Complex multi-enhancer contacts captured by Genome Architecture Mapping (GAM)

Robert A. Beagrie, Antonio Scialdone, Markus Schueler, Dorothee C.A. Kraemer, Mita Chotalia, Sheila Q. Xie, Mariano Barbieri, Inês de Santiago, Liron-Mark Lavitas, Miguel R. Branco, James Fraser, Josée Dostie, Laurence Game, Niall Dillon, Paul A.W. Edwards, Mario Nicodemi, Ana Pombo

Supplementary Notes 1 to 3

Supplementary Note 1

Contents

List of Figures

1 The SLICE model

GAM measures raw co-segregation frequencies that include both specific chromatin interactions and the "background" co-segregation of non-interacting loci. The rate at which non-interacting loci fall randomly into the same nuclear profile (NP) - i.e., their co-segregation frequency - is predictable and directly dependent on their nuclear distance. Hence, unlike other methods, the nature of GAM data allows to dissect and quantify significant interactions above the expected random behavior by use of statistical models. This was accomplished by the development and application of the SLICE (**S**tatistica**L** **I**nference of **C**o-s**E**gregation) model that infers non-random contacts from the expected random behavior of loci at a given genomic distance, taking into account a variety of confounding effects, such as the detection efficiency of genomic loci and genomic resolution. SLICE also dissects the probability of significant, simultaneous three-way interactions (triplets), and more generally multivalent contacts between several loci.

SLICE explicitly considers that a given pair of loci can be (a) specifically interacting, or (b) non-interacting (and thus randomly located) across the cell population of interest. As the co-segregation frequency of a locus pair observed in a collection of randomly-sectioned nuclear profiles depends directly on the interaction state (interacting/non-interacting) of the locus pair, the expected proportions of nuclear profiles in each state, containing two loci, one locus or neither locus, can be determined by mathematical analysis and expressed as a function of π (i.e., the interaction probability, in the main text referred to as Pi; see schematic in main text, Fig. 3a). SLICE predicts the expected proportion of GAM nuclear profiles containing the two loci of a given pair, one of the loci or no locus, to find the value of π which best describes the experimental GAM data. By fitting the experimental data, π is derived as the proportion of the cell population where the loci are in a state of direct interaction. Mathematical modeling of the GAM data can also be exploited to find the most prominent interactions and estimate GAM sensitivity. The SLICE method can be extended to identify multivalent interactions, as shown in the last section of this document for triplets of interacting loci.

1.1 Introduction

We introduced the SLICE model to extract genome-wide interaction probabilities from GAM co-segregation data. Consider the case where the GAM procedure is applied to N cells. From each cell, a single "slice" or "nuclear profile" (NP) is taken. For each NP, the GAM method provides information on presence/absence of essentially all loci of a given genome: e.g., we know whether the NP k (for all k's) includes or not a locus A, B, C, and so on. By collecting the statistics of co-

segregation of loci in the NPs, we can estimate their interaction probabilities via a mathematical model. We first focus on the statistics of occurrence of one locus and pairs of loci in a single NP. We then introduce the corresponding parameters and their estimation from GAM data. Finally, the last section is devoted to the determination of the interaction probabilities of triplets of loci.

1.2 Distribution of loci in a nuclear profile

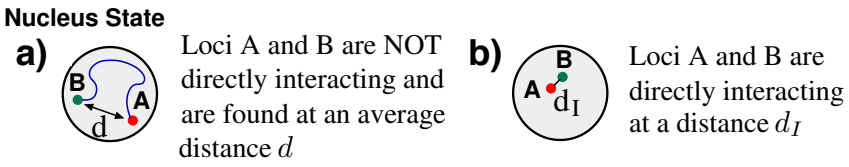

Figure S1.1: Schematic representation of non-interacting loci A and B separated by genomic distance d (**a**) and interacting loci A and B at interacting distance d_I (**b**)

Let us consider two loci A and B. We suppose that they can be found in two possible states (see Fig. S1.1): **a)** A and B do not interact directly with each other and are found in the nucleus at an average distance, d; **b)** they interact directly (they are "tied") and are found at a much smaller distance, the interaction distance d_I. To test such a scenario against experimental data, we have to compute for each NP the probabilities:

- to find or not a single locus (v_1, v_0);
- to find 2, 1 or 0 loci of a pair of non-interacting (*"untied"*) loci (u_2, u_1, u_0);
- to find 2, 1 or 0 loci of a pair of interacting (*"tied"*) loci (t_2, t_1, t_0).

A schematic representation of each probability is shown in Fig. S1.2. Each of these probabilities are functions of the NP and the locus positions. Since the NPs are taken at random orientations and include both equatorial and apical sections, if we seek the average number of NPs including, e.g., a pair of interacting loci, we must average the value of t_2 over all possible configurations of the NPs and the loci. The position of the NP can be indicated by the coordinate z, i.e., by its distance from the equatorial plane (see fig. S1.2.a).

In the paragraphs below, we discuss possible methods to calculate the averages (which we will indicate with the symbol $\langle ... \rangle$) of the probabilities listed above. Importantly, the probabilities t_i, u_i, v_i obey a set of normalization relations. For instance, as a locus must be present in the nucleus, we can state that: $v_1 + v_0 = 1$.

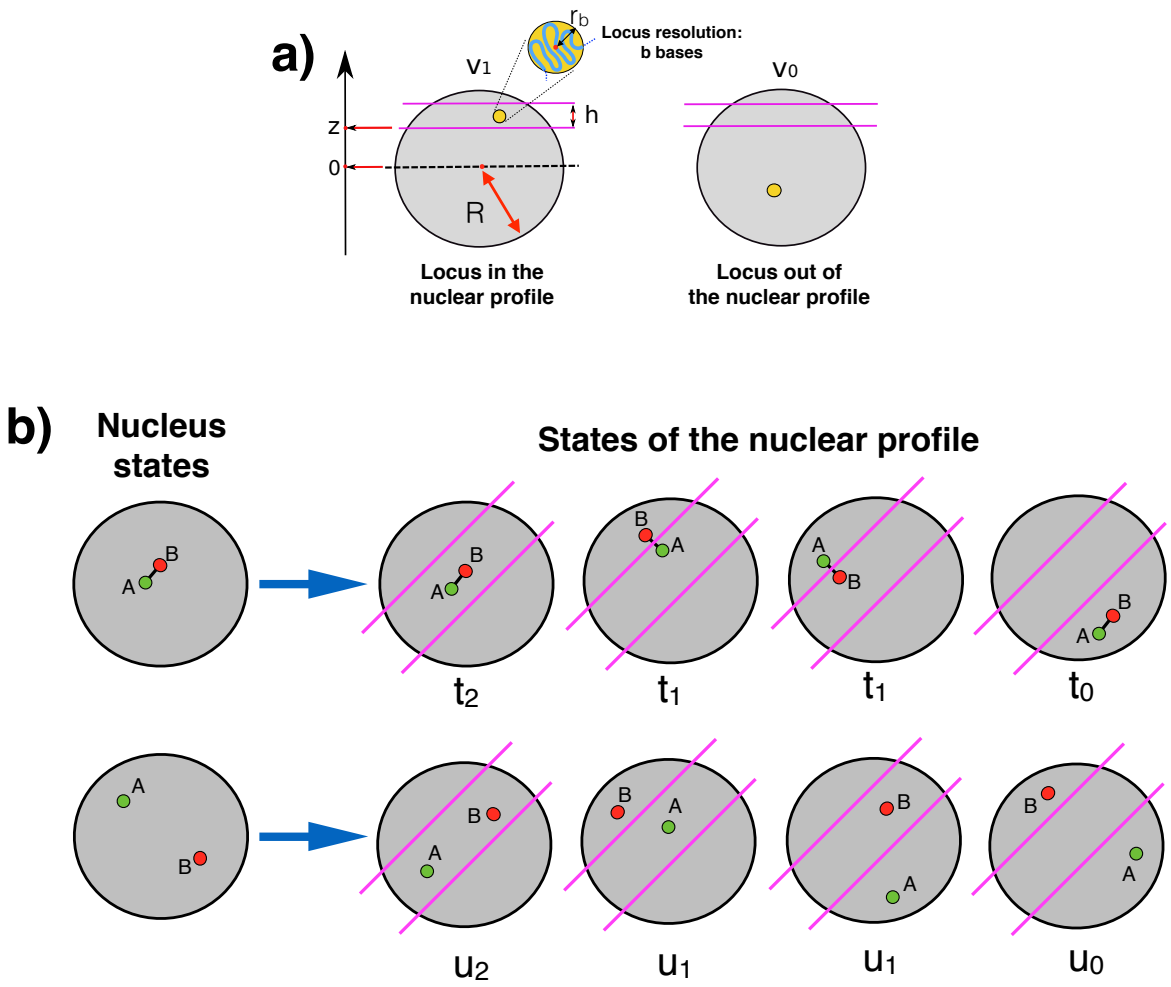

Figure S1.2: **(a)** Schematic representation of the probabilities of finding a single locus in nuclear profiles sectioned at random orientations. z is the distance of the NP from the equatorial plane, h is the slice thickness and R the nuclear radius. A DNA locus can be represented in our model by a sphere having a radius r_b which depends on the considered resolution b (see section 1.4 for a discussion on how the resolution b can be accounted for in the model). **(b)** Probability of finding a pair of loci (interacting or non-interacting) in a nuclear profile.

