



HAL
open science

Spectral-based detection of chromatin loops in multiplexed super-resolution FISH data

Michaël Liefsoens, Timothy Földes, Maria Barbi

► **To cite this version:**

Michaël Liefsoens, Timothy Földes, Maria Barbi. Spectral-based detection of chromatin loops in multiplexed super-resolution FISH data. *Nature Communications*, 2024, 15 (7670), <https://doi.org/10.1038/s41467-024-51650-w>. 10.1038/s41467-024-51650-w . hal-04595813v2

HAL Id: hal-04595813

<https://hal.science/hal-04595813v2>

Submitted on 13 Sep 2024

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

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

Spectral-based detection of chromatin loops in multiplexed super-resolution FISH data

Received: 11 December 2023

Accepted: 14 August 2024

Published online: 04 September 2024

 Check for updatesMichael Liefsoens^{1,2,3}✉, Timothy Földes^{2,4}✉ & Maria Barbi²

Involved in mitotic condensation, interaction of transcriptional regulatory elements and isolation of structural domains, loop formation has become a paradigm in the deciphering of chromatin architecture and its functional role. Despite the emergence of increasingly powerful genome visualization techniques, the high variability in cell populations and the randomness of conformations still make loop detection a challenge. We introduce an approach for determining the presence and frequency of loops in a collection of experimental conformations obtained by multiplexed super-resolution imaging. Based on a spectral approach, in conjunction with neural networks, this method offers a powerful tool to detect loops in large experimental data sets, both at the population and single-cell levels. The method's performance is confirmed on experimental FISH data where Hi-C and other loop detection results are available. The method is then applied to recently published experimental data, where it provides a detailed and statistically quantified description of the global architecture of the chromosomal region under study.

Loop formation is central to understanding chromatin architecture and its functional role. During mitosis, chromatin adopts a compact structure composed of loops, forming a rod-like configuration¹. SMC (structural maintenance of chromosome) proteins like condensin and cohesin play a pivotal role in organizing these loops². Recent research reveals that loop formation, mediated by proteins such as CCCTC-binding factor (CTCF) and cohesin, is also critical in interphase for gene regulation by facilitating interactions between distant enhancers and promoters in mammals^{3,4}, *Drosophila*⁵, and yeast⁶. The identification of chromatin loops has become central to unraveling gene regulation complexities and spatial genome organization⁷. Furthermore, cohesin-dependent loops are involved in the segmentation of interphase chromosomes into topologically associating domains (TADs), defined as sub-Mb self-interacting regions, often delimited by CTCF-binding. Depletion of CTCF disrupts both TAD loops and insulation of neighboring TADs⁸. Key questions arise regarding loop formation mechanisms, their prevalence, determinants of their position and sizes, and biological functions.

The loop extrusion mechanism⁹ performed by most SMC family proteins can explain loop formation. In particular, in interphase,

cohesins perform loop extrusion by binding to DNA as dimers, after which they act as motors, sliding in opposite directions and enlarging the loop by pulling along the chromatin fiber¹⁰. Looping by SMC complexes is observed in various cell types, including mammalian and bacterial cells¹¹. As insulator proteins, CTCF and cohesin regulate chromatin loop stability, probably as a 'dynamic complex' that frequently breaks and reforms throughout the cell cycle¹².

Visualizing the dynamics of loop extrusion in single living cells remains challenging. Fluorescence microscopy tracking two loop anchors has been explored^{7,13}, but requires prior anchor position knowledge and, thus, a strategy to identify the loop. In high-throughput genomic techniques like Hi-C¹⁴, stable loops manifest as discrete points in contact maps. Data analysis tools detecting DNA loops in contact maps, based on contact count enrichment or specific patterns, are available¹⁵⁻¹⁷.

Hi-C methods, however, lack the ability to reconstruct the polymer's spatial trajectory, only quantifying contact frequencies between monomers. These limitations might be overcome by combining fluorescence in situ hybridization (FISH) and super-resolution microscopy to achieve high-resolution imaging of individual genomic

¹Department of Mathematics, KU Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium. ²LPTMC, Sorbonne Université, CNRS, F-75005 Paris, France.

³Department of Physics and Astronomy, KU Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium. ⁴Institute for Medical Engineering and Science, MIT, Cambridge, MA 02139, USA. ✉e-mail: michael.liefsoens@kuleuven.be; foldes@mit.edu

regions (Hi-M¹⁸, ORCA¹⁹, OligoFISSEQ²⁰, MERFISH^{21,22}). These are high-throughput, high-resolution, microscopy-based technologies that, for the first time, allow the visualization of the spatial trajectory of the polymer by sequential labeling and imaging multiple loci along a single chromosome region, in fixed cells. This results in collections of configurations, sampled with a resolution up to 30 kb, which are, for the moment, difficult to fully exploit, especially to the aim of loop determination. The most frequent approaches are based on the reconstruction of distance maps, then interpreted as contact maps^{21,23,24}. However, this approach restricts the information to a level already obtainable with previous techniques.

Innovative methods are clearly needed to fully exploit this new data. In this study, we address the possibility of characterizing chromatin loops through an alternative approach based on the spectral representation of chain configurations, thereby leveraging the whole information of chain 3D spatial arrangement offered by sequential FISH methods.

Loops represent a distinctive aspect of chromosome folding, which must be considered within the broader context of the stochastic chromatin architecture. At a macroscopic level, heterochromatin is denser and transcriptionally repressed, while euchromatin is lighter and active, akin to polymers adopting globule and coil conformations, respectively^{25–31}. More specifically, super-resolution imaging of epigenetic domains in *Drosophila*²⁸ seems to indicate that their structure is compatible with the behavior of a self-attracting polymer close to the coil-globule transition³⁰. This transition, governed by the monomer-monomer interaction parameter ε , manifests through state-dependent scaling properties of the mean squared radius of gyration $\langle R_g^2 \rangle$ or, equivalently, the end-to-end distance $\langle R^2 \rangle$ as a function of monomer number N . Scaling laws, thus, enable the identification of the folding state. Nevertheless, this method necessitates the comparison of polymers with varying lengths, which may not always be feasible.

In prior work³¹, we developed an approach for analyzing fluorescent imaging data that overcomes this obstacle. Our method employs spectral analysis of configurations, focusing on long-distance features. Specifically, we apply a discrete cosine transform (DCT) to spatial coordinates, i.e. performing a Rouse mode decomposition, and, by taking the mean squared amplitudes of the DCT coefficients X_p , we construct a power spectral density (PSD) as $\langle X_p^2 \rangle = \langle \hat{\varepsilon}_p \rangle$. For low ε , in the coil state, PSDs follow the expected scaling $\langle \hat{\varepsilon}_p \rangle \propto p^{-(1+2\nu)}$, where the exponent $\nu \approx 0.588$ is the Flory exponent³²: this scaling law is indeed the spectral counterpart of Flory's scaling, $\langle R^2 \rangle \sim N^{2\nu}$. However, as ε increases above a critical value $\varepsilon_\theta(N)$, the strong attraction induces a second-order phase transition to curled-up conformations, called globules^{31,33,34}. Globules have a roughly spherical volume and uniform density, yielding the typical scaling $\langle R^2 \rangle \propto N^{2/3}$. Now, this state has a characteristic spectrum that becomes constant for the smallest p modes, making it possible to use the PSD to characterize the coil-globule state of a polymer by identifying its low p spectral scaling³¹.

These findings emphasize the significance of examining large-scale features, namely the first spectral modes, when probing overall polymer organization. They motivate further exploration to determine if this spectral approach can detect loops in chromosomal regions. In this work, we develop a spectral-based technique to detect loops in chromatin conformations acquired by multiplexed sequential FISH data. Additionally, we construct a neural network (NN) to segregate looped and non-looped conformations at the single-cell level.

Results

Power spectral density differentiates between looped and non-looped fBm-based polymer models

As a first step, we extend the PSD analysis to circular polymers, to examine the impact of looping on the spectrum. We employ a minimal, yet instructive, model of polymer configurations represented as 3D

correlated random walks γ_n , using fractional Brownian motion (fBm). The degree of correlation of the fBm is determined by the Hurst exponent H :

$$C_{\gamma\gamma}(i,j) = \frac{1}{2} \sigma_\gamma^2 (i^{2H} + j^{2H} - |i-j|^{2H}), \quad (1)$$

where $\sigma_\gamma^2 = \langle \gamma_1^2 \rangle$ is the variance of the first step. For our theoretical description, we consider polymer conformations with a uniform Hurst exponent H . Following ref. 35 and as detailed in Supplementary Note 1, we define a looped fBm as

$$\lambda_n = \gamma_n - \mathcal{B}_n^{(H)} \mathbf{R} : \quad (2)$$

here, $\mathbf{R} = \gamma_N - \gamma_1$ represents the fBm end-to-end vector and $\mathcal{B}_n^{(H)} = N^{-2H} C_{\gamma\gamma}(n, N)$ is the appropriate bridge function needed to connect the two ends of the fBm to construct an fBm loop.

