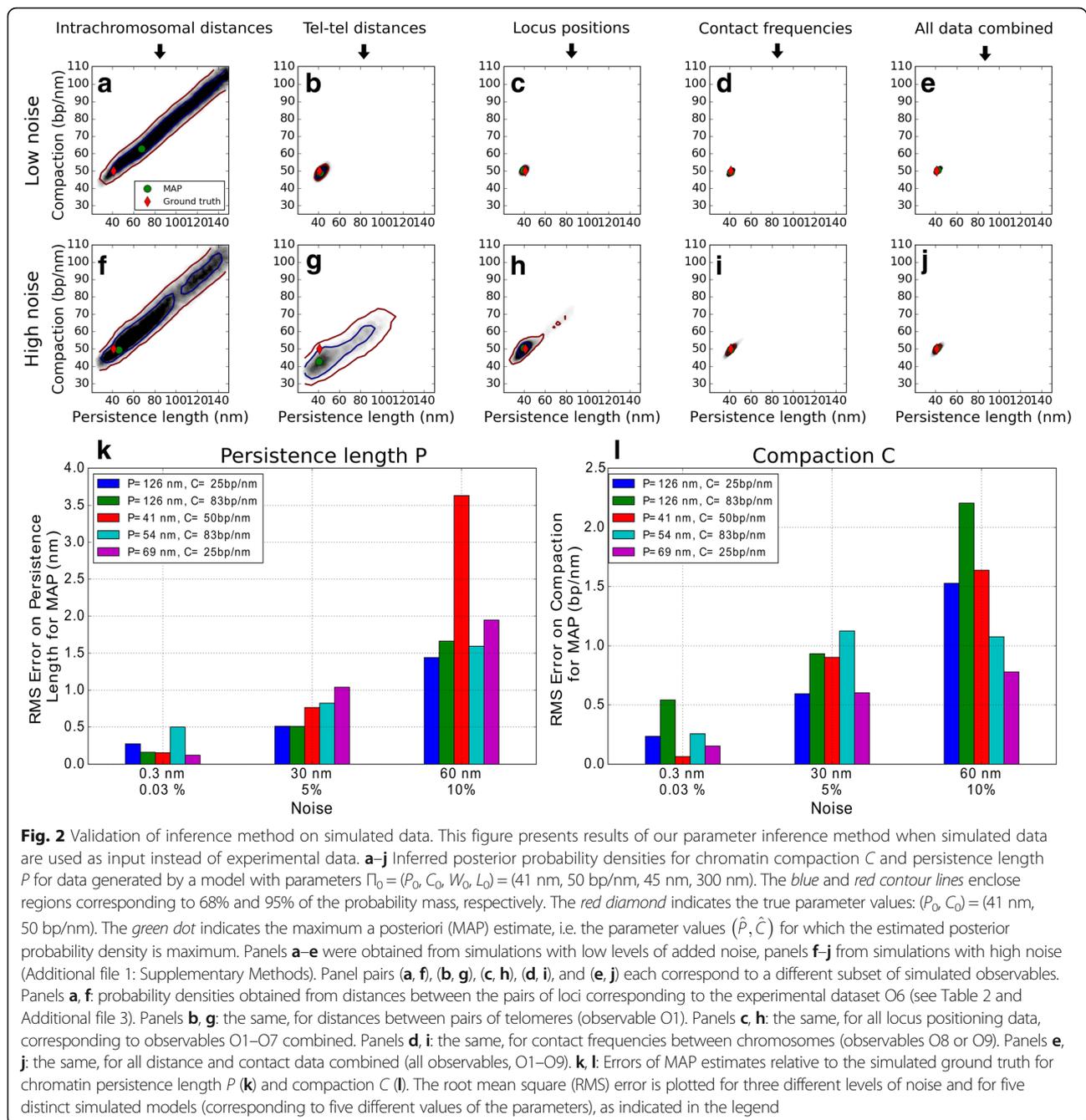
Figure 2a–j shows  $p(P, C|D^S)$ , the joint posterior probability for chromatin compaction  $C$  and rigidity  $P$ , computed for different subsets of observables  $D^S$  or all observables together, for  $\Pi_0 = (P_0, C_0, W_0, L_0) = (41 \text{ nm}, 50 \text{ bp/nm}, 45 \text{ nm}, 300 \text{ nm})$ , and assuming either low (Fig. 2a–e) or high levels of noise (Fig. 2f–j). These plots highlight how different observables constrain chromatin compaction and rigidity. For example, it is apparent from the elongated probability density in Fig. 2a that intrachromosomal distances can be used to determine  $C$  if  $P$  is known, or vice versa, but do not suffice to determine both parameters simultaneously. This is consistent with the well-known behavior of ideal or real polymer chains, where the mean square distance  $\langle R^2 \rangle$  between two loci separated by genomic distance  $s$  depends on  $P$

and  $C$  only via their combination  $P^{1-\nu}/C^\nu$  through the relation  $\langle R^2 \rangle \propto (P^{1-\nu}/C^\nu)^2 s^{2\nu}$ , (where  $\nu = 0.5$  and  $\nu = 0.6$  for an ideal and real chain, respectively [59]). However, our analysis can also reveal less obvious constraints. For example, Fig. 2b shows that unlike intrachromosomal distances, distances between telomeres do in fact allow to simultaneously determine  $P$  and  $C$ . The same holds true for contact frequencies between chromosomes (Fig. 2d). The probability densities obtained by combining all distances, or all distances and contact frequencies are shown in Fig. 2c and e, respectively. For low levels of noise, the computed posterior probability density based on the combined data is sharply peaked at the true parameter value  $\Pi_0$  (Fig. 2e). Picking the MAP parameters, we obtain an accurate match to the true values  $C_0$  and  $P_0$ , with root mean squared (RMS) errors  $< 1 \text{ nm}$  for  $P$  and  $< 1 \text{ bp/nm}$  for  $C$ , based on six independent simulations, and for each of the five parameter values  $\Pi_0$  (Fig. 2e, k, l). At the highest level of noise we considered, the posterior probability densities  $p(P, C|D^S)$  broadened for all sets of observables (Fig. 2f–j), as expected, implying that the parameters were less well constrained by the data, but still contained the true value ( $P_0, C_0$ ) within the 95% percentile region. This was true for all five parameter values  $\Pi_0$  tested. We further quantified the MAP estimation error as function of noise (Fig. 2k, l). While the RMS error increased with noise, as expected, we found that it always remained  $< 4 \text{ nm}$  for  $P$  and  $< 2.5 \text{ bp/nm}$  for  $C$  (Fig. 2k, l). We obtained similar errors, when instead of using a core simulation to generate the noisy input data, we used an independent replica that had not been employed for the interpolation in Fig. 1 (for  $(P_0, C_0) = (41 \text{ nm}, 50 \text{ bp/nm})$ ) (Additional file 1: Figure S4a, b, "replica").