In the following sections, we analyse in more detail each of the probabilities t_i, u_i, v_i, and illustrate how they can be estimated. We used both analytical and numerical techniques, alongside Monte Carlo simulations to verify our results (see below and Fig.s S1.3 and S1.5).

Single locus probability: v_1, v_0

For clarity, we first discuss the case where only a single locus is considered (see Fig. S1.2 top panel). We define:

- $\langle v_1 \rangle$ = probability that the locus is found in an NP
- $\langle v_0 \rangle = 1 - \langle v_1 \rangle$ = probability that the locus is not found in an NP

We can derive $\langle v_0 \rangle$ and $\langle v_1 \rangle$ from the geometric properties of the system. If the locus is randomly located in the nucleus, $\langle v_1 \rangle = V_{NP}/V_{Nucleus}$, where $V_{Nucleus}$ is the volume of the nucleus and V_{NP} the average volume of an NP. For the sake of simplicity, we approximate the nuclear shape with a sphere, and consider first the case where all nuclei have a radius R. Under these approximations, $V_{Nucleus}$ is constant and equal to $(4/3)\pi R^3$. Similarly, the volume of a NP at height z, $V_{NP}(z)$, can be computed in a sphere.

By averaging $V_{NP}(z)/V_{Nucleus}$ over z, we obtain (see fig. S1.3):

$$\langle v_1 \rangle \left(\frac{h}{R}\right) = \frac{\frac{h}{R}}{2 + \frac{h}{R}}, \qquad \langle v_0 \rangle \left(\frac{h}{R}\right) = 1 - \langle v_1 \rangle \left(\frac{h}{R}\right) = \frac{2}{2 + \frac{h}{R}} \tag{1}$$

Note that $\langle v_1 \rangle$ and $\langle v_0 \rangle$ are functions of the ratio h/R. This is due to their scaling properties: they remain unchanged if the radius and the slice thickness are scaled by the same factor. In the more general case when the nuclei have different radii across the sampled population, the previous eq.s can be easily generalized to include the probability distribution of nuclear radii $P(R)$. Alternatively, in a mean-field-like approximation [1], the radius R in the eq.s (1) can be replaced by the average nuclear radius $\langle R \rangle$.

Estimation of average nuclear radius

The average nuclear radius can be estimated from the radii of the NPs. Indeed, when NPs are randomly sampled from spherical nuclei with a radius R, the average NP radius is equal to $(\pi/4)R$. From this equation, if we consider the case of nuclei with a generic probability distribution of radii $P(R)$, we obtain: $\langle R \rangle = (4/\pi)\langle R_{NP} \rangle$, where $\langle R \rangle$ is the average nuclear radius and $\langle R_{NP} \rangle$ the average radius of the NPs. We measured the radii of mESC NPs from images of cryosections and, by using the formula above, we obtained a value for the average nuclear radius of $4.5\mu m$.

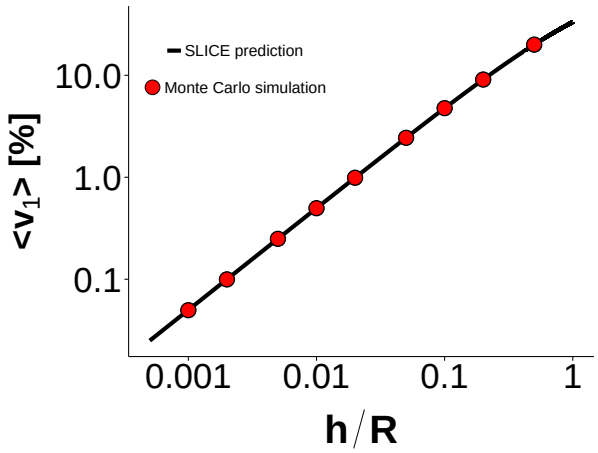

Figure S1.3: Average probability that the locus is found in a NP (as calculated from the theoretical distribution of NP volumes), $\langle v_1 \rangle$, as function of the ratio between the slice thickness h and the nuclear radius R. The SLICE prediction (black line) perfectly matches the estimations from Monte Carlo simulations (red circles). In the present work, the average nuclear radius is $R = 4.5\mu m$ and the average width of a nuclear profile is $h = 0.22\mu m$, so $h/R \sim 0.05$.

Two loci in a non-interacting state: u_i probabilities

Let us consider now the case of two non-interacting loci. Under the current model, we describe below the general equations for u_i (note the normalisation condition: $\langle u_2 + 2u_1 + u_0 \rangle = 1$).

Equations for u_i **-** Let us assume that the relative vector distance of the two loci, $\vec{d}$, has a probability distribution $\Phi(\vec{d}, \vec{\theta})$, with $\vec{\theta}$ being the parameters of the distribution. The probability that two loci at a distance $\vec{d}$ co-segregate in the same NP at height z is:

$$u_2(z, \vec{d}) = \frac{\int_{\vec{x_1}, \vec{x_2} \in V_{NP}(z)} \delta\left(\vec{x_1} - \vec{x_2} - \vec{d}\right) d\vec{x_1} d\vec{x_2}}{\int_{\vec{x_1}, \vec{x_2} \in V_{Nucleus}} \delta\left(\vec{x_1} - \vec{x_2} - \vec{d}\right) d\vec{x_1} d\vec{x_2}} \tag{2}$$

Then, $\langle u_2 \rangle$ is given by:

$$\langle u_2 \rangle = \frac{\int dz \int d(\vec{d})\, \Phi(\vec{d}, \vec{\theta})\, u_2(z, \vec{d})}{\int dz \int d(\vec{d}) \Phi(\vec{d}, \vec{\theta})} \tag{3}$$

The equations for u_1 and u_0 can be analogously derived. Depending on specific assumptions on $\Phi(\vec{d}, \vec{\theta})$, the u_i can be calculated analytically or by numerical methods

such as Monte Carlo simulations.
The probability distribution of the non-interaction distance $\Phi(\vec{d},\vec{\theta})$ is determined by complex factors such as the confinement of the chromatin into territories, the interaction with other genomic loci and/or with nuclear elements (e.g., membrane, nucleolus), and many more. While the SLICE model is versatile enough to accommodate these complexities, the simplest approach for pairs of loci on different chromosomes would be to assume that their nuclear positions are random and independent from each other, which leads to:

$$\langle u_0 \rangle = \langle v_0^2 \rangle, \qquad \langle u_1 \rangle = \langle v_1 v_0 \rangle, \qquad \langle u_2 \rangle = \langle v_1^2 \rangle \tag{4}$$

As for loci on the same chromosome, instead of calculating the probabilities u_2, u_1, u_0 for a particular $\Phi(\vec{d},\vec{\theta})$ that is fixed a priori, we can instead estimate them directly from the data, as we describe in the following paragraph.

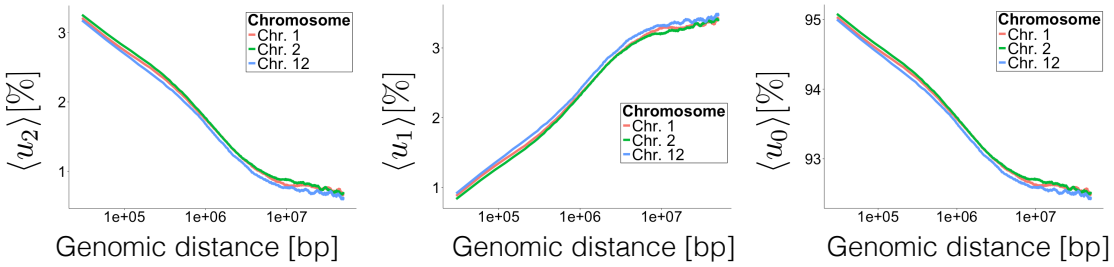

Figure S1.4: Co-segregation probabilities u_2, u_1, u_0 for non-interacting loci estimated from the data for chromosomes 1, 2 and 12 (see below section 1.5). Note that different chromosomes have very similar values of u_i's.