For our simplified fBm model, the PSD of the looped chain can be obtained analytically. Thanks to the linearity of the DCT, the difference between looped and linear (i.e. non-looped) fBm is indeed simply the DCT of the bridge function $\mathcal{B}_n^{(H)} \mathbf{R}$. The symmetry properties of this function then ensure that (i) the even modes for looped fBm remain asymptotically unchanged compared to those of the corresponding non-looped; and (ii) the odd modes systematically decrease, with the extent of reduction diminishing as the mode number p increases. These results are proven in Supplementary Note 1. Additionally, we demonstrate that the latter property is a general consequence of the condition that the first and last monomer coincide, and thus applies to any looped conformation.

It's interesting to observe that the difference between looped and non-looped configurations primarily impacts the first modes, emphasizing the pivotal role of large-scale features in defining polymer structure. The behavior of the PSD for non-looped and looped fBm polymer configurations is visually depicted in Fig. 1a.

Log-spectral ratio $\Lambda(x)$ as an effective observable for loops in fBm signals

These spectral features offer a method to distinguish between looped and non-looped configurations. Consider indeed a statistical ensemble of 3D signal realizations \mathbf{x}_n . We introduce the log-spectral ratio $\Lambda(\mathbf{x}_n)$ for \mathbf{x}_n , defined as the logarithmic difference between the observed amplitude of the first mode and its amplitude predicted on the basis of a power-law extrapolation from the second and fourth modes. Some manipulation (detailed in Supplementary Note 2) yields the following expression for the log-spectral ratio:

$$\Lambda(\mathbf{x}_n) = \log \left(\frac{\langle \hat{\varepsilon}_2 \rangle^2}{\langle \hat{\varepsilon}_1 \rangle \langle \hat{\varepsilon}_4 \rangle} \right). \quad (3)$$

Fig. 1a provides an illustration of the definition of $\Lambda(\mathbf{x}_n)$ based on the PSD of linear and looped fBm related by Equation (2). Based on our fBm model, we can demonstrate that the log-spectral ratio for a non-looped random walk scales as N^{-2} for $N \rightarrow \infty$. In contrast, for a looped fBm, it converges to a finite limit of approximately 1.66, which clearly distinguishes the two configurations (see Supplementary Note 2).

For finite chains, we determine a discrimination metric by computing the absolute difference between the spectra of looped and non-looped random walks, and then normalizing it by the same difference for N at infinity (≈ 1.66). This results in a discriminability level ranging between 0 and 1. This quantity can be calculated analytically and converges remarkably fast: having $N > 6$ is sufficient to achieve a 90 percent discriminability level; $N > 20$ guarantees a 99 percent discriminability level.

To ensure the robustness and applicability of $\Lambda(\mathbf{x}_n)$ for signals with varying degrees of correlation, we calculate and display in Fig. 1b

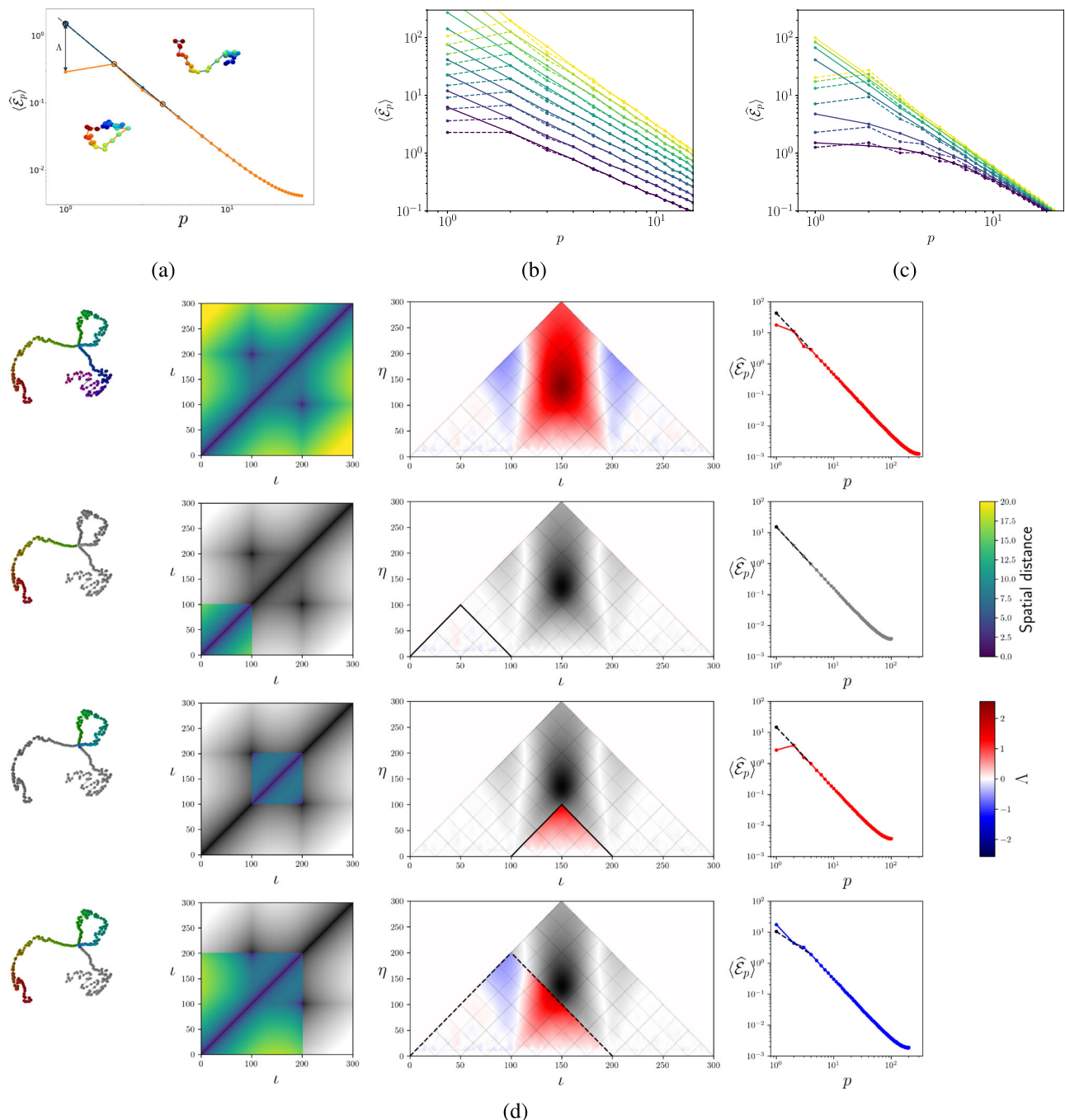


Fig. 1 | Principle of the log-spectral ratio and Λ -plot exemplified on a simulated single loop dataset. **a** Theoretical PSD for an $H = 0.5$ fBm γ_n (Equation (S8), blue curve) and the corresponding looped λ_n (Equation (S9), orange curve). Snapshots show one specific conformation before (upper) and after (lower) looping by means of Equation (2). The parameter $\Lambda(x)$ is the difference between the observed first mode (here for the looped conformations) and the expected first mode extrapolated from the second and fourth modes (see black dotted line and circles). **b** Estimated PSD for looped (dashed lines) and non-looped (solid lines) fBm signals with varying Hurst exponents ($H = 0.3, 0.35, \dots, 0.75$, from yellow to purple), each from samples of 2000 signals of length $N = 512$. **c** Estimated PSD of self-intersecting looped (dashed lines) and non-looped (solid lines) polymers for $\varepsilon = 0, 0.1, 0.2, 0.3, 0.4, 0.49$ (from yellow to purple). Each spectrum is obtained from

samples of 20000 equilibrium conformations for a polymer size $N = 512$, simulated by the on-lattice Monte Carlo approach. **d** Λ -plot for an ensemble of 2000 samples of a (random walk) polymer with $N = 300$ monomers, all containing an internal loop of size 100 in the middle (from index 100 to 200). Different rows focus on distinct sub-regions of the same polymer: whole polymer; first third; inner loop; first two thirds (including the loop). The first column displays a mean polymer configuration, (similar to the ShRec3D algorithm⁴⁵, see Methods). Sub-regions are colored accordingly. The second column shows the distance map of the polymer, where coloring focus on the selected region. The third column shows the Λ -plot for this polymer ensemble, with colored triangles highlighting regions corresponding to the selected sub-chain. The spectrum for the selected sub-region is shown in column 4.

the PSD of fBm signals with different Hurst exponents H . Clearly, the behavior theoretically described above and shown in Fig. 1a is always observed, regardless of the value of H .

The log-spectral ratio $\Lambda(x_n)$ proves, therefore, to be a robust observable that allows us to determine whether a polymer is in a linear or looped configuration, independently of its degree of correlation.

However, our aim is to investigate the presence of loops in chromosomes. This implies two additional issues, which will be addressed in the following sections. First, as discussed in the introduction, chromatin domains are expected to be near the coil-globule transition³⁰ and exhibit more or less collapsed, globule-like conformations, depending on epigenetics and transcription activity^{28,29}. Therefore, it is crucial to verify whether the log-spectral ratio remains reliable across the coil-globule transition. Second, chromatin loops can vary in size and position along the chromosome. Consequently, we need to adapt our approach to this more general case.