Thus, realistic levels of noise in the experimental data should entail only moderate errors in determining chromatin compaction and rigidity by our method.

### Chromatin parameter inference is robust to moderate model mismatch

To further assess the robustness of our parameter inference method, we tested the effect of a model mismatch, i.e. when the data are not strictly consistent with any of the assumed models (even discounting noise). To do this, we generated synthetic data from a new simulation with parameter values  $\Pi$  outside of the range explored by our core simulations. Specifically, we increased or decreased the parameter  $W_{\text{rDNA}}$  that specified the diameter of the chromosome region representing the rDNA, in such a way that the net nucleolar volume  $V$  changed from  $V = 14\%$  of the nuclear volume (its value in all core simulations) to either  $V = 5\%$ ,  $10\%$ ,  $16\%$ ,  $20\%$ , or  $30\%$  (Table 1). In absence of noise, the MAP estimation of  $P$  and  $C$ , as computed from the combined data, fell within  $7 \text{ nm}$  and  $11 \text{ bp/nm}$  of the ground truth (RMS error),



respectively, except for the simulation with  $V = 30\%$  (Additional file 1: Figure S4a, b). In presence of added noise, the RMS errors of MAP estimations increased, but remained within 16 nm and 10 bp/nm even for high noise, again except for  $V = 30\%$  (Additional file 1: Figure S4a, b). For  $V = 30\%$ , errors became unacceptably high ( $\approx 40 \text{ nm}$  and  $\approx 30 \text{ bp/nm}$  for  $P$  and  $C$ , respectively) (Additional file 1: Figure S4a, b). However, this case corresponds to a nucleolus effectively occupying almost  $2V = 60\%$  of the nuclear volume, which is

unreasonably large and leads to a considerably degraded agreement of model predictions with the experimental data (Additional file 1: Figure S4c).

These results suggest that our method to infer mechanical chromatin parameters is robust both to realistic levels of noise and to reasonable parameter mismatch.

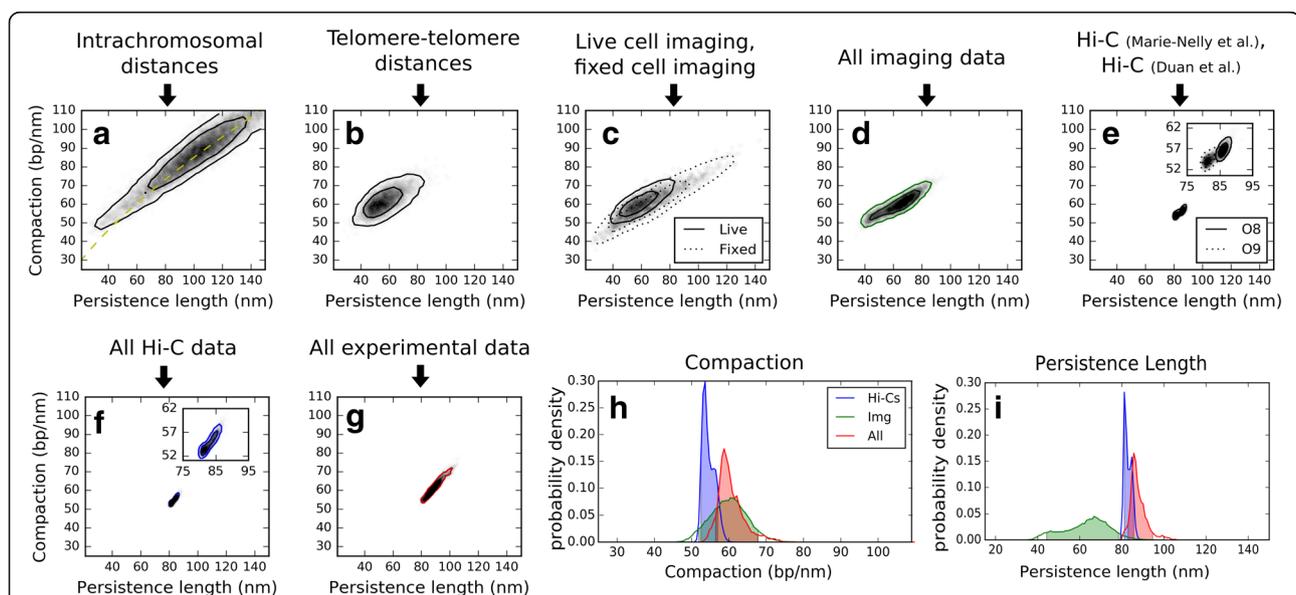
#### New constraints on mechanical chromatin properties

We then applied our inference method to the experimental data described above (Table 2 and Additional file 3).

The posterior probability densities for persistence length  $P$  and compaction  $C$  estimated by our method are shown in Fig. 3, for different subsets of the data (Fig. 3a–f), and all datasets combined (Fig. 3g). Intrachromosomal distances alone constrained the compaction  $C$  and persistence length  $P$  to the vicinity of a curve given by:  $C \propto P^{0.7}$  (Fig. 3a), but these data could not separately determine  $C$  and  $P$  on their own, as expected from the above simulation results (Fig. 2a). By contrast, also as expected (Fig. 2b), distances between telomeres allowed to constrain both parameters to a much smaller range, namely  $C = 61 \pm 6$  bp/nm and  $P = 56 \pm 10$  nm (mean  $\pm$  standard deviation) (Fig. 3b). When all imaging data were combined, a similar posterior probability density was obtained, with  $C = 60 \pm 5$  bp/nm and  $P = 63 \pm 11$  nm (Fig. 3d). Interestingly, imaging data obtained in live and fixed cells resulted in similar probability densities, despite the potential artefacts caused by fixation [50, 60] (Fig. 3c).

