Article

Heterogeneous Spatial Distribution of Transcriptional Activity in Budding Yeast Nuclei

Naoko Tokuda[1] and Masaki Sasai[1,2,*]

[1]Department of Computational Science and Engineering and [2]Department of Applied Physics, Nagoya University, Nagoya, Japan

ABSTRACT   Recent microscopic and simulation studies have shown that the genome structure fluctuates dynamically in the nuclei of budding yeast *Saccharomyces cerevisiae*. This genome-wide movement should lead to the fluctuations of individual genes in their territorial regions. This raises an intriguing question of whether the resulting distribution of genes is correlated to their transcriptional activity. An effective method for examining this correlation is to analyze how the spatial distribution of genes and their transcriptional activity are modified by mutation. In this study, we analyzed the modification observed in a budding yeast mutant in which genes necessary for anchoring telomeres to the nuclear envelope, *yku70* and *esc1*, are silenced. Taddei et al. reported that 60 genes are clearly misregulated by this mutation, with 28 and 32 genes downregulated and upregulated, respectively. We calculated the probability density maps of the misregulated genes using a model of dynamical movement of the yeast genome in both wild-type (WT) and *yku70 esc1* mutant and showed that the density of downregulated genes is larger near the nucleolus, whereas the density of upregulated genes is larger at the opposite side of the nucleus. By comparing these genes with those highly (top 200 of transcriptome) and lowly (bottom 200) expressed, we showed that the simulated distribution of 28 downregulated (12 out of 32 upregulated) genes has a distinctly larger overlap with the distribution of lowly (highly) expressed genes in the mutant than in the WT. The remaining 20 upregulated genes are localized near the nuclear envelope both in the WT and in the mutant. These results showed that the transcriptional level of genes is affected by their spatial distribution, thus highlighting the importance of the structural regulation in the yeast genome.

INTRODUCTION

The eukaryotic genome is compactly packed in the nucleus. For example, the haploid genome of budding yeast, with a total length of ~4 mm, is packed in the nucleus with a diameter of 2 μm. Understanding how the genome is organized in such a narrow space has been a long-standing problem in cell biophysics (1–3). In particular, the quantitative understanding of the genome structure and dynamics is important because they should regulate the access of proteins to DNA sequences (4). Recently developed experimental techniques shed new light on this problem. Particularly, the Hi-C method, which combines next-generation sequencing and chromosome conformation capture techniques, is important for inferring the average three-dimensional (3D) genomic structure (5–8). The dynamics of chromosome structural fluctuation has been monitored for budding yeast using the LacO/LacI or TetO/TetR system marked with green fluorescent protein (9–11). Here, we focus on haploid budding yeast *Saccharomyces cerevisiae* as a model organism. Using

the Hi-C results (12) and the results from fluorescence monitoring (10,11,13) as reference data, we develop a computational model to simulate the 3D structure and dynamics of the yeast genome to improve upon the model we previously reported (14). With this improved version of the model, we show that the largely fluctuating genome determines the spatial distribution of genes in the nucleus, and we discuss how the spatial distribution of genes is correlated to their transcriptional regulation.

Experimental and computational evidence suggests relationships between the spatial distribution of genes and their transcriptional regulation. For example, telomeres and subtelomeric regions in yeast are distributed around the nuclear envelope to coexist with silent information regulators (SIR proteins; Sir2, Sir3, and Sir4), which form complexes with other factors at around the nuclear envelope (4,15). The effect of SIR proteins to repress genes located next to telomeres is referred to as the "telomere position effect" (16–18). Some part of the repeats of ribosomal DNA (rDNA) in yeast is repressed through localization of rDNA near the nuclear envelope, which is mediated by inner nuclear membrane proteins including Heh1 and others (19,20). RNA Pol III-transcribed genes, such as transfer

CrossMark

RNA (tRNA) genes in yeast, are localized around the nucle-olus (21), and RNA Pol II-transcribed genes nearby the tRNA genes are repressed, which is a phenomenon termed "tRNA gene-mediated silencing" (22). In mammalian cells, actively transcribed genes are localized in discrete foci as transcription factories, which are associated with the local-ization of RNA Pol II (23–25). For fission yeast (26,27) and budding yeast (27), a similar localization of highly ex-pressed genes was suggested using computational models of 3D genome structures. Moreover, it was suggested that functionally related genes belonging to particular gene ontology groups tend to colocalize (26,27). Localization of the functionally related genes was also assessed through statistical analyses of the Hi-C data (28–30).

One approach to examine the hypothesis that the spatial distribution of genes is correlated with transcriptional regu-lation is to analyze mutants in which the genome structure is arranged in a different way from the wild-type (WT). In the WT budding yeast nucleus, telomeres are localized around the nuclear envelope. Taddei et al. (31) generated a mutant in which Yku70 and Esc1, which are proteins necessary for anchoring telomeres to the nuclear envelope, are knocked out. They showed that the pattern of gene expression in the microarray data was significantly changed by this mutation (31). In particular, 60 genes were misregulated by >1.5-fold change in the transcription level, wherein 28 genes were downregulated and 32 genes were upregu-lated (31). They considered that genes located near telo-meres, which detect the telomere position effect in the WT, are upregulated because they do not interact with the SIR proteins when telomeres are not anchored to the nuclear envelope in the mutant. They showed that seven upregulated genes are indeed located within 20 kb from telomeres, and argued that this portion (7/32 = 22%) is higher than the probability of finding randomly selected genes near telo-meres. However, the reason for the upregulation of the remaining 25 genes is not clear. The authors also demon-strated a statistical tendency for the downregulation of genes whose promoters contained RNA polymerase A and C pro-moters, ribosomal RNA processing elements, or Abf1 mo-tifs. They considered that SIR proteins are sequestered in the SIR complex in the WT, which is formed together with Yku70 and Esc1, and that SIR proteins are released to the nucleoplasm and operate on these target promoters to downregulate the corresponding genes when the SIR complex is lost in the mutant. However, not all the downre-gulated genes have the above RNA polymerase A and C pro-moters, ribosomal RNA processing elements, or Abf1 motifs. A possible explanation is that many genes are mis-regulated by the action of genes that directly interact with SIR proteins; however, there is as yet no evidence for such an indirect mechanism. Taddei et al. (31) also argued that there is no evidence for the bias in the one-dimensional (1D) positions of misregulated genes distributed along chromosomes. In this way, the reason for the observed mis-

regulation has not yet been fully elucidated, and other mech-anisms than the one suggested in Taddei et al. (31) may also occur in the mutant. In this article, we consider the possibil-ity that the shifted spatial distribution of genes in the mutant may lead to the shift in the pattern of gene expression, and this hypothesis is tested using a computer simulation.

When initiating a computer simulation, some basic prob-lems must be considered: One is about the approach using Hi-C data in modeling. In the model of Duan et al. (12), the Hi-C data were used as major constraints to simulate a genome structure together with some minor constraints rep-resenting localization of centromeres and confinement of rDNA in the nucleolus. In other models, the Hi-C data were not used for modeling, and structures were simulated under the constraints to represent interactions between chro-mosomes and nuclear structures such as the spindle pole body (SPB), nuclear envelope, and nucleolus (27,32,33). In our previous model, constraints generated from the Hi-C data were used as moderate forces to restrict structural fluctuation. The Hi-C based interactions to represent these constraining forces and interactions to represent nuclear structures were used in a combined way in our previous model (14). Because of the small size of the yeast nucleus, interactions between chromosomes and nuclear structures have significant effects on the constraining of the genome structure; therefore, these interactions should be carefully considered during the construction of a model. Even without using the Hi-C data, the simulated genome structures were in reasonable agreement with the observed experimental features when interaction potentials between chromosomes and nuclear structures were suitably introduced (27,32,33). However, interactions between chromosome loci mediated by proteins binding on chromatins are not represented by the interactions between chromosomes and nuclear struc-tures. We will show later in this article that interactions be-tween tRNA genes and other chromosome loci, which were estimated from the Hi-C data (12), play important roles in determining the genome structure in yeast; therefore, the careful consideration of interactions between chromatin loci appears indispensable. Therefore, in this study, we use the combined interactions including the Hi-C based in-teractions by extending our previous model (14). The other problem to consider is the approach to represent fluctuations in the genome structure. Microscopic observations using of the fluorescent markers have shown that the budding yeast genome fluctuates dynamically in the nucleus (9,11,34). An equation of motion of the genome to simulate these dy-namics has not yet been established; however, we here as-sume that a method analogous to the one used to describe protein folding dynamics can be used in this problem of genome fluctuation, although the physical scale of the two problems differs largely. In protein dynamics, potentials that stabilize the observed average structure (i.e., the native protein structure) can describe the fluctuating structural dy-namics of proteins even for the process of large structural

change, including unfolding/folding. Similarly, we expect that in genome folding, potentials that stabilize the observed average structure are useful to describe the large-amplitude liquidlike fluctuation of genome structure around the average structure inferred from the Hi-C data. We use the Langevin dynamics method to describe fluctuations under these potentials and to calculate the spatial distribution of genes in the nucleus.

In this simulation of Langevin dynamics, 16 chromosomes in haploid budding yeast are described as 16 polymer chains, where each chain is composed of beads representing 3 kb segments of chromatin connected by springs. These coarse-grained chains of chromosomes are assumed to be subjected to the Langevin equation of motion under the potentials representing the Hi-C based interactions and interactions between chromosomes and nuclear structures. This model is an improved version of the previous model (14). Three aspects of this version represent major improvements to the previous version: (1) interactions between tRNA genes and 5S rRNA genes were introduced to explain the localization of tRNA genes around the nucleolus as was observed by monitoring using fluorescence markers (21) and shown by the Hi-C data analysis (12,30); (2) the previous semilunar shape of potential function to represent the constraint for confinement of rDNA in the nucleolus was changed to a more realistic crescent shape; and (3) we removed biases from the Hi-C data by filtering out the noise from the data. The details of the filtering method are described in the Methods section. The *yku70 esc1* mutant in which telomeres are not anchored to the nuclear envelope is represented in the model by turning off the attractive interactions between telomeres and nuclear envelope while these interactions are turned on in the WT model. In this way, we describe the *yku70 esc1* mutation as a perturbation to the WT.

