

Precise measurements of chromatin diffusion dynamics by modeling using Gaussian processes

Guilherme M. Oliveira¹, Attila Oravecz¹, Dominique Kobi¹, Manon Maroquenne¹, Kerstin Bystricky², Tom Sexton^{1,*}, and Nacho Molina^{1,*}

¹Institute of Genetics and Molecular and Cellular Biology (IGBMC)

CNRS UMR7104, INSERM U1258, University of Strasbourg, Illkirch, France

²Molecular Cellular and Developmental Biology unit (MCD), Centre de Biologie Integrative (CBI)
UPS, CNRS, 31062 Toulouse, France

*Equal contribution and corresponding authors: molinan@igbmc.fr, sexton@igbmc.fr

The spatiotemporal organization of chromatin influences many nuclear processes: from chromosome segregation to transcriptional regulation. To get a deeper understanding of these processes it is essential to go beyond static viewpoints of chromosome structures, and to accurately characterize chromatin mobility and its diffusion properties. Here, we present GP-FBM: a new computational framework based on Gaussian processes and fractional Brownian motion to analyze and extract diffusion properties from stochastic trajectories of labeled chromatin loci. GP-FBM is able to optimally use the higher-order correlations present in the data and therefore outperforms existing methods. Furthermore, GP-FBM is able to extrapolate trajectories from missing data and account for substrate movement automatically. Using our method we show that diffusive chromatin diffusion properties are surprisingly similar in interphase and mitosis in mouse embryonic stem cells. Moreover, we observe surprising heterogeneity in local chromatin dynamics, which correlates with transcriptional activity. We also present GP-Tool, a user-friendly graphical interface to facilitate the use of GP-FBM by the research community for future studies of nuclear dynamics.

Introduction

The spatiotemporal organization of chromatin plays a crucial role in several nuclear processes: from cell division, where chromatin is compacted during mitosis facilitating chromosome segregation, to gene regulation, where precise control of transcription correlates with specific long-range chromatin contacts [1]. Chromosome conformation capture techniques and imaging approaches have revealed

fundamental structural features of chromatin at different resolutions. Of special interest are topological associated domains (TADs) which are characterized by an increased frequency of interactions between genomic loci within the same domain, and reduced interactions across domains [2, 3]. Remarkably, it has been shown that TAD organization can influence regulation of transcription [4] and that TADs are dismantled during mitosis where chromatin density dramatically increases [5, 6]. However, much less is known about the diffusion and mobility properties of chromatin and how they depend on the genomic context. For instance, it is not clear whether or how transcriptional activation affects chromatin mobility, with previous studies giving seemingly conflicting results [7, 8]. Insights into the dynamic properties of chromatin motion are required to understand how gene regulatory elements explore the nuclear space to “find” their distal targets and/or to build regulatory hubs [9].

The simplest model to describe the diffusion of microscopic systems is Brownian motion, whereby movements are caused by random collisions of small particles within the system [10, 11]. However, as a bulky polymer making interactions with the nuclear environment, chromatin frequently displays sub-diffusive behavior, more constrained than classical Brownian motion [12]. Therefore, the mean squared displacement (MSD) of chromatin is expected to follow this relationship with time: $MSD \propto Dt^\alpha$. Two parameters thus describe the diffusion properties of chromatin: the apparent diffusion coefficient D , indicating the speed of motion, and the anomalous diffusion exponent α , which for sub-diffusive behavior is < 1 which indicates greater confinement (henceforth we refer to it as confinement parameter). The traditional method used to estimate the diffusion parameters is based on calculating the MSD over time from measured trajectories and fitting the above theoretical expression to the data. More sophisticated methods, such as can be derived from Spot-On [13], use the whole distribution of observed displacements (henceforth referred to as DDB (displacement distribution-based) methods). However, they do not use all the information contained in the trajectories as higher-order correlations are discarded. Furthermore, errors due to measurement noise cannot easily be included into the analysis, and it is not possible to recover missing data points due to misdetection or occlusions.

Here, we propose a novel computational method, GP-FBM, based on Gaussian Processes (GP) [14] and fractional Brownian motion (FBM) [15, 16]. Importantly, GP provides a consistent probabilistic framework that considers entire trajectories and thus utilizes all the available information. Trials on simulated data demonstrate a greater precision of GP-FBM in measuring diffusion parameters over existing techniques. Furthermore, GP-FBM naturally takes into account imprecision in localization and occlusions without the need to establish a fitted MSD curve or displacement distributions, as it is applied directly on trajectories. We further extend this model to account for external sources of movement (e.g. displacement of whole nuclei or chromosomes) using underlying correlations between

multiple trajectories, without the need to further develop substrate motion models and experiments for calibration. Finally, we applied GP-FBM to two experimental systems to study chromatin diffusion properties in different contexts. First, we characterized chromatin dynamics in interphase and mitosis using tagged arrays inserted at random genomic locations in mouse embryonic stem (ES) cells. Although chromatin density increases by a factor of three during mitosis [17], our results surprisingly indicate that there are no significant differences on average in the diffusion coefficient or confinement parameter. Second, to compare the diffusion properties of different specific genomic regions, we performed double-labeling and live tracking experiments around the *HoxA* locus in mouse ES cells before and after induction of the genes with retinoic acid. We discover that, instead of having homogeneous diffusion properties across euchromatin, genomic loci significantly differ in both their apparent diffusion coefficients and confinement parameters. In some cases, altered chromatin diffusion properties correlate with underlying functions such as gene activity changes. The methods we have developed have been integrated into a user-friendly package, GP-Tool, for use in the scientific community. Chromatin mobility has been largely overlooked in previous studies of genome functions, and we anticipate that GP-FBM will greatly facilitate research in that area.

Results

GP-FBM: A Gaussian process with a fractional Brownian kernel to model particle diffusion dynamics

Current methods to analyze particle diffusion dynamics rely on two-point statistics as only particle displacements calculated between two frames at different time intervals are used. This has important drawbacks: higher order correlations within the trajectories are discarded; errors due to measurement noise cannot easily be included into the analysis; and missing data points due to misdetection or occlusions are ignored. To address these problems we built a consistent probabilistic framework based on Gaussian Processes (GP). Briefly, a GP is defined as a collection of random variables such that every finite subset of them follows a multivariate normal distribution which is fully determined by its mean and kernel functions $\mu(t)$ and $\Sigma(t, t')$ [14]. We assume that a stochastic diffusion trajectory $x(t)$ of a given chromatin locus can be modeled as a Gaussian process with the following fractional Brownian kernel:

$$\Sigma_{D,\alpha}(t, t') = D (|t|^\alpha + |t'|^\alpha - |t - t'|^\alpha), \quad (1)$$

where D is the apparent diffusion coefficient and α is the confinement parameter defined in the range $0 < \alpha < 2$. This kernel produces a generalized Brownian motion with a mean squared displacement

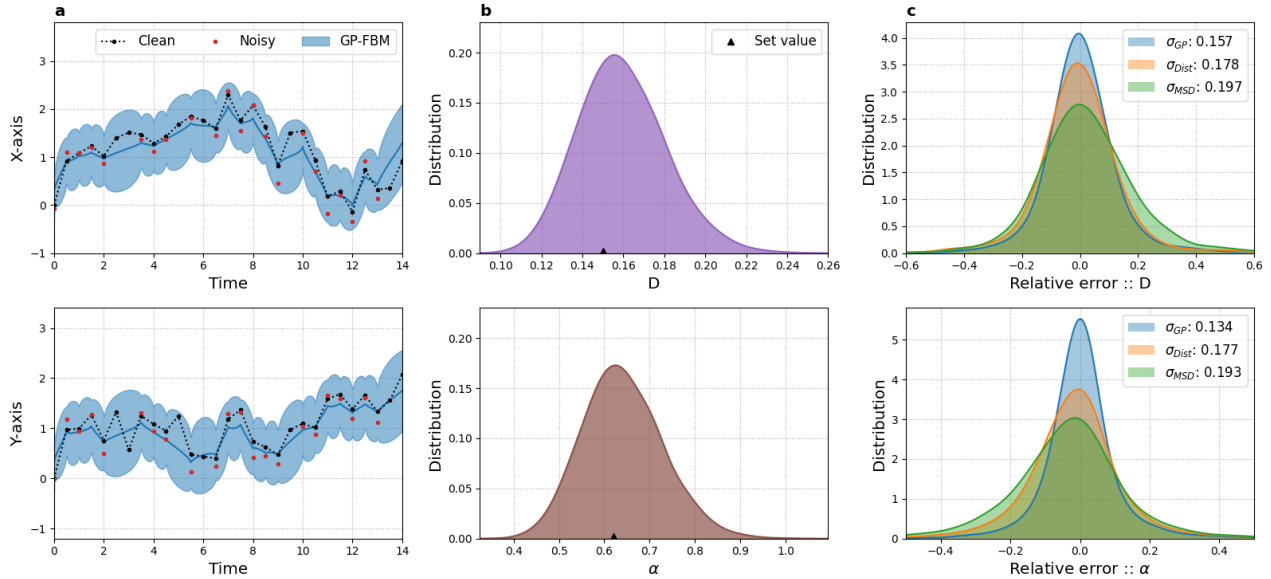


Figure 1: GP-FBM outperforms existing methods on simulated data. (a) Simulated 2D trajectory and GP-FBM inference: black dots are simulated values, red dots are the values after inclusion of noise and 10% occlusions, while the blue line is the most probable trajectory with shading representing the 95% credible interval as inferred by GP-FBM. (b) Posterior probability distributions for D (top) and α (bottom) inferred for trajectory in (a). Triangles denote the values set in the simulation. (c) Comparing inference relative errors as obtained by GP-FBM (blue), MSD (green) and DDB (orange) for 2000 simulated trajectories with random diffusion parameters.