Λ detects loops across the simulated coil-globule transition

To validate the log-spectral ratio approach for identifying loop structures in polymeric molecules, regardless of their state along the coil-globule transition, we performed Monte Carlo simulations of a self-avoiding walk on a cubic lattice with an energy gain of $-\varepsilon$ (in units of $k_B T$) associated with nearest-neighbor “contact”, simulating monomer-monomer effective attraction. Linear and circular polymers were simulated separately, with reptation moves in the former case³⁰ and Crankshaft rotation, wedge flip, and kink-translocation techniques³⁶ in the latter, which enhanced simulation efficiency. For the circular polymer, the initial configuration was obtained by the *growing SAW*'s algorithm outlined in ref.³⁶.

Spectra were then estimated and compared for linear and looped polymers across a range of ε values from 0 to 0.5. As shown in Fig. 1c, the difference between the looped and non-looped configurations of the simulated polymers reproduces the expected behavior. Consequently, $\Lambda(\mathbf{x})$ can be defined and used in the same way as theoretically predicted.

Efficient identification of internal loops with the Λ -plot

We can introduce an internal loop within a random walk by extending the procedure outlined in Eq. (2) to an inner segment, which generalizes the definition of the bridge function. This enables us to generate sets of fBm-based polymer configurations $\{\mathbf{x}_n\}$ that incorporate one or more internal loops. These loops are defined by their positions ι and lengths η , meaning that monomers $\iota - \eta/2$ and $\iota + \eta/2$ are brought together.

We used these synthetic configurations with internal loops to develop and validate a loop-detection technique, named the Λ -plot and based on the computation of the log-spectral ratio. For each set of N -length signals $\{\mathbf{x}_n\}$, we consider all the sub-signals for any length η and any center ι , defined as $\{\mathbf{x}^{(\iota, \eta)}\} = (x_{\iota-\eta/2}, \dots, x_{\iota+\eta/2-1})$. We calculate the log-spectral ratio $\Lambda(\mathbf{x}^{(\iota, \eta)})$ for each of these sub-signals and represent the results on a color-scale on the plane (ι, η) . In Fig. 1d, we provide typical examples of the expected outcomes when identifying a single inner loop, and compare these results with corresponding distance maps and relevant spectra of sub-polymers.

As showcased in Fig. 1d, Λ -plots show distinct maxima indicating the presence of a loop. A careful inspection reveals that the ι coordinate of these maxima precisely corresponds to the midpoint of the loop, while the η coordinate is systematically slightly larger than the actual loop size. Thanks to our straightforward loop modeling, we can derive analytical results, as outlined in Supplementary Note 3.

For a given fBm signal containing an internal loop centered at ι_0 with a size of θ , the Λ -plot restricted to the $\iota = \iota_0$ line is indeed given by Equation (S10). This allows a precise determination of the loop position and size starting from the detected maxima (ι, η) . As mentioned earlier, we have $\iota_0 = \iota$, and from Equation (S10), the loop size θ is related to η by $\theta = \eta/\mu_0$, where $\mu_0 \approx 1.34767$ is a universal constant. Finally, in a typical experiment, only a fraction of the configurations will exhibit a specific loop. In Supplementary Note 4, we derive an expression for the log-spectral ratio Λ (for fixed $\iota = \iota_0$) for mixed populations, and show that the position of the maximum is independent of p , while its amplitude depends on it.

With these results, we can formulate a method for detecting loops in signals. Given a set of signals $\{\mathbf{x}_n\}$ containing internal loops, to estimate their position and size, follow these steps:

1. Calculate the estimated Λ -plot from the available samples;
2. Find the position $(\iota = \iota_0, \eta)$ of any maximum;
3. Divide η by $\mu_0 \approx 1.34767$ to find the approximate size of the corresponding loop;
4. The estimated loop falls then between monomers $\iota_0 - \eta/(2\mu_0)$ and $\iota_0 + \eta/(2\mu_0)$.

Note that, taken a point (ι, η) on the Λ -plot, the triangle of which it is the vertex corresponds to the Λ -plot of the region $[\eta - \iota/2, \eta + \iota/2]$, as shown by the multiple examples given in Fig. 1d.

Λ -plot loop-detection benchmark

To assess the performance of the Λ -plot method on experimental data, we first apply it to selected regions of mouse embryonic stem cells, for which Hi-C³⁷ and DNA seqFISH+³⁸ are available. Our comparison involves two aspects. First, we assess our results against the identification of Hi-C data loops using the HiCCUPS method³⁹, as conducted by Lee et al. Second, we compare our findings with loops detected using the SnapFISH method they introduced based on DNA seqFISH+ data. We analyzed this publicly available DNA seqFISH+ dataset where the authors selected one region from each chromosome, with region length ranging from 1.5 Mb to ~2 Mb. Among the 19 chromosomes, we then focus on eight of them where HiCCUPS detected loops are also considered in ref. 24, so to compare our findings with both HiCCUPS and snapFISH outputs. To identify potential loops, we automatically detect Λ -plot maxima, and then filter maxima above a certain threshold, followed by a width-based filtering (see Methods, Supplementary Note 6).

The detection filtering parameters, listed in Supplementary Table 1, can be adjusted to set the detection rate. Here, we have intentionally chosen them to be less restrictive. With this choice, we identify 22 out of 35 loops previously identified by HiCCUPS, indicating the detection of 7 additional loops compared to snapFISH on the same dataset, as detailed in Table 1. All the details and plots of these benchmark results are given in Supplementary Note 7. Our method, thanks to the fact that it analyzes the entire conformation between two loci, enables the detection of more loops, some of which would be impossible to detect by a simple evaluation of the end-to-end distance between loop ends. However, it's important to note that our method is prone to localization errors in regions with complex architectures, where, for example, several loops are nested. All in all, this demonstrates the usefulness of the Λ -plot analysis in identifying loops in FISH data, motivating us to continue our investigation.

Application to an independent multiplexed FISH dataset

We then apply our approach to MERFISH datasets by Bintu et al., acquired from HCT116 cells of human chromosome 21²¹. The examined genomic region hosts the *RUNX1* gene, known for encoding a highly conserved transcription factor featuring a complex regulatory mechanism⁴⁰. The region spans from 34.6 Mb to 37.1 Mb and is sampled at a genomic resolution of 30 kb. The 3D position of the 30 kb segment's center of mass is determined as the center of a Gaussian fit of the fluorescence spot imaged using diffraction-limited microscopy²¹.

Two variants were analyzed: an untreated, wild-type variant (WT), and an auxin-treated variant. Cohesin depletion, resulting from auxin treatment, leads to the removal of TADs at the population level³, without altering the occurrence of TAD-like structures in individual cells. However, it does disrupt the typical positioning of domain boundaries, often associated with CTCF-binding sites, which explains the loss of TADs at the population level²¹.

Table 1 | Comparison between Λ plot loop detections and HiCCUPS and snapFISH loop detections

Chrm. #	Total Λ loop detections	Matching Λ loops/HiCCUPS loops	Matching SnapFISH loops/HiCCUPS loops
2	11	2/4	1/4
3	5+(9)	3/4	4/4
6	12	6/7	4/7
9	5	2/4	0/4
10	7	2/4	0/4
11	3	0/1	1/1
14	4	3/7	1/7
16	8	4/4	4/4
Tot.	55+(9)	22/35	15/35

We compare the detections obtained by our approach to those discussed by Lee et al.²¹. For each of the considered chromosomes the total number of loops detected by the Λ -plot, and the fraction of HiCCUPS loops detected respectively by Λ -plot and snapFISH are given. Numbers in brackets correspond false positives due to the presence of missing lines in the dataset (see Supplementary Note 7).

In Fig. 2a, we present the Λ -plots we obtained for the two datasets, along with the identified maxima. For comparison, the median distance maps for the same datasets are shown in Fig. 2b, and the corresponding Hi-C maps³ in Fig. 2c. In the distance map for WT data, two large TADs are evident, along with numerous sub-TADs. However, identifying specific loops is challenging. In the auxin-treated variant, the (sub-)TADs are less pronounced, and a significant loss of structural detail is observed at the ensemble average level. No distinct loop can be identified from the distance map.

The log-spectral ratio detects numerous loops in the conformations, listed in Table 2. Specifically, we detected 14 loops in the WT (labeled 1–14) and 8 loops in the auxin-treated variant (labeled A1 to A8). In the wild-type cells, loop extremities align with regions rich in RAD21, a cohesin protein complex component, indicating they are likely due to cohesin-mediated loop extrusion halting at a physical barrier, likely CTCF (Figs. 2 and 3f).

In auxin-treated cells, some cohesins remain (as shown by the RAD21 peaks) and may account for certain loops, namely A2 and A3 (Figs. 2 and 3g). Other loops seem, however, unrelated to cohesin-mediated loop extrusion. We then explored whether these loops could be due to enhancer-promoter or promoter-promoter interactions, known to persist despite cohesin loss⁴¹ and be invisible in regular Hi-C maps⁴². Using a 30 kb threshold for colocalization with ChIP-seq epigenetic peaks associated with active enhancers and promoters, we found three potential enhancer-promoter loops and two potential promoter-promoter loops (Supplementary Fig. 5). The remaining loops had ends that colocalized with neither epigenetic marks (see Supplementary Notes 8 for more details). Yet, given the genomic resolution of the data and the uncertainty as to the location of the loops, we believe that the definite ascription of a biological origin for loops detected in the cohesin-depleted cells requires further validation against ultra-high-resolution 3C data, such as Micro-C, which, to the best of our knowledge, are not available for the specific cell line under study.