With this model, we simulate the genome structure and fluctuation in both the WT and the *yku70 esc1* mutant and calculate the spatial distribution of the misregulated genes. We also calculate spatial distributions of highly and lowly expressed genes by extracting 200 genes showing the highest and lowest transcription levels, respectively (top 200 and bottom 200 genes, respectively) from the ArrayExpress database (Accession number: E-TABM-630) (31,35). By comparing these distributions, we discuss how the spatial distribution of genes is correlated to their transcriptional activity.

MATERIALS AND METHODS

In a previous study, we developed a model to simulate dynamic genome movement in interphase haploid budding yeast (Tokuda-Terada-Sasai (TTS) model) (14). Here, we use the improved version of the model to better reproduce the observed spatial distributions of centromeres, telomeres, rDNA, and other genes.

We describe 16 chromosomes as polymer chains that are composed of beads connected by finitely extensible springs. One bead corresponds to a

3 kb segment of chromatin, so that each chromosome has 78 (chromosome I) to 806 (chromosome XII) beads and the entire genome comprises 4460 beads. We simulate movement of these 16 chains by numerically solving the following Langevin equation:

$$m \frac{d^2 \mathbf{r}_i^\mu}{dt^2} = -\frac{\partial}{\partial \mathbf{r}_i^\mu} U - \zeta \frac{d\mathbf{r}_i^\mu}{dt} + \mathbf{w}_i^\mu. \quad (1)$$

Here, m is the mass of a bead; $\mathbf{r}_i^\mu$ is the position of the ith bead of the μth chain with $\mu = 1$–16; U is the potential representing interactions between chromosomes and those between chromosomes and inner structures of the nucleus; ζ is the friction coefficient; and the vector $\mathbf{w}_i^\mu$ is Gaussian white noise with

$$\left\langle \mathbf{w}_{i\alpha}^\mu(t)\mathbf{w}_{j\beta}^\nu(t') \right\rangle = 2\zeta T \delta(t - t')\delta_{ij}\delta_{\mu\nu}\delta_{\alpha\beta}, \quad (2)$$

where α and β represent the x, y, or z component of the vector, and T is the effective temperature with the unit of $k_B = 1$. Although the movement of chromosomes in the nucleus is out of equilibrium and the rate of structural change depends on the ATP consumption (4,15), we expect that the rapidly fluctuating movement of local segments to be much faster than the global movement of chains. Therefore, we use white noise to represent such rapid fluctuation and its strength is represented here by the effective temperature T. We use T as a unit to define interaction strengths in U. The potential U in Eq. 1 comprises three terms:

$$U = U_{\text{chain}} + U_{\text{HiC}} + U_{\text{nucleus}}, \quad (3)$$

where the first term describes interactions that define the basic physical features of the chain, the second term represents interactions derived from the Hi-C data (12), and the third term represents interactions between chromosomes and nuclear structures.

Interactions to define the physical features of the chain

The first term in Eq. 3 is composed of two terms as $U_{\text{chain}} = U_{\text{ex}} + U_{\text{spring}}$, where U_{ex} represents the exclusive interactions between beads, and U_{spring} represents a finitely extensible spring potential that connects the neighboring beads in chains. Explicit functional forms of U_{ex} and U_{spring} are explained in the Supporting Material.

In addition to the above terms, a kinkable bending potential was included in U_{chain} of the previous TTS model. This bending potential constrains the angle fluctuation between neighboring segments, which results in the chain of a 20–30 kb persistent length. We used this bending potential in the previous model because the fluorescence-in-situ-hybridization observation showed that chromatin chains in yeast cells should be kept straight for this distance (36). This persistent length was also consistent with the assumption that chromatin forms a regular 30-nm fiber. However, we should note that in the yeast nucleus, chromosomes tend to show stretched forms because centromeres are anchored to the SPB and telomeres are anchored to the nuclear envelope (4,15). Therefore, chains in the in situ condition are far from free polymer chains in solution, and chromosomes in the nucleus should be kept rather straight for 20–30 kb even when the persistent length, which should be defined in free solution, is sufficiently small. Because the regular 30-nm fiber is unstable against perturbations (such as the heterogeneous length distribution of linker DNA between nucleosomes, partial depletion of histone H1, or histone acetylation (37–40)), it is as yet uncertain whether the regular 30-nm fiber is indeed formed in the yeast nucleus. With the recent high resolution Micro-C analyses, no evidence for the regular nucleosome arrangement was found in yeast chromatin (41,42). Existence or absence of regular 30-nm fibers is under debate because the

accumulating evidence shows that irregularly folded 10-nm fibers, rather than the regular 30-nm fibers, dominate the nuclei of various organisms (43). Here, we do not use the bending potential to constrain angles between segments, which corresponds to the assumption that the persistence length is <3 kb. Even with this assumption, the simulation results are not significantly different from those obtained with the bending potential because chromosomes tend to be stretched through interactions with nuclear structures. To define U_{ex}, we use 30 nm as a measure of thickness of chromosomes by assuming that an irregularly folded 10-nm fiber should typically have the effective thickness of 30 nm (44) particularly in a coarse-grained representation, although the chain is more flexible than the regular 30-nm fiber.

Interactions derived from the Hi-C data

The potential U_{HiC} in Eq. 3 represents interactions derived from the Hi-C data obtained by Duan et al. (12). In the previous TTS model, the potential representing the Hi-C based interactions was termed the Gō-like potential because a similar structure-based potential termed the Gō-like potential was used in protein folding studies as the potential to stabilize the observed average structure. In this study, we emphasize the effectiveness of the use of such structure-based potential to describe structural fluctuations. However, unlike the protein folding problem, the simulated genome does not fold into a unique structure, but rather continues fluctuating largely around the average structure. In this study, to avoid confusion, we refrain from referring to our potential as Gō-like potential, and instead refer to Hi-C based potential.

We first remove biases from the observed Hi-C data. An interaction frequency obtained by the Hi-C measurement represents frequency demonstrating that two chromatin fragments in the genome are in physical proximity in an ensemble of cells. However, it has been shown that Hi-C data contain some biases because various features such as the fragment length, the guanine-cytosine content of the fragment, and the mappability of fragment affect the interaction frequencies (45,46). These biases depend on the definition of fragments used, and the fragments are determined by which restriction enzyme is used to excise the fragments from the genome. Therefore, we filter out these biases by comparing two sets of the Hi-C data obtained from two experiments using different restriction enzymes, *Hin*dIII and *Eco*RI, and extracting the common signals from them. From the libraries of fragments defined by the Tables S5 and S6 for *Hin*dIII and Tables S7 and S8 for *Eco*RI in Duan et al. (12), we first map each fragment to a bead in the model by assigning a bead nearest to the midpoint of the fragment. From this mapping, the matrix of interaction frequency $f_{ij}^{\mu\nu}$ between the ith bead in the μth chromosome and the jth bead in the νth chromosome is defined for both *Hin*dIII and *Eco*RI, and we select the matrix elements that satisfy the condition

$$\frac{1}{1+x} < \frac{f_{ij}^{\mu\nu}(Hin\text{dIII})\big/f_{\text{total}}(Hin\text{dIII})}{f_{ij}^{\mu\nu}(Eco\text{RI})\big/f_{\text{total}}(Eco\text{RI})} < 1+x, \qquad (4)$$

where $f_{\text{total}}(Hin\text{dIII}) = \sum_{\mu\nu}\sum_{ij} f_{ij}^{\mu\nu}(Hin\text{dIII})$ and $f_{\text{total}}(Eco\text{RI}) = \sum_{\mu\nu}\sum_{ij} f_{ij}^{\mu\nu}(Eco\text{RI})$ are total frequencies. Using the threshold value $x = 4$ for both intra- and interchromosome frequencies, we use $f_{ij}^{\mu\nu}$ that satisfies Eq. 4 for estimating the average distance between beads by maintaining the value of $f_{ij}^{\mu\nu}$ as it is, and set the other matrix elements that do not satisfy Eq. 4 as $f_{ij}^{\mu\nu} = 0$.

As discussed by Yaffe and Tanay (45), the effects of the filtering can be assessed by comparing the 1D coverage. Here we use the normalized 1D coverage, $g_i^{\mu\ \text{intra}} = \sum_j f_{ij}^{\mu\mu}/\sum_{i,j} f_{ij}^{\mu\mu}$ for the intrachromosome comparison in the chromosome μ, and $g_i^{\mu\ \text{inter}} = \sum_{\nu(\neq\mu)}\sum_j f_{ij}^{\mu\nu}/\sum_{\mu}\sum_{\nu(\neq\mu)}\sum_{i,j} f_{ij}^{\mu\nu}$ for the genomewide comparison of the interchromosome interactions.

If biases are removed through filtering, two values of 1D coverage, $g_i^{\mu\ \alpha}(Hin\text{dIII})$ and $g_i^{\mu\ \alpha}(Eco\text{RI})$, where $\alpha = $ "intra" or "inter" should show a large correlation. Fig. S1 shows the 1D coverage $g_i^{\mu\ \text{intra}}$ with $\mu = 1$ (chromosome I) as a function of i. We find that data for two 1D coverages, $g_i^{1\ \text{intra}}(Hin\text{dIII})$ and $g_i^{1\ \text{intra}}(Eco\text{RI})$, show a small correlation before filtering, as indicated by a Spearman correlation coefficient of $\rho = 0.2$; however, this correlation coefficient increases to $\rho = 1.0$ after filtering. In Fig. S2, $g_i^{\mu\ \text{intra}}(Hin\text{dIII})$ and $g_i^{\mu\ \text{intra}}(Eco\text{RI})$ are compared for $\mu = 1–16$ (the normalized intrachromosome 1D coverage for chromosomes I–XVI) for the data after filtering. Also for the interchromosome interactions, comparing two $g_i^{\mu\ \text{inter}}$ s, it shows that the correlation is largely increased from $\rho = 0.5$ to $\rho = 1.0$ as shown in Figs. S3 and S4.