$\langle r^2 \rangle = 2nDt^\alpha$, where n corresponds to the number of degrees of freedom. Then the probability of observing a discrete trajectory $\mathbf{r} = \{r_i\}$ measured at a set of times $\mathbf{t} = \{t_i\}$ is given by the multivariate Gaussian distribution,

$$\mathcal{N}(\mathbf{r}|D, \alpha, \boldsymbol{\mu}) \propto \exp \left[-\frac{1}{2}(\mathbf{r} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{r} - \boldsymbol{\mu}) \right], \quad (2)$$

where the covariance matrix is defined as $\Sigma_{ij} \equiv \Sigma_{D,\alpha}(t_i, t_j)$ and we take a constant $\boldsymbol{\mu}$ without loss of generality. Furthermore, we can easily incorporate localization errors by adding the diagonal term $\sigma_i^2 \delta_{ij}$ to the covariance matrix Σ , which assumes that errors are decorrelated and normally distributed with standard deviations $\boldsymbol{\sigma} = \{\sigma_i\}$ (see Methods). Ultimately, providing a trajectory \mathbf{r} and the localization errors $\boldsymbol{\sigma}$, the likelihood (2) can be used to calculate estimates of the diffusion parameters D and α .

To test the performance of the GP-FBM method, we generated 2000 synthetic trajectories using uniformly distributed random values of D and α in the range $0 < D < 1.5$ and $0 < \alpha < 2$. As real trajectories present occlusions and localization imprecision, we removed 10% of the points and introduced positional noise of about 1/10 of a pixel (an example is shown in Fig. 1a). We then obtain a posterior distribution over the parameters D and α given the trajectory and the localization errors by combining the likelihood (2) with flat priors over the parameters, which can be sampled using Markov Chain Monte Carlo (MCMC) (see Fig. 1b and Methods). Interestingly, once the diffusion parameters are estimated, the power of the GP framework can be used to infer the most probable trajectory of

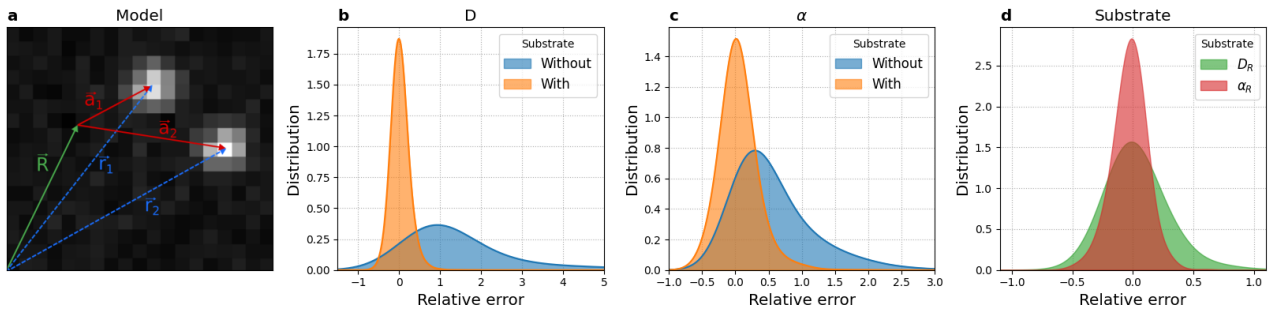


Figure 2: GP-FBM can remove substrate movement to improve estimation of diffusion parameters. (a) Scheme showing the measured trajectories of the particles r_1 and r_2 as the combination of substrate motion R and the particles' diffusion with respect to the substrate, a_1 and a_2 . (b-c) Distributions of relative errors for the estimation of diffusion parameters on 2000 pairs of simulated trajectories with random parameters when substrate movement is accounted for (orange) or not (blue). (d) Distributions of relative errors for the estimation of diffusion parameters of substrate motion.

the particle by removing measurement noise and predicting the localization of the particle position where occlusions or misdetections occurred (Fig. 1a). We compared the results obtained using our approach on the simulated data with the traditional MSD method and a DDB method inspired by Spot-On [13]. GP-FBM clearly outperforms both methods, producing much smaller relative errors of parameter estimation (Fig. 1c). Unlike GP-FBM, both MSD and DDB methods require trajectories to be split into individual displacements, thus breaking higher order correlations that the trajectories may contain. Therefore GP-FBM method can optimally infer diffusion parameters from single trajectories, using all the information contained in the data and thus achieving greater precision.

Accounting for substrate movement with GP-FBM

Often a particle may be subject to secondary movement that is entangled with its diffusion dynamics. In chromatin dynamics, this movement is frequently associated with the substrate in which the particle is diffusing, such as cell displacement, membrane fluctuations or chromatin reallocation, as well as technical considerations such as thermal drift and undesired media flow. If overlooked, this may result in over-estimation of the diffusive properties. However, when two or more particles are measured in the same cell, this substrate movement can be accounted for by analyzing the correlation that introduces between the particle trajectories. To that end, we developed a correlation covariance model that takes advantage of the GP-FBM framework to quantify and handle the substrate movement and the possible correlations that may introduce into the movement of particles (see Methods). In the case of two particles, we obtain the probability distribution,

$$\rho(\mathbf{r}_1, \mathbf{r}_2 | \alpha, \mathbf{D}) \propto \exp \left\{ -\frac{1}{2} \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{pmatrix}^T \begin{pmatrix} \Sigma_1 + \Sigma_R & \Sigma_R \\ \Sigma_R & \Sigma_2 + \Sigma_R \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{pmatrix} \right\}, \quad (3)$$

where Σ_1 , Σ_2 and Σ_R are FBM covariance matrices for the two particles and substrate respectively with diffusion parameters $\mathbf{D} = \{D_1, D_2, D_R\}$ and $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \alpha_R\}$. This method can easily be extended for higher numbers of particles, limited only by required computational power in practice. In this study, we restrict our analysis to five particles per cell.

To demonstrate the utility of this approach, we generated 2000 synthetic trajectories as before, but now including substrate movement, generating vectors \mathbf{r}_i as a combination of substrate displacement \mathbf{R} and the actual particle displacement \mathbf{a}_i (Fig. 2a). Again, simulations are generated with 10% occlusion rate and localization errors. As expected, without applying the correlation covariance model, D and α tend to be overestimated; however, the parameters are precisely measured when the substrate correction is incorporated into the model (Fig. 2b,c). The method is also able to precisely estimate the dynamic properties of the substrate and to estimate, albeit with less precision, the substrate movement (see Fig. 2d, Fig. S1 and Methods). Furthermore, we tested the performance of the method depending on the number of tracked particles subjected to the same substrate movement. Precision is increased with use of more particles, but the bulk of the error is already removed with only two particles (see Fig. S2). Finally, we showed that GP-FBM outperforms a DDB method where the substrate movement is taken into account (see Methods and S3). In conclusion, GP-FBM has the ability to remove the substrate movement from the analysis which is demonstrably important for precise measurement of diffusion parameters. Other approaches can be applied to estimate cell movements and correct trajectories [18], but GP-FBM has the advantage of being able to derive this information directly from the trajectories themselves, provided that two or more particles are tracked per cell, thanks to the correlation that the substrate movement introduces between the particles.

Analyzing chromatin dynamics in interphase and mitosis with GP-FBM

Due to chromosome compaction and condensation, chromatin density increases by a factor of three during mitosis [17]. Although, the structure of mitotic chromatin has been intensively studied [5,19,20], it is unknown if or how the higher density and rearrangement of chromatin fibers affects chromatin diffusion properties. To measure chromatin dynamics in interphase and mitosis, we used a mouse ES cell line carrying approximately 20 *TetO* arrays of 7 kb length inserted at random genomic locations [21]. GFP::TetR is stably expressed in these cells, where it binds to the *TetO* arrays for the simultaneous visualization of several chromatin loci in each cell. We performed confocal live-imaging of these cells, distinguishing interphase and mitotic cells by DNA staining using Hoechst 33342, recording images at 4 frames/s for 75 s. To increase the number of mitotic cells we also performed live-imaging experiments on cells arrested in prometaphase with nocodazole (see Fig 3a and Methods). Before