We next analyzed two independent Hi-C datasets [29, 30] and found that they again led to very similar probability densities (Fig. 3e). This is remarkable, given that these data were obtained using different experimental protocols and that both led to rather narrow densities. The probability

density obtained when combining both Hi-C datasets is shown in Fig. 3f. Although the joint density obtained from Hi-C does not strictly overlap with that determined by imaging, it is relatively close (Fig. 3d, f). For the compaction  $C$ , the posterior density derived from the Hi-C data is  $55 \pm 2$  bp/nm and overlaps substantially with that derived from the imaging data (Fig. 3h). For the persistence length  $P$ , the Hi-C data yielded an estimate of  $P = 83 \pm 2$  nm, an approximately 30% increase relative to the range  $P = 63 \pm 11$  nm determined by imaging (Fig. 3i). This discrepancy can potentially arise from multiple sources, including estimation errors due to mismatch of other parameters (see above), measurement errors exceeding our estimates, differences in data processing, or biological differences related to the different experimental protocols used in imaging or Hi-C. For example, a fixation-induced shrinkage of the nucleus without alteration of chromatin properties should result in an underestimation of compaction and an overestimation of persistence length. This might explain why the Hi-C data alone predicted lower  $C$  and higher  $P$  than the imaging data alone (Fig. 3h, i). If we ignored these discrepancies and combined all 266 experimental data points, we obtained the posterior density shown in Fig. 3g, based on which the



**Fig. 3** Yeast chromatin parameters inferred from imaging and Hi-C data. This figure shows the posterior probability densities of chromatin compaction  $C$  and persistence length  $P$  as inferred from a variety of experimental datasets. **a–g** joint posterior probability densities for  $(P, C)$  obtained for different subsets of experimental data (**a–f**) or the whole experimental dataset (**g**), as detailed below. The two contour lines shown enclose 68% and 95% of the probability mass. Individual panels correspond to the following experimental datasets: (**a**) modes of 3D distances between eight pairs of loci on chromosome 14 measured by imaging in fixed cells (observable O6, Table 2). The dashed yellow curve has  $C \propto P^{0.7}$ . (**b**) Median 3D distances between 62 pairs of telomeres measured by imaging in live cells [50] (O1). (**c**) Combined set of data from imaging, in live cell or fixed cell experiments. Solid lines: combined data from live cell imaging (111 data points, O1–O4); dotted lines: combined data from imaging fixed cells (28 data points, observables O5–O7). (**d**) All imaging data from fixed and live cells pooled together (O1–O7; 139 data points). (**e**) Summary statistics from genome-wide contact frequencies measured by Hi-C data (see Table 2 and Additional file 1: Supplementary Methods). Solid lines: Hi-C data from [29] (56 data points; O8). Dotted lines: Hi-C data from [30] (56 data points; O9). (**f**) Combination of the two Hi-C datasets [29, 30] (116 data points; O8, O9). (**g**) Combination of all experimental data from imaging and Hi-C (266 data points; O1–O9). **h, i** Probability densities for  $C$  (**h**) and  $P$  (**i**), obtained either from all the imaging data as in (**d**) (green), from the two Hi-C datasets as in (**f**) (blue) or from the combination of imaging and Hi-C data as in (**g**) (red)

compaction and persistence lengths are most likely to fall within the following ranges:  $C = 61 \pm 4$  bp/nm,  $P = 88 \pm 4$  nm (Fig. 3h, i). More conservative estimations were obtained when taking into consideration the discrepancy between the Hi-C and the imaging data, which our method does not account for, and using parameter ranges that span the above confidence intervals determined from both datasets separately. This led to an increase of the range for the persistence length to  $P = 52$ – $85$  nm, while that for the compaction remained similar at  $C = 53$ – $65$  bp/nm. These numbers provide new constraints on the average properties of chromatin in yeast and constitute a main result of this study (see “Discussion”). We note that in contrast to the compaction and rigidity parameters  $C$  and  $P$ , the diameter of the chromatin fiber,  $W$ , was not strongly constrained by the data (Additional file 1: Figure S5).

#### Improved predictions of relative and nuclear locus positions

In addition to constraining chromatin parameters, our analysis yielded a model of the yeast nucleus that matches experimental data significantly better than our previous model [35]. To assess this, we chose the core model with the parameters closest to MAP estimates obtained from the combined data, namely the model with  $C = 50$  bp/nm and  $P = 69$  nm (Table 1, “best model”). With this model, we obtained a much better agreement between predicted and observed distances for 62 pairs of telomeres, as reflected by a Pearson correlation of  $r = 0.75$  and an RMS error of 86 nm, compared to  $r = 0.64$  and 173 nm, respectively, with the previous model [35] (Additional file 1: Figure S6). Furthermore, our improved model also faithfully recapitulated a new set of 55 distance measurements, including 12 new measurements between pairs of loci along the right arm of chromosome 4 (Fig. 4a). Interestingly, for a given genomic separation between the loci, our model predicted consistently larger distances for the pericentromeric region than for the internal region and this was indeed observed experimentally (Fig. 4b). This differential stretching of the pericentromeric chromosome region can be attributed in part to repulsion between the 32 chromosome arms, which are all confined by their centromere to the vicinity of the SPB. Indeed, in simulations of chromosome 4 where all other chromosomes were removed, the difference between pericentromeric and internal regions was reduced to less than half (Additional file 1: Figure S7). This behavior is qualitatively similar to the stretching of polymers in a polymer brush, where chains are grafted at one of their extremities on a common planar surface [59, 61].

We further compared the entire distributions of predicted distances to experimental measurements, for five intrachromosomal and four interchromosomal locus pairs [50] (Additional file 4). As shown in Additional file 1:

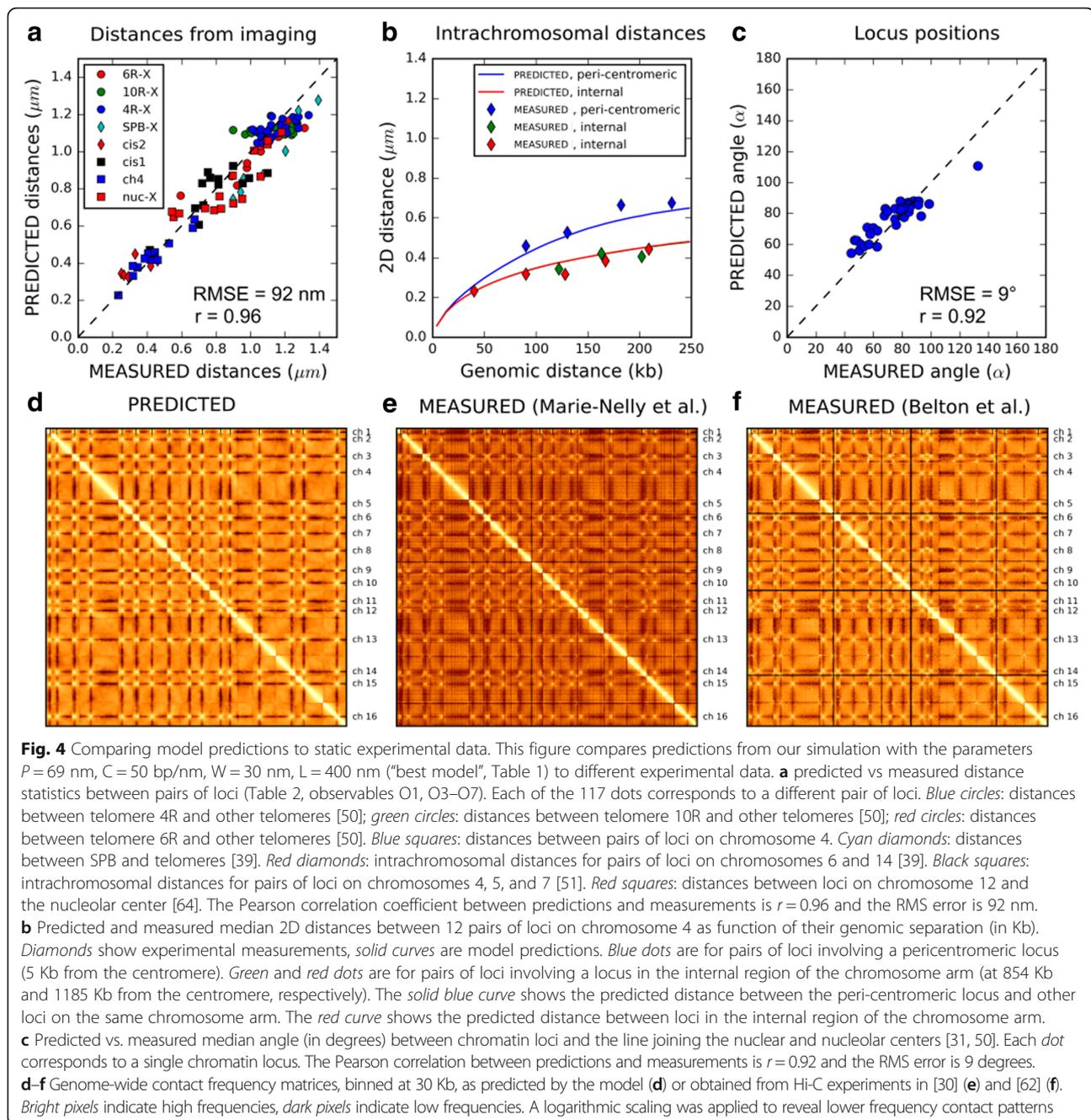
Figure S8, the agreement between predictions and measurements was good or very good, even though only the median distances were used for parameter inference. This further highlights the model’s ability to accurately predict the entire statistical distributions of locus positions.

Our improved model also provides a better fit to the nuclear territories occupied by individual loci, as measured by the median angle between the locus, the nuclear center, and the nucleolar centroids [31, 50]. In our previous model, the predicted angles correlated well with the measured angles (Pearson’s  $r = 0.87$ ), but were systematically underestimated by  $\sim 18^\circ$  [35]. With our new model, the shift reduced to  $-3^\circ$  and the correlation slightly improved to  $r = 0.92$  with a RMS error of  $8.6^\circ$  (Fig. 4c).

#### Improved predictions of genome-wide contact maps

We also compared the predicted genome-wide contact frequency maps to the two Hi-C datasets used for parameter inference above [29, 30] (Fig. 4d, e, Additional file 1: Figure S9a–c). When the contacts were binned at the highest resolution of 5 Kb, our improved model achieved a correlation of  $r = 0.65$  or  $r = 0.85$ , with these data [29, 30] (Additional file 1: Figure S9f). This exceeded the correlation between the two Hi-C datasets themselves ( $r = 0.6$ ) (Additional file 1: Figure S9f). The correlation of predicted contacts with either Hi-C dataset further improved when increasing the genomic bin, and exceeded  $r = 0.9$  for bins of 70 Kb (Additional file 1: Figure S9f). To better assess our model’s predictive power, we further tested it against two additional genome-wide contact maps: a distinct Hi-C study by Belton et al. [62] (Fig. 4f), and a dataset obtained using Micro-C XL, a recently described variation of the Hi-C technique that uses alternative cross-linking agents instead of formaldehyde and DNA digestion by micrococcal nuclease instead of restriction enzymes [63] (Additional file 1: Figure S9e). Correlations between model and data were even higher for the Belton et al. Hi-C data and were only slightly lower for the Micro-C XL data (Additional file 1: Figure S9f). Since these two datasets were not used for parameter inference, this comparison further demonstrates the predictive power of our model.

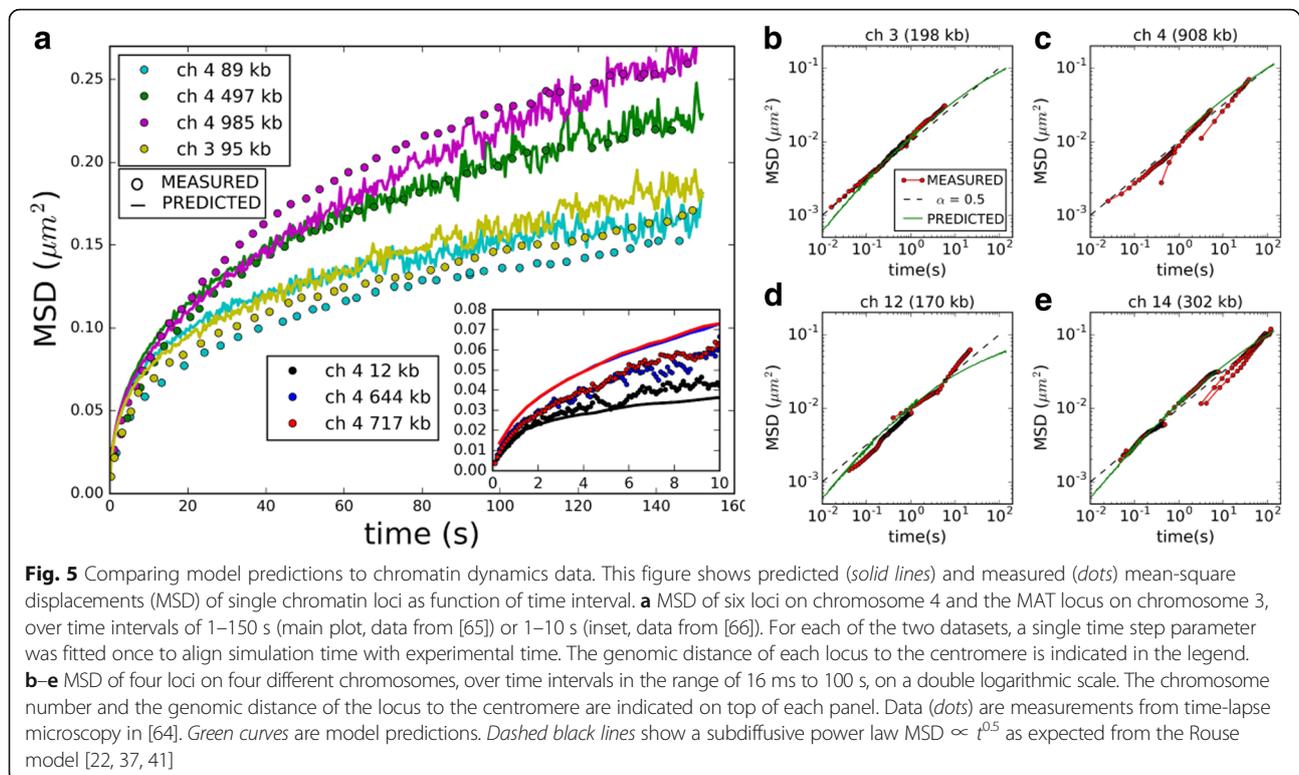
The predicted mean contact frequencies among the 16 chromosomes also agreed very well with measurements from the three Hi-C data (Additional file 1: Figure S10a–d, f–h). We also analyzed the intrachromosomal contact frequencies as function of genomic distance. We separately considered the genome-wide average or an average restricted to contacts involving a centromere (blue and green curves in Additional file 1: Figure S11, respectively). The measured contact frequencies decay faster in the centromeric region than elsewhere in the genome and this was also predicted by the