Estimation of u_2***,***u_1***,***u_0 ***from the co-segregation data -*** The average co-segregation frequencies of pairs of loci on the same chromosome at a given genomic distance can be used to estimate u_2, u_1, u_0 (see also below section 1.5) under the assumption that the average interaction probability between loci at a fixed genomic distance is small, which is supported by our analysis of GAM data (e.g., the average interaction probability of loci at a genomic distance of $1Mb$ is $\sim 7.5\%$). In order to account for differences between chromosomes due to, e.g., different lengths, the u_2, u_1, u_0 were estimated separately for each chromosome (see Fig. S1.4; as discussed in section 1.5). We adopted this simplified approach with the current dataset, but as the size of the dataset increases this could be refined by using, e.g., an iterative procedure: the u_i's can be first estimated by using all pairs of loci, then they can be fed into the model to find the most weakly interacting loci, which in turn will be used to improve the estimations of the u_i's.

Two loci in an interacting state: t_i probabilities

Let's now consider the case of two specifically interacting loci. We derive here the probabilities t_i (satisfying the normalisation condition $\langle t_2 + 2t_1 + t_0 \rangle = 1$).

Equations and estimation of t_i **-** Name $\vec{d_I}$ the relative distance vector of the two interacting loci and $\Psi(\vec{d_I}, \vec{\theta})$ its probability distribution. The general equations for the t_i's are similar to those used for the u_i. Specific probability distributions, Ψ, can be chosen for different cases. Monte Carlo simulations were used to estimate the t_i's with fixed values of d_I and different values of h (Fig. S1.5). Note that from the Monte Carlo simulations, for $d_I \lesssim h$, the t_i can be well approximated by:

$$\langle t_2 \rangle \sim \langle v_1 \rangle, \qquad \langle t_1 \rangle \sim 0, \qquad \langle t_0 \rangle \sim \langle v_0 \rangle = 1 - \langle v_1 \rangle \tag{5}$$

with $\langle v_1 \rangle, \langle v_0 \rangle$ as predicted by the SLICE formulas above (eq.s (1))

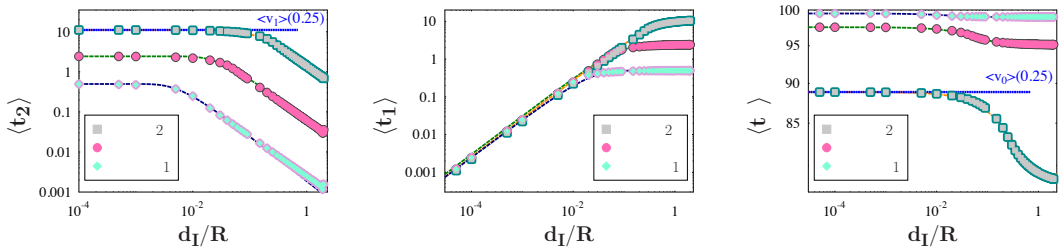

Figure S1.5: The values of $\langle t_2 \rangle, \langle t_1 \rangle, \langle t_0 \rangle$ from Monte Carlo simulations are plotted as function of d_I/R, i.e., the ratio between the interaction distance and the nuclear radius, for different values of h/R, h being the slice thickness. In the present study, $h/R \sim 0.05$. The dashed blue lines indicate the asymptotic values of $\langle t_2 \rangle$ and $\langle t_0 \rangle$ for small values of d_I/R corresponding to $\langle v_1 \rangle$ and $\langle v_0 \rangle$. The superimposed continuous lines are fits with exponential functions.

1.3 Occurrence probability of pairs of loci in single nuclear profiles

States of loci in the nucleus

By using the results of the previous section, we now compute the statistics of the presence of loci in a NP. We consider a pair of loci, A and B: the two alleles are A_1 and A_2, and B_1 and B_2. We describe, for brevity, only the case when A and B are on the same chromosome. The case of loci on different chromosomes is derived with analogous calculations. For the sake of simplicity, let us consider that:

- the interaction probabilities of loci on different chromosomes (e.g., the interaction between locus A_1 and locus B_2) is negligible with respect to that of A and B loci on the same chromosome (e.g., of A_1 and B_1);

- the two homologous pairs A_1, B_1 and A_2, B_2 interact independently of one another with the same probability π.

Under these assumptions, the loci can exist in three different types of conformations having probabilities P_2, P_1, P_0 (see fig. S1.6, upper panel):

- P_2 = probability that two pairs of loci are present in an interacting state in the nucleus;
- P_1 = probability that one pair interacts while the other does not;
- P_0 = probability that no interacting pairs are present;

with the normalization $P_2 + P_1 + P_0 = 1$. The probabilities P_i can be written as:

$$P_2 = \pi^2, \quad P_1 = 2\pi(1-\pi), \quad P_0 = (1-\pi)^2 \tag{6}$$

where π is the probability that A_1 is interacting with B_1 or A_2 with B_2.

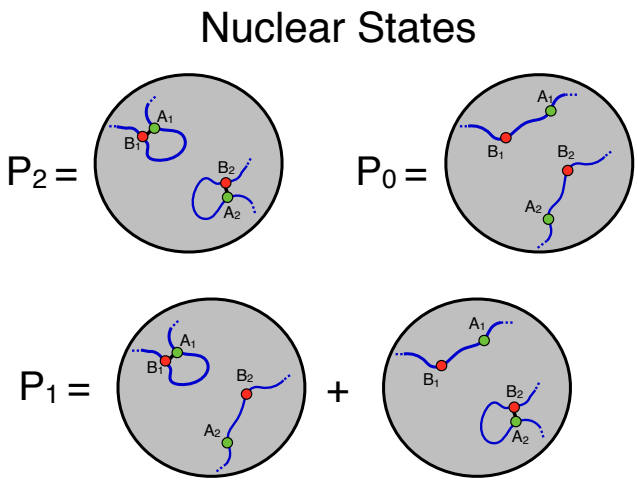

Figure S1.6: Illustration of the three nuclear states considered here.

Probability of locus segregation in single nuclear profiles

Consider a set of N independent NPs taken from N nuclei, and let us consider the case in which we know the number of times that loci A and B are present in each NP across the ensemble of NPs. By using the above probabilities, we can derive the expected number of NPs, $N_{i,j}$, with a number i of A loci and a number j of B loci (i and j are equal to 0,1 or 2 loci). For instance, the number of slices with one A locus and one B locus is:

$$\frac{N_{1,1}}{N} = \left(2\langle t_2 t_0\rangle + 2\langle t_1^2\rangle\right) P_2 + \left(\langle t_2 u_0\rangle + \langle t_0 u_2\rangle + 2\langle t_1 u_1\rangle\right) P_1 + \left(2\langle u_2 u_0\rangle + 2\langle u_1^2\rangle\right) P_0 \tag{7}$$

The intervening (auto)correlation terms, like $\langle u_i^{n_u} t_j^{n_j}\rangle$, can be evaluated by Monte Carlo techniques as those in section 1.2. In Statistical Mechanics "mean-field" approximations are often exploited to simplify calculations (see, e.g., [1]) where $\langle u_i^{n_u} t_j^{n_t}\rangle$ is taken equal to $\langle u_i\rangle^{n_u}\langle t_j\rangle^{n_t}$, with $i, j = 0, 1, 2$. We adopt this simplifying assumption here and we will omit the average symbols ($\langle ...\rangle$) from now on to simplify the notation.

The equations can be simplified by using eq.s (6) for the P_i and the following expressions:

$$c_2 = \pi t_2 + (1-\pi)u_2, \qquad c_1 = \pi t_1 + (1-\pi)u_1, \qquad c_0 = \pi t_0 + (1-\pi)u_0 \tag{8}$$

where c_2, c_1, c_0 represent the probabilities that $2, 1$ or 0 loci of an AB couple are segregated in a NP. From the normalisation equations, it can be shown that the following relations hold (not only for the averages over the NPs, but also in each single NP):

$$c_2 + 2c_1 + c_0 = 1, \qquad c_1 + c_0 = v_0, \qquad c_2 + c_1 = 1 - v_0 \tag{9}$$

By using eq.(8) and (9) and the mean-field approximations, the $N_{i,j}$ can be written as:

$$\begin{cases} \frac{N_{0,0}}{N} = c_0^2 \\ \frac{N_{1,1}}{N} = 2[(1 - 2v_0 + c_0)c_0 + (v_0 - c_0)^2] \\ \frac{N_{2,2}}{N} = (1 - 2v_0 + c_0)^2 \\ \frac{N_{1,0}}{N} = \frac{N_{0,1}}{N} = 2(v_0 - c_0)c_0 \\ \frac{N_{2,0}}{N} = \frac{N_{0,2}}{N} = (v_0 - c_0)^2 \\ \frac{N_{2,1}}{N} = \frac{N_{1,2}}{N} = 2(1 - 2v_0 + c_0)(v_0 - c_0) \end{cases} \tag{10}$$

1.4 Coping with finite detection efficiency and resolution

In practical applications of GAM, we have also to consider within our statistical description the effects related to the limited experimental detection efficiency of loci and finite resolution. This section expands the model above to accommodate these additional complications.