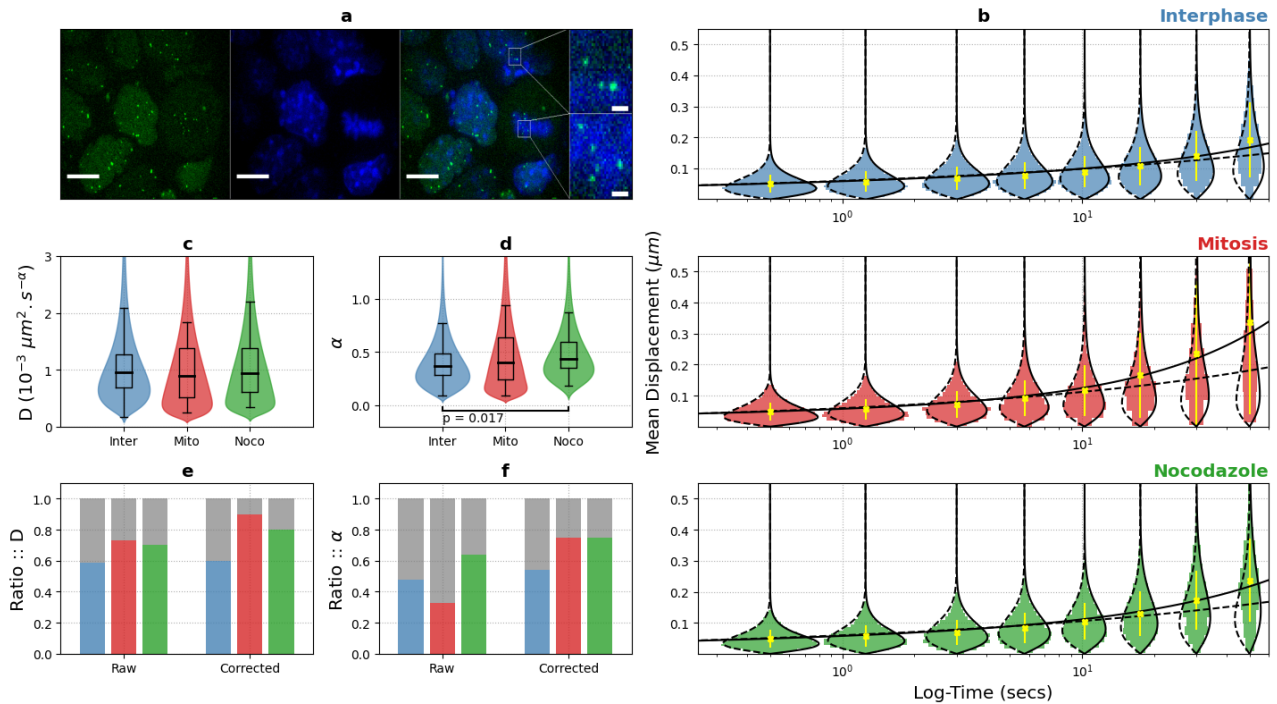


Figure 3: Chromatin dynamics are similar in interphase and mitosis, and are highly variable across loci. (a) Maximum projection images of ES cells containing spots of TetR::GFP (green) bound to *TetO* arrays, with DNA (blue) stained with Hoechst (scale bar is $10 \mu\text{m}$). Inset shows magnification of the selected area (scale bar $1 \mu\text{m}$). (b) Displacement distributions for loci in interphase (blue), mitosis (red) or mitotic arrest (green), plotted for different time intervals. The observed mean displacement is shown with yellow crosses and theoretical curves with and without substrate movement are shown with black continuous and dashed lines, respectively. The model with substrate movement fits better the data specially at greater time intervals, suggesting a greater confounding effect, especially for active mitotic cells. (c-d) Distribution of estimated D and α in the three conditions, correcting for substrate movement. (e-f) Estimations of the inter-(grey) and intra-cell (color) proportions of total variance for D and α in interphase (blue), mitosis (red) and mitotic arrest (green).

applying the GP-FBM probabilistic framework, we enhanced particle localization precision by fitting a 2D Gaussian function to the intensity of the tracked spots (see Methods and S4). Comparing the performance of GP-FBM with and without substrate movement correction, it was apparent that actively dividing mitotic cells had greater substrate movement (presumably due to coordinated alignment and movement of chromosomes by the mitotic spindle), but that appreciable correction was required for precise chromatin dynamics measurements in all conditions (Fig. 3b). Surprisingly, we observed no significant differences in the mean diffusion coefficient or confinement parameter between interphase and mitotic chromosomes, suggesting that condensation may not necessarily affect the average local diffusion dynamics of chromatin (Fig. 3c and d). We observed a small but significant increase in the confinement parameter of mitotic-arrested cells compared to interphase, which might be related to the effect that nocodazole has on microtubule formation and thus mitotic chromosome stability [22].

Interestingly, we obtained a wide range of estimated D and α coefficients indicating a remarkable spot-to-spot variability in their diffusion dynamics, even when correcting for substrate movement. This variability could partially be caused by differences in the state of the analyzed cells (inter-cell variability) leading to different overall chromatin dynamics. Alternatively, differences in the chromatin context of the genomic loci could lead to specific diffusion dynamics (intra-cell variability). Applying the law of total variance, we quantified the contribution of inter-cell vs intra-cell variability (see Methods) (Fig. 3e and f). Strikingly, as much as 75% of the variability in D and 65% in α could be explained by differences *within* the same cells, and this estimate is even higher when substrate movement is taken into account. Together, this suggests that different genomic loci may have characteristic diffusion dynamic properties due to their specific chromatin or nuclear context.

GP-FBM distinguishes locus- and cell-specific chromatin diffusion properties

Except for a tendency for chromatin mobility to be reduced at centromeric or telomeric locations in yeast [23], little is known about how different genomic contexts may affect dynamics of the underlying chromatin. Further, previous studies give conflicting views on whether transcriptional activation can increase local confinement of a gene (as observed in the same cell before and after estrogen stimulation [24]) and/or increase gene mobility (as observed comparing cells before and after differentiation [25]). To compare the diffusion properties of different specific genomic regions, we performed double-labeling and live tracking experiments with the ANCHOR system [26] around the HoxA locus in mouse ES cells before and after induction of Hox genes with retinoic acid. We engineered the ANCH1 and ANCH3 [24] labels into different locations within the same allele to generate two ES lines with equidistant probes assessing inter-TAD (T1-T2) or intra-TAD (T2-T3) associations (Fig. 4a), and imaged at 2 frames/s

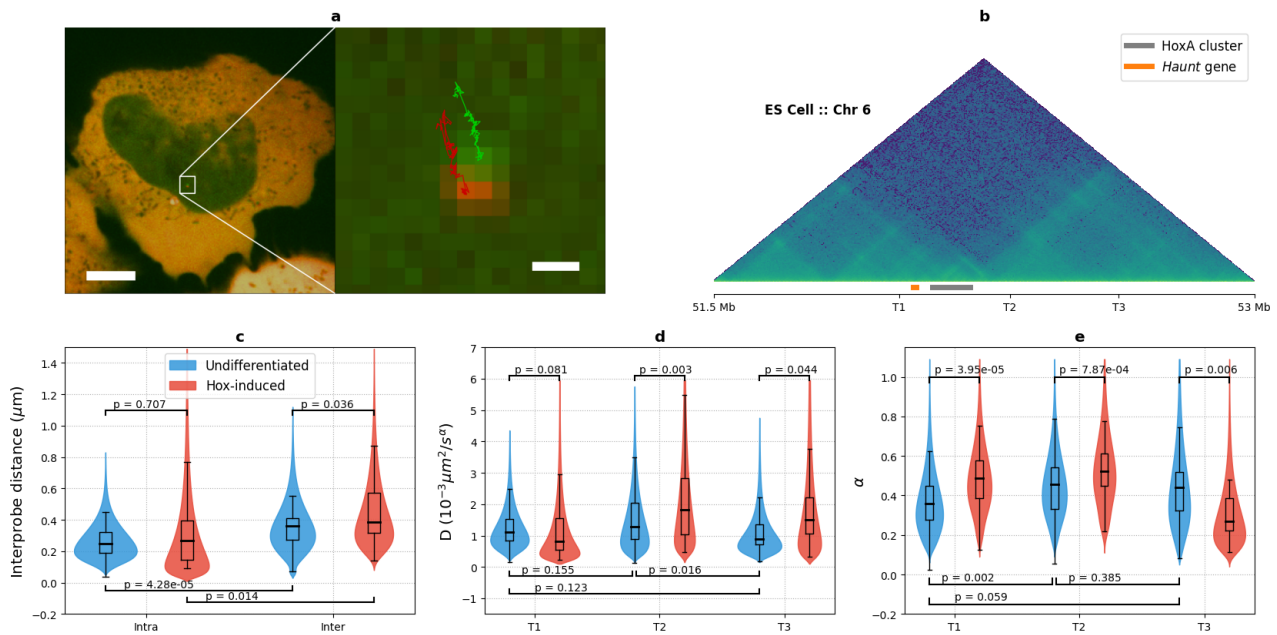


Figure 4: Chromatin dynamics of three chromatin loci within the HoxA genomic region. (a) Example of microscopy image obtained on ANCHOR cell line (scale $5 \mu\text{m}$) with tracked trajectories shown in amplified subpanel (scale $0.3 \mu\text{m}$). (b) ES Hi-C map around the HoxA cluster, illustrating the locus TAD structure. The positions of ANCHOR probes (T1, T2 and T3) and the genes *Haunt* and the HoxA cluster are shown underneath the map. (c) Inter-probe distances measured for ES cells before (blue) and after (red) treatment with retinoic acid. (d-e) Comparisons of diffusion coefficients and confinement parameters between the three labeled loci, before (blue) and after (red) treatment with retinoic acid.