Examining the relative positions of loops, with the help of the graphical representation in Fig. 3f, g, is also interesting. First of all, we can notice a rather complex region, with many overlapping loops. This observation is in agreement with the complexity of the corresponding Hi-C map. Some loops overlap or are included within larger loops. Notably, some loops are closely adjacent to each other, such as loops A4 and A8 or loops 12 and 14, forming what appears to be the two “petals” of a flower-like shape. It’s interesting to note that ref. 5 suggests that flower-like looping is a fundamental mechanism in chromatin folding, leading to hubs or clusters of interacting cis-regulatory modules including enhancers and promoters. This shows that our

algorithm is capable of detecting such structures. However, it’s important to confirm that these loops are present simultaneously in unique configurations, rather than being a result of averaging across the entire dataset. To address this question, we need to determine in which specific samples a detected loop is present. This will be explored in the next subsection.

Using neural networks to segregate looped and non-looped configurations

To extend and enhance the log-spectral method, it is of particular interest to be able to sort, for each specific loop, the looped and non-looped conformations from the dataset, enabling further analysis of separate datasets. A naive attempt would be to simply rely on the distance between the two extremities of the identified loop. However, it is important to notice that the position of each probe is determined by fitting a Gaussian to the spot representing the entire 30 kb segment. While this method accurately determines the segment’s center of mass, the distance between the specific loci that interact at the loop ends is provided with a limited resolution, of the order of the standard deviation of the Gaussian spot (see Fig. 5). Consequently, while two loci may indeed be in direct physical contact, the center of mass of the 30 kb regions containing them might appear distant in the processed data. This resolution effect is illustrated in the original publication by Bintu et al. (see Fig. 1C in ref. 21), where 3D STORM images of two pairs of chromatin segments show different degrees of overlap but similar distances between their center positions, of the order of 300 nm. Moreover, this issue was encountered in a recent analysis of the dynamics of CTCF sites, which required sophisticated inference-based analysis to establish contacts⁷. We show in Supplementary Note 9 and particularly Supplementary Fig. 7, based on simulations, that the probe-probe distance is indeed, rather counter-intuitively, a very bad discriminant for looped conformations.

To overcome this difficulty, we developed a NN approach. Once a loop is identified by locating a maximum (l, η) in the Λ -plot of FISH data, we train a NN to recognize loops of size η/μ_0 in a chain of size η (see Equation (S11)). To this aim, we generate synthetic data, namely artificial looped and non-looped random walks at a scale of 1 kb, and then coarse-grained by taking the center of mass of consecutive 30 kb segments in order to mimic the experimental procedure. Then, the trained NN is applied to each individual experimental conformation to ascertain whether it contains a loop at the specified position. This, in turn, partitions the dataset into two distinct sub-populations: one with looped configurations and the other with non-looped configurations. In Methods, we give more details about the NN approach and the training data, while in Supplementary Note 9 we assess its efficiency and robustness by applying it on artificial configuration ensembles of varying parameters.

For illustration, in Fig. 4, we present a comparison of average distance maps and mean configuration reconstructions for the whole auxin-treated dataset and those derived from its sub-populations - one with loop A4 and the other without, as determined by our NN approach. Strikingly, in the looped sub-population distance map, a local minimum appears at the position of the predicted loop. Correspondingly, the Λ -plots show a strong enhancement of the A4 maximum in the looped population, where it overcomes all other maxima, while it is clearly suppressed in the plot for the non-looped population. While the loop extremities are closer in the looped population (see Fig. 4), the mean end-to-end distance remains clearly as large as the standard deviation of the Gaussian spot, and therefore, too large to be used as a discriminating parameter.

End-to-end distance distributions for looped and non-looped populations overlap

Thanks to NN, the distance between the two extremities of the loop (for the experimental datasets) can be investigated at the sub-

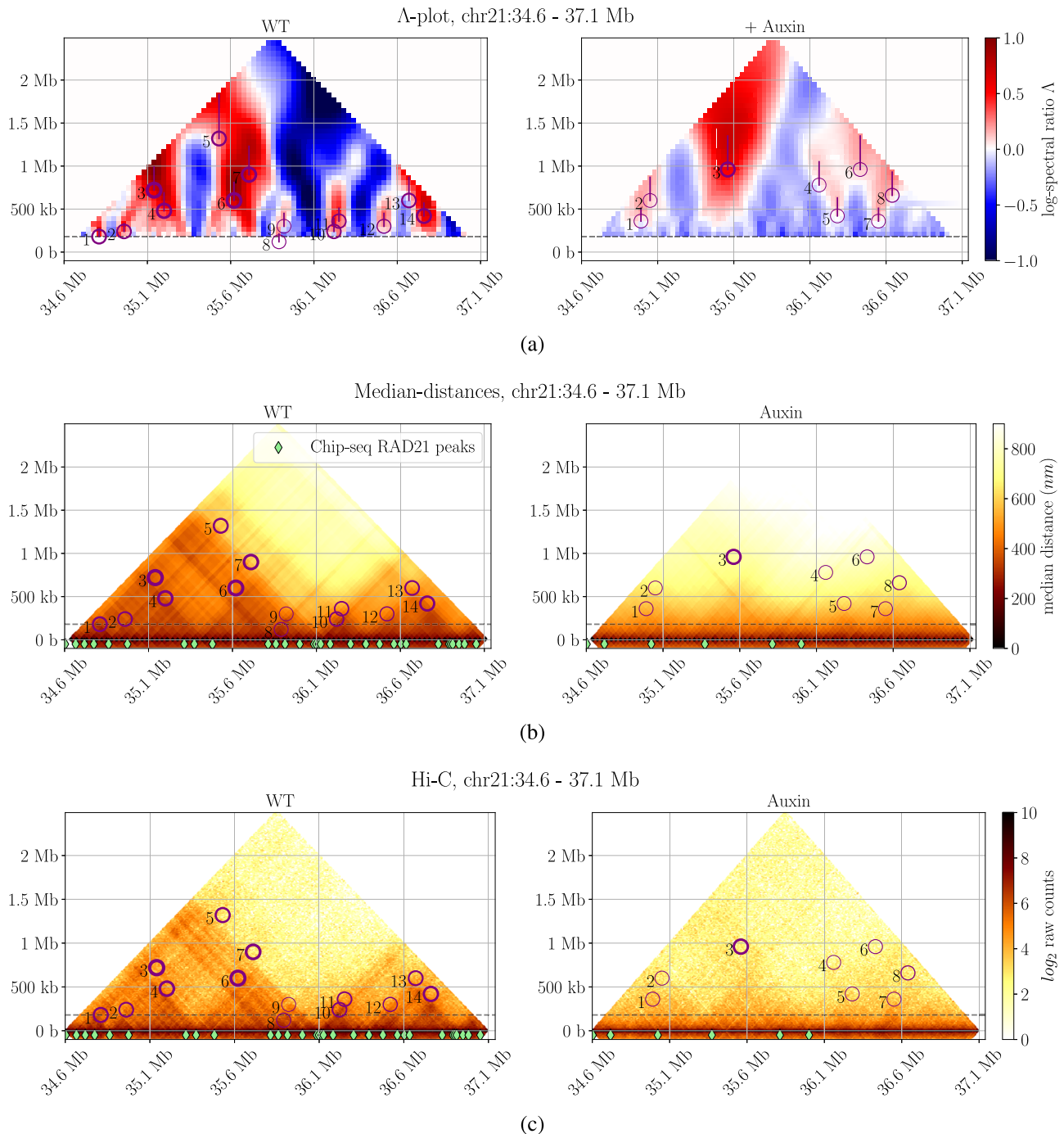


Fig. 2 | A-plot based loop-detection on an HCT116 human chromosome region, wildtype and auxin-treated. The A-plots (a), distance maps (b), and Hi-C maps³ (c) for the experimental data of Bintu et al. for both the wild-type variant (left) and an auxin-treated variant (right). **a** Purple lines link the positions of Δ maxima and detected loops. In each plot, purple circles indicate the location of loops detected

by the A-plot procedure. The symbol linewidth is proportional to the corresponding maximum Δ intensity. Each detected maximum is given a loop id. On the bottom distance and Hi-C plots, green diamonds represent Rad21 main ChIP-seq peaks³.

population level. The distributions of end-to-end distances are shown in Fig. 3a, b. As predicted, there is a significant overlap in the end-to-end distance distributions between the two sub-populations. Based on simulations, we were able to show that the overlap is largely due to the resolution effect previously discussed.

Despite their inefficiency at the single conformation level, the distributions in Fig. 3a, b reveal some variations between looped and non-looped populations for both the wild-type and auxin-treated cases, with a systematically shorter mean distance in looped populations with respect to non-looped ones.

More quantitatively, the mean end-to-end distance is on average 1.8 times larger in the non-looped than in the looped population for the WT dataset, and 2.6 larger for the auxin-treated dataset. When the non-looped conformations only are taken into account, all the mean end-to-end distances (within the loop region) fit well to a power-law with slightly varying exponents (see Table 3), as expected for linear polymers.