Detection efficiency

We consider the situation where the detection efficiency of a DNA locus in a NP is less than 1, i.e., where there is a certain probability that the locus is present in the NP but not detected by DNA sequencing (see Fig. S1.7). For simplicity, we assume that both A and B loci have the same detection efficiency ϵ (the generalization to the case of different efficiencies for different loci is straightforward). Name $N^{\epsilon}_{\alpha,\beta}$ the expected number of NPs where a number α of A loci and a number β of B loci is detected. If $\epsilon < 1$, we should take into account the possibility that in a NP there could be additional undetected loci with a probability $1-\epsilon$. The $N^{\epsilon}_{\alpha,\beta}$ are a linear combination of the $N_{i,j}$ introduced above in the ideal case (i.e., with $\epsilon = 1$):

$$N^{\epsilon}_{\alpha,\beta} = \epsilon^{\alpha+\beta} \sum_{i=\alpha,j=\beta}^{2} (1-\epsilon)^{(i+j)-(\alpha+\beta)} \left(\delta_{\alpha,1}\delta_{i,2}+1\right)\left(\delta_{\beta,1}\delta_{j,2}+1\right) N_{i,j} \qquad (11)$$

δ being the Kronecker indicator-function.

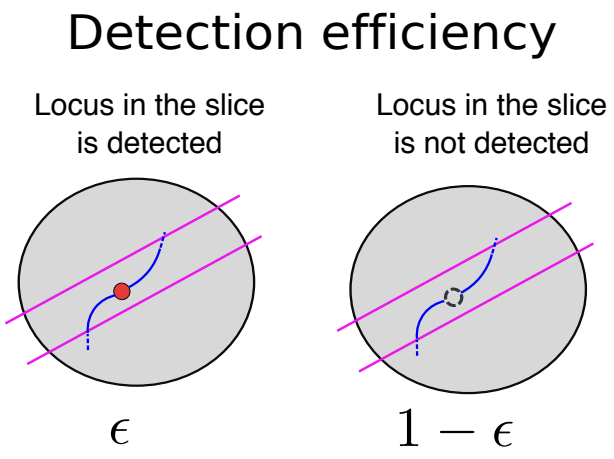

Figure S1.7: An illustration of the effects of the detection efficiency.

Genomic finite resolution

In the calculations above, DNA loci were approximated as dimensionless points. In real situations, a DNA locus includes several bases, and can be better approximated as a sphere having a finite radius. For instance, if the experiment aims to detect locus contacts at a given resolution of b DNA bases, a DNA locus can be represented in our model by a sphere having a radius r_b, which contains b DNA bases around it (see Fig. S1.8). The above equations for v_0, v_1 (eq. (1)) can be adapted to this more general case that considers genomic resolution:

$$\langle v_1 \rangle = \frac{h + 2r_b}{h + 2R}, \qquad \langle v_0 \rangle = \frac{2(R - r_b)}{h + 2R} \tag{12}$$

which, if $r_b \ll R$, are well approximated by:

$$\langle v_1 \rangle = \frac{h_{eff}}{h_{eff} + 2R}, \qquad \langle v_0 \rangle = \frac{2R}{h_{eff} + 2R} \tag{13}$$

with $h_{eff} = h + 2r_b$. These equations are identical to eq.s (1), except for the presence of h_{eff}. This shows that the equations given in the previous sections still hold true when the finite size of the DNA loci is taken into account, provided that the slice thickness h is replaced by the effective slice thickness h_{eff} (see Fig. S1.8). We can obtain an approximation of r_b as function of the DNA resolution b: $\frac{r_b^3}{R^3} = \frac{b}{G}$, where G is the size of the genome multiplied by the ploidy. This can be derived from the assumption of an homogeneous distribution of the genetic material in the nucleus.

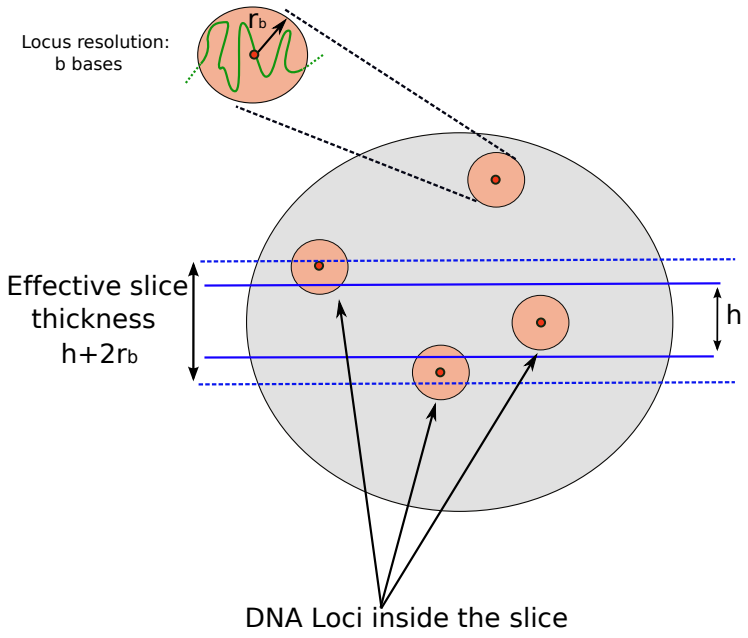

Figure S1.8: Schematic representation of DNA loci with a resolution b in the cell nucleus.

1.5 Co-segregation of pairs of loci

In the GAM method described, DNA is amplified from a single NP. The DNA in each NP is then sequenced and the presence/absence of each locus is determined. Below we compute the probability for each NP to include $0, 1, 2$ loci of a pair and show how their interaction probability can be estimated.

Equations for the NP content

The expected number of NPs with $2, 1$ or 0 loci from a given pair of loci (m_2, m_1 and m_0, respectively) can be calculated by using the equations shown in the previous section. If m is the total number of NPs, and each locus is detected with an efficiency ϵ, the following equations hold:

$$\begin{aligned}
\frac{m_0}{m} &= \text{Expected fraction of NPs with neither A, nor B} = N^{\epsilon}_{0,0}/N \qquad (14)\\
\frac{m_1}{m} &= \text{Expected fraction of NPs with A or B} = \\
&= 2\left(N^{\epsilon}_{1,0}/N + N^{\epsilon}_{2,0}/N\right)\\
\frac{m_2}{m} &= \text{Expected fraction of NPs with both A and B} = 1 - \frac{m_0}{m} - \frac{m_1}{m}\\
&= 1 - N^{\epsilon}_{0,0}/N - 2\left(N^{\epsilon}_{1,0}/N + N^{\epsilon}_{2,0}/N\right)
\end{aligned}$$

where the $N^{\epsilon}_{\alpha,\beta}$ are given by eq. (11) and the symmetry between the two types of loci (i.e., $N^{\epsilon}_{0,i} = N^{\epsilon}_{i,0}$) has been used.
These equations allow us, in particular, to calculate the expected value of the co-segregation ratio, i.e., the ratio between the number of NPs that include both A and B loci divided by the number of non-empty NPs, $m_2/(m_1 + m_2)$.
In addition to π, the free parameters of these equations are the detection efficiency ϵ, the co-segregation probabilities for interacting loci t_2, t_1, t_0 and the co-segregation probabilities for non-interacting loci u_2, u_1, u_0. In the following paragraphs we show how these parameters can be estimated.