for 2 min. As may be expected, the average inter-probe distance was higher for the inter-TAD than intra-TAD combination, but with large heterogeneity in the distance distributions (Fig. 4b; [27]). Interestingly, Hox gene induction had no effect on intra-TAD distances within the neighboring domain, but increased inter-TAD distances, supporting the idea of general TAD reinforcement as cell differentiation is induced [28]. We performed GP-FBM for the three loci and found that, in undifferentiated ES cells, although they have equivalent diffusion rates, region T1 is significantly more confined than T2 or T3 (Fig. 4c,d). Closer inspection of the ES (and differentiated neuronal precursor cell) epigenomic profiles around these regions showed that T1 is close ($<15 \text{ kb}$) to a putative active enhancer of *Halr1* (Fig. S5). This gene encodes the long non-coding RNA *Haunt*, whose specific expression in ES cells is linked to suppression of the HoxA genes [29]. Active histone modifications around T1, compared to the silent T2 and T3 regions, thus appears linked to a greater confinement of the chromatin, in line with a previous study of an estrogen-induced gene [24]. Hox gene induction by retinoic acid had no significant effect on the diffusion rate of T1 but did reduce locus confinement (Fig. 4c,d). In contrast, the region T2, which lacks any known epigenomic or regulatory features, had increases in D and α , perhaps indicative of general chromatin remodeling caused on onset of differentiation. Curiously, T3 became more confined on retinoic acid treatment, with a concomitant increase in diffusion rate. This region contains sites bound by the architectural protein CTCF, whose

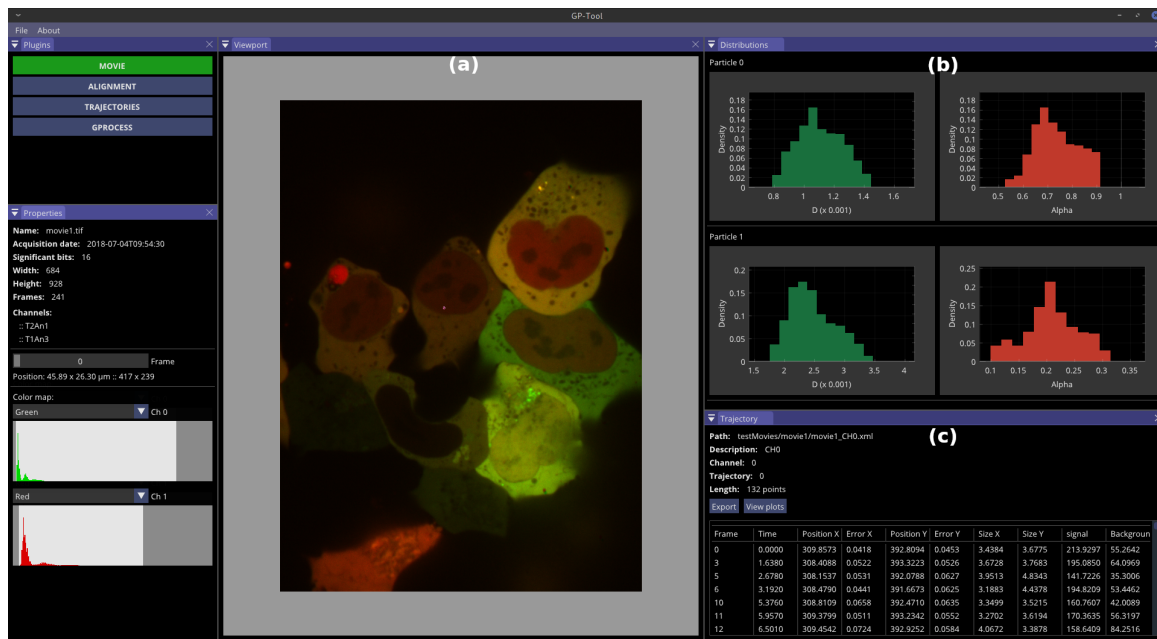


Figure 5: GP-Tool program: A graphical user interphase to apply GP-FBM on microscopy movies. (a) Main viewport (b) GProcess plugin also calculates the posterior distribution for D and α using MCMC. (c) The plugin for trajectories enhances the localization of tracked trajectories, estimate error, particle size and signal.

binding is either lost or reduced on differentiation to neuronal precursors (Fig. S5). CTCF is proposed to form a roadblock for cohesin-mediated loop extrusion processes [30,31], and this may be expected to play out in alterations to local chromatin dynamics, although this has been largely unexplored. Overall, these results show a correlation between activity or function of a gene locus and its local chromatin dynamics, a feature that has been largely overlooked in most previous studies.

GP-Tool: A user-friendly graphical interface to apply GP-FBM

To facilitate use of GP-FBM by the community, we developed a freely available graphical user interface called GP-Tool (Fig. 5; github.com/guilmont). GP-Tool contains 4 plugins: movie, alignment, trajectories and g-process. The movie plugin allows the user to open TIFF files, display basic ImageJ and OME metadata, define colormaps for each channel and manually correct for contrast. The alignment plugin runs the algorithm described in Methods to digitally correct chromatin aberration and possible camera calibration issues. Alternatively, the user can manually modify each of the parameters. The program can perform the analysis of several cells in the same movie. The trajectories plugin provides the ROI utility which allows the user to select spots of interest and extract trajectories. Finally, the g-process plugin allows to infer the diffusion and confinement parameters correcting for substrate movement if two or more particles are tracked. It is also possible to use a MCMC sampler to obtain the posterior probability distribution associated with each of these parameters. Once the analysis is complete, the tool provides the possibility to save the results into two file formats: JSON and HDF5.

It also provides export functions to save tables in CSV format. All these formats are easily parsed in all major computing languages, such as C/C++, Python and R.

Discussion

We developed GP-FBM, a Bayesian framework that combines the inference power of Gaussian processes with fractional Brownian motion, a flexible model to describe diffusion dynamics. GP-FBM treats stochastic trajectories as a whole without preprocessing or extracting limited statistics from them. Therefore, this approach utilizes optimally all the information contained in the data by incorporating higher-order correlations into the analysis. Furthermore, the Gaussian process framework allows us to easily integrate spot-dependent localization errors which translates into a consistent weighting of the time point depending on the precision at which the spot position is determined. Furthermore, missing data due to spot misdetection or occlusion does not hinder the analysis and, on the contrary, GP-FBM can be used to infer spot positions at time points when it was not measured. Finally, when two or more particles are tracked inside the cell, GP-FBM uses possible cross-correlation between the trajectories to characterize substrate movement and, therefore, remove it from the analysis of the diffusion dynamics of particles. Benchmarking shows improved precision of GP-FBM against two existing methods: the traditional MSD curve fitting and the more advanced DDB approach. This increase in accuracy can potentially be crucial to study changes in diffusion dynamic properties in different conditions, and can even be applied to single trajectories.

We applied GP-FBM to two ES cell systems to ask different questions about chromatin diffusion properties. Overall, we observe a large variability in chromatin dynamics when comparing individual cells and comparing different loci. A fraction of this variability can be explained by differences across cells, especially in interphase cells, indicating that cell state related to cell cycle or metabolism may globally influence chromatin dynamics. However, the majority of the observed variability is related to differences across loci. On the other hand, chromatin exhibits on average similar diffusion dynamics in interphase and mitosis despite a large difference in chromatin density. This result may be related with recent findings showing that mitotic chromatin is not as inaccessible and inert as previously thought. Indeed, several studies have shown that mitotic chromatin is bound by transcription factors [32,33] and some genes are even transcribed during mitosis [34]. In contrast, different genomic loci can have striking differences from average chromatin dynamic properties, which appear to correlate with underlying functions such as transcriptional activity or binding of insulator proteins. More widespread application of GP-FBM to labeled transcribed loci and other specific regulatory elements, such as enhancers or TAD borders, will likely uncover more interesting functional links between genome function and the

dynamics of its component chromatin.

Finally, we present GP-Tool a graphical user interface that helps to perform GP-FBM analysis on microscopy movies with only a few mouse clicks. Importantly, this tool and the GP-FBM framework can be applied to study not only chromatin dynamics but potentially any labeled particle that can be tracked over time. Moreover, the FBM kernel used in this study can easily be replaced in principle by alternative kernels that may better describe dynamics of other systems. We thus anticipate that GP-FBM and GP-Tool will greatly facilitate the analysis of diffusion dynamics in biology.

Methods

Cell lines, culture and treatments

Transgenic TetO ES line

The mouse ES cell line was kindly provided by Dr. Luca Giorgetti. It is derived from an X0 clone of the PGKT2 subclone of the feeder-independent PGK12.1 mouse ES line which was engineered by co-transfection with pBROAD3-TetR-ICP22NLS-eGFP and pcDNA3.1Hygro to stably express the TetR-eGFP recombinant protein after random integration and hygromycin selection (250 $\mu\text{g}/\text{ml}$) as described in [21,35]. The piggyBac transposon system was then used to generate cells with 20-25 stable random integrations of a 150 TetO binding site array as described in [21]. Cells were cultured on 0.1% gelatin-coated culture plates in DMEM (4.5g/l glucose) supplemented with GLUTAMAX-I, 15% fetal calf serum (ES cell culture tested), 0.1 mM beta-mercaptoethanol, 1,500 U/ml leukemia inhibitory factor (LIF; produced in house) and 0.1 mM non-essential amino acids in 5% CO₂ at 37° C. Mitotic arrest was performed by treating the cells for 5 h with 100 ng/ml Nocodazole (Sigma, M1404-2MG).