Note that the presence of two coexisting phases for certain positions of the same data recently suggested by Remini et al.⁴³ can be justified by this result, and indeed the distributions obtained here can

Table 2 | Specifics of the loops detected in wild-type and auxin-treated HCT116 regions

Loop id	1	2	3	4	5	6	7	
Loop range (probe #)	3–9	7–15	5–29	11–27	8–52	23–43	21–51	
Starting point (Mb)	34.69	34.81	34.75	34.93	34.84	35.29	35.23	
Ending point (Mb)	34.87	35.05	35.47	35.41	36.16	35.89	36.13	
Loop length (Mb)	0.18	0.24	0.72	0.48	1.32	0.60	0.90	
Frequency (%)	60 ± 10	39 ± 3	65 ± 4	67 ± 4	53 ± 4	61 ± 5	44 ± 3	
Loop id	8	9	10	11	12	13	14	
Loop range (probe #)	40–44	38–48	49–57	48–60	58–68	58–78	64–78	
Starting point (Mb)	35.80	35.74	36.07	36.04	36.34	36.34	36.52	
Ending point (Mb)	35.92	36.04	36.31	36.40	36.64	36.94	36.94	
Loop length (Mb)	0.12	0.30	0.24	0.36	0.30	0.60	0.42	
Frequency (%)	56 ± 8	35 ± 2	41 ± 3	49 ± 4	34 ± 2	58 ± 5	59 ± 4	
Loop id	A1	A2	A3	A4	A5	A6	A7	A8
Loop range (probe #)	6–18	4–24	15–47	38–64	48–62	44–76	58–70	56–78
Starting point (Mb)	34.78	34.72	35.05	35.74	36.04	35.92	36.34	36.28
Ending point (Mb)	35.14	35.32	36.01	36.52	36.46	36.88	36.70	36.94
Loop length (Mb)	0.36	0.60	0.96	0.78	0.42	0.96	0.36	0.66
Frequency (%)	41 ± 3	36 ± 3	32 ± 2	25 ± 2	33 ± 3	27 ± 2	39 ± 4	34 ± 3

Inferred loop ranges $l_0 \pm \eta/(2\mu_0)$, genomic coordinates, and estimated proportions of looped conformations (Frequency) as inferred by neural networks (see main text), for the loops identified in Fig. 2a in WT and auxin-treated samples.

be fitted, albeit not perfectly, by the theoretical expressions they proposed.

Correlation between nested loops detected

The detection of loops at the single conformation level also enables us to investigate the relationships between loops, specifically the joint probabilities of each loop pair. In Fig. 3c, d, we present the Pearson correlations for the loops detected at the single-cell level in the experimental data.

In both datasets, the majority of loops show positive correlations with all other loops. No significant anti-correlations appear. The former aligns with the finding by Bintu et al. that two loci A and B are more likely to be in contact if locus B is in contact with any other locus C located upstream of B²¹. We extend this observation to suggest that loci A and B have a higher contact probability if any other loci C and D are in contact. Intriguingly, this effect is stronger in the auxin-treated cells, which was also noted by Bintu et al. for 3-point cooperation. In this dataset, due to the absence of cohesin contacts, the proportion of contacts due to random chain fluctuations is expected to be proportionally higher than in WT. The higher pairwise correlation between loops in the absence of cohesin may, therefore, be due to a general physical feature of chromatin architecture, such as large-scale conformational changes, whose effect may be masked in WT due to the superposing effect of cohesin. This aligns with the view of chromatin being at the critical point of the coil-globule phase transition, where large-scale density fluctuations are expected, as already suggested^{27,30}.

Both in the WT and auxin-treated samples, some loop pairs appear more correlated than the average. Notably, (5, 7), (6, 7), (10, 11), (12, 13), (13, 14); and (A1, A2), (A4, A5), (A4, A6), (A5, A6), (A6, A8), (A7, A8), all exhibit correlations larger than 0.15.

In the auxin-treated case, at least one group of three mutually correlated loops, (A4, A5, A6), exists. The probability of observing at least one of them in the total dataset is here 54%, of which 6% of the coexistence of the three of them and 18% of the coexistence of two of them. However, their relative organization is hard to interpret.

More generally, it is challenging to find a common origin for the observed correlations. In our search for a common rationale, we have carried out various statistical analyses. These analyses revealed a consistent feature among all significantly correlated loop pairs: as

shown in Fig. 3e, for both datasets and for almost all correlated loop pairs, one loop is always contained within the other, and their extremities are either coincident or very close (≤ 2 probes). If we accept to say that loop A8 is included in A6 (within the 2 probes) and that the left ends of loop A5 and A6 are close enough (distance = 4 probes), this is indeed the case for all the WT and auxin-treated loops mentioned above, with the unique exception of (A4, A6).

Flower-like loop architecture

A particular architectural feature is showcased by loops 12, 13, and 14. Loops 12 and 14 seem to combine together to form loop 13, like two petals in a flower-like shape. Indeed, loops 12 and 13, and loops 13 and 14 are positively correlated, while loops 12 and 14 are not. More explicitly, when loop 13 is present (58% of the total), it is associated with 12 or 14 in about 48% of cases, but only 17% with both of them. Also, while the presence of loop 12 or 14 is up to 72% of the total, they only coexist in 21% of the cases, and only 4% of the time in the absence of loop 13. This suggests a more complex organization than a flower shape, for which the simultaneity of loops 12 and 13 will necessarily bring the three ends close to each other and, therefore, imply the presence of 14. In fact, loops 12 and 14 overlap in part in the middle of loop 13. One compelling explanation is the existence of 2 additional loops within loop 13 that haven't been detected. One of these forms a flower-like structure with loop 12 (designated as 12'), and the other with loop 14 (designated as 14'). Both pairs 12 and 12', and 14 and 14' cooperate, independently from one another, to form loop 13. This claim is coherent with the landscape of RAD21 peaks within loop 13 (see Fig. 3f), suggesting an intricate loop structure that can't be resolved at this resolution. Moreover, considering the previously mentioned resolution effect, it is clear that loop coordinates at the 30 kb scale are not necessarily a clear indication for common extremities.

Insights into chromatin local architecture

We can use the analogy of fBm to gain further insights into chromatin architecture features in TADs. Let's consider the two large TADs in the wild-type (chr21:346000000:361000000, region (1), and chr21:361000000:371000000, region (2)) and the entire region in the auxin-treated dataset (Region 3) as a potential third TAD. If we treat

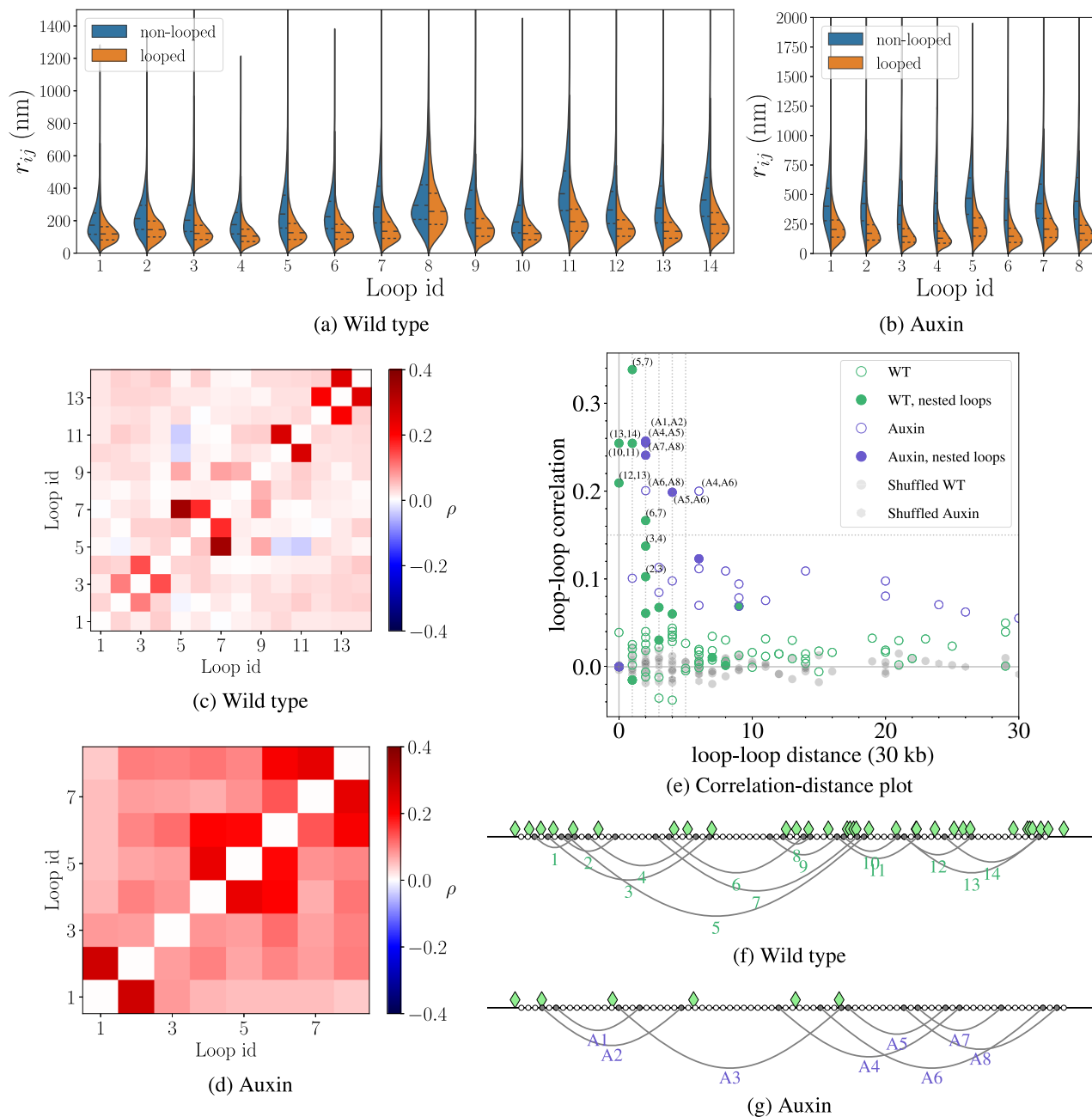


Fig. 3 | Single-cell-based loop statistics and correlations for the wild-type and auxin-treated HCT116 regions. a, b Distributions of end-to-end distances r_{ij} which measure the separation between the two extremities $i = \iota_0 - \eta / (2\mu_0)$ and $j = \iota_0 + \eta / (2\mu_0)$ for looped (orange) and non-looped (blue) configurations across all loops identified in the FISH data. **c, d** Pearson correlation (see Methods) between loops in both WT and auxin-treated datasets. **e** Pairwise correlation for the presence of loops plotted against the minimum distance between loop ends considering all