From the filtered interaction matrix, $f_{ij}^{\mu\nu}(Hin\text{dIII})$, we estimate the average spatial distance $r(\mu,i;\nu,j)$ between beads (μ,i) and (ν,j). The relation between the intrachromosome frequency $f_{ij}^{\mu\mu}(Hin\text{dIII})$ and the average distance $r(\mu,i;\mu,j)$ was estimated as summarized in Supplementary Fig. 17 of Duan et al. (12). Here, we use the same relation in this study. In addition, for interchromosome interactions, as was proposed by Duan et al. (12), the same relationship is used to estimate the average interchromosome distances, $r(\mu,i;\nu,j)$. We expect that the estimation could be further improved if we used the interchromosome fluorescence-in-situ-hybridization data for budding yeast as a reference, as was used for fission yeast by Tanizawa et al. (26).

Here, the Hi-C based potential has the following terms:

$$
\begin{aligned}
U_{HiC} = &\sum_{\mu>\nu}\sum_{i,j} U_{HiC}\big(r_{ij}^{\mu\nu}\big) + \sum_{\mu}\sum_{j>i} U_{HiC}\big(r_{ij}^{\mu\mu}\big) \\
&+ \sum_{\mu}\sum_{i\in\text{tRNA}}\sum_{j\in 5\text{S rRNA}} U_{tRNA}\big(r_{ij}^{\mu 12}\big),
\end{aligned}
\qquad (5)
$$

where $r_{ij}^{\mu\nu} = \left|\mathbf{r}_i^{\mu} - \mathbf{r}_j^{\nu}\right|$. The first term in Eq. 5 represents the interchromosome interactions, whereas the second term represents the intrachromosome interactions. The third term represents the effective interactions between tRNA genes and 5S rRNA genes. The first and second terms have the same functional form as

$$
\begin{aligned}
U_{HiC}\big(r_{ij}^{\mu\nu}\big) = &-\frac{\xi r_0}{\sqrt{2\pi(r(\mu,i;\nu,j))^2}} \\
&\times \exp\left(-\frac{(r_{ij}^{\mu\nu} - r(\mu,i;\nu,j))^2}{2(r(\mu,i;\nu,j))^2}\right).
\end{aligned}
\qquad (6)
$$

Because the number of interacting pairs becomes smaller after filtration, we use a larger ξ than used in the previous TTS model before filtration. Here, we use $r_0 = 10$ nm as a unit of distance, and $\xi/T = 50$. Notably, the overall features of the simulated genome structure, such as the distributions of telomere positions, are not altered greatly from those obtained before filtration because they are mostly determined by interactions between chromosomes and nuclear structures; however, the spatial distributions of individual genes are largely modified because they are affected by the interactions between chromatins.

Further attention should be paid to the interactions of tRNA genes. Although the sequence positions of tRNA genes in budding yeast are scattered over the genome, some are spatially localized around the nucleolus and the region where centromeres are gathered (4,12,15,21). Such localization would be natural when we consider that Pol III-transcribed genes such as tRNA and 5S rRNA genes, where the latter is confined in the nucleolus, may share a concentration of RNA polymerase III and enter close proximity. Because the frequency of interactions of 5S rRNA genes is not explicitly included in the Hi-C data, we additionally introduce the following

attractive interaction between the ith bead corresponding to the tRNA gene and the jth bead corresponding to the 5S rRNA gene as

$$U_{\text{tRNA}}\left(r_{ij}^{\mu 12}\right) \;=\; -\frac{\xi_t r_0}{\sqrt{2\pi r_t^2}} \exp\left(-\left(r_{ij}^{\mu 12}\right)^2 \middle/ \left(2r_t^2\right)\right), \quad (7)$$

where we use $\xi_t/T = 10$. We consider that the potential Eq. 7 does not represent interaction in the atomic scale, but rather a mean-field-like effective interaction for a statistical tendency of gene gathering; therefore, we use a relatively long distance as a parameter to define the potential as $r_t = 200$ nm.

By analyzing the interacting frequencies in the Hi-C data, Duan et al. showed that tRNA genes are clustered into three groups; one localized around the nucleolus, one localized around the centromeres, and one not showing a distinct localization, Details of the clustered results of the Hi-C data are found on the web site (47). For the beads (μ, i) in Eq. 7, tRNA genes that were identified as genes localized around the nucleolus by the Hi-C analysis (47) are assumed to be targets of the potential in this model; however, the tRNA genes in the vicinity of telomeres are not considered as targets because they should be more attracted to the nuclear envelope in the WT. For $(12, j)$ in Eq. 7, we use beads representing 5S rRNA genes, which are located in chromosome XII. We assume that 5S rDNA has 150 repeats of sequences as $j = 154 + 3k$ with $k = 0$–149.

Interactions between chromosomes and nuclear structures

In budding yeast, 16 chromosomes are packed in the nucleus, and their motion is constrained by interactions between chromosomes and nuclear structures such as the nuclear envelope, nucleolus, and SPB. We represent these interactions by U_{nucleus} in Eq. 3, which consists of the following three terms;

$$U_{\text{nucleus}} \;=\; U_{\text{envelope}} + U_{\text{cen}} + U_{\text{nucleolus}}, \qquad (8)$$

where the first term U_{envelope} represents interaction between chromosomes and the nuclear envelope, and the second term U_{cen} represents interactions between centromeres and the SPB. The last term $U_{\text{nucleolus}}$ represents interactions that confine rDNA in the nucleolus and exclude other parts of chromosomes from the nucleolus. In this model, the semilunar-shaped functional form for $U_{\text{nucleolus}}$ in the previous TTS model is replaced by the more realistic crescent-shape functional form. To describe interaction terms in U_{nucleus}, we define the coordinate system as shown in Figs. S5 and S6. With this coordinate system, the center of the nucleus is $\mathbf{r}_{\text{center}} = (1000, 1000, 1000)$ in units of nanometers, the nucleus is a sphere of 1000 nm radius, and the SPB is placed on $(1000, 1000, 10)$.

U_{envelope} in Eq. 8 is composed of three terms,

$$U_{\text{envelope}} \;=\; U_{\text{tel}-\text{env}} + U_{\text{rDNA}-\text{env}} + U_{\text{chr}-\text{env}}, \qquad (9)$$

where the first term $U_{\text{tel}-\text{env}}$ represents interactions between telomeres or subtelomeric regions and the nuclear envelope, the second term $U_{\text{rDNA}-\text{env}}$ represents interactions between the rDNA region and the nuclear envelope, and the last term $U_{\text{chr}-\text{env}}$ represents interactions between the other parts of chromosomes and the nuclear envelope. Telomeres and subtelomeric regions attractively interact with the nuclear envelope in the WT because they form a complex at the nuclear envelope in association with several factors: the inner membrane-associated protein Esc1 (48), the Sad1-UNC-84 domain protein Mps3 (49), the end-binding complex Yku70/80 (50), and the Rap1 (51) and SIR proteins (52,53). In particular, SIR proteins in this complex promote heterochromatin formation near telomeres and repress genes in that location (52,53). Here, we represent the attractive interactions arising from the formation of

this complex as follows: $U_{\text{tel}-\text{env}} = \sum_{\mu}\sum_{i \,\in\, \text{telom, subtelom}} U_{\text{tel}-\text{env}}(R_i^{\mu})$ with

$$\begin{aligned} U_{\text{tel}-\text{env}}(R_i^{\mu}) &= 2\epsilon\left[\left(\frac{R_i^{\mu} - u_0}{u - u_0}\right)^{12} - c_1\right], \text{ for } R_i^{\mu} > u_3, \\ &= -yc_0\epsilon, \text{ for } u_2 < R_i^{\mu} \leq u_3, \\ &= -yc\epsilon\left(\frac{R_i^{\mu} - u_1}{u_2}\right), \text{ for } u_1 < R_i^{\mu} \leq u_2, \\ &= 0, \text{ for } R_i^{\mu} \leq u_1, \end{aligned} \qquad (10)$$

where $R_i^{\mu} = \left|\mathbf{r}_i^{\mu} - \mathbf{r}_{\text{center}}\right|$. As a unit of interaction strength, we use $\epsilon = T$. The values c_0 and c_1 are determined to smoothly connect the parts of $U_{\text{tel}-\text{env}}$. For simulating the *yku70 esc1* mutant, we set $y = 0$ and $c_1 = 0$ (turned off), whereas $y = 1$ and $c_1 \neq 0$ (turned on) for simulating the WT. Parameters are chosen to set the telomere behaviors simulated with $y = 1$ to be consistent with the experimental observations of the WT (10,13) as $u_0 = 800$ nm, $u_1 = 400$ nm, $u_2 = 800$ nm, $u_3 = 950$ nm, $u = 1000$ nm, and $c = 3$.

The rDNA repeats in budding yeast can be tethered to the nuclear envelope by several proteins such as Heh1, Lrs4, and Sir2 (4,15,19). Here, we consider that beads $i = 151$–606 in the chain $\mu = 12$ represent the rDNA repeats. We represent the attractive interactions due to this tethering as $U_{\text{rDNA}-\text{env}} = \sum_{i \,\in\, \text{rDNA},\, \mu=12} U_{\text{rDNA}-\text{env}}(R_i^{\mu})$, where $\sum_{i \,\in\, \text{rDNA},\, \mu=12}$ is a sum over the randomly chosen 30% of beads in the rDNA repeat region, and we use the same function for $U_{\text{rDNA}-\text{env}}(R_i^{\mu})$ as $U_{\text{tel}-\text{env}}(R_i^{\mu})$ by maintaining $y = 1$ and $c_1 \neq 0$ in both the WT and the mutant.

The potential $U_{\text{chr}-\text{env}} = \sum_{\mu}\sum_i U_{\text{chr}-\text{env}}(R_i^{\mu})$ represents interactions between other parts of chromosomes and the nuclear envelope, which confine chromosomes inside the nucleus,

$$\begin{aligned} U_{\text{chr}-\text{env}}(R_i^{\mu}) &= 2\epsilon\left(\frac{R_i^{\mu} - u_0}{u - u_0}\right)^{12}, \text{ for } R_i^{\mu} > u_0, \\ &= 0, \text{ for } R_i^{\mu} \leq u_0, \end{aligned} \qquad (11)$$

where $u_0 = 800$ nm and $u = 1000$ nm are used.