Transgenic ANCHOR ES lines

J1 mouse ES cells were grown on gamma-irradiated mouse embryonic fibroblast cells under standard conditions (4.5 g/L glucose-DMEN, 15% FCS, 0.1 mM non-essential amino acids, 0.1 mM beta-mercaptoethanol, 1 mM glutamine, 500 U/mL LIF, gentamicin), then passaged onto feeder-free 0.2% gelatin-coated plates for at least two passages to remove feeder cells before subsequent transfections. The two (“inter-TAD” and “intra-TAD”) ANCHOR transgenic lines were generated by sequential CRISPR/Cas9-mediated knock-in experiments in the following manner. First, flanking homology arms (mm9 chr6: 52,320,061-52,321,144, and chr6: 52,321,145-52,322,244) were introduced by PCR amplification and Gibson assembly into a vector containing ANCH1 sequence [26]. This vector (1 μg) was co-transfected with 3 μg of a vector containing Cas9-GFP, a puromycin resistance marker and the

scaffold to transcribe the sgRNA specific to the T2 insertion site (CGGCGCGCACTTAACACCAA; vector generated by the IGBMC Molecular Biology platform) in 1 million cells with Lipofectamine-2000. Two days after transfection, the cells were cultured for 24 h with 3 $\mu\text{g}/\text{ml}$ puromycin, then 48 h with 1 $\mu\text{g}/\text{ml}$ puromycin to enrich for transfected cells, before sorting individual GFP-positive cells on to feeders to amplify individual clones. Clones with the correct sequence were screened by PCR and sequencing, then the CRISPR knock-in process was repeated to insert the ANCH3 sequence [24] into either the T1 site (“inter-TAD” line; homology arms at chr6: 52,013,471-52,014,370 and chr6: 52,014,371-52,015,270; gRNA sequence AATCGAGCTCACGCCATTAG) or the T3 site (“intra-TAD” line; homology arms at chr6: 52,622,955-52,623,855 and chr6: 52,623,856-52,624,755; gRNA sequence TATGCTGAGGCGTGTCGCAA). Final clones were verified for maintained pluripotency by qRT-PCR to assess Oct4, Nanog (e.g. Fig. S5) and Sox2 expression. Subsequent microscopy experiments (see below) confirmed heterozygous incorporation of the ANCH sequences (detection of one specific spot per ANCH sequence per cell) within the same allele (two spots were always in close proximity).

OR transfection

150,000 cells are plated two days prior to imaging off feeder cells onto laminin-511-coated 35 mm glass bottom petri dishes, and transfected with 3 μg OR1-EGFP and 3 μg OR3-IRFP plasmids (vectors available from NeoVirTech (contact@neovirtech.com); were modified from original source by changing the C-terminal fluorescent protein sequence, introducing Kozak sequence before the translation start site and replacing the CMV promoter with EF-1 α) using Lipofectamine-2000. After two days, the medium is changed to remove dead cells, before passing directly to microscopy.

Hox induction

ES cells were passaged without feeders and cultured on laminin-511 for two days without LIF, then for a subsequent three days without LIF and with the addition of 5 μM retinoic acid. One day after the addition of retinoic acid, the cells are transfected with the OR proteins as previously.

Microscopy

Live cell imaging of TetO ES cells

35 mm glass-bottom dishes (Ibidi 81158) were coated with 10 $\mu\text{g}/\text{ml}$ fibronectin human plasma (Sigma, F2006-1MG) in PBS for 45 minutes at room temperature. $3\text{-}5 \times 10^5$ cells were seeded one day before imaging, then the medium was replaced by phenol-red-free medium containing 500 ng/ml Hoechst 33342 (Invitrogen, H3570). Cells arrested in mitosis were collected on the day of imaging by “shake-

off", incubated with 0.25% Trypsin-1 mM EDTA (Invitrogen, 25200-072) for 1 min at 37°C and washed, and placed on fibronectin-coated glass-bottom dishes in phenol-red-free medium containing 100 ng/ml Nocodazole and 500 ng/ml Hoechst 33342. Confocal live-cell imaging was performed on a Nikon Eclipse Ti-E inverted widefield microscope (Perfect Focus System) equipped with a CSU-X1 confocal scanner unit and an Evolve back-illuminated EMCCD camera (Photometrics). Images were recorded using 100x HC Plan APO oil immersion objective (Leica, NA 1.4). Intensities were set to 10% for the 405 nm and 30% or 50% for the 491 nm lasers, with exposure times of 100 ms and 50 ms or 25 ms, respectively. 5 z-stacks with 0.5 μm distances were recorded for each channel. 301 time-lapse images were recorded only in the 491 channel.

Live cell imaging of ANCHOR ES cells

Imaging experiments were performed on an inverted Nikon Eclipse Ti microscope equipped with a PFS (perfect focus system), a Yokogawa CSU-X1 confocal spinning disk unit, two sCMOS Photometrics Prime 95B cameras for simultaneous dual acquisition to provide 95% quantum efficiency at 11 μm x 11 μm pixels and a Leica 100x oil objective (HC PL APO 1,4 oil immersion). We excited EGFP and IRFP with a 491 nm (100mw) and a 635-nm laser (>28mW), respectively. We detected green and far red fluorescence with an emission filter using a 525/50 nm and a 708/75 nm detection window, respectively. A thermostated heater (Tokai Hit Stage Top Incubator) allowed for heating at 37° C, humidity and CO₂ control (5%). Time-lapse analysis of GFP and IRFP foci was performed in 2D acquiring 241 time points at a 0.5 s time interval. The system was controlled using Metamorph 7.10 software. Time-lapse was concatenated into single TIFF file.

RT-qPCR

RNA was extracted from cells using the Nucleospin RNA extraction kit (Machery-Nagel), then cDNA was prepared with SuperScript IV (Invitrogen), following the manufacturer's instructions and using random hexanucleotides as primers. The cDNA was quantified by qPCR on a LC480 LightCycler (Roche), using QuantitTect SYBR Green PCR kit (Qiagen). Amplification was normalized to GAPDH. Primer sequences are given in Supplemental Table 1.

Image pre-processing

Spot detection and tracking

Spot detection and tracking for all movies was performed with ICY, an image analysis software [36]. Localization precision was then enhanced by assuming that the spots have the shape of a 2D Gaussian

function. We optimize its localization using the NM-Simplex method [37] and estimate localization precision using MCMC [38]. This method is implemented and automatically runs while loading trajectories in GP-Tool. For more information see figure S4.

Multi-channel alignment correction

For the ANCHOR ES cell line experiments, we used a spinning disk microscope setup with 2 cameras, i.e., one per channel. Even though these cameras were aligned manually using fluorescent beads, we could still observe non-negligible differences between images captured in both cameras. Furthermore, even in rare situations when both cameras were in perfect alignment, we could observe effects of chromatic aberrations towards the edges of the image due the different wavelength used. To correct for such problems, we performed digital post-alignment using a generic set of affine transformations including translation, rotation and scaling as defined in

$$\Omega = \begin{pmatrix} s_x & 0 & (1-s_x)W/2 \\ 0 & s_y & (1-s_y)H/2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & d_x + c_x \\ 0 & 1 & d_y + c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -c_x \\ 0 & 1 & -c_y \\ 0 & 0 & 1 \end{pmatrix}, \quad (4)$$

where, s_i accounts for scaling in directions x and y, d_i accounts for translation in both directions and θ is the angle of rotation between both channels in relation to point c_i .

To infer optimal parameters for correction, we used 10 frames from all the movies recorded in the session and maximize the following likelihood using the Nelder-Mead simplex model [37]

$$\log P \propto -\frac{WH}{2} \log \left\{ \sum_{k,l} [I_2(k,l|\Omega) - I_1(k,l|\mathbb{1})]^2 \right\}, \quad (5)$$

where W and H correspond to width and height of images and $I_r(k,l|A)$ is the value of pixel (k,l) in channel r given transformation A. Fig. S6 shows examples of misaligned images and how the alignment improves greatly after applying our algorithm.