simulation (Additional file 1: Figure S11). This effect is consistent with the peri-centromeric stretching observed on intrachromosomal distances above (Fig. 4b). The model’s agreement with the MicroC-XL data was also significant, though considerably less good (Additional file 1: Figure S10e, i), because this protocol appears to overestimate interchromosomal and long-range intrachromosomal contacts relative to Hi-C, perhaps due to an excess of random ligations (Additional file 1: Figure S10j, S11).

#### Model recapitulates subdiffusive chromatin dynamics for multiple loci and time scales

Finally, although the model parameters were entirely inferred from static data—i.e. measurements from fixed cells or single snapshots of live cells—we decided to test our model’s ability to predict chromatin dynamics. We therefore compared the mean-squared displacements (MSD) of single loci, predicted by our simulation, to time-lapse microscopy data acquired in three recent studies [64–66] (Additional file 5) (Fig. 5).



In order to map simulation time to physical time, we fitted a single scaling parameter once for each of the three datasets [64–66]. Figure 5a shows the predicted and measured MSD between 1 s and 150 s or between 1 s and 10 s for seven different loci on the right arm of chromosome 4 (4R) and for the MAT locus on chromosome 3 [65, 66]. The distances of loci to the centromere were in the range of 12–985 Kb, spanning over 90% of the 1050 Kb long chromosome arm 4R except for the immediate vicinity of the centromere and telomere. As apparent from Fig. 5a, the model approximately reproduced the measured MSD and recapitulated the observed ordering, i.e. the fact that loci at larger genomic distance from the centromere moved over longer distances during the same time interval. Figure 5b–e shows measured MSD for four loci on four different chromosomes (including the rDNA-carrying chromosome 12, whose spatial configuration is quite different from the other chromosomes [64]) over time intervals in the range of 16 ms to 100 s. As previously pointed out, the MSD obeyed a subdiffusive power law with an exponent  $\sim 0.5$ , roughly consistent with the Rouse polymer model [18, 37, 41, 44, 64]. The model reproduced this behavior relatively well over almost four orders of magnitude of time. Discrepancies between predictions and measurements were similar to or smaller than internal discrepancies within the experimental datasets (Fig. 5b–e). These results confirm the ability of our model to

correctly predict the dynamic behavior of chromatin, even though its structural parameters were inferred exclusively from static data.

## Discussion

In this paper, we described a new approach to infer mechanical parameters of chromatin using a whole nucleus simulation of chromosomes. Compared to methods that estimate hundreds or thousands (or more) parameters to reproduce Hi-C contact data or other genome-wide data [17, 24, 25, 67, 68], our model assumes only a small number of structural parameters, giving it high predictive power. Our Bayesian formulation and sampling approach allow to determine the region of parameter space consistent with the data and to assign probabilities to parameter values. This approach thereby provides information about the uncertainties associated to the inferred parameters for each dataset or, equivalently, about how strongly the data constrain the parameters. We applied our method to a large set of experimental data on yeast nuclear architecture. This analysis led to two main outcomes: new constraints on yeast chromatin and a better model of yeast nuclear organization.

First, we derived new constraints on yeast chromatin compaction and rigidity, with implications for chromatin structure. Although several previous studies have attempted to determine these parameters from imaging or Hi-C data [5, 39–42], they typically used either a single experimental dataset (e.g. only distances between

loci, which are insufficient to unambiguously determine compaction and rigidity simultaneously, as shown in Figs. 2a and 3a) and/or used simpler polymer models that ignored important aspects of nuclear organization, such as tethering of centromeres and telomeres, or entropic repulsion between chromosomes. Accordingly, estimates of compaction  $C$  and persistence length  $P$  from previous studies spanned a wide range: 30–150 bp/nm and <30–200 nm, respectively. By contrast, our analysis is based on a much more varied and complete dataset compiled from multiple independent studies, which include distances between loci in *cis* and *trans*, nuclear locus territories, and two Hi-C datasets. We analyzed these data using a chromosome model that accounts for nuclear confinement, steric hindrance among chromosomes, tethering at centromere and telomeres, and the nucleolar compartment. This more elaborate model accounts for non-trivial features of chromosome organization, such as stretching of peri-centromeric chromatin (Fig. 4b). Most importantly, our analysis supports a relatively narrow range of values for chromatin compaction  $C$  (53–65 bp/nm) and significantly constrains the bending persistence length  $P$  (52–85 nm) in yeast chromatin.