Explicit functional forms of other terms, U_{cen} and $U_{\text{nucleolus}}$ in Eq. 8 are explained in the Supporting Material.

Simulation

We set $m = T = 1$. With this unit, t has the dimension of length L and the friction constant ζ has the dimension of L^{-1}. The friction constant ζ was set to be $\zeta = 10^{-5}$ to allow efficient sampling in the allowed computation time. Values of other parameters are explained in the Supporting Material and summarized in Table S1. Starting from the structure inferred by Duan et al. (12), the structure was relaxed under the potential U with the steepest descent method. Then, using the Langevin equation of motion of Eq. 1, the high temperature run with $T' = 2T$ was performed, and the last structure of this high-temperature run was used as the initial structure of our simulation. Equation 1 was discretized with the step length $\Delta t = 0.01$. We calculated up to 1×10^5 steps, and from the last 5×10^4 steps we sampled data in every 200 steps to obtain 250 structures. By calculating 40 simulation runs with different random number realizations, $250 \times 40 = 10^4$ structures in total were sampled for each analysis.

RESULTS

Simulated genome organization in the WT

To assess the performance of this model, we compare the simulated and observed (13) probability density maps of

centromeres, rDNA, and the left telomere of chromosome VII in the WT. As shown in Fig. 1 a, the 3D distribution is projected onto a two-dimensional (2D) plane by defining a grid of 0.1 × 0.1 μm spacing. The probability density maps are then calculated using the method explained in the caption of Fig. 1. Hereafter, this method of projection is used throughout the Results to show the calculated probability density maps of genes on the 2D plane. The simulated probability density maps in Fig. 1 b show an improved agreement with those of the observed data when compared with the previous version of the model (14). In Fig. S7 and Table S2, we compared the simulated and observed (11) distributions of the telomere-telomere distance in the WT, which shows that the calculated results explain the observed features of distributions. Two telomeres of a short chromosome, such as chromosome III, tend to reside near to each other, whereas two telomeres of a long chromosome, such as chromosome XIV, tend to be separate from each other. Therizols et al. (54) reported the median distances from subtelomeric regions of chromosomes IV, VI, and X to the other subtelomeric regions, which

are compared with the simulation data in Fig. S8. The Pearson correlation coefficient is 0.52, which is similar to the value obtained by the model of Wong et al. (33) created without using the Hi-C based potential, suggesting that the distribution of telomeres is almost determined by the interactions between chromosomes and the nuclear envelope.

The simulation results for the WT model show intriguing features. An example of the simulated movement of the genome structure is shown in Movie S1. This movie shows that the genome largely fluctuates in the nucleus. Dependences of the motion on the friction coefficient ζ and the other parameters that can affect fluctuation of chains are discussed in the Supporting Material and Table S2. Through this genomewide movement, individual genes fluctuate in their territorial region. Therefore, the probability density maps of genes should provide useful information. The simulated probability density maps of 5S rRNA genes and several groups of tRNA genes are shown in Fig. 2. Near the nucleolus, we find the localized population of these tRNA genes, which is consistent with the cluster analysis of the Hi-C data (12) and the observed fluorescence localization of tRNA
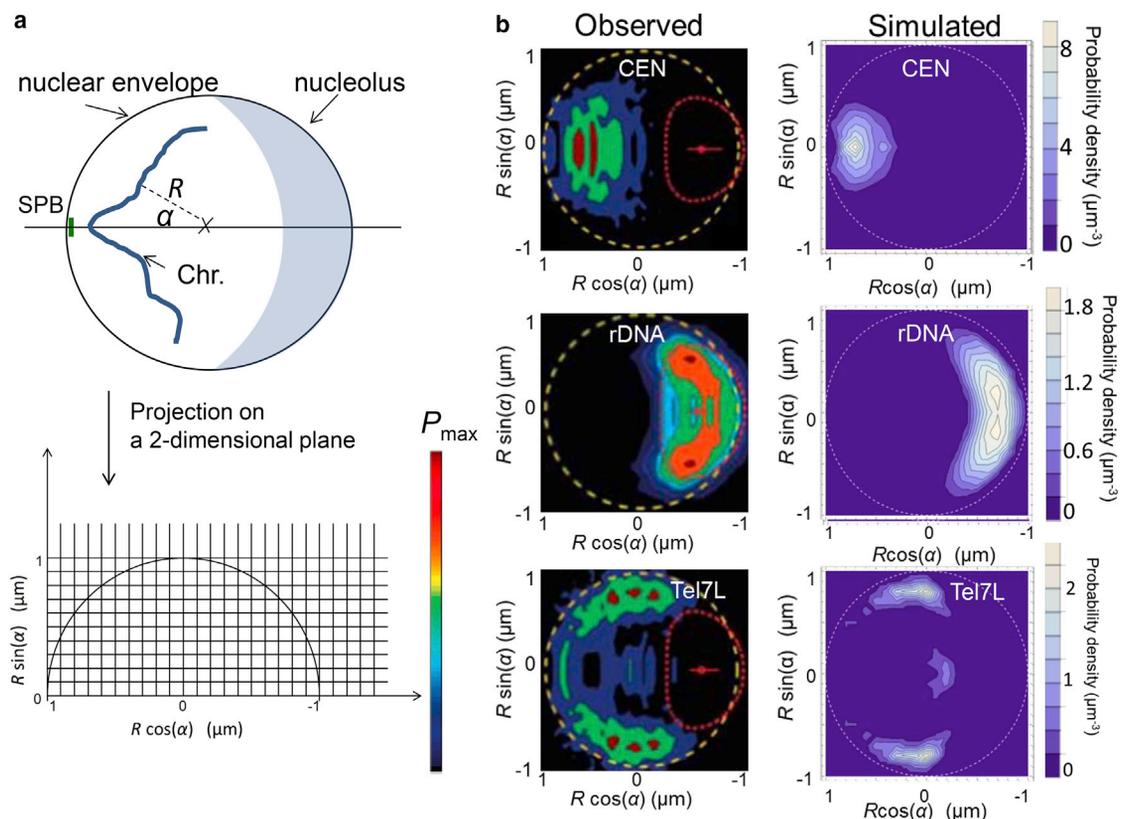
FIGURE 1 2D representation of probability density maps of various parts of the yeast genome. (a) A locus on an illustrated chromosome is represented by the cylindrical coordinate (R, α). The nucleus is represented as a sphere of radius 1 μm, and the z axis is defined as the axis running through the SPB (spindle pole body) and the center of the nucleus. A grid of 0.1 × 0.1 μm spacing is defined on the 2D plane and a cylindrical tube running around the z axis is defined. This tube has the radius $R \sin \alpha$ and the 0.1 × 0.1 μm^2 square intersection. The probability density of beads at the position ($R \cos \alpha$, $R \sin \alpha$) on the 2D plane is calculated by the number of beads found in this cylindrical tube divided by the volume of the tube, which is further normalized by dividing by the total number of corresponding beads within the nucleus, and then averaged over 10^4 structures. (b) Comparison of the simulated and observed probability density maps of each part in the wild-WT yeast nucleus; centromeres of 16 chromosomes (top), rDNA (middle), and the left telomere of chromosome VII (bottom). (Panels of the observed data were reproduced from Berger et al. (13) with permission.) To see this figure in color, go online.
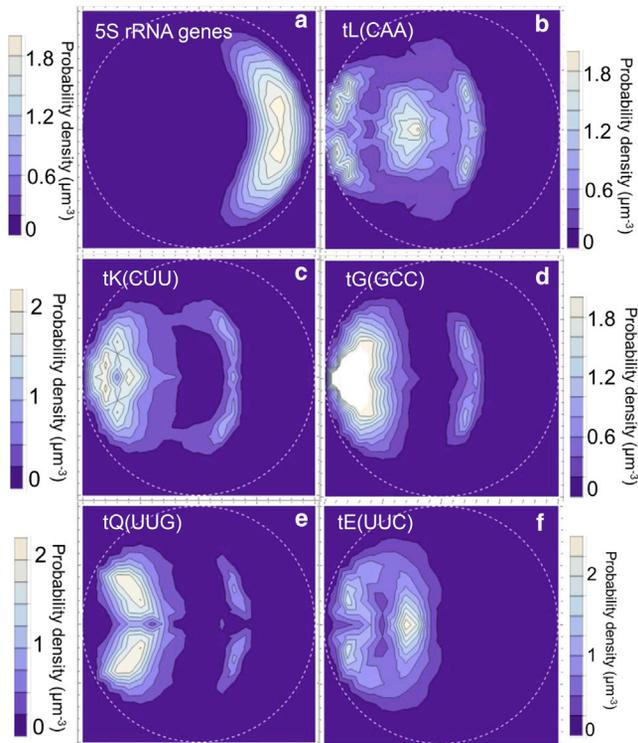
FIGURE 2 Probability density maps of RNA polymerase III-transcribed genes in the simulated WT yeast nucleus. (*a*) Probability density maps of 5S rRNA genes. In this model, 150 copies of 5S rRNA genes were assumed to be located on chromosome XII. (*b–f*) Probability density maps of tRNA genes observed in Thompson et al. (21): (*b*) tRNA[Leu](CAA) genes (10 genes are found in the genome), (*c*) tRNA[Lys](CUU) genes (14 genes), (*d*) tRNA[Gly](GCC) genes (16 genes), (*e*) tRNA[Gln](UUG) genes (9 genes), and (*f*) tRNA[Glu](UUC) genes (14 genes). In (*d*), the region with a density larger than 2 μm^{-3} is colored white. To see this figure in color, go online.

genes (21). Here, each group comprises 9–16 copies of genes, and the probability distribution of the summed density of these multiple genes are shown in each panel of Fig. 2. Among these genes, only a few were identified by the cluster analysis of the Hi-C data (12) to localize around the nucleolus. Therefore, the attractive interactions between tRNA and 5S rRNA genes were assumed in the model only for those small number of genes. The ratio of the number of genes attracted to 5S rRNA genes to the number of copies of genes in the genome is 2/10, 3/14, 2/16, 2/9, and 4/14 for Fig. 2, *b–f*, respectively. Therefore, the major portion of the density was not attracted to the region near the nucleolus but concentrated around the center of the nucleus or around the region centromeres are gathered. It would be interesting to see whether the localized density around the nucleolus increases when a larger number of tRNA genes would have the attractive interactions; however, we use the assumption derived from the clustering analysis of the Hi-C data in a straightforward way. Thus, the study of various attractive interactions is not addressed in this article.