Estimating the interaction probabilities of pairs

Importantly, we can use the eq.s (14) to estimate the detection efficiency ϵ. Indeed, from our simplifying assumptions about homologous loci (see above section 1.3), and by using the u_2, u_1, u_0 given by eq.s (4), it is found that the expected average fraction of NPs, f, which includes a given single locus is:

$$f = -\epsilon^2(1 - v_0)^2 + 2\epsilon(1 - v_0) \qquad (15)$$

By solving for ϵ:

$$\epsilon = \frac{1 - \sqrt{1 - f}}{1 - v_0} \tag{16}$$

where v_0 is given by eq. (13) and f is measured from the data. In the present study, loci present in an NP are detected by next-generation sequencing (NGS). Due to methodological and bioinformatic limitations, there are a number of genomic regions that are too repetitive to be reliably detected by NGS (so called "unmappable" regions). To avoid the skewing effect due to these unmappable regions, the loci that were never detected in any NPs were removed from the analysis.
Fig. S1.9 shows the detection efficiency obtained for different genomic resolutions, and that we detect $30kb$ windows with $\sim 83\%$ efficiency and $50kb$ windows with $\sim 98\%$ efficiency, at the current sequencing depth.

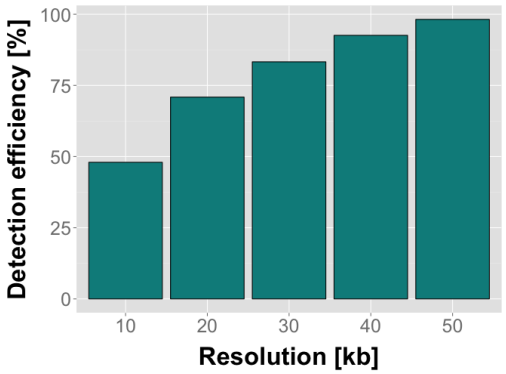

Figure S1.9: Detection efficiency as function of resolution computed from our data (slice thickness $h = 0.22\mu m$, total number of nuclear profiles $N = 408$).

As we mentioned above, also the u_2, u_1, u_0 probabilities can be estimated from the data (see section 1.2). In order to compute these probabilities for loci at a given genomic distance g, we compared the experimental value of the average co-segregation ratio $m_2/(m_1 + m_2)$ for all couples at a distance g with the theoretical value we get from eq.s (14) at $\pi = 0$. This was done separately for each chromosome to take into account different folding properties of chromosomes (e.g., different sizes of chromosome territories) that could influence the values of u_2, u_1, u_0.
Finally, the co-segregation probabilities t_2, t_1, t_0 for interacting loci can be derived by the v_1, v_0 for $d_I \lesssim h$ (section 1.2).

Once the above parameters are determined, the equations (14) describing the NP content depend only on the interaction probability π, that can be estimated by

fitting the co-segregation ratio $m_2/(m_1+m_2)$ (see Fig. S1.10).

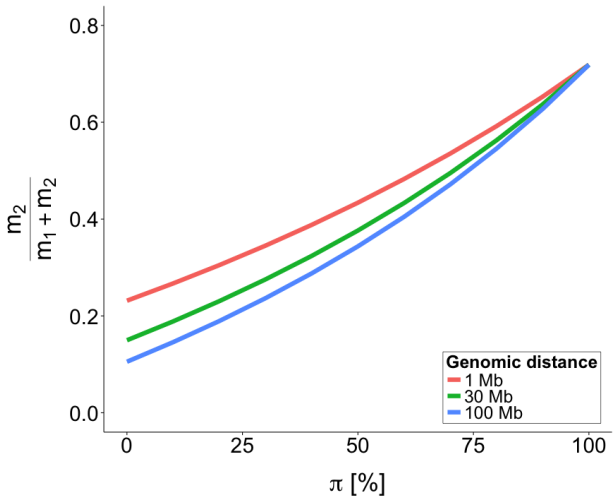

Figure S1.10: The expected co-segregation ratio, $m_2/(m_1+m_2)$ as function of the interaction probability π at three genomic distances $(1, 30, 100Mb)$. We used the parameters corresponding to our current experimental setting: slice thickness $h = 0.22\mu m$, resolution $30kb$, detection efficiency $\epsilon = 83.3\%$, u_i's estimated from chromosome 1.

1.6 Co-segregation of triplets of loci

GAM allows us to measure the co-segregation frequencies of any set of loci. The SLICE model can also exploit these co-segregation frequencies to extract multivalent interactions between 3 or more loci. In this section, we show how to estimate the probability of formation of a triplet of loci. In the first sub-section, we derive some relations between the interaction probabilities of triplets and the interaction probabilities of pairs of loci. The second sub-section will introduce the co-segregation probabilities of triplets of loci (the analogous of the probabilities t_i and u_i introduced above for pairs). Finally, we show the equations for the number of NPs in the third sub-section. For the sake of simplicity, we discuss in detail only the case of loci located on the same chromosome.

Relations between the interaction probabilities of pairs and triplets

Let's consider three loci A, B, C and name $\pi_{2,AB}, \pi_{2,BC}, \pi_{2,AC}$ the total probabilities that $A-B$, $B-C$ or $A-C$ are interacting (in pairs, triplets or otherwise). Under

the same assumptions used for pairs about the interactions between homologous loci (section 1.3), the two triplets of homologous loci can be treated as independent. When two loci are interacting, they can be either in a pair or a triplet, therefore:

$$\begin{aligned} \pi_{2,AB} &= \pi'_{2,AB} + \pi_3 \\ \pi_{2,BC} &= \pi'_{2,BC} + \pi_3 \\ \pi_{2,AC} &= \pi'_{2,AC} + \pi_3 \end{aligned} \tag{17}$$

where π'_2 is the probability that two loci are interacting in a pair, and π_3 the probability that the triplet $A-B-C$ is formed (Pi_3 in the main text). The pairwise interaction probabilities, $\pi_{2,AB}, \pi_{2,BC}, \pi_{2,AC}$, can be estimated by the method described in the previous section. We can use the equations (17) to find upper and lower bounds for π_3. Indeed, by imposing that all probabilities vary between 0 and 1, it can be obtained that:

$$max\left(\frac{\Sigma - 1}{2}, 0\right) \leq \pi_3 \leq min(\pi_{2,AB}, \pi_{2,BC}, \pi_{2,AC}) \tag{18}$$

where $\Sigma = \pi_{2,AB} + \pi_{2,BC} + \pi_{2,AC}$.

The co-segregation probabilities for triplet interactions

Given three loci A, B, C, there are five classes of conformations in which they can be found in the nucleus, and for each of them there is a set of co-segregation probabilities in a NP, i.e., probabilities that 0,1,2,3 loci are segregated in a randomly taken NP (see Fig. S1.11):

1. none of them is interacting
   probability π_0, co-segregation probabilities: $\omega_0, \omega_{1,(i)}, \omega_{2,(ij)}, \omega_3$;
2. A and B are interacting
   probability $\pi'_{2,AB}$, co-segregation probabilities: $\xi_0^{AB}, \xi_{1,(i)}^{AB}, \xi_{2,(ij)}^{AB}, \xi_3^{AB}$;
3. B and C are interacting
   probability $\pi'_{2,BC}$, co-segregation probabilities: $\xi_0^{BC}, \xi_{1,(i)}^{BC}, \xi_{2,(ij)}^{BC}, \xi_3^{BC}$;
4. A and C are interacting
   probability $\pi'_{2,AC}$, co-segregation probabilities: $\xi_0^{AC}, \xi_{1,(i)}^{AC}, \xi_{2,(ij)}^{AC}, \xi_3^{AC}$;
5. all three loci are interacting
   probability π_3, co-segregation probabilities: $\tau_0, \tau_1, \tau_2, \tau_3$.

where $i \in \{A, B, C\}$ and $(ij) \in \{AB, BC, AC\}$ are the loci that are segregated in the NP. Each of these sets of co-segregation probabilities satisfies a normalization condition. For instance:

$$\omega_0 + \sum_{(i)\in\{A,B,C\}} \omega_{1,(i)} + \sum_{(ij)\in\{AB,BC,AC\}} \omega_{2,(ij)} + \omega_3 = 1 \tag{19}$$

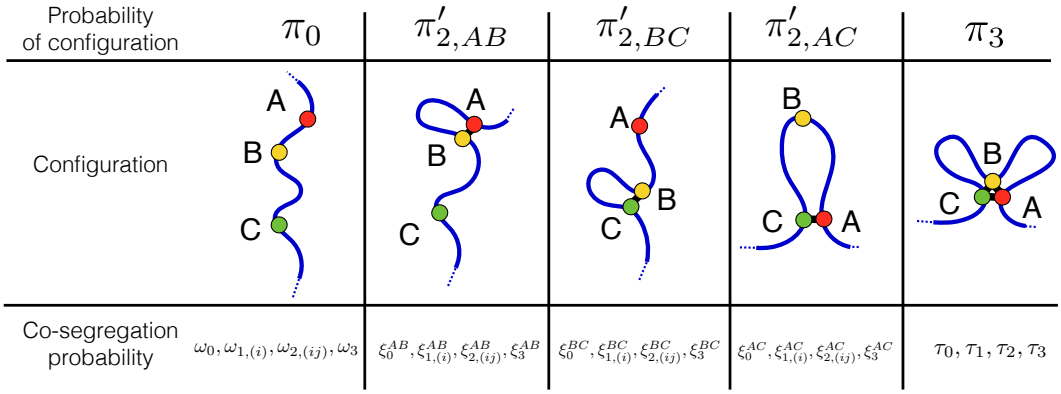

Figure S1.11: Interaction configurations of a triplet of loci with their co-segregation probabilities indicated at the bottom.