Given a nucleosome repeat length of 167 bp in yeast [69], this estimated compaction corresponds to approximately four nucleosomes per 11 nm turn of the DNA double helix. This has implications for the possible structures of the chromatin fiber and can be confronted to two textbook structures: the 30 nm fiber, in which the  $\approx$  11 nm diameter nucleosomes are tightly packed together in a compact 3D structure of  $\sim$ 30 nm diameter, and the beads-on-a-string structure, where consecutive nucleosomes are arranged in a much looser  $\sim$ 10 nm diameter fiber. The classical 30 nm fiber has a compaction of  $\sim$ 11 nucleosomes per 11 nm, much larger than our estimate above [70]. Because of its short nucleosomal repeat length, the yeast chromatin fiber can be expected to adopt a somewhat looser  $\sim$ 20 nm (rather than 30 nm) diameter structure, with a compaction of approximately six nucleosomes per 11 nm [69, 70]. This is, however, still too compact to agree with our estimates. Thus, our results cast doubt on the existence of a compact 20–30 nm fiber throughout the yeast genome. They complement previous studies questioning the existence of 30 nm fibers in other organisms [4, 71, 72]. On the other hand, a stretched 10 nm fiber, with a compaction of approximately one nucleosome per 11 nm, is insufficiently compact to fit our estimates [73]. Thus, the possibility arises that the chromatin fiber has a structure quite different from both textbook structures.

As a caveat, since our model rests on the assumption of a homogeneous chromatin fiber throughout the genome (except for the rDNA region), our estimates only pertain to the average properties of the fiber. The

chromatin fiber is potentially heterogeneous and may consist of alternating stretches with different compaction and rigidity. Hi-C data on the human genome, for example, indicate that chromosomes are partitioned into “close” and “open” regions of chromatin on the order of 10 Mb, most readily identified as blocks of negative and positive values in contact correlation matrices (Additional file 1: Figure S12a) [15]. In contrast, correlation matrices of budding yeast Hi-C data [62] at genomic resolutions of 10 Kb remain quite homogeneous at high positive correlations, except in the pericentromeric region where correlations are strongly negative (Additional file 1: Figure S12b). Although this pericentromeric pattern might first be interpreted as a signature of chromatin heterogeneity, it is in fact recapitulated by our homogeneous fiber model (Additional file 1: Figure S12c), confirming that homogeneity is a valid assumption at this genomic scale. Nevertheless, heterogeneities at smaller scales, such as rapid alternations between compact and less compact chromatin might in principle lead to detectable structures in the contact matrix (Additional file 1: Figure S13a, c). Such signatures are hard to identify in the Hi-C data [62] (Additional file 1: Figure S13d), although they may be present in the raw Micro-C XL data (Additional file 1: Figure S13e). However, reliably distinguishing these signatures from potential biases [54, 74, 75] is challenging, and none of the observables used for parameter inference in our study is sensitive to such small-scale heterogeneities (Additional file 1: Figure S14). Although large-scale heterogeneities are not supported by the Hi-C data, we therefore cannot rule out small scale heterogeneities, e.g. rapidly alternating stretches of compact 20–30 nm fibers and loose 10 nm fibers. For such a heterogeneous structure, assuming the above compaction values, we can estimate that roughly half of the linear length of the chromatin fiber (ignoring the non-rDNA yeast genome) would be structured as a compact 20–30 nm fiber and the other half as a 10 nm beads-on-a-string fiber. Accounting for the different compactions, these proportions correspond to 86% and 14% of the genome, respectively. Our modeling approach cannot currently map these potential regions to the genome. Extending our method to heterogeneous fibers at small (<5 Kb) genomic scales would imply estimating thousands of parameters at the risk of overfitting and loss of predictive power. Nonetheless, future heterogeneous models that avoid overfitting might leverage state-of-the-art genomic contact data and high resolution imaging data to shed more light on potential structural variations of the chromatin fiber [12, 24, 42, 60, 76].

In addition to new insights into average yeast chromatin structure, our analysis yielded a set of models that provide a better explanation of experimental data than our previous model [35]. In particular, for parameters that maximize the

posterior probability, the model agrees well with measured distances between loci (means and distributions), intranuclear gene territory positions, and several aspects of Hi-C data, including Hi-C data that were not used for parameter estimation. Moreover, unlike Monte-Carlo simulations or structural ensembles obtained by optimization [24, 36], our model can predict the dynamic behavior of chromatin, in good agreement with observations. Based on these results, it is now possible to make accurate predictions of absolute and relative locus positions, movements and contact frequencies throughout the yeast genome. This in turn has implications for a quantitative understanding of functional processes such as DNA repair by homologous recombination and mating type switching in yeast [37, 38, 77, 78].

## Conclusion

In summary, our work provides new insights into yeast chromatin fiber structure, and proposes a substantially improved predictive model of yeast nuclear architecture and dynamics, both of which will help to advance our understanding of structure-function relations in the nucleus. The computational analysis method proposed here should also be applicable to characterizing chromatin structure and chromosome organization in many other organisms.

## Additional files

**Additional file 1:** Supplementary information. Includes: Supplementary Figure S1–S19; Supplementary Methods A–F; Supplementary References. (PDF 19000 kb)

**Additional file 2:** Table listing all 144 models used for inference and their parameter values. (XLSX 40 kb)

**Additional file 3:** Table listing all image-based distance measurements used for parameter inference. (XLSX 77 kb)

**Additional file 4:** Entire set of measured distances between nine pairs of loci used to analyze distance distributions. (XLSX 160 kb)

**Additional file 5:** Mean square displacements measured on multiple chromatin loci in independent studies. (XLSX 79 kb)

**Additional file 6:** Zipped archive containing the scripts to run the best model simulation. (ZIP 241 kb)

## Acknowledgements

We thank the IT service of Institut Pasteur for access to the computing cluster and relevant assistance. We thank O. Gadal, A. Bancaud, and K. Bystrycky for providing data, J. Parmar for help with Hi-C data analysis and proofreading, and M. Woringer for comments on the manuscript. We thank the anonymous reviewers for their constructive criticism of our manuscript, which led to important improvements. C.Z. acknowledges support from the Siebel Stem Cell Foundation during a visit to UC Berkeley, where a part of this paper was completed.

## Funding

This work was funded by Institut Pasteur, Agence Nationale de la Recherche (grant ANR-11-MONU-020–02 to C.Z.), Fondation pour la Recherche Médicale (Equipe FRM DEQ20150331762 grant to C.Z.) and a Labex Who am I grant (EE-2013 to E.F.). J.-M.A. was recipient of a Pasteur-Roux fellowship from Institut Pasteur.

## Availability of data and materials

The experimental data analyzed during this study are either available online (see Additional file 1: Supplementary Information) or included as Additional

files 3, 4 and 5. The simulation libraries are available at <https://doi.org/10.5281/zenodo.292972> and the LAMMPS [79] scripts to run our “best model” are available in Additional file 6.