The simulated localization of the tRNA genes around either the nucleolus or the center of the nucleus is weakened

when the attractive interactions between tRNA genes and 5S rRNA genes are turned off by setting $U_{\text{tRNA}}(r_{ij}^{\mu12}) = 0$ in Eq. 7; however, the density around the center of nucleus remains (Fig. S9), which is in clear contrast to the results of the model in which the Hi-C based potential is not used (27); in the latter model, the tRNA genes strongly concentrate around the region where the centromeres are gathered. Thus, the interactions between chromosome loci in the model do not affect the distributions of telomeres but considerably affect the distribution of genes.

Probability density maps of the highly expressed genes (top 200 genes) and the lowly expressed genes (bottom 200 genes) were calculated in the WT by extracting those genes from the microarray data deposited in ArrayExpress (E-TABM-630) (31,35). Notably, five upregulated and one downregulated genes were lowly expressed in the WT, and two upregulated and four downregulated genes were lowly expressed in the *yku70 esc1* mutant; these genes were removed from the list of the above bottom 200 genes, so that there is no overlap of the gene names between the misregulated genes and the top 200 or bottom 200 genes. Lists of the top 200 genes and bottom 200 genes derived in this way are shown in Table S3 for both the WT and the mutant. Fig. 3 shows maps of differential probability density for the top 200 genes and bottom 200 genes, $\Delta P =$ (probability density of the genes) $-$ (probability density of randomly selected
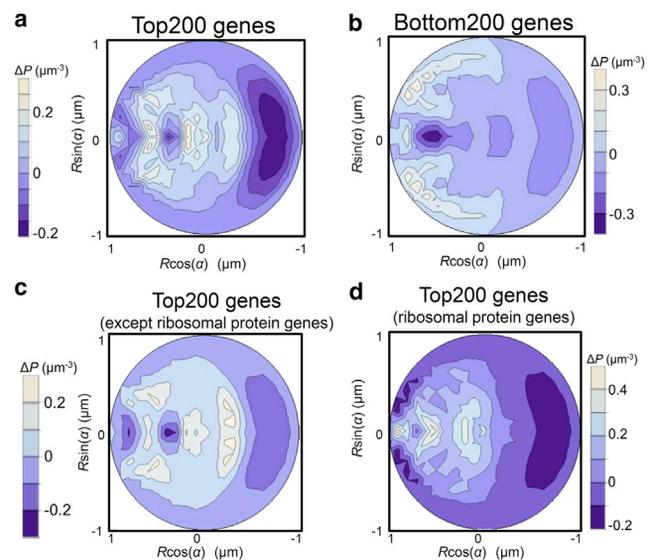
FIGURE 3 Differential probability density maps, ΔP, of highly and lowly expressed genes in the simulated WT yeast nucleus. Differential probability density maps of (*a*) highly expressed genes (the top 200 genes in the transcription level), (*b*) lowly expressed genes (the bottom 200 genes), (*c*) highly expressed genes excluding the ribosomal protein genes from the top 200 genes, and (*d*) the ribosomal protein genes within the list of the top 200 genes. The differential probability density is defined by subtracting the probability density of randomly distributed sites from the corresponding probability density (see text for more explanation). The top 200 and bottom 200 genes were extracted from the microarray data deposited in ArrayExpress (E-TABM-630) (31,35). To see this figure in color, go online.

Tokuda and Sasai

sites). Here, the density of randomly selected sites was generated by selecting 100 beads randomly from every 10^4 structures; therefore, the density of randomly selected sites represents the density of chromatins in nucleus. In Fig. S10, the probability density maps before this subtraction are shown together with the density map of random sites. As shown in Fig. 3, highly expressed genes show a higher density in the inner regions of the nucleus and the lowly expressed genes show a higher density at the nuclear periphery. This is similar to the simulation results of Gong et al. (27), and is consistent with the localization of repressive factors such as SIR proteins near the nuclear envelope. It should be noted that 67 of the top 200 genes are ribosomal protein genes, which may be regulated in a different way from remaining 133 genes; many of the latter genes are related to basic metabolism. As evident in Fig. 3, it is interesting that the ribosomal protein genes are localized around the center of nucleus, whereas the other highly expressed genes are more localized around the region near the nucleolus and the outside part of the region where the centromeres are gathered with some additional population around the center of nucleus. This result suggests the existence of a mechanism of gene regulation that depends on the spatial position of genes.

Spatial distributions of transcriptional activities in the *yku70 esc1* mutant

The *yku70 esc1* mutant was simulated by setting $y = 0$ and $c_1 = 0$ in Eq. 10 to turn off the attractive interactions between telomeres or subtelomeric regions and the nuclear envelope. The validity of this calculation can be confirmed by comparing the simulated and observed radial distributions of telomere positions. In Hediger et al. (10) and Taddei et al. (55), the peripheral-most zone with $0.82~\mu m < R < 1~\mu m$ was defined as zone I, where R is the radial distance from the center of the nucleus as defined in Fig. 1 *a*. The probability to find the left telomere of chromosome XIV in zone I in the mutant is 40% in cells observed in the G1 phase (55) and 19% in the simulated data, and probability to find the right telomere of chromosome VI in zone I in the mutant is 38% in the observed data (55) and 44% in the simulated data. Therefore, the model has the ability to reasonably reproduce the observed radial distributions of telomeres in the mutant. Further discussion on the telomere residence probability in zone I is given in the Supporting Material.

Fig. 4 is a snapshot of the simulated genome structure in the *yku70 esc1* mutant, and the misregulated genes are drawn as spheres. We get the impression that the downregulated 28 genes (*blue spheres*) and the 32 upregulated genes (*red spheres*) do not appear to be homogeneously mixed, rather each group of genes shows some clustered features with heterogeneous density distributions. These misregulated genes are listed in Table S4. An example of the simu-
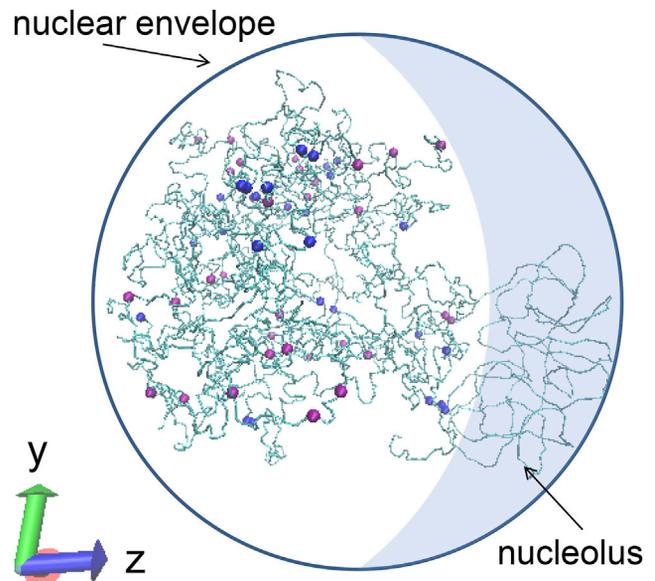
FIGURE 4 A snapshot of the simulated genome structure of the *yku70 esc1* mutant yeast. Sixty misregulated genes in the mutant are shown with spheres; 32 upregulated genes (*red*) and 28 downregulated genes (*blue*). This image was constructed with Visual Molecular Dynamics software (University of Illinois at Urbana-Champaign, Champaign, IL). To see this figure in color, go online.

lated movement of the genome structure in the *yku70 esc1* mutant is shown in Movie S2. In this motion, the misregulated genes fluctuate largely based on their characteristic distributions.

The tendency of the distributions found in the snapshot and the movie is confirmed by plotting the probability density maps obtained from the average of 10^4 structures. In Fig. 5, the differential probability density maps are shown for upregulated (Fig. 5 *a*) and downregulated (Fig. 5 *b*) genes, which are compared with the density maps of the top 200 (Fig. 5 *c*) and the bottom 200 (Fig. 5 *d*) genes. Notably, 93 of the top 200 genes are ribosomal protein genes in the mutant. In Fig. 5 *e*, we show the differential probability density map of highly expressed genes by subtracting ribosomal protein genes from the top 200 genes. From Fig. 5, we get the impression that the distribution of downregulated genes is similar to the distribution of the lowly expressed genes in the mutant.