From the eq.s (17) and the normalization $\pi_3 + \pi'_{2,AB} + \pi'_{2,BC} + \pi'_{2,AC} + \pi_0 = 1$:

$$\begin{aligned} \pi'_{2,AB} &= \pi_{2,AB} - \pi_3 \\ \pi'_{2,BC} &= \pi_{2,BC} - \pi_3 \\ \pi'_{2,AC} &= \pi_{2,AC} - \pi_3 \\ \pi_0 &= 1 - \Sigma + 2\pi_3 \end{aligned} \tag{20}$$

The general equations for the co-segregation probabilities are very similar to those written for the u_i and t_i (i.e., the co-segregation probabilities for pairs of loci), and the model can include any available information on the spatial organization of loci by using the numerical/analytical approaches described above.
Consistently with the above calculations for pairs, we consider here interaction distances $\lesssim h$ for loci interacting in couples and triplets. In this case, the ξ_i and τ_i can be well approximated by using the v_i and u_i. For example, the configuration where only A and B are interacting is equivalent to a single pair of loci at a genomic distance equal to d_{BC} (see Fig. S1.11), and therefore the probability that, e.g., the C locus is in the same NP as the pair AB is $\xi_3^{AB} \sim u_2(d_{BC})$. By similar arguments, all the ξ co-segregation probabilities can be computed. As we saw for the t_i's, the τ's are: $\tau_0 \sim v_0, \quad \tau_1 \sim \tau_2 \sim 0, \quad \tau_3 \sim v_1$.

The ω's also can be written as functions of u_0 and v_0, by using the marginal probabilities that pairs of loci or a single locus are not segregated in the NP. For instance, the probability that neither A nor B are segregated, $u_0(d_{AB})$, can be written as $\omega_{1,C} + \omega_0$. Similarly, the probability that C is not found in the NP, which is v_0 by definition, is also equal to $\omega_{2,AB} + \omega_{1,A} + \omega_{1,B} + \omega_0$. Moreover, when the central locus B is in the NP, the two external loci A and C can be approximated as positioned independently from each other in the nucleus (yet still constrained by B), which implies that $\omega_3 = \frac{\omega_{2,AB}\omega_{2,BC}}{\omega_{1,B}}$.

All these relationships, along with the normalization conditions (eq. (19)), allow to compute all the probabilities ω as function of u_0 and v_0:

$$
\begin{aligned}
\omega_0 &= \frac{[u_0(d_{AB}) - v_0]\,[u_0(d_{BC}) - v_0] + u_0(d_{AC})(v_0 - 1)}{v_0 - 1} \qquad (21)\\
\omega_{1,A} &= -u_0(d_{AC}) + u_0(d_{BC}) + \frac{[u_0(d_{AB}) - v_0]\,[-u_0(d_{BC}) + v_0]}{v_0 - 1}\\
\omega_{1,B} &= \frac{[u_0(d_{AB}) - v_0]\,[-u_0(d_{BC}) + v_0]}{v_0 - 1}\\
\omega_{1,C} &= -u_0(d_{AC}) + u_0(d_{AB}) + \frac{[u_0(d_{BC}) - v_0]\,[-u_0(d_{AB}) + v_0]}{v_0 - 1}\\
\omega_{2,AB} &= \frac{[1 + u_0(d_{AB}) - 2v_0]\,[u_0(d_{BC}) - v_0]}{v_0 - 1}\\
\omega_{2,BC} &= \frac{[1 + u_0(d_{BC}) - 2v_0]\,[u_0(d_{AB}) - v_0]}{v_0 - 1}\\
\omega_{2,AC} &= 1 - 2u_0(d_{AB}) + u_0(d_{AC}) - 2u_0(d_{BC}) + \frac{[u_0(d_{AB}) - 1]\,[u_0(d_{BC}) - 1]}{v_0 - 1} + 2v_0\\
\omega_3 &= -\frac{[1 + u_0(d_{AB}) - 2v_0]\,[1 + u_0(d_{BC}) - 2v_0]}{v_0 - 1}
\end{aligned}
$$

where d_{AB}, d_{BC}, d_{AC} are the genomic distances of the loci in the triplet.

Probability of locus segregation in single NPs

As we did for pairs of loci, we can define the following co-segregation probabilities for each of the two homologous triplets:

$$
\begin{aligned}
c_0 &= \pi_0\,\omega_0 + \pi'_{2,AB}\,\xi_0^{AB} + \pi'_{2,BC}\,\xi_0^{BC} + \pi'_{2,AC}\,\xi_0^{AC} + \pi_3\,\tau_0 \qquad (22)\\
c_{1,A} &= \pi_0\,\omega_{1A} + \pi'_{2,AB}\,\xi_{1,A}^{AB} + \pi'_{2,BC}\,\xi_{1,A}^{BC} + \pi'_{2,AC}\,\xi_{1,A}^{AC} + \pi_3\,\tau_1/3\\
c_{1,B} &= \pi_0\,\omega_{1B} + \pi'_{2,AB}\,\xi_{1,B}^{AB} + \pi'_{2,BC}\,\xi_{1,B}^{BC} + \pi'_{2,AC}\,\xi_{1,B}^{AC} + \pi_3\,\tau_1/3\\
c_{1,C} &= \pi_0\,\omega_{1C} + \pi'_{2,AB}\,\xi_{1,C}^{AB} + \pi'_{2,BC}\,\xi_{1,C}^{BC} + \pi'_{2,AC}\,\xi_{1,C}^{AC} + \pi_3\,\tau_1/3\\
c_{2,AB} &= \pi_0\,\omega_{2,AB} + \pi'_{2,AB}\,\xi_{2,AB}^{AB} + \pi'_{2,BC}\,\xi_{2,AB}^{BC} + \pi'_{2,AC}\,\xi_{2,AB}^{AC} + \pi_3\,\tau_2/3\\
c_{2,BC} &= \pi_0\,\omega_{2,BC} + \pi'_{2,AB}\,\xi_{2,BC}^{AB} + \pi'_{2,BC}\,\xi_{2,BC}^{BC} + \pi'_{2,AC}\,\xi_{2,BC}^{AC} + \pi_3\,\tau_2/3\\
c_{2,AC} &= \pi_0\,\omega_{2,AC} + \pi'_{2,AB}\,\xi_{2,AC}^{AB} + \pi'_{2,BC}\,\xi_{2,AC}^{BC} + \pi'_{2,AC}\,\xi_{2,AC}^{AC} + \pi_3\,\tau_2/3\\
c_3 &= \pi_0\,\omega_3 + \pi'_{2,AB}\,\xi_3^{AB} + \pi'_{2,BC}\,\xi_3^{BC} + \pi'_{2,AC}\,\xi_3^{AC} + \pi_3\,\tau_3
\end{aligned}
$$

where c_0 and c_3 are the probabilities that none or all the loci of a triplet are co-segregated in the NP; $c_{1,i}$ the probability that only the locus i is in the NP; $c_{2,ij}$ the probability that the loci i and j are both in the NP. These probabilities obey the normalization condition $c_0 + \sum_{i\in\{A,B,C\}} c_{1,i} + \sum_{(ij)\in\{AB,BC,AC\}} c_{2,ij} + c_3 = 1$.
Under the assumption that the two homologous triplets are independent, the equations describing the NP content will be products of the c_i's, similar to those obtained for pairs of loci. For instance, if we call $N_{i,j,k}$ the expected number of NPs with a number i, j, k of A, B, C loci, the probability that a NP does not include any of the loci is $N_{0,0,0}/N = c_0^2$, while the probability that a single A locus is intercepted by the NP is $N_{1,0,0}/N = 2\,c_{1,A}\,c_0$.
Similarly to what we saw with pairs of loci, if we consider a detection efficiency ϵ, the expected number of NPs where no loci are detected is:

$$
\begin{aligned}
N^{\epsilon}_{0,0,0} &= N_{0,0,0} + (1-\epsilon)\,N_{[1,0,0]} + (1-\epsilon)^2\,(N_{[2,0,0]} + N_{[1,1,0]}) \qquad (23)\\
&\quad +(1-\epsilon)^3\,(N_{1,1,1} + N_{[1,2,0]})\\
&\quad +(1-\epsilon)^4\,(N_{[2,2,0]} + N_{[2,1,1]}) + (1-\epsilon)^5\,(N_{[1,2,2]}) + (1-\epsilon)^6\,N_{2,2,2}
\end{aligned}
$$

where the notation $[i, j, k]$ indicates the sum over all possible permutations of the indices (for instance $N_{[1,0,0]} = N_{1,0,0} + N_{0,1,0} + N_{0,0,1}$). By analogous calculations, the expected number of NPs with 0,1,2 or 3 loci (m_0, m_1, m_2, m_3 respectively) can be written.