Derivation of GP-FBM models

Covariance function of fractional Brownian Motion

The covariance function of FBM can easily be derived from the assumption of two basic properties: stationary increments $B(t) - B(s) \propto B(t - s)$ and a power-law variance, $\langle B(t)^2 \rangle \propto |t|^\alpha$. Then, the

off-diagonal terms of the covariance function can be determined as follows:

$$\begin{aligned}
 \langle B_\alpha(t)B_\alpha(s) \rangle &\propto \frac{1}{2} \langle [B_\alpha(s) - B_\alpha(s) + B_\alpha(t)]B_\alpha(s) + B_\alpha(t)[B_\alpha(t) - B_\alpha(t) + B_\alpha(s)] \rangle \\
 &= \frac{1}{2} \langle B_\alpha(s)^2 + B_\alpha(t-s)B_\alpha(s) - B_\alpha(t)B_\alpha(t-s) + B_\alpha(t)^2 \rangle \\
 &= \frac{1}{2} \langle B_\alpha(t)^2 + B_\alpha(s)^2 + B_\alpha(t-s)(B_\alpha(s) - B_\alpha(t)) \rangle \\
 &= \frac{1}{2} \langle B_\alpha(t)^2 + B_\alpha(s)^2 - B_\alpha(t-s)^2 \rangle \\
 &= \frac{1}{2} (|t|^\alpha + |s|^\alpha - |t-s|^\alpha).
 \end{aligned} \tag{6}$$

Finally, the apparent diffusion coefficient D is introduced as a proportionality factor to re-scale the movement leading to the final kernel as presented in the main text:

$$\Sigma_{D,\alpha}(t, s) = 2D_\alpha \langle B_\alpha(t)B_\alpha(s) \rangle. \tag{7}$$

Bayesian inference of diffusion parameters

The GP provides the probability of observing a trajectory \mathbf{r} given D and α . Then, we applied Bayes theorem [39] to obtain the posterior distribution over the diffusion parameters given the measured trajectory:

$$P(D, \alpha, \boldsymbol{\mu}|\mathbf{r}) = \frac{P(\mathbf{r}|D, \alpha, \boldsymbol{\mu}) P(D, \alpha, \boldsymbol{\mu})}{\int P(\mathbf{r}|D, \alpha, \boldsymbol{\mu}) P(D, \alpha, \boldsymbol{\mu}) dD d\alpha d\boldsymbol{\mu}}, \tag{8}$$

where $P(D, \alpha, \boldsymbol{\mu})$ represents the prior distribution of the model parameters. Assuming a flat prior on $\boldsymbol{\mu}$, D and α , the log-posterior can be expressed as

$$\log(P(D, \alpha, \boldsymbol{\mu}|\mathbf{r})) \propto -\frac{1}{2}(\mathbf{r} - \boldsymbol{\mu})^T \Sigma_{D,\alpha}^{-1}(\mathbf{r} - \boldsymbol{\mu}) - \frac{1}{2} \log |\Sigma_{D,\alpha}| - \frac{N}{2} \log(2\pi), \tag{9}$$

where N represents the number of points measured and $|\cdot|$ is the determinant function. To obtain maximum posterior estimates, we optimized (9) using the Nelder-Mead Simplex method [37]. In addition, to calculate confidence intervals for our estimations we used the MCMC method [38] to sample from the posterior probability distribution. Note that, thanks to this Bayesian approach, available prior knowledge of the diffusion parameters can easily be incorporated into the analysis.

Incorporation of substrate movement in the GP-FBM framework

In the main text, we introduced an extended GP-FBM model to deal with external sources of movement that may distort the analysis of particle diffusion. Here, we present the derivation for two particles subject to a common substrate movement, however it can be extended for an arbitrary number of

particles using the marginalization rule of multi-variate Gaussian distributions [39]. The key idea is to assume that the movement of the particles with respect to the substrate as well as the movement of the substrate itself can be described by independent fractional Brownian motions. Therefore, the probability of observing the particle trajectories \mathbf{a}_1 and \mathbf{a}_2 with respect to a given frame of reference that moves with the substrate and the trajectory \mathbf{R} of the moving frame is:

$$\rho(\mathbf{a}_1, \mathbf{a}_2, \mathbf{R} | \boldsymbol{\alpha}, \mathbf{D}) \propto \exp \left(-\frac{1}{2} \mathbf{a}_1^T \Sigma_1^{-1} \mathbf{a}_1 - \frac{1}{2} \mathbf{a}_2^T \Sigma_2^{-1} \mathbf{a}_2 - \frac{1}{2} \mathbf{R}^T \Sigma_R^{-1} \mathbf{R} \right) \quad (10)$$

where Σ_1 , Σ_2 and Σ_R are FBM covariance matrices that are fully characterized given the diffusion parameters $\mathbf{D} = \{D_1, D_2, D_R\}$ and $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \alpha_R\}$.

Next, to obtain the probability distribution over the particle trajectories \mathbf{r}_1 and \mathbf{r}_2 with respect to the microscope reference frame, we applied the the change of coordinates $\mathbf{r}_i = \mathbf{a}_i + \mathbf{R}$ (see scheme in Fig. 2a) leading to the matrix expression,

$$\rho(\mathbf{r}_1, \mathbf{r}_2, \mathbf{R} | \boldsymbol{\alpha}, \mathbf{D}) \propto \exp \left(-\frac{1}{2} \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{R} \end{pmatrix}^T \begin{pmatrix} \Sigma_1^{-1} & 0 & -\Sigma_1^{-1} \\ 0 & \Sigma_2^{-1} & -\Sigma_2^{-1} \\ -\Sigma_1^{-1} & -\Sigma_2^{-1} & \Sigma_1^{-1} + \Sigma_2^{-1} + \Sigma_R^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{R} \end{pmatrix} \right), \quad (11)$$

Then, to marginalize out the unobserved trajectory of the moving frame \mathbf{R}_i , we need to calculate the inverse of the block matrix in equation 11. To do so, we use results from [40] on inverting a 2x2 block matrix such as

$$\Lambda = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \quad (12)$$

according to the following result

$$\Lambda^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix} \quad (13)$$

Taking $A = [(\Sigma_1^{-1}, 0); (0, \Sigma_2^{-1})]$, $B = -[\Sigma_1^{-1}; \Sigma_2^{-1}]$, $C = B^T$ and $D = \Sigma_1^{-1} + \Sigma_2^{-1} + \Sigma_R^{-1}$, it is easily shown that the top-left corner of Λ^{-1} is given by $[(\Sigma_1 + \Sigma_R, \Sigma_R); (\Sigma_R, \Sigma_2 + \Sigma_R)]$. Using this result, we can marginalize out \mathbf{R} in equation 11 giving the expression,

$$\rho(\mathbf{r}_1, \mathbf{r}_2 | \boldsymbol{\alpha}, \mathbf{D}) \propto \begin{vmatrix} \Sigma_1 + \Sigma_R & \Sigma_R \\ \Sigma_R & \Sigma_2 + \Sigma_R \end{vmatrix}^{-1} \exp \left(-\frac{1}{2} \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{pmatrix}^T \begin{pmatrix} \Sigma_1 + \Sigma_R & \Sigma_R \\ \Sigma_R & \Sigma_2 + \Sigma_R \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{pmatrix} \right). \quad (14)$$

This result clearly shows that the substrate movement induces a correlation between the particle

trajectories as the off-diagonal elements of the block matrix in (14) are non zero. In addition, the covariance matrix of the substrate movement appears also in the diagonal terms, increasing the overall variance of the particles and their total movement. Therefore, if this correction is ignored the diffusion parameters are over-estimated.

Inference of substrate movement

Similarly as before, the diffusion parameters of the particles as well as the substrate can be estimated using equation 14 and Bayesian inference. Furthermore, we can also estimate how large the movement of the substrate is. For that, we can calculate the conditional distribution of \mathbf{R} given the particle trajectories as

$$\langle \mathbf{R} \rangle = - (\Sigma_1^{-1} + \Sigma_2^{-1} + \Sigma_R^{-1})^{-1} (\Sigma_1^{-1} \quad \Sigma_2^{-1}) \begin{pmatrix} \mathbf{r}_1 - \boldsymbol{\mu}_1 \\ \mathbf{r}_2 - \boldsymbol{\mu}_2 \end{pmatrix} \quad (15)$$

with co-variance matrix given by $\Sigma_{\langle \mathbf{R} \rangle} = (\Sigma_1^{-1} + \Sigma_2^{-1} + \Sigma_R^{-1})^{-1}$.

Unfortunately, we could not find an analytical solution for this problem. Nonetheless, we can solve it numerically. In Fig. S1a we show an example of the estimated movement for the substrate with one standard deviation compared to the real simulated trajectory for a system with two particles. In Fig. S1b-c we display the overall accuracy when working with two or more particles.

Benchmarking GP-FBM

Simulated trajectories

We simulated single trajectories with 250 time points using the aforementioned Gaussian process with FBM kernel. Similarly to experimental setup, trajectories are simulated using time steps of 0.5 seconds and positional noise was included. To benchmark a system of N particles affected by substrate movement, we generate $N+1$ trajectories and add the latter to all the others. Finally, 10% of points are randomly removed from each trajectory to simulate for experimental occlusions.

Mean squared displacement (MSD) implementation

To calculate D and α for single trajectories, we estimate a MSD using a sliding window method. This method is mathematically defined as follows

$$\langle r_n^2 \rangle = \frac{1}{N-n} \sum_{i=1}^{N-n} (\mathbf{r}_{i+n} - \mathbf{r}_i)^2, \quad (16)$$

for a trajectory with N points and step interval n . Due to implicit correlations present in single trajectories, we use only initial 10% step intervals. To improve accuracy, we also estimate an average localization error σ . Finally, this experimental curve is approximated by the theoretical mean squared displacement equation

$$\langle r^2 \rangle = 4D t^\alpha + 2\sigma^2, \quad (17)$$

from which diffusion parameters are inferred using linear regression. For more information [41].