The similarity or difference between spatial distributions can be quantified by calculating the overlap of 3D distributions of genes in different groups. We calculate the following degree of overlap between the spatial distributions of the bottom 200 genes and the 28 downregulated genes as

$$I_\kappa^{bottom-down} = \frac{1}{N_{bottom}} \frac{1}{N_{down}} \sum_i^{N_{bottom}} \sum_j^{N_{down}} \exp\left(-\frac{\left|\mathbf{r}_i^\kappa - \mathbf{r}_j^\kappa\right|^2}{2\delta^2}\right),$$

(12)
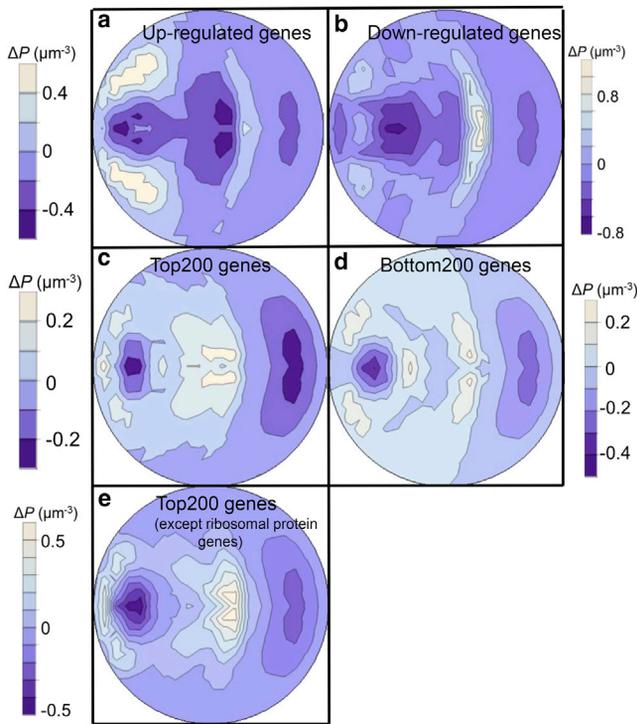
FIGURE 5 Simulated differential probability density maps, ΔP, of misregulated, highly expressed, and lowly expressed genes in the *yku70 esc1* mutant yeast nucleus. Differential probability density maps of (*a*) 32 upregulated genes, (*b*) 28 downregulated genes, (*c*) highly expressed genes (the top 200 genes in the transcription level), (*d*) lowly expressed genes (the bottom 200 genes), and (*e*) highly expressed genes excluding the ribosomal protein genes from the top 200 genes. The top 200 and bottom 200 genes were extracted from the microarray data deposited in ArrayExpress (E-TABM-630) (31,35). Probability density maps of the corresponding genes before the subtraction of the probability density of random sites are shown in Fig. S11. To see this figure in color, go online.

where $N_{bottom} = 200$, $N_{down} = 28$, and $\delta = 100$ nm; $\mathbf{r}_i^\kappa$ is the position of the *i*th bead of the bottom 200 genes in the κth structure sampled from simulations; and $\mathbf{r}_j^\kappa$ is the position of the *j*th bead of the downregulated genes in the κth structure. The term $I_\kappa^{\text{bottom}-\text{down}}$ is calculated for each of 10^4

structures with $\kappa = 1-10^4$. In Fig. 6 *a*, the distribution of the calculated 10^4 values of $I_\kappa^{\text{bottom}-\text{down}}$ is shown by plotting the fraction, $f^{\text{bottom}-\text{down}}(I)$, which is the probability to find a structure κ in the range $I - \Delta I/2 < I_\kappa^{\text{bottom}-\text{down}} < I + \Delta I/2$ with a suitable bin size ΔI. In addition, similar quantities, $f^{\text{top}-\text{down}}(I)$ and $f^{\text{random}-\text{down}}(I)$, are plotted together. Fig. 6 *a* shows that the overlap between the lowly expressed genes and the downregulated genes in the mutant are distinctly larger than the overlap between the random sites and the downregulated genes (*p* value $< 10^{-4}$) and that the overlap between the highly expressed genes and the downregulated genes tends to be slightly smaller than the overlap between the random sites and the downregulated genes (*p*-value $< 10^{-4}$), where the *p* values were calculated with the Wilcoxon signed-rank test. A similar argument holds for the 32 upregulated genes. In Fig. 6 *b*, distributions, $f^{\text{top}-\text{up}}(I)$, $f^{\text{top(ex.rib.)}-\text{up}}(I)$, $f^{\text{bottom}-\text{up}}(I)$, and $f^{\text{rand}-\text{up}}(I)$, are plotted. Fig. 6 *b* shows that the upregulated genes in the mutant tend to have a larger overlap with the highly expressed genes than with the random sites (*p* value $< 10^{-4}$), and they tend to have a smaller overlap with the lowly expressed genes than with the random sites (*p* value $< 10^{-4}$).

Comparisons among these degrees of overlap suggest that there are spatial actively transcribing region and inactively transcribing region in the yeast nucleus. The genes should be downregulated when they show higher probability of remaining in the inactive region in the mutant than in the WT. Similarly, the genes could be upregulated when they show higher probability of remaining in the active region in the mutant. This hypothesis, however, should be carefully examined because the difference between $f^{\text{top(ex.rib.)}-\text{up}}(I)$ and $f^{\text{rand}-\text{up}}(I)$ shown in Fig. 6 *b* is not large. The reason for this small difference becomes evident when the 32 upregulated genes are classified into the following three groups: group 1, which comprises seven genes located within 20 kb from any of the telomeres; group 2, which comprises 12 genes with a tendency to reside in the active region; and group 3, which comprises 13 genes with a tendency to
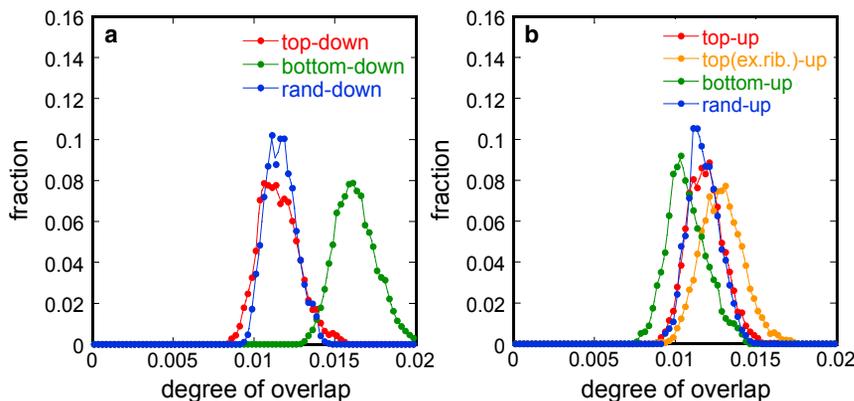
FIGURE 6 Distributions $f(I)$ of the degree of 3D overlap I_κ between groups of genes in the *yku70 esc1* mutant yeast nucleus. Each distribution was calculated by sampling 10^4 different structures, $\kappa = 1-10^4$, and plotted with the bin size $\Delta I = 2.5 \times 10^{-4}$. (*a*) Distributions $f^{\text{top}-\text{down}}(I)$ of the overlap between the top 200 genes and the downregulated genes (*red*), $f^{\text{bottom}-\text{up}}(I)$ of the overlap between the bottom 200 genes and the downregulated genes (*green*), and $f^{\text{rand}-\text{down}}(I)$ of the overlap between the randomly selected 1000 sites and the downregulated genes (*blue*). (*b*) Distributions $f^{\text{top}-\text{up}}(I)$ between the top 200 genes and the upregulated genes (*red*), $f^{\text{top(ex.rib.)}-\text{up}}(I)$ between the highly expressed genes excluding the ribosomal protein genes from the top 200 genes and the upregulated genes (*orange*), $f^{\text{bottom}-\text{up}}(I)$ between the bottom 200 genes and the upregulated genes (*green*), and $f^{\text{rand}-\text{up}}(I)$ between the randomly selected 1000 sites and the upregulated genes (*blue*). To see this figure in color, go online.

be outside of the active region in the mutant. Group 1 genes were first selected, and then the remaining 25 genes were classified into groups 2 and 3 according to their average degree of overlap with the highly expressed genes. The average degree of overlap between gene i and the top 200 genes excluding the ribosomal genes was calculated as $\overline{I}^{\text{top(ex.rib.)}-i} = (1/N_\kappa) \sum_{\kappa=1}^{N_\kappa} I_\kappa^{\text{top(ex.rib.)}-i}$, where $N_\kappa = 10^4$ is the number of sampled structures. The average degree of overlap, $\overline{I}^{\text{rand}-i}$, was calculated in a similar way. Then, the gene i was classified into group 2 when $\overline{I}^{\text{top(ex.rib.)}-i}/\overline{I}^{\text{rand}-i} > 1$, and group 3 when $\overline{I}^{\text{top(ex.rib.)}-i}/\overline{I}^{\text{rand}-i} \leq 1$. These groups of genes are listed in Table S5. Distributions of degree of overlap between the group-i genes and the top 200 genes except for the ribosomal protein genes, $f^{\text{top(ex.rib.)}-\text{up}(i)}(I)$ with $i = 1$–3, are shown in Fig. 7. We find that group 2 genes largely overlap with the highly expressed genes, whereas group 1 and group 3 genes show less overlap with the highly expressed genes in the mutant. The distribution $f^{\text{top(ex.rib.)}-\text{up}}(I)$ in Fig. 6 b is obtained by averaging three distributions, $f^{\text{top(ex.rib.)}-\text{up}(i)}(I)$ with $i = 1$–3, each of which has the large difference from $f^{\text{random}-\text{up}}(I)$ in opposite directions, resulting in the small difference between $f^{\text{top(ex.rib.)}-\text{up}}(I)$ and $f^{\text{random}-\text{up}}(I)$.

The difference among three groups of upregulated genes is visualized with 2D maps of differential probability densities (Fig. 8). In the WT, the group 1 genes, which are located near telomeres, are strongly localized near the nuclear envelope, particularly around the SPB. Group 3 genes are also localized at the region near the nuclear envelope, whereas the group 2 genes are in various locations in the nucleus. In the mutant, group 1 and group 3 genes largely remain near the nuclear envelope. Therefore, the activity of group 1 and group 3 genes may be repressed in the WT due to their interactions with SIR proteins or other factors localized on the nuclear envelope, and the resolution of SIR complexes or loss of interactions with other repressive factors in the mutant should upregulate group 1 and group 3 genes. Group 2 genes do not show extensive interactions with the nuclear envelope in the WT, and largely shift toward the active region in the inside of the nucleus in the mutant, leading to their upregulation.

Such effects of spatial distributions can be further analyzed by examining changes in the overlap induced in the mutant. In Fig. 9, distributions of the degree of overlap in the WT and in the mutant are compared. Fig. 9, a–c, shows that the overlap between the downregulated genes and the inactive region is distinctly larger in the mutant than in the WT (Fig. 9 b), whereas the overlap between the downregulated genes and the active region is smaller

FIGURE 7 Distributions $f(I)$ of the degree of 3D overlap I_κ between the highly expressed genes excluding the ribosomal protein genes from the top 200 genes and each of three groups, group 1, 2, and 3, of the upregulated genes in the *yku70 esc1* mutant yeast nucleus, $f^{\text{top(ex.rib.)}-\text{up}(1)}(I)$, (*green*); $f^{\text{top(ex.rib.)}-\text{up}(2)}(I)$ (*red*); and $f^{\text{top(ex.rib.)}-\text{up}(3)}(I)$ (*cyan*), respectively. The distributions $f^{\text{top(ex.rib.)}-\text{up}}$ (*orange*) and $f^{\text{rand}-\text{up}}$ (*blue*), which are the same as those drawn in Fig. 6 b, are shown for comparison. Each $f(I)$ was calculated by sampling 10^4 different structures, $\kappa = 1$–10^4, and plotted with the bin size $\Delta I \approx 4.3 \times 10^{-4}$. To see this figure in color, go online.