Estimation of triplet interaction probabilities - Given the interaction probabilities of all pairs in the triplets $\pi_{2,AB}, \pi_{2,BC}, \pi_{2,AC}$ (that can be estimated by using the method outlined in the previous section) and the eq.s (20), the probability of triplet formation π_3 will be the only unknown parameter in the NP content equations. Along the lines of what we did with the pairs, we define a triplet co-segregation ratio as $m_3/(m_3 + m_2 + m_1)$ and fit this quantity to find π_3. The probabilities of all the other possible configurations can then be found by using eq.s (20).

1.7 Conclusion

In this Supplementary Note 1, we described the SLICE model and showed how it can estimate the probabilities of specific interactions genome-wide from GAM raw co-segregation data. Possible confounding factors, such as the effects of genomic distance, limited experimental detection efficiency and finite genomic resolution, can be fully taken into account. SLICE can be applied to unravel pairwise interactions as well as the complexity of multivalent interactions to extract specific interaction probabilities.
We discussed a first version of the model, yet refinements can be made by including further information where available, e.g., on chromatin folding and preferential nuclear positions of specific genomic loci.

References

[1] G. Parisi. *Statistical Field Theory.* Frontiers in Physics. Addison Wesley, 1988.

Supplementary Note 2

Contents

List of Figures

1 Detection of the most prominent interactions with the SLICE model

1.1 Introduction

The SLICE model can be used to identify loci interacting more than expected by random effects such as spurious contacts occurring by chance or by mere sequence proximity. We discuss first the case of detection of interactions between pairs of loci and, in a later section, triplets and multiplets.

1.2 A statistical test for detection of interacting pairs of loci

We consider in this section the case where GAM data is used to infer the interaction probability π of a pair of loci, A and B, through the SLICE model described in Supplementary Note 1.
To set the notation, we name m_0 the measured number of nuclear profiles (NPs) with neither locus, m_1 the number of NPs with only one locus (either A or B) and m_2 the number of NPs with both A and B. We define the co-segregation ratio as the ratio $m_2/(m_1+m_2)$, which depends on the interaction probability value, π (Pi in the main text; see Supplementary Note 1). We derive the probability distribution of the co-segregation ratio and test the null hypothesis $H_0 : \{\pi = 0\}$ against the alternative hypothesis $H_1 : \{\pi > 0\}$.

Probability distribution of the co-segregation ratio $m_2/(m_1+m_2)$

As described in Supplementary Note 1, the SLICE model allows us to compute the expected values of m_0, m_1, m_2, which are functions of the GAM tuneable parameters (the total number of NPs, m, and the slice thickness h), the detection efficiency ϵ and the genomic distance g between the loci A and B. The theoretical probability distribution of m_2, m_1, m_0 is multinomial with event probabilities $\frac{m_2}{m}, \frac{m_1}{m}, \frac{m_0}{m}$: $(m_2, m_1, m_0) \sim \mathcal{M}\left(\frac{m_2}{m}, \frac{m_1}{m}, \frac{m_0}{m}\right)$. As the co-segregation ratio $m_2/(m_1+m_2)$ is a stochastic variable dependent on m_0, m_1 and m_2, its probability distribution, $\mathcal{D}$, can be derived from $\mathcal{M}$. In Fig. S2.1, we show the probability distribution of the co-segregation ratio for different values of the interaction probability (for this figure and the following, we used parameters corresponding to the experimental data presented in the main part of the paper, as derived in Supplementary Note 1, specifically: slice thickness $h = 0.22\mu m$, number of NPs $m = 408$, nuclear radius $R = 4.5\mu m$, resolution $b = 30kb$, detection efficiency $\epsilon = 83.3\%$ and we took loci at a genomic distance of $1Mb$, unless otherwise specified).

By using the probability distribution of the co-segregation ratio at $\pi = 0$, we can test the null hypothesis $H_0 : \{\pi = 0\}$ against the alternative hypothesis $H_1 : \{\pi > 0\}$ for a given pair of loci. In order to do so, we computed t, the 95% quantile of the probability distribution of $\frac{m_2}{m_1+m_2}$ for $\pi = 0$. We then selected the most prominent interactions by taking all the pairs with a co-segregation ratio above t. The matrices of the interaction probabilities of pairs shown in the main text (Fig.s 3 and 4) include only the π's of these above threshold pairs of loci.

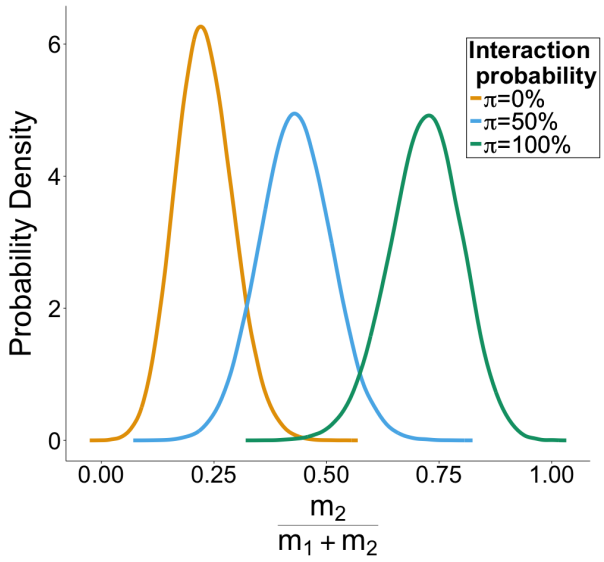

Figure S2.1: Probability density functions of the co-segregation ratio $m_2/(m_1 + m_2)$ for three values of the interaction probability π for loci at a $1Mb$ genomic distance.

Sensitivity of pairwise interaction detection

Having derived a statistical test for identifying prominent interactions, we then used the SLICE model to explore our power to detect interactions of a certain strength and genomic distance. Here we mean "power" (or "sensitivity") in the formal statistical sense, i.e., the probability that the test correctly rejects the null hypothesis H_0 when the alternative hypothesis H_1 is true. In other words, we will use SLICE to simulate the existence of a particular interaction with a probability $\pi > 0$ and we calculate the probability $P_R(\pi)$ of successfully rejecting the null hypothesis $H_0 : \{\pi = 0\}$ after measuring the co-segregation of the two interacting loci over m NPs:

$$P_R(\pi) = 1 - Cdf[\mathcal{D}(\pi), t] \tag{1}$$

where $Cdf[\mathcal{D}(\pi), t]$ is the cumulative distribution associated to $\mathcal{D}(\pi)$ calculated in t, which is the 95% quantile of $\mathcal{D}(\pi = 0)$. $P_R(\pi)$ is a measure of GAM sensitivity, as it returns the proportion of actual interactions that are identified by the method within the given confidence threshold. We calculated the behaviour of $P_R(\pi)$ for several combinations of the parameters. In Fig. S2.2 left panel, P_R is plotted as function of the number of NPs for 3 different values of the interaction probability. As expected, a higher sensitivity corresponds to a larger number of NPs. Moreover, for a fixed value of m, P_R is an increasing function of π (Fig. S2.2 left panel) and the genomic distance (Fig. S2.2 right panel). Indeed, loci that are close on the genome can have a high chance of being co-segregated in the same NP even if they do not interact, making it more challenging to identify a possible interaction.

In the specific case considered in our current study, where the dataset is composed of around 400 NPs, SLICE predicts that we are able to detect a fraction $\geq 85\%$ of all the pairwise interactions with $\pi \geq 50\%$ at or above $1Mb$.