Displacement distribution based (DDB) implementation

The theoretical expressions for the displacement distribution is obtained as a solution of the Fokker-Planck equation with localization error σ . In polar coordinates it takes the form

$$\rho(r, \theta | D, \alpha, t, \sigma) drd\theta = \frac{r}{2\pi(2D t^\alpha + \sigma^2)} e^{-\frac{r^2}{4Dt^\alpha + 2\sigma^2}} drd\theta. \quad (18)$$

In order to calculate experimental distributions for single cells, we resort to a sliding window method similar to the one present for MSD. Differently, we calculate normalized histograms with all the absolute displacement values. As before, we calculate an average localization error σ to improve localization and use only histograms calculated for initial 10 step intervals. With these measurements, we optimize the equation above for D and α using Bayes approach with non-informative priors for both parameters.

Statistical analysis

To compare diffusive properties or inter-probe distances across different loci or conditions (Figs 3c and d, and 4b-d), we performed Wilcoxon rank sum tests. For inter-probe distances, the distributions of the median distances for each movie were used. For Figs 4c and d, where fifteen pairwise comparisons are possible, the p-values were corrected for multiple testing with the Benjamini-Hochberg method.

Use of published epigenomic datasets

ES Hi-C sequence data from [28] were taken from Gene Expression Omnibus (GSE96107), and mapped to mm9 and normalized with FAN-C [42]. The normalized submatrix (chr6:51500000-53000000) was then extracted for visualization. ES and neuronal precursor cell H3K27ac and CTCF ChIP-seq data were taken as bigWig files from Gene Expression Omnibus; GSE96107 for all except ES H3K27ac, taken from GSE49847) and visualized in R using the package *rtracklayer*.

Acknowledgements

We thank Luca Giorgetti for providing the TetO ES line and for critical reading of the manuscript. This work was possible thanks to funding from grants by LabEx INRT (ANR-10-LABX-0030-INRT, a French State fund managed by the Agence Nationale de la Recherche under the frame program Investissements d’Avenir ANR-10-IDEX-0002-02), CNRS «Osez l’interdisciplinarité !», ERC (Starting Grant 678624 - CHROMTOPOLOGY) and ATIP-Avenir. The microscopy was performed at the Imaging Center of the IGBMC.

References

- [1] Tom Sexton and Giacomo Cavalli. The role of chromosome domains in shaping the functional genome. *Cell*, 160(6):1049 – 1059, 2015.
- [2] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380, Apr 2012.
- [3] Tom Sexton, Eitan Yaffe, Ephraim Kenigsberg, Frédéric Bantignies, Benjamin Leblanc, Michael Hoichman, Hugues Parrinello, Amos Tanay, and Giacomo Cavalli. Three-dimensional folding and functional organization principles of the drosophila genome. *Cell*, 148(3):458 – 472, 2012.
- [4] Darío G. Lupiáñez, Katerina Kraft, Verena Heinrich, Peter Krawitz, Francesco Brancati, Eva Klopocki, Denise Horn, Hülya Kayserili, John M. Opitz, Renata Laxova, Fernando Santos-Simarro, Brigitte Gilbert-Dussardier, Lars Wittler, Marina Borschiwer, Stefan A. Haas, Marco Osterwalder, Martin Franke, Bernd Timmermann, Jochen Hecht, Malte Spielmann, Axel Visel, and Stefan Mundlos. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 161(5):1012 – 1025, 2015.
- [5] Natalia Naumova, Maxim Imakaev, Geoffrey Fudenberg, Ye Zhan, Bryan R. Lajoie, Leonid A. Mirny, and Job Dekker. Organization of the mitotic chromosome. *Science*, 342(6161):948–953, 2013.
- [6] H. Zhang, D. J. Emerson, T. G. Gilgenast, K. R. Titus, Y. Lan, P. Huang, D. Zhang, H. Wang, C. A. Keller, B. Giardine, R. C. Hardison, J. E. Phillips-Cremins, and G. A. Blobel. Chromatin structure dynamics during the mitosis-to-G1 phase transition. *Nature*, 576(7785):158–162, 12 2019.
- [7] T. Germier, S. Kocanova, N. Walther, A. Bancaud, H. A. Shaban, H. Sellou, A. Z. Politi, J. Ellenberg, F. Gallardo, and K. Bystricky. Real-Time Imaging of a Single Gene Reveals Transcription-Initiated Local Confinement. *Biophys J*, 113(7):1383–1394, Oct 2017.
- [8] Bo Gu, Tomek Swigut, Andrew Spencley, Matthew R. Bauer, Mingyu Chung, Tobias Meyer, and Joanna Wysocka. Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science*, 359(6379):1050–1055, 2018.
- [9] J. Li, A. Hsu, Y. Hua, G. Wang, L. Cheng, H. Ochiai, T. Yamamoto, and A. Pertsinidis. Single-gene imaging links genome topology, promoter-enhancer communication and transcription control. *Nat Struct Mol Biol*, 27(11):1032–1040, 11 2020.

- [10] Leonid Mirny, Michael Slutsky, Zeba Wunderlich, Anahita Tafvizi, Jason Leith, and Andrej Kosmrlj. How a protein searches for its site on DNA: The mechanism of facilitated diffusion. *Journal of Physics A: Mathematical and Theoretical*, 2009.
- [11] Ohad Givaty and Yaakov Levy. Protein Sliding along DNA: Dynamics and Structural Characterization. *Journal of Molecular Biology*, 2009.
- [12] Maxime MC Tortora, Hossein Salari, and Daniel Jost. Chromosome dynamics during interphase: a biophysical perspective. *Current Opinion in Genetics & Development*, 61:37 – 43, 2020. Genome Architecture and Expression.
- [13] Anders S Hansen, Maxime Woringer, Jonathan B Grimm, Luke D Lavis, Robert Tjian, and Xavier Darzacq. Robust model-based analysis of single-particle tracking experiments with Spot-On. *eLife*, 7:e33125, jan 2018.
- [14] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, 2006.
- [15] Jae-Hyung Jeon and Ralf Metzler. Fractional Brownian motion and motion governed by the fractional Langevin equation in confined geometries. *Physical Review E*, 81(2):021103, feb 2010.
- [16] Felix Höfling and Thomas Franosch. Anomalous transport in the crowded world of biological cells. *Reports on Progress in Physics*, 76(4), 2013.
- [17] Paola Vagnarelli. Mitotic chromosome condensation in vertebrates. *Experimental Cell Research*, 318(12):1435 – 1441, 2012. Experimental Cell Research Special Review Issue: Chromosome Biology, 2012.
- [18] Deqing Sun, Stefan Roth, and Michael J. Black. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *Int. J. Comput. Vision*, 106(2):115–137, January 2014.
- [19] Horng D. Ou, Sébastien Phan, Thomas J. Deerinck, Andrea Thor, Mark H. Ellisman, and Clodagh C. O’Shea. Chromemnt: Visualizing 3d chromatin structure and compaction in interphase and mitotic cells. *Science*, 357(6349), 2017.
- [20] Johan H. Gibcus, Kumiko Samejima, Anton Goloborodko, Itaru Samejima, Natalia Naumova, Johannes Nuebler, Masato T. Kanemaki, Linfeng Xie, James R. Paulson, William C. Earnshaw, Leonid A. Mirny, and Job Dekker. A pathway for mitotic chromosome formation. *Science*, 359(6376), 2018.
- [21] Josef Redolfi, Yinxiu Zhan, Christian Valdes-Quezada, Mariya Kryzhanovska, Isabel Guerreiro, Vytautas Iesmantavicius, Tim Pollex, Ralph S Grand, Eskeatnaf Mulugeta, Jop Kind, et al. Damc reveals principles of chromatin folding in vivo without crosslinking and ligation. *Nature structural & molecular biology*, 26(6):471–480, 2019.
- [22] Assaf Amitai, Andrew Seeber, Susan M. Gasser, and David Holcman. Visualization of chromatin decompaction and break site extrusion as predicted by statistical polymer modeling of single-locus trajectories. *Cell Reports*, 18(5):1200–1214, 2017.
- [23] Patrick Heun, Thierry Laroche, Kenji Shimada, Patrick Furrer, and Susan M Gasser. Chromosome dynamics in the yeast interphase nucleus. *Science*, 294(5549):2181–2186, 2001.
- [24] Thomas Germier, Silvia Kocanova, Nike Walther, Aurélien Bancaud, Haitham Ahmed Shaban, Hafida Sellou, Antonio Zaccaria Politi, Jan Ellenberg, Franck Gallardo, and Kerstin Bystricky. Real-time imaging of a single gene reveals transcription-initiated local confinement. *Biophysical journal*, 113(7):1383–1394, 2017.
- [25] Bo Gu, Tomek Swigut, Andrew Spencley, Matthew R Bauer, Mingyu Chung, Tobias Meyer, and Joanna Wysocka. Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science*, 359(6379):1050–1055, 2018.