FIGURE 8 (*a–f*) Simulated differential probability density maps, ΔP, of three groups of upregulated genes, i.e., group 1 (*top*), group 2 (*middle*), and group 3 (*bottom*), in the WT (*left column*) and the *yku70 esc1* mutant (*right column*) yeast nucleus. To see this figure in color, go online.
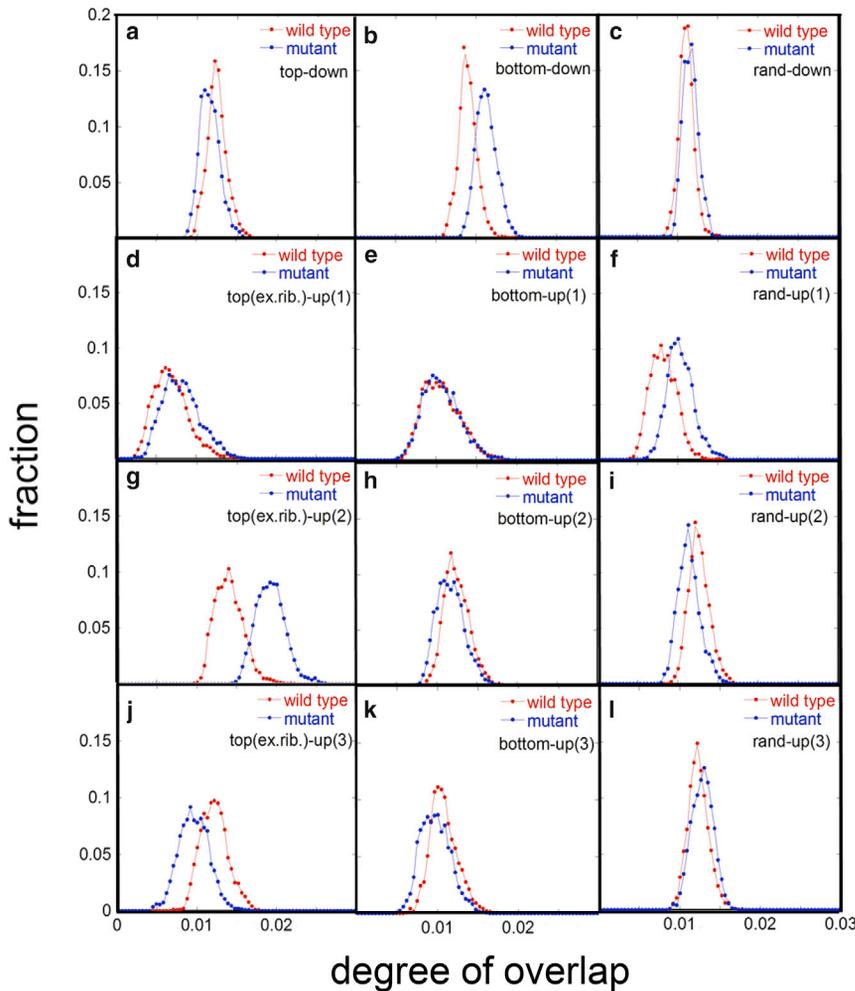
FIGURE 9 Comparisons between the distribution $f(I)$ of the degree of 3D overlap I_κ in the WT (*red*) and that in the *yku70 esc1* mutant (*blue*). Distributions (*a*) $f^{\text{top−down}}$, (*b*) $f^{\text{bottom−down}}$, (*c*) $f^{\text{rand−down}}$, (*d*) $f^{\text{top(ex.rib.)−up(1)}}$, (*e*) $f^{\text{bottom−up(1)}}$, (*f*) $f^{\text{rand−up(1)}}$, (*g*) $f^{\text{top(ex.rib.)−up(2)}}$, (*h*) $f^{\text{bottom−up(2)}}$, (*i*) $f^{\text{rand−up(2)}}$, (*j*) $f^{\text{top(ex.rib.)−up(3)}}$, (*k*) $f^{\text{bottom−up(3)}}$, and (*l*) $f^{\text{rand−up(3)}}$ are compared between the WT and the mutant. Each $f(I)$ was calculated by sampling 10^4 different structures, $\kappa = 1–10^4$, and plotted with the bin size $\Delta I \approx 4.3 \times 10^{-4}$. The *p* value for the Mann-Whitney U test is 0.277 in (*e*), and $<10^{-4}$ in all other panels, (*a–d*) and (*f–l*). To see this figure in color, go online.

in the mutant than in the WT (Fig. 9 *a*). This result indicates that the spatial distribution of downregulated genes shifts toward the inactive region in the mutant.

Fig. 9, *d–l*, shows the shift of upregulated genes. Group 1 genes, which are located near telomeres, show a small shift toward the active region (Fig. 9 *d*) but not the inactive region (Fig. 9 *e*). The overlap between group 1 genes and the random sites is somewhat increased (Fig. 9 *f*) because of the shift of group 1 genes from the nuclear periphery in the WT toward the inner region in the mutant due to the loss of telomere anchoring by the mutation. The transcriptionally repressive SIR complexes are formed near telomeres in the WT; thus the release of group 1 genes from the SIR complexes in the mutant can be sufficient for the upregulation as suggested by Taddei et al. (31). Group 2 genes, on the other hand, show a distinct shift toward the active region (Fig. 9 *g*) and a shift away from the inactive region (Fig. 9 *h*). Therefore, we hypothesize that these genes are upregulated because of their spatial shift toward the active region. Group 3 genes show a different behavior; they shift away from both the active (Fig. 9 *j*) and inactive (Fig. 9 *k*) regions. As shown in Fig. 8, group 3 genes are kept localized around the region near the nuclear envelope in the mutant, which separates

these genes from both the active and inactive regions. The release of SIR proteins or other factors from the envelope in the mutant may explain the upregulation of these genes, but further examination is necessary to identify the interactions responsible for the upregulation of group 3 genes.

We showed that the simulated 28 downregulated genes shifted toward the inactive region and 12 out of the 32 upregulated genes shifted toward the active region in the *yku70 esc1* mutant. Seven genes located near the telomeres were strongly localized around the nuclear envelope, and 13 genes were also localized near the envelope. These 20 genes did not show a distinct shift toward either the active or the inactive regions in the mutant; however, their localized distributions suggest that the loss of interactions with suppressive factors on the envelope leads to upregulation. Combining these simulation results, we hypothesize that the spatial distribution of genes and its shift induced by the *yku70 esc1* mutation explain the observed misregulation.

DISCUSSION

We developed a computational model of a dynamically fluctuating genome structure for simulating the genome

of the WT and the *yky70 esc1* mutant of budding yeast. With this model, we showed that the loss of the anchoring of telomeres to the nuclear envelope changes the spatial distribution pattern of genes. We calculated the distribution of the genes misregulated by the mutation and the distribution of lowly (the bottom 200 genes in the transcriptome) and highly (the top 200 genes) expressed genes in both the mutant and the WT models and found that the downregulated genes are localized around the nucleolus, and the overlap of distributions of the downregulated genes and the lowly expressed genes is larger in the mutant when compared with the WT. We also found that the overlap of distributions for 12 out of 32 upregulated genes and the highly expressed genes is larger in the mutant when compared with the WT. The remaining 20 upregulated genes are localized near the nuclear envelope; therefore they may be affected by the change of interactions with factors on the nuclear envelope. These simulation results suggest that the transcriptional activity of genes is regulated by their spatial position in the nucleus. This mechanism of regulation could arise from the localization of RNA polymerases, transcription factors, or repressive factors such as SIR proteins in the nucleus. Further computational and experimental tests are required to verify this hypothesis. Particularly, the genomewide test for the generic genes is needed. This type of analysis is discussed in the Supporting Material.

To develop efficient computational tests, it is important to further improve the simulation model. In particular, a more accurate treatment of interactions between chromatin loci is important. As shown by previous computational models (27,32,33), simulations that do not consider the specific interactions between chromosome loci reproduce many aspects of the overall genome structure such as the distribution of telomere positions. Indeed, those simulated features do not differ significantly in this model, which explicitly considers the specific chromosome interactions. However, the distribution pattern of individual genes is more sensitive to the interactions between chromosome loci, which can, for example, be found in the simulated distribution of the tRNA genes. Given the interactions between tRNA genes and 5S rRNA genes in this model, tRNA genes tend to localize around the nucleolus consistently with the observation (12,21), and without these interactions, tRNA genes distribute more around the center of the nucleus or the centromere region. As shown in Fig. S12, the distribution of other genes is also affected by these interactions; the distribution of the bottom 200 genes shifts, particularly in the mutant, from the region near the nucleolus to the opposite side of the nucleus if the interactions between tRNA genes and 5S rRNA genes are turned off.

The results of our simulation showed that the budding yeast genome fluctuates dynamically in the nucleus, which determines the regions in which individual genes are localized. An interesting finding is that the ribosomal protein genes are highly expressed; however, their position is localized in a region different from other highly expressed genes. As shown in Fig. 3, the lowly expressed genes show the smallest density around the region of ribosomal protein gene accumulation, suggesting that the repressive regulatory factors are excluded or RNA polymerases are accumulated in this region. This result is a significant feature of the position-dependent regulation in the yeast nucleus, and should be examined by monitoring the distribution and dynamics of these genes and factors in the nucleus using fluorescent marker techniques.