possible pairs of ends for both WT (green) and auxin-treated (purple), filled symbols correspond to nested loop pairs where one loop is contained in the other. In gray, as a reference, the result obtained by shuffling the presence/absence elements, loop by loop (see Methods). **f, g** For the two sets of loops, a graphical representation is given, with Rad21 main ChIP-seq peaks represented by green diamonds³. White circular points represent the 30 kb probes. Gray points correspond to loop anchors.

these regions as non-looped, we can fit the internal end-to-end distance $\langle R^2(s) \rangle$ with a power-law $f(s) = A(s/30 \text{ kb})^{2H}$, for each of these regions. The fitted exponents H are given in the first row of Table 3. It's worth noting that these three values are quite close to each other, and they are not significantly distant from $1/3$, which is the typical exponent expected for the crumpled globule model^{25,44}.

However, our previous results allows us to potentially determine the effects of the presence of loops on the exponent H . In particular, if we only select the population with two loops or fewer for the wild-type, and the population without any loops for the auxin-treated variant, we

find different exponents, as shown in the second row of Table 3. By excluding looped populations, the fitted exponents change notably, becoming closer to 0.4 rather than 0.3. This suggests that an incorrect interpretation of $R(s)$ behavior in experimental data might result from the influence of undetected loops in chromatin.

It's important to note that an exponent of about 0.4 in non-looped chromatin is consistent with the hypothesis that chromatin conformations can be described as polymers at the coil-globule phase transition^{30,31}, which indeed leads to a wider range of possible exponents.

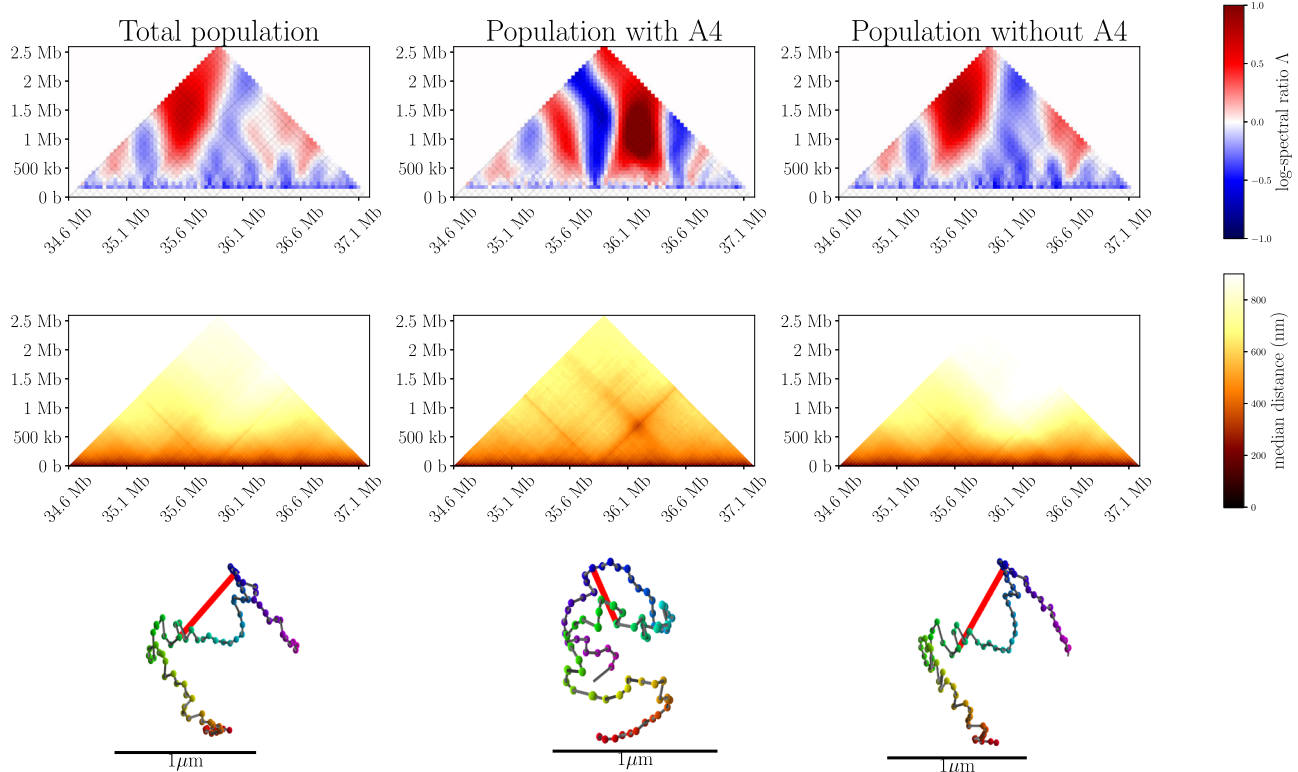


Fig. 4 | Typical example of output of neural network segregation of looped and non-looped populations. We focus on loop A4 (detected in 24% of the population, see Table 2). The first column shows the Δ -plot, distance map, and mean configuration (similar to the ShRec3D algorithm⁴⁵, see Methods) for all data. The second

and third columns show Δ -plot, distance map, and mean configuration for measurements that the NN recognized as containing or lacking loop A4, respectively. The end-to-end distance (length of the red segment) in the mean configuration is 640 nm, 547 nm, and 710 nm, respectively.

Table 3 | Comparison of scaling laws for mixed configurations compared to mainly non-looped sub-populations

Region	(1)	(2)	(3)
All conformations	0.284 ± 0.014	0.314 ± 0.014	0.300 ± 0.003
Low loop content	0.41 ± 0.03	0.436 ± 0.011	0.372 ± 0.004

Values of the exponent H obtained by fitting $\langle R^2(s) \rangle = A(s/30 \text{ kb})^{2H}$ for the two large TADs in the wild-type, region (1) and (2), and the entire region in the auxin-treated dataset (3) (see also main text) while considering all the conformations (upper row) or only conformations with two loops or less (for the WT, 171 samples) or without any loops (for the auxin-treated variant, 834 samples) (lower row).

Discussion

Despite their high accuracy in determining the three-dimensional structure of chromosomes, multiplexed FISH experiments can only visualize the center of mass of probe segments. Gauging the proximity of loci contained within these segments is therefore challenging. As our study on synthetic data clearly confirms, FISH-based distance maps cannot detect the presence of all loops in experimental multiplexed super-resolution FISH data, due to this limited resolution. Yet, FISH experimental data offers more comprehensive information than distance maps by encompassing the complete 3D configuration, information that has been overlooked until now.

Using all the information on the 3D configuration, the Δ -plot approach introduced in this work provides a reliable and fast method to detect loops in synthetic as well as experimental multiplexed FISH data, irrelevant of size and position, and sensitive to small looped populations.

Based on the analysis of the conformation's low-frequency modes to characterize the presence of loops in a chain, the method takes great advantage of the large-scale characteristics of polymers and is, therefore, more robust than indeterminate small-scale properties.

The presence of loops is confirmed via a NN approach, which further results in the opportunity to classify chromatin conformations as looped or non-looped in each cell by assessing the presence of specific loops in each measurement. We have demonstrated the feasibility, speed, and reliability of this process by mimicking both the 3D chromatin arrangement and the limited resolution effect in simulated configurations used to train the NN. A significant portion of the success of this NN approach is attributed to the initial guidance provided by the Δ -plot, which also provides a physically interpretable and mathematically sound basis for the detection pipeline. Thanks to the ease with which we can generate artificial training data based on an fBm model, we can furthermore avoid wasting valuable measurements.

The proposed detection pipeline was tuned to have a fairly high sensitivity, as the main purpose is to provide a tool with which potential loops can be detected. These potential loops may then be studied further by examining additional information coming from other biological signals. The higher sensitivity of the detection pipeline enabled the detection of several loops either in DNA seqFISH+ mouse embryonic stem cells³⁸ and human chromosome 21 HCT116 cells²¹. For the former, our approach outperforms distance-based approaches on loops already detected through Hi-C maps, and detects additional loops that need further confirmation.