- [26] Hicham Saad, Franck Gallardo, Mathieu Dalvai, Nicolas Tanguy-le Gac, David Lane, and Kerstin Bystricky. Dna dynamics during early double-strand break processing revealed by non-intrusive imaging of living cells. *PLoS Genet*, 10(3):e1004187, 2014.
- [27] Elizabeth H Finn, Gianluca Pegoraro, Hugo B Brandão, Anne-Laure Valton, Marlies E Oomen, Job Dekker, Leonid Mirny, and Tom Misteli. Extensive heterogeneity and intrinsic variation in spatial genome organization. *Cell*, 176(6):1502–1515, 2019.
- [28] Boyan Bonev, Netta Mendelson Cohen, Quentin Szabo, Lauriane Fritsch, Giorgio L Papadopoulos, Yaniv Lubling, Xiaole Xu, Xiaodan Lv, Jean-Philippe Hugnot, Amos Tanay, et al. Multiscale 3d genome rewiring during mouse neural development. *Cell*, 171(3):557–572, 2017.
- [29] Yafei Yin, Pixi Yan, Jinlong Lu, Guang Song, Yangyang Zhu, Zhaohui Li, Yi Zhao, Bin Shen, Xingxu Huang, Heng Zhu, et al. Opposing roles for the lncrna haunt and its genomic locus in regulating hoxa gene activation during embryonic stem cell differentiation. *Cell stem cell*, 16(5):504–516, 2015.
- [30] Adrian L Sanborn, Suhas SP Rao, Su-Chen Huang, Neva C Durand, Miriam H Huntley, Andrew I Jewett, Ivan D Bochkov, Dharmaraj Chinnappan, Ashok Cutkosky, Jian Li, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences*, 112(47):E6456–E6465, 2015.
- [31] Geoffrey Fudenberg, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A Mirny. Formation of chromosomal domains by loop extrusion. *Cell reports*, 15(9):2038–2049, 2016.
- [32] Cédric Deluz, Elias T. Friman, Daniel Streibinger, Alexander Benke, Mahé Raccaud, Andrea Callegari, Marion Leleu, Suliana Manley, and David M. Suter. A role for mitotic bookmarking of sox2 in pluripotency and differentiation. *Genes & Development*, 2016.
- [33] Nicola Festuccia, Agnès Dubois, Sandrine Vandormael-Pournin, Elena Gallego Tejada, Adrien Mouren, Sylvain Bessonard, Florian Mueller, Caroline Proux, Michel Cohen-Tannoudji, and Pablo Navarro. Mitotic binding of Esrrb marks key regulatory regions of the pluripotency network. *Nature Cell Biology*, 18(11):1139–1148, nov 2016.
- [34] Katherine C. Palozola, Greg Donahue, Hong Liu, Gregory R. Grant, Justin S. Becker, Allison Cote, Hongtao Yu, Arjun Raj, and Kenneth S. Zaret. Mitotic transcription and waves of gene reactivation during mitotic exit. *Science*, 358(6359):119–122, 2017.
- [35] Osamu Masui, Isabelle Bonnet, Patricia Le Baccon, Isabel Brito, Tim Pollex, Niall Murphy, Philippe Hupé, Emmanuel Barillot, Andrew S Belmont, and Edith Heard. Live-cell chromosome dynamics and outcome of x chromosome pairing events during es cell differentiation. *Cell*, 145(3):447–458, 2011.
- [36] Fabrice de Chaumont, Stéphane Dallongeville, Nicolas Chenouard, Nicolas Hervé, Sorin Pop, Thomas Provoost, Yannary Meas-Yedid, Praveen Pankajakshan, Timothée Lecomte, Yoann Le Montagner, Thibault Lagache, Alexandre Dufour, and Jean-Christophe Olivo-Marin. Icy: an open bioimage informatics platform for extended reproducible research. *Nature Methods*, 9(7):690–696, jul 2012.
- [37] John A Nelder and Roger Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313, jan 1965.
- [38] Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to MCMC for machine learning. *Machine learning*, 50(1):5–43, 2003.
- [39] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

- [40] Tzon-Tzer Lu and Sheng-Hua Shiou. Inverses of 2 x 2 block matrices. *Computers & Mathematics with Applications*, 43(1-2):119–129, jan 2002.
- [41] Xavier Michalet. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Physical Review E*, 82(4):041914, 2010.
- [42] Kai Kruse, Clemens B Hug, and Juan M Vaquerizas. Fan-c: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. *Genome biology*, 21(1):1–19, 2020.

Supplementary Information

Primer Name	Sequence
GAPDH S	5'- CATCACTGCCACCCCAGAAGACTG
GAPDH R	5'- ATGCCAGTGAGCTTCCCGTTCAG
Nanog S	5'- AGCAGAAGTACCTCAGCCTC
Nanog R	5'- CCGCTTGCACTTCATCCTTT
Hoxa1 S	5'- CACCAAGAAGCCTGTCGTTC
Hoxa1 R	5'- CGTAGCCGTACTCTCCAAC
Hoxa4 S	5'- GGTGTACCCCTGGATGAAGA
Hoxa4 R	5'- AAGACTTGCTGCCGGGTATA
Hoxa7 S	5'- AAGCCAGTTTCCGCATCTAC
Hoxa7 R	5'- CTTCTCCAGTTCAGCGTCT
Hoxa9 S	5'- CAATGCCGAGAATGAGAGCG
Hoxa9 R	5'- CTGGTGT TTTGTGTAGGGGC
Hoxa13 S	5'- GCCAAATG TACTGCCCCAAA
Hoxa13 R	5'- CTATAGGAGCTGGCGTCTGA

Table S1: Primers

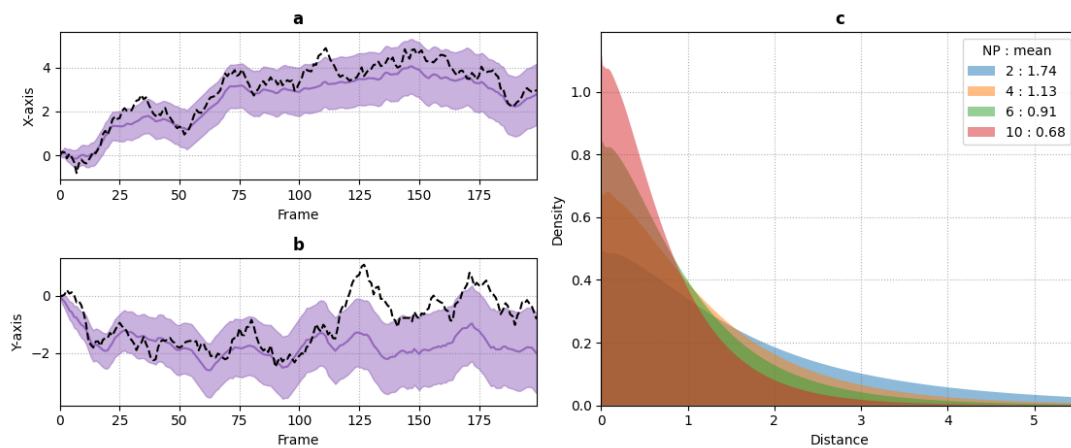


Figure S1: Estimating background movement from tracked particles. (a) Black dashed lines correspond to simulated substrate movement, while purple line is the path estimated for the substrate using GP-FBM. Shaded area correspond to one standard deviation. (b) Positional error distribution in pixel units calculated between substrate simulated and estimated trajectories for systems with 2, 4, 6 and 10 particles.

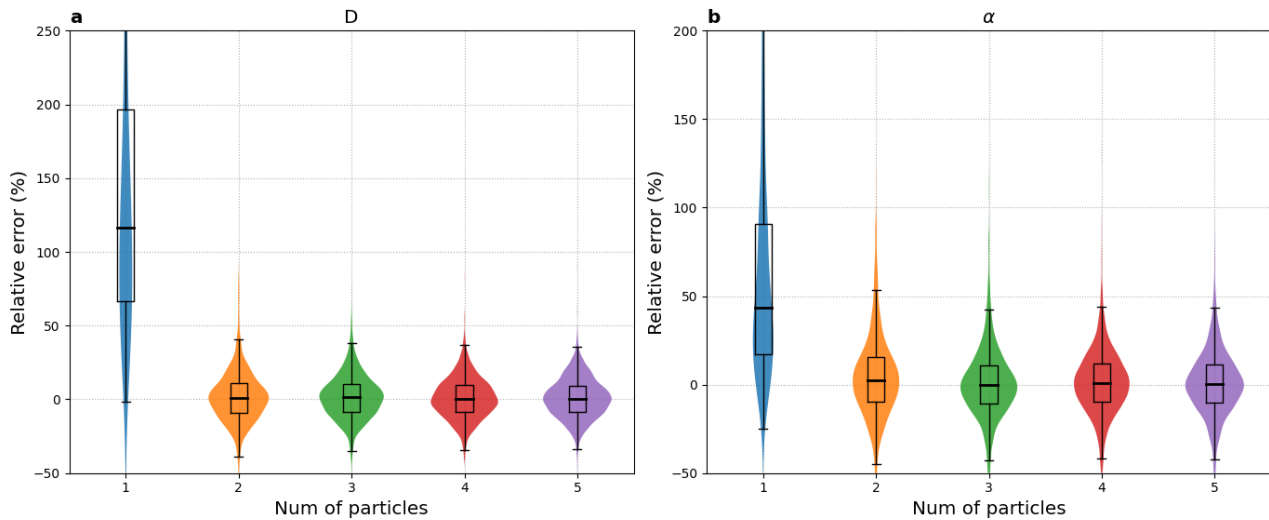


Figure S2: Accuracy gain with number of particles when correcting for substrate movement. If more than one particle under similar context are present, we can use implicit correlations to correct inference. As we can see, there is little gain in precision when using more than 3-4 particles.