## Authors' contributions

Conceived the study: J-MA, CZ. Performed simulations and analyses: J-MA. Provided experimental data: SH, EF. Wrote the paper: CZ with input from J-MA. All authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Ethics approval and consent to participate

Not applicable.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Unité Imagerie et Modélisation, Institut Pasteur, 25 rue du Docteur Roux, 75015 Paris, France. <sup>2</sup>UMR 3691, CNRS; C3BI, USR 3756, IP CNRS, Paris, France. <sup>3</sup>Université Paris Diderot, Sorbonne Paris Cité, Cellule Pasteur, 75015 Paris, France. <sup>4</sup>Chromosome Biology and Dynamics, Hôpital Saint Louis, Paris, France.

Received: 16 September 2016 Accepted: 23 March 2017

Published online: 03 May 2017

## References

- Luger K, Dechassa ML, Tremethick DJ. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol.* 2012;13:436–47.
- Cavalli G, Misteli T. Functional implications of genome topology. *Nat Struct Mol Biol.* 2013;20:290–9.
- Quénet D, McNally JG, Dalal Y. Through thick and thin: the conundrum of chromatin fibre folding in vivo. *EMBO Rep.* 2012;13:943–4.
- Joti Y, Hikima T, Nishino Y, Kamada F, Hihara S, Takata H, et al. Chromosomes without a 30-nm chromatin fiber. *Nucleus.* 2012;3:404–10.
- Sanborn AL, Rao SSP, Huang S-C, Durand NC, Huntley MH, Jewett AJ, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci USA.* 2015;112:201518552.
- Tremethick DJ. Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell.* 2007;128:651–4.
- Song F, Chen P, Sun D, Wang M, Dong L, Liang D, et al. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science.* 2014;344:376–80.
- Fussner E, Strauss M, Djuric U, Li R, Ahmed K, Hart M, et al. Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. *EMBO Rep.* 2012;13:992–6.
- Grigoryev SA, Woodcock CL. Chromatin organization - the 30 nm fiber. *Exp Cell Res.* 2012;318:1448–55.
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature.* 2012;485:376–80.
- Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature.* 2012;485:381–5.
- Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell.* 2014;159:1665–80.
- Smith EM, Lajoie BR, Jain G, Dekker J. Invariant TAD boundaries constrain cell-type-specific looping interactions between promoters and distal elements around the CFTR locus. *Am J Hum Genet.* 2016;98:185–201.
- Flavahan WA, Drier Y, Liao BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature.* 2015;529:110–4.
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragozcy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 2009;326:289–93.