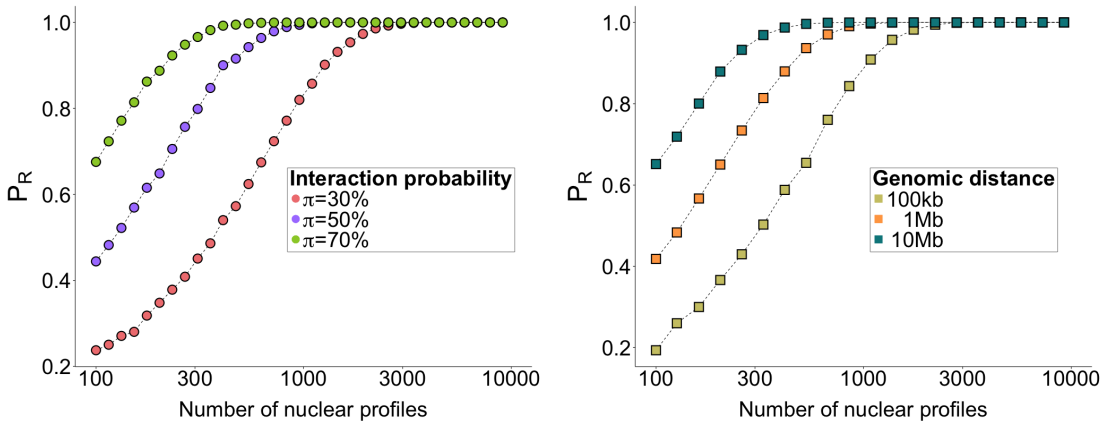

Figure S2.2: GAM sensitivity P_R as function of the number of NPs for different values of interaction probabilities π (**left panel**) and for different values of the genomic distances of the loci pair for $\pi = 50\%$ (**right panel**; the purple circles in the left panel and the orange squares in the right panel are equivalent). The parameters we used here correspond to the experimental data presented in the main part of the paper, as derived in Supplementary Note 1, specifically: slice thickness $h = 0.22\mu m$, number of NPs $m = 408$, nuclear radius $R = 4.5\mu m$, resolution $b = 30kb$, detection efficiency $\epsilon = 83.3\%$.

1.3 Detection of triplets

In this section, we analyse the sensitivity of GAM in the estimation of the interaction probability of triplets of loci. We will show how the sensitivity P_R changes with the probability of three loci specifically interacting in a simultaneous fashion (the triplet interaction probability or π_3; referred to as Pi_3 in the main text) and is affected by the existence of pairwise interactions.

Extending our statistical test to triplet interactions

We proceed here along the same lines followed before for the detection of pairwise interactions. We use the statistics of co-segregation ratio $m_3/(m_3 + m_2 + m_1)$ (see Supplementary Note 1) to estimate the probability P_R to reject the null hypothesis $H_0 : \{\pi_3 = 0\}$ against the alternative hypothesis $H_1 : \{\pi_3 > 0\}$.

Sensitivity of triplet interaction detection

The possible presence of pairwise interactions between the three considered loci influences P_R, therefore we discuss two distinct cases describing two opposite extreme situations: Case 1) the three loci do not engage in pairwise interactions; Case 2) the two loci are always directly interacting either in a pairwise interaction or in the triplet state (see Fig. S2.3 left panel). The investigation of such extreme cases allows us to identify an upper and a lower bound for P_R. In case the loci are engaged in pairwise interactions, the number of "false positives" for triplets (i.e., co-segregation in the same NP of three loci that are not forming a triplet) increases. Hence,

in Case 2, P_R tends to be comparatively lower than in the general case since at least one pair is interacting. Conversely, as in Case 1 interactions can only occur in triplets, P_R is expected to reach a maximum.

The right panel in Fig. S2.3 shows P_R in the two cases described above as function of π_3. As expected, P_R in Case 1 (blue squares) is always larger than in Case 2 (pink triangles).
The two curves identify a range in which the actual value of P_R lies. More precise estimation of sensitivity can be obtained in specific cases by leveraging the knowledge of pairwise interaction probabilities that reduces the number of unknown parameters (see Supplementary Note 1).

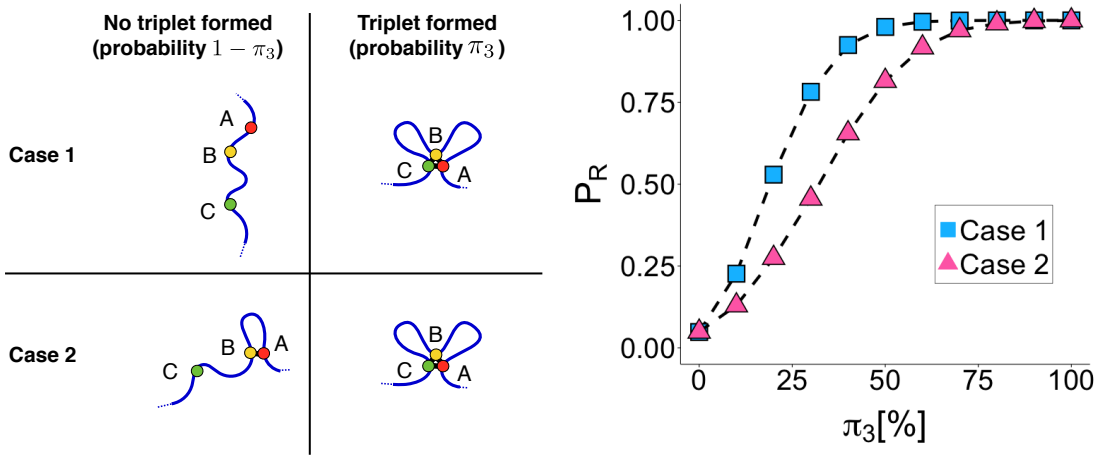

Figure S2.3: **Left Panel -** Schematic picture illustrating the two extreme situations considered to obtain a range in which sensitivity varies. **Right Panel -** GAM sensitivity P_R for detection of triplets as function of their interaction probability π_3 calculated for Case 1 (blue squares) and Case 2 (pink triangles) discussed in the text and illustrated in the left panel. Three loci equally spaced at a genomic distance of $1Mb$ were considered. The other parameters are the same as in Fig. S2.1.

1.4 Conclusion

The SLICE model can extract quantitative information about loci interactions from GAM raw data under different specific situations. As such it can be used to optimise the application of GAM to different cell types or organisms. In particular, the discussion of this document illustrates how the SLICE model can be used to find the most strongly interacting loci and estimate GAM sensitivity under different specific cases of interest. We showed in detail how the detection sensitivity can be calculated for both pairs and triplets of interacting loci as a function of different parameters, such as their genomic separation. With larger GAM datasets, SLICE can be extended to accommodate additional biological details, such as different nuclear or chromosomal shapes, association of loci with the nuclear periphery and other features of nuclear architecture. These further developments will increase the resolution of detection of

interacting loci across genomic scales and aid further optimization of the experimental protocol under different conditions.

Supplementary Note 3: Mathematical equations

3.1 Calculating sequencing depth saturation point:

We classified samples as saturated or unsaturated by calculating a coverage estimator C_n by analogy to species accumulation curves in ecology[53]. C_n estimates coverage at a given point in the saturation curve as:

$$C_n = 1 - \frac{f_1}{n}\left[\frac{(n-1)f_1}{(n-1)f_1 + 2f_2}\right]$$

where n is the total coverage over positive windows divided by the threshold coverage, f_1 is the number of positive windows with between 1 and 2 times the threshold coverage and f_2 is the number of positive windows with between 2 and 3 times the threshold coverage.

3.2 Estimating non-interacting distances for FISH probe pairs

The median distance measured between two non-interacting TADs (d_{ni}) was used to estimate the non-interacting distance for a TAD pair with a different genomic separation (d_{est}) using the following equation:

$$d_{est} = d_{ni}\left(\frac{g_{est}}{g_{ni}}\right)^{1/3}$$

where g_{ni} is the genomic separation between the measured TADs and g_{est} is the genomic separation between the TADs whose non-interacting distance is to be estimated. For derivation see Supplementary Note 1.4.

3.3 Linkage disequilibrium

The linkage D between two genomic windows A and B is defined as:

$$D = f_{AB} - f_A f_B$$

where f_A is the detection frequency of a given locus "A" (i.e. the number of nuclear profiles in which A is detected divided by the total number of nuclear profiles) and f_{AB} is the co-segregation of a pair of loci "A" and "B" (i.e. the number of nuclear profiles in which both A and B are detected divided by the total number of nuclear profiles).

3.4 Normalized linkage disequilibrium

Normalized linkage is calculated by dividing the linkage D by its theoretical maximum (D_{max}). D_{max} is defined as:

$$D_{\max=} \begin{cases} \min\langle f_a f_b, (1-f_a)(1-f_b)\rangle & when\ D < 0 \\ \min\langle f_b(1-f_a), f_a(1-f_b)\rangle & when\ D > 0 \end{cases}$$