The latter includes two datasets, one wild-type and one cohesin-depleted. We detected loops in both datasets with comparable occurrence frequencies, in different locations. Although determining the biological determinants at the origin of these loops is beyond the scope of this work, we have been able to establish strong correlations with Rad21, characteristic of the presence of cohesin, in the wild-type, while more varied determinants, including enhancer-promoter contacts, are suggested for the cohesin-depleted system.

Another intriguing question involves the potential existence of clusters of adjacent loops, resulting in flower-like structures reminiscent of cis-regulatory module hubs⁵. The ability to examine loops at the single-cell level now allows for a quantitative investigation of correlations between different loops for the first time. In the datasets studied here, one candidate flower-like structure was studied. The correlations point toward a more complex behavior than a simple coexistence of two “petal” loops, this being potentially related to the observation that these loops are not adjacent but partially overlap in the center of the larger loop. Moreover, we evidenced the frequent presence of correlation between loops that are nested one in the other and sharing one end. This rather unexpected finding surely deserves to be further investigated on alternative datasets.

The method introduced here broadens data processing possibilities and strengthens the foundation for advancing chromatin’s theoretical understanding through precise and comprehensive experimental data analysis. We discovered that the corresponding critical exponents of 1/3, frequently encountered in experimental data, may result from averaging looped and non-looped configurations within a dataset. This effect may have remained unnoticed due to the need for a prior looped conformations segregation.

Multiple future research endeavors are possible with the developed method, and we highlight some topics that warrant further exploration. The study of loop extrusion in interphase by cohesin has been a hot topic since its (re)-discovery³. Notwithstanding, a surprising amount of questions remain open. For example, the interaction of cohesin with other cohesins, physical barriers and regulatory proteins (WAPL and NIPBL) remains poorly understood, as does its modality of translocation (symmetric or one-sided). In this context, the detection of cohesin-CTCF loops and the ability to classify conformations with respect to the presence of a specific loop at the single-cell level could potentially contribute to elucidating certain issues. Similarly, the identification of potential enhancer-promoter and promoter-promoter loops at the single-cell level could forward the understanding of how regulatory elements interact to orchestrate gene expression regulation.

In addition, our confidence in the versatility of the spectral-based technique developed in this study encourages its application to investigate a broader range of phenomena. For instance, the method can be adapted for detecting plectonemes in supercoiled DNA or for identifying density variations across the genome or in spatial arrangements, such as alternating coils and globules, or alternating A and B compartments. These structures are predominantly characterized by their large-scale behaviors, where low-mode spectral features prove to be particularly suitable for in-depth investigation. Preliminary investigations of data from ref. 22, which is on a much larger scale than the one considered here, seems to indicate indeed that the same analysis can readily identify AB-compartments and their corresponding boundaries. These findings are consistent with the conclusions drawn in the original paper. Additionally, loops were also detected and warrant further investigation in future research.

Method

We confirm that our research complies with all relevant ethical regulations.

Mean polymer configuration: ShRec3D-like approach

The ShRec3D algorithm⁴⁵ is aimed to reconstruct spatial distances and three-dimensional genome structures from observed contacts between genomic loci. In the data from multiplexed super-resolution, the single configurations are known. However, we follow a simplified approach in the spirit of the ShRec3D algorithm in order to have a representation of the average features of an entire dataset. To this aim, we calculate individual distance maps for each configuration, then

average over all these maps. This average map will be invariant to translations and rotations of each individual polymer. Moreover, the averaged map will still be a distance map (i.e. be symmetric and satisfy the triangle inequalities).

Maxima detection

The detection of maxima on the Λ -plots is automated as follows. First, the `peak_local_maxima` function from the python library Scipy is applied to identify all local maxima greater than a predetermined threshold, that was fixed to `threshold=0.1`, see Supplementary Note 5. Due to the correlations between overlapping sub-chains, loop maxima are expected to decay progressively. Consequently, loop-related maxima appear as blobs rather than isolated maxima in the Λ -plot. We therefore gauge the width of each detected maxima by examining the Λ value of its nearest neighbors and only retain the maxima based on the number of nearest neighbors that exceed a certain threshold (see Supplementary Table 1 and Supplementary Fig. 4). In particular, we required that at most two nearest neighbors of the maximum are lower than half the detection threshold, i.e. `nn_threshold=0.05`. More details are given in Supplementary Note 6.

NN specifics

The NN used in this work has five layers. The input layer takes the 3D-coordinates of the N probes and reduces this to N nodes. Prior to input, the polymer data is normalized so that its center of mass is in the origin and the radius of gyration becomes one. The following layers have thirty, five, and two nodes, respectively. Finally, the output layer has one node. The network has the ReLU-activation function on hidden layers, and a sigmoid-activation function on the output layer, see for example ref. 46. We use the binary cross-entropy as loss function.

Each time the position of a maximum (t, η) of the Λ -plot is found, the NN is trained to segregate random walks with and without an internal loop for loop sizes lying uniformly in the range

$$\frac{\eta}{\mu_0} \pm \max\left(0.1 \frac{\eta}{\mu_0}, 1\right), \quad (4)$$

where μ_0 is given in Equation (S11). This range is arbitrarily chosen as to give enough variability in the training data so that the NN can more easily generalize to unseen data. Training is done using a batch size of 20; 200 epochs; and with the loss on the validation data as overfitting check (with callback to best parameters after loss of validation data starts to increase).

We train our NN on looped random walks artificially generated following the procedure detailed in the next section—with balanced training data—with 25000 learning samples, 7500 validation samples, and 5000 test samples for both the looped (Equation (S1)) and non-looped random walk. The validation samples are used to monitor and prevent overfitting, and the test samples give an estimate of the intrinsic accuracy of the NN, following the procedure delineated in Supplementary Note 9.

We want to remark that it is an enormous benefit that we can use synthetic training data, as we do not need to waste any experimental data on training the networks.

Reproduction of the limited spatial resolution

As discussed, the experimental resolution of 30 kb means that consecutive strands of DNA of 30 kb are lit up, and then, via Gaussian fitting, the center of mass of each strand is determined with high accuracy. However, there is no guarantee that this center of mass is always part of the DNA strand, and, more importantly, even if two DNA loci come together to form a loop, the centers of mass of the larger segments may be quite far apart. We reproduced this effect explicitly

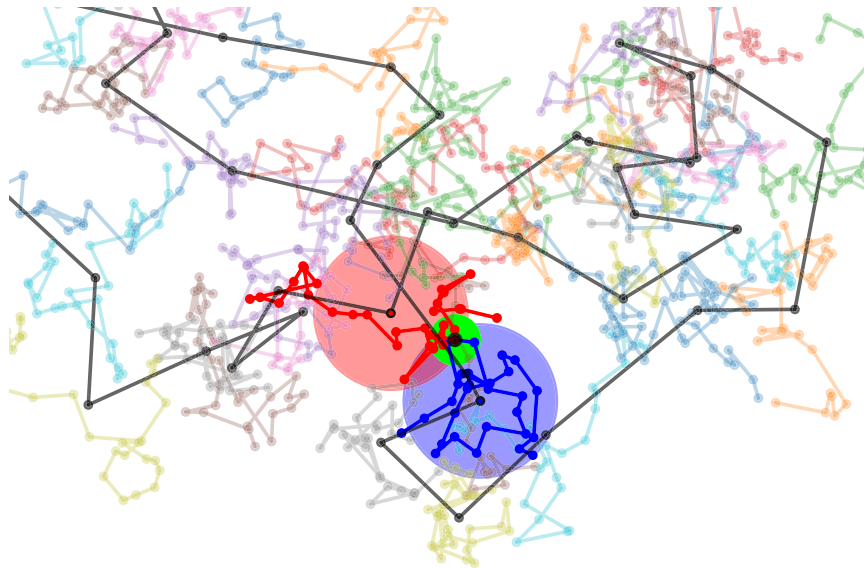


Fig. 5 | Simulated effect of the limited imaging resolution on loop-detection. Sample random walk of length 1500 (including a loop from index 500 to 1000) projected to xy -plane, in colors, and coarse-grained walk obtained by taking centers of mass of consecutive sub-chains of length 30, in black. Each colored region is represented by one black point. The larger black dot highlighted in green

represents the extremities of the loop in the high-resolution random walk, and the two points with blue- and red-shaded regions show the loop anchors for the coarse-grained walk. Moreover, the radius of the shaded disks is the radius of the gyration of the sub-chain.

in Fig. 5, by simulating a polymer chain at the resolution of a Kuhn length (of the order of 1 kb), then considering 30 kb long sub-chains. The centers of mass of consecutive sub-chains are then calculated. It is clear from the figure that, while the anchor points of the loops in the 1 kb resolution walk are, as expected, close together, the corresponding positions in the coarse-grained walk are quite far apart. Hence, it is reasonable to think that naive techniques—such as looking at the distance between two FISH markers—can miss these loops. Remarkably, the log-spectral ratio Λ singles out the low-frequency modes and thus take more advantage of the large-scale polymer features.