In this study, we used the *yku70 esc1* mutant genome as a perturbed genome from the WT. A similar analysis can be used for other mutants or perturbed cells. In particular, an artificial manipulation to chemically change the genome architecture is important (56); such genome engineering methods should provide a useful test platform to analyze the effect of perturbations. We consider that computational simulation models will play an important role in the elucidation of genome organization and function as tools for integrating results from different approaches, including a biochemical approach like the Hi-C method, a biophysical approach like the fluorescence monitoring of chromatin, and the genetic and chemical methods used for the genome organization.

SUPPORTING MATERIAL

Twelve figures, five tables, and two movies are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(16)34270-9.

AUTHOR CONTRIBUTIONS

N.T. and M.S. designed research; N.T. performed research; N.T. and M.S. analyzed data; and N.T. and M.S. wrote the article.

ACKNOWLEDGMENTS

REFERENCES

1. Zimmer, C., and E. Fabre. 2011. Principles of chromosomal organization: lessons from yeast. *J. Cell Biol.* 192:723–733.

2. Rosa, A., and C. Zimmer. 2014. Computational models of large-scale genome architecture. *Int. Rev. Cell Mol. Biol.* 307:275–349.

3. Wang, R., J. Mozziconacci, …, O. Gadal. 2015. Principles of chromatin organization in yeast: relevance of polymer models to describe nuclear organization and dynamics. *Curr. Opin. Cell Biol.* 34:54–60.

4. Taddei, A., and S. M. Gasser. 2012. Structure and function in the budding yeast nucleus. *Genetics.* 192:107–129.

5. Tanizawa, H., and K. Noma. 2012. Unravelling global genome organization by 3C-seq. *Semin. Cell Dev. Biol.* 23:213–221.

6. Dekker, J., M. A. Marti-Renom, and L. A. Mirny. 2013. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat. Rev. Genet.* 14:390–403.

7. Denker, A., and W. de Laat. 2016. The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev.* 30:1357–1382.

8. Schmitt, A. D., M. Hu, and B. Ren. 2016. Genome-wide mapping and analysis of chromosome architecture. *Nat. Rev. Mol. Cell Biol.* 17:743–755.

9. Heun, P., T. Laroche, …, S. M. Gasser. 2001. Chromosome dynamics in the yeast interphase nucleus. *Science.* 294:2181–2186.

10. Hediger, F., F. R. Neumann, …, S. M. Gasser. 2002. Live imaging of telomeres: yKu and Sir proteins define redundant telomere-anchoring pathways in yeast. *Curr. Biol.* 12:2076–2089.

11. Bystricky, K., T. Laroche, …, S. M. Gasser. 2005. Chromosome looping in yeast: telomere pairing and coordinated movement reflect anchoring efficiency and territorial organization. *J. Cell Biol.* 168:375–387.

12. Duan, Z., M. Andronescu, …, W. S. Noble. 2010. A three-dimensional model of the yeast genome. *Nature.* 465:363–367.

13. Berger, A. B., G. G. Cabal, …, C. Zimmer. 2008. High-resolution statistical mapping reveals gene territories in live yeast. *Nat. Methods.* 5:1031–1037.

14. Tokuda, N., T. P. Terada, and M. Sasai. 2012. Dynamical modeling of three-dimensional genome organization in interphase budding yeast. *Biophys. J.* 102:296–304.

15. Taddei, A., H. Schober, and S. M. Gasser. 2010. The budding yeast nucleus. *Cold Spring Harb. Perspect. Biol.* 2:a000612.

16. Tham, W. H., and V. A. Zakian. 2002. Transcriptional silencing at *Saccharomyces* telomeres: implications for other organisms. *Oncogene.* 21:512–521.

17. Perrod, S., and S. M. Gasser. 2003. Long-range silencing and position effects at telomeres and centromeres: parallels and differences. *Cell. Mol. Life Sci.* 60:2303–2318.

18. Wellinger, R. J., and V. A. Zakian. 2012. Everything you ever wanted to know about *Saccharomyces cerevisiae* telomeres: beginning to end. *Genetics.* 191:1073–1105.

19. Mekhail, K., J. Seebacher, …, D. Moazed. 2008. Role for perinuclear chromosome tethering in maintenance of genome stability. *Nature.* 456:667–670.

20. Kim, Y. H., D. Ishikawa, …, S. Harashima. 2006. Chromosome XII context is important for rDNA function in yeast. *Nucleic Acids Res.* 34:2914–2924.

21. Thompson, M., R. A. Haeusler, …, D. R. Engelke. 2003. Nucleolar clustering of dispersed tRNA genes. *Science.* 302:1399–1401.

22. Wang, L., R. A. Haeusler, …, D. R. Engelke. 2005. Silencing near tRNA genes requires nucleolar localization. *J. Biol. Chem.* 280:8637–8639.

23. Cook, P. R. 1999. The organization of replication and transcription. *Science.* 284:1790–1795.

24. Osborne, C. S., L. Chakalova, …, P. Fraser. 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* 36:1065–1071.

25. Sutherland, H., and W. A. Bickmore. 2009. Transcription factories: gene expression in unions? *Nat. Rev. Genet.* 10:457–466.

26. Tanizawa, H., O. Iwasaki, …, K. Noma. 2010. Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic Acids Res.* 38:8164–8177.

27. Gong, K., H. Tjong, …, F. Alber. 2015. Comparative 3D genome structure analysis of the fission and the budding yeast. *PLoS One.* 10:e0119672.

28. Ben-Elazar, S., Z. Yakhini, and I. Yanai. 2013. Spatial localization of co-regulated genes exceeds genomic gene clustering in the *Saccharomyces cerevisiae* genome. *Nucleic Acids Res.* 41:2191–2201.

29. Homouz, D., and A. S. Kudlicki. 2013. The 3D organization of the yeast genome correlates with co-expression and reflects functional relations between genes. *PLoS One.* 8:e54699.

30. Witten, D. M., and W. S. Noble. 2012. On the assessment of statistical significance of three-dimensional colocalization of sets of genomic elements. *Nucleic Acids Res.* 40:3849–3855.

31. Taddei, A., G. van Houwe, …, S. M. Gasser. 2009. The functional importance of telomere clustering: global changes in gene expression result from SIR factor dispersion. *Genome Res.* 19:611–625.

32. Tjong, H., K. Gong, …, F. Alber. 2012. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Res.* 22:1295–1305.

33. Wong, H., H. Marie-Nelly, …, C. Zimmer. 2012. A predictive computational model of the dynamic 3D interphase yeast nucleus. *Curr. Biol.* 22:1881–1890.

34. Dultz, E., H. Tjong, …, K. Weis. 2016. Global reorganization of budding yeast chromosome conformation in different physiological conditions. *J. Cell Biol.* 212:321–334.

35. https://www.ebi.ac.uk/arrayexpress/experiments/E-TABM-630/.

36. Bystricky, K., P. Heun, …, S. M. Gasser. 2004. Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc. Natl. Acad. Sci. USA.* 101:16495–16500.

37. Collepardo-Guevara, R., and T. Schlick. 2014. Chromatin fiber polymorphism triggered by variations of DNA linker lengths. *Proc. Natl. Acad. Sci. USA.* 111:8061–8066.

38. Kepper, N., D. Foethke, …, K. Rippe. 2008. Nucleosome geometry and internucleosomal interactions control the chromatin fiber conformation. *Biophys. J.* 95:3692–3705.

39. Routh, A., S. Sandin, and D. Rhodes. 2008. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proc. Natl. Acad. Sci. USA.* 105:8872–8877.

40. Dekker, J. 2008. Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction. *J. Biol. Chem.* 283:34532–34540.

41. Hsieh, T. H., A. Weiner, …, O. J. Rando. 2015. Mapping nucleosome resolution chromosome folding in yeast by micro-C. *Cell.* 162:108–119.

42. Hsieh, T. S., G. Fudenberg, …, O. J. Rando. 2016. Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. *Nat. Methods.* 13:1009–1011.

43. Maeshima, K., S. Ide, …, M. Sasai. 2016. Liquid-like behavior of chromatin. *Curr. Opin. Genet. Dev.* 37:36–45.

44. Ricci, M. A., C. Manzo, …, M. P. Cosma. 2015. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell.* 160:1145–1158.

45. Yaffe, E., and A. Tanay. 2011. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat. Genet.* 43:1059–1065.

46. Cournac, A., H. Marie-Nelly, …, J. Mozziconacci. 2012. Normalization of a chromosomal contact map. *BMC Genomics.* 13:436.

47. https://noble.gs.washington.edu/proj/yeast-architecture/sup.html.

48. Andrulis, E. D., D. C. Zappulla, …, R. Sternglanz. 2002. Esc1, a nuclear periphery protein required for Sir4-based plasmid anchoring and partitioning. *Mol. Cell. Biol.* 22:8292–8301.

49. Bupp, J. M., A. E. Martin, …, S. L. Jaspersen. 2007. Telomere anchoring at the nuclear periphery requires the budding yeast Sad1-UNC-84 domain protein Mps3. *J. Cell Biol.* 179:845–854.

50. Fisher, T. S., and V. A. Zakian. 2005. Ku: a multifunctional protein involved in telomere maintenance. *DNA Repair (Amst.).* 4:1215–1226.

51. Moretti, P., K. Freeman, …, D. Shore. 1994. Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1. *Genes Dev.* 8:2257–2269.

52. Gasser, S. M., and M. M. Cockell. 2001. The molecular biology of the SIR proteins. *Gene.* 279:1–16.

53. Kueng, S., M. Oppikofer, and S. M. Gasser. 2013. SIR proteins and the assembly of silent chromatin in budding yeast. *Annu. Rev. Genet.* 47:275–306.

54. Therizols, P., T. Duong, …, E. Fabre. 2010. Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proc. Natl. Acad. Sci. USA.* 107:2025–2030.

55. Taddei, A., F. Hediger, …, S. M. Gasser. 2004. Separation of silencing from perinuclear anchoring functions in yeast Ku80, Sir4 and Esc1 proteins. *EMBO J.* 23:1301–1312.

56. Keung, A. J., J. K. Joung, …, J. J. Collins. 2015. Chromatin regulation at the frontier of synthetic biology. *Nat. Rev. Genet.* 16:159–171.