16. Imakaev MV, Fudenberg G, Mirny LA. Modeling chromosomes: Beyond pretty pictures. *FEBS Lett.* 2015;589:3031–6.
17. Rosa A, Zimmer C. Computational models of large-scale genome architecture. *Int Rev Cell Mol Biol.* 2014;307:275–349.
18. Rosa A, Everaers R. Structure and dynamics of interphase chromosomes. *PLoS Comput Biol.* 2008;4:e1000153.
19. Grosberg AY. How two meters of DNA fit into a cell nucleus: Polymer models with topological constraints and experimental data. *Polym Sci Ser C.* 2012;54:1–10.
20. Huet S, Lavelle C, Ranchon H, Carrivain P, Victor J-M, Bancaud A. Relevance and limitations of crowding, fractal, and polymer models to describe nuclear architecture. *Int Rev Cell Mol Biol.* 2014;307:443–79.
21. Langowski J. Polymer chain models of DNA and chromatin. *Eur Phys J E Soft Matter.* 2006;19:241–9.
22. Doi M, Edwards AM. The theory of polymer dynamics. Oxford: Oxford University Press; 1988.
23. Chiariello AM, Annunziatella C, Bianco S, Esposito A, Nicodemi M. Polymer physics of chromosome large-scale 3D organisation. *Sci Rep.* 2016;6:29775.
24. Giorgetti L, Galupa R, Nora EP, Piolot T, Lam F, Dekker J, et al. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell.* 2014;157:950–63.
25. Brackley CA, Brown JM, Waithe D, Babbs C, Davies J, Hughes JR, et al. Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models. *Genome Biol.* 2016;17:59.
26. Rosa A, Becker NB, Everaers R. Looping probabilities in model interphase chromosomes. *Biophys J.* 2010;98:2410–9.
27. Zimmer C, Fabre E. Principles of chromosomal organization: lessons from yeast. *J Cell Biol.* 2011;192:723–33.
28. Taddei A, Schober H, Gasser SM. The budding yeast nucleus. *Cold Spring Harb Perspect Biol.* 2010;2:a000612.
29. Duan Z, Andronescu M, Schutz K, McIlwain S, Kim YJ, Lee C, et al. A three-dimensional model of the yeast genome. *Nature.* 2010;465:363–7.
30. Marie-Nelly HH, Marbouty M, Cournac A, Liti G, Fischer G, Zimmer C, et al. Filling annotation gaps in yeast genomes using genome-wide contact maps. *Bioinformatics.* 2014;30:2105–13.
31. Berger AB, Cabal GG, Fabre E, Duong T, Buc H, Nehrbass U, et al. High-resolution statistical mapping reveals gene territories in live yeast. *Nat Methods.* 2008;5:1031–7.
32. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science (80-).* 2002;295:1306–11.
33. Tokuda N, Terada TP, Sasai M. Dynamical modeling of three-dimensional genome organization in interphase budding yeast. *Biophys J.* 2012;102:296–304.
34. Gehlen LR, Gruenert G, Jones MB, Rodley CD, Langowski J, O'Sullivan JM. Chromosome positioning and the clustering of functionally related loci in yeast is driven by chromosomal interactions. *Nucleus.* 2012;3(4):370–83.
35. Wong H, Marie-Nelly H, Herbert S, Carrivain P, Blanc H, Koszul R, et al. A Predictive computational model of the dynamic 3D interphase yeast nucleus. *Curr Biol.* 2012;22:1881–90.
36. Tjong H, Gong K, Chen L, Alber F. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Res.* 2012;22:1295–305.
37. Wong H, Arbona J-M, Zimmer C. How to build a yeast nucleus. *Nucleus.* 2013;4:361–6.
38. Agmon N, Liefshitz B, Zimmer C, Fabre E, Kupiec M. Effect of nuclear architecture on the efficiency of double-strand break repair. *Nat Cell Biol.* 2013;15:694–9.
39. 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;101:16495–500.
40. Dekker J. Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction. *J Biol Chem.* 2008;283:34532.
41. Hajjoul H, Mathon J, Ranchon H, Goiffon I, Mozziconacci J, Albert B, et al. High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. *Genome Res.* 2013;23:1829–38.
42. Hsieh T-HS, Weiner A, Lajoie B, Dekker J, Friedman N, Rando OJ. Mapping nucleosome resolution chromosome folding in yeast by micro-C. *Cell.* 2015;162:108–19.
43. Cui Y, Bustamante C. Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure. *Proc Natl Acad Sci. USA.* 2000;97:127–32.
44. Cabal GG, Genovesio A, Rodriguez-Navarro S, Zimmer C, Gadal O, Lesne A, et al. SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature.* 2006;441:770–3.
45. Egecioglu D, Brickner JH. Gene positioning and expression. *Curr Opin Cell Biol.* 2011;23:338–45.
46. Jaspersen SL, Winey M. The budding yeast spindle pole body: structure, duplication, and function. *Annu Rev Cell Dev Biol.* 2004;20:1–28.
47. Hediger F, Neumann FR, Van Houwe G, Dubrana K, Gasser SM. Live imaging of telomeres: yKu and Sir proteins define redundant telomere-anchoring pathways in yeast. *Curr Biol.* 2002;12:2076–89.
48. Léger-Silvestre I, Trumtel S, Noaillic-Depeyre J, Gas N. Functional compartmentalization of the nucleus in the budding yeast *Saccharomyces cerevisiae*. *Chromosoma.* 1999;108:103–13.
49. O'Toole ET, Winey M, McIntosh JR. High-voltage electron tomography of spindle pole bodies and early mitotic spindles in the yeast *Saccharomyces cerevisiae*. *Mol Biol Cell.* 1999;10:2017–31.
50. Thérizols P, Duong T, Dujon B, Zimmer C, Fabre E. Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proc Natl Acad Sci. USA.* 2010;107:2025.
51. Kimura H, Shimooka Y, Nishikawa J-I, Miura O, Sugiyama S, Yamada S, et al. The genome folding mechanism in yeast. *J Biochem.* 2013;154:137–47.
52. Ober RJ, Ram S, Ward ES. Localization accuracy in single-molecule microscopy. *Biophys J Biophysical Soc.* 2004;86:1185–200.
53. Zimmer C. From microbes to numbers: extracting meaningful quantities from images. *Cell Microbiol.* 2012;14:1828–35.
54. Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods.* 2012;9:999–1003.
55. Yaffe E, Tanay A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat Genet.* 2011;43:1059–65.
56. Rieping W, Habeck M, Nilges M. Inferential structure determination. *Science (80-).* 2005;309:303.
57. Box GEP, Tiao GC. Bayesian inference in statistical analysis. Chichester: John Wiley & Sons; 2011.
58. Foreman-Mackey D, Hogg DW, Lang D, Goodman J. Emcee: The MCMC Hammer. *Publ Astron Soc Pacific.* 2013;125:306–12.
59. Rubinstein M, Colby R. Polymer physics. Oxford: Oxford University Press; 2003.
60. Boettiger AN, Bintu B, Moffitt JR, Wang S, Beliveau BJ, Fudenberg G, et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature.* 2016;529:418–22.
61. Milner ST. Polymer brushes. *Science (80-).* 1991;251:905–14.
62. Belton J-M, Lajoie BR, Audibert S, Cantaloube S, Lassadi I, Goiffon I, et al. The conformation of yeast chromosome III is mating type dependent and controlled by the recombination enhancer. *Cell Rep.* 2015;13:1855–67.
63. Hsieh T-HS, Fudenberg G, Goloborodko A, Rando OJ. Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. *Nat Methods.* 2016;13:1009–11.
64. Albert B, Mathon J, Shukla A, Saad H, Normand C, Léger-Silvestre I, et al. Systematic characterization of the conformation and dynamics of budding yeast chromosome XII. *J Cell Biol.* 2013;202:201–10.
65. Strecker J, Gupta GD, Zhang W, Bashkurov M, Landry M-C, Pelletier L, et al. DNA damage signalling targets the kinetochore to promote chromatin mobility. *Nat Cell Biol.* 2016;18:281–90.
66. Spichal M, Brion A, Herbert S, Cournac A, Marbouty M, Zimmer C, et al. Evidence for a dual role of actin in regulating chromosome organization and dynamics in yeast. *J Cell Sci.* 2016;129:681–92.
67. Baù D, Sanyal A, Lajoie BR, Capriotti E, Byron M, Lawrence JB, et al. The three-dimensional folding of the  $\alpha$ -globin gene domain reveals formation of chromatin globules. *Nat Struct Mol Biol.* 2011;18:107–14.
68. Kalhor R, Tjong H, Jayatilaka N, Alber F, Chen L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat Biotechnol.* 2012;30:90–8.
69. Woodcock CL, Skoultchi AI, Fan Y. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res.* 2006;14:17–25.
70. Routh A, Sandin S, Rhodes D. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proc Natl Acad Sci U S A.* 2008;105:8872–7.
71. Fussner E, Ching RW, Bazett-Jones DP. Living without 30 nm chromatin fibers. *Trends Biochem Sci.* 2011;36:1–6.

72. Eltsov M, MacLellan KM, Maeshima K, Frangakis AS, Dubochet J. Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. *Proc Natl Acad Sci. USA*. 2008;105:19732.
73. Thoma F, Koller T. Unravalled nucleosomes, nucleosome beads and higher order structures of chromatin: Influence of non-histone components and histone H1. *J Mol Biol*. 1981;149:709–33.
74. Cournac A, Marie-Nelly H, Marbouty M, Koszul R, Mozziconacci J. Normalization of a chromosomal contact map. *BMC Genomics*. 2012;13:436.
75. Marie-Nelly H, Marbouty M, Cournac A, Flot J-F, Liti G, Parodi DP, et al. High-quality genome (re)assembly using chromosomal contact data. *Nat Commun*. 2014;5:5695.
76. Ricci MA, Manzo C, García-Parajo MF, Lakadamyali M, Cosma MP. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell*. 2015;160:1145–58.
77. Lee C-S, Wang RW, Chang H-H, Capurso D, Segal MR, Haber JE. Chromosome position determines the success of double-strand break repair. *Proc Natl Acad Sci U S A*. 2016;113:E146–54.
78. Haber JE, Abraham J, Nasmyth KA, Strathern JN, Klar AJ, Hicks JB, et al. Mating-type genes and MAT switching in *Saccharomyces cerevisiae*. *Genetics*. 2012;191:33–64.
79. Plimpton S. Fast parallel algorithms for short-range molecular dynamics. *J Comput Phys*. 1995;117:1–19.
80. Bystricky K, Laroche T, van Houwe G, Blaszczyk M, Gasser SM. Chromosome looping in yeast: telomere pairing and coordinated movement reflect anchoring efficiency and territorial organization. *J Cell Biol*. 2005;168:375–87.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