Accordingly, the NN procedure has to take these local errors into account, and has to learn to look at the large-scale features. Therefore, following the previous idea, we chose to generate (looped and non-looped) random walk data at a scale of 1 kb, coarse-grain them by taking the center of mass of consecutive 30 kb segments, and then train the NN on the resulting data. When tested on synthetic data, the NN obtains remarkable accuracy in finding the underlying looped configuration, indicating that it only takes into account the large-scale features of being looped or not. This is discussed in more detail in Supplementary Note 9, where we also study the effect of this center of mass sampling on the measured end-to-end distance distributions.

Time complexity

Creating the Λ -plot requires studying all the sub-polymers at all possible positions, which can be quite time extensive at first glance. Luckily, due to the application of the Fast Fourier Transform to compute the Discrete Cosine Transform and by using the fast vectorizing abilities of numerical software like NumPy, this is actually not a problem. Without performing a detailed analysis—since the timing results were satisfactory—we can report that the creation of the two experimental Λ -plots of Fig. 2a only took about 30 seconds, which is for around 20,000 configurations of 83 3D-points each. This timing is for a MacBook pro with apple M1 MAX chip and 32 GB RAM. The training and application of each NN to each separate loop takes about 25 minutes in total (training one after the other).

Pearson correlation and randomized datasets

The Pearson correlation used in Fig. 3c, d is classically defined as the covariance of the two variables divided by the product of their standard deviations:

$$\rho_{XY} = \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sqrt{(\mathbb{E}[X^2] - (\mathbb{E}[X])^2)(\mathbb{E}[Y^2] - (\mathbb{E}[Y])^2)}} \quad (5)$$

In Fig. 3e, the Pearson correlation is calculated using Equation (5), where X and Y are the binary vectors of length N indicating the absence (0) or presence (1) of two different loops across the N configurations. To get a statistical estimate of the fluctuations, X and Y are then shuffled to randomize the loop presence, while keeping its overall presence probability unchanged. The same procedure is repeated, separately, for the WT and auxin-treated datasets.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The main data used is publicly available at <https://github.com/BogdanBintu/ChromatinImaging/tree/master> and has been collected in the context of ref. 21. The data used for the benchmarking process is publicly available at <https://zenodo.org/records/3735329> and has been collected in the context of ref. 38. Finally, additional. Source data are provided with this paper.

Code availability

The code to make the Λ -plots and apply the NNs is publicly available at <https://github.com/michael-liefsoens/LambdaPlot> under the GNU GPLv3 license.

References

- Paulson, J. R., Hudson, D. F., Cisneros-Soberanis, F. & Earnshaw, W. C. Mitotic chromosomes. *Semin. Cell Dev. Biol.* **117**, 7–29 (2021).

2. Swedlow, J. R. & Hirano, T. The making of the mitotic chromosome: modern insights into classical questions. *Mol. Cell* **11**, 557–569 (2003).
3. Rao, S. S. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 (2017).
4. Karpinska, M. A. & Oudelaar, A. M. The role of loop extrusion in enhancer-mediated gene activation. *Curr. Opin. Genet. Dev.* **79**, 102022 (2023).
5. Espinola, S. M. et al. Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early drosophila development. *Nat. Genet.* **53**, 477–486 (2021).
6. Costantino, L., Hsieh, T.-H. S., Lamothe, R., Darzacq, X. & Koshland, D. Cohesin residency determines chromatin loop patterns. *eLife* **9**, e59889 (2020).
7. Gabriele, M. et al. Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* **376**, 496–501 (2022).
8. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944.e22 (2017).
9. Alipour, E. & Marko, J. F. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res.* **40**, 11202–11212 (2012).
10. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
11. Banigan, E. J., van den Berg, A. A., Brandão, H. B., Marko, J. F. & Mirny, L. A. Chromosome organization by one-sided and two-sided loop extrusion. *eLife* **9**, e53558 (2020).
12. Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife* **6**, e25776 (2017).
13. Mach, P. et al. Cohesin and CTCF control the dynamics of chromosome folding. *Nat. Genet.* **54**, 1907–1918 (2022).
14. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
15. Roayaei Ardakany, A., Gezer, H. T., Lonardi, S. & Ay, F. Mustache: multi-scale detection of chromatin loops from Hi-C and micro-C maps using scale-space representation. *Genome Biol.* **21**, 1–17 (2020).
16. Salameh, T. J. et al. A supervised learning framework for chromatin loop detection in genome-wide contact maps. *Nat. Commun.* **11**, 3428 (2020).
17. Matthey-Doret, C. et al. Computer vision for pattern detection in chromosome contact maps. *Nat. Commun.* **11**, 5795 (2020).
18. Cardozo Gizzi, A. M. et al. Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms. *Mol. Cell* **74**, 212–222.e5 (2019).
19. Mateo, L. et al. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).
20. Nguyen, H. Q. et al. 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. *Nat. Methods* **17**, 822–832 (2020).
21. Bintu, B. et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783 (2018).
22. Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B. & Zhuang, X. Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell* **182**, 1641–1659.e26 (2020).
23. Huang, H. et al. Ctfc mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nat. Genet.* **53**, 1064–1074 (2021).
24. Lee, L. et al. SnapFISH: a computational pipeline to identify chromatin loops from multiplexed DNA FISH data. *Nat. Commun.* **14**, 4873 (2023).
25. Mirny, L. A. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.* **19**, 37–51 (2011).
26. Grosberg, A. Y. How two meters of DNA fit into a cell nucleus: polymer models with topological constraints and experimental data. *Polym. Sci. Ser. C.* **54**, 1–10 (2012).
27. Barbieri, M. et al. Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl. Acad. Sci.* **109**, 16173–16178 (2012).
28. Boettiger, A. N. et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).
29. Szabo, Q. et al. TADs are 3D structural units of higher-order chromosome organization in *Drosophila*. *Sci. Adv.* **4**, eaar8082 (2018).
30. Lesage, A., Dahirel, V., Victor, J.-M. & Barbi, M. Polymer coil-globule phase transition is a universal folding principle of *Drosophila* epigenetic domains. *Epigenet. Chromatin* **12**, 28 (2019).
31. Földes, T., Lesage, A. & Barbi, M. Assessing the polymer coil-globule state from the very first spectral modes. *Phys. Rev. Lett.* **127**, 277801 (2021).
32. Grosberg, A. Y. & Khokhlov, A. R. Statistical Physics of Macromolecules. In: AIP series in polymers and complex materials (AIP Press, 1994).
33. Grassberger, P. & Hegger, R. Simulations of three-dimensional θ polymers. *J. Chem. Phys.* **102**, 6881–6899 (1995).
34. Vogel, T., Bachmann, M. & Janke, W. Freezing and collapse of flexible polymers on regular lattices in three dimensions. *Phys. Rev. E* **76**, 061803 (2007).
35. Gasbarra, D., Sottinen, T., Valkeila, E. Gaussian bridges. In: *Stochastic Analysis and Applications: The Abel Symposium 2005*, pp. 361–382 (Springer, Berlin, Heidelberg, 2007).
36. Vettorel, T., Reigh, S. Y., Yoon, D. Y. & Kremer, K. Monte-carlo method for simulations of ring polymers in the melt. *Macromol. Rapid Commun.* **30**, 345–351 (2009).
37. Bonev, B. & Cavalli, G. Organization and function of the 3d genome. *Nat. Rev. Genet.* **17**, 661–678 (2016).
38. Takei, Y. et al. Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).
39. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
40. Owens, D. D. G. et al. Dynamic *RunX1* chromatin boundaries affect gene expression in hematopoietic development. *Nat. Commun.* **13**, 773 (2022).
41. Hsieh, T.-H. S. et al. Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat. Genet.* **54**, 1919–1932 (2022).
42. Hsieh, T.-H. S. et al. Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell* **78**, 539–553.e8 (2020).
43. Remini, L. et al. Chromatin structure from high resolution microscopy: Scaling laws and microphase separation. *Phys. Rev. E* **109**, 024408 (2024).
44. Grosberg, A. Y., Nechaev, S. & Shakhnovich, E. The role of topological constraints in the kinetics of collapse of macromolecules. *J. Phys.* **49**, 2095–2100 (1988).
45. Lesne, A., Riposo, J., Roger, P., Cournac, A. & Mozziconacci, J. 3d genome reconstruction from chromosomal contacts. *Nat. Methods* **11**, 1141–1143 (2014).
46. Aggarwal, C. C. Neural networks and deep learning: a textbook <http://link.springer.com/10.1007/978-3-319-94463-0>. (Springer International Publishing, Cham, 2018).

Acknowledgements

The authors thank Enrico Carlon, Loucif Remini, and Andrea Parmeggiani for fruitful discussions. This work was supported by the ANR project ANR-19-CE45-0016. M.L. acknowledges support from the Research Foundation Flanders (FWO) with project 11PG324N.

Author contributions

T.F. and M.B. introduced the polymer spectral analysis and defined the problem. M.L. conceived the Λ parameter and the Λ -plot, performed the theoretical calculations, and constructed the NN. M.L. and T.F. performed the simulations and experimental data analysis. T.F. supervised the data analysis and interpreted the biological significance of the results. M.B., M.L., and T.F. wrote the manuscript. All authors discussed the results, commented on the manuscript and have given approval to the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-51650-w>.

Correspondence and requests for materials should be addressed to Michaël. Liefsoens or Timothy Földes.

Peer review information *Nature Communications* thanks Jia-Ming Chang, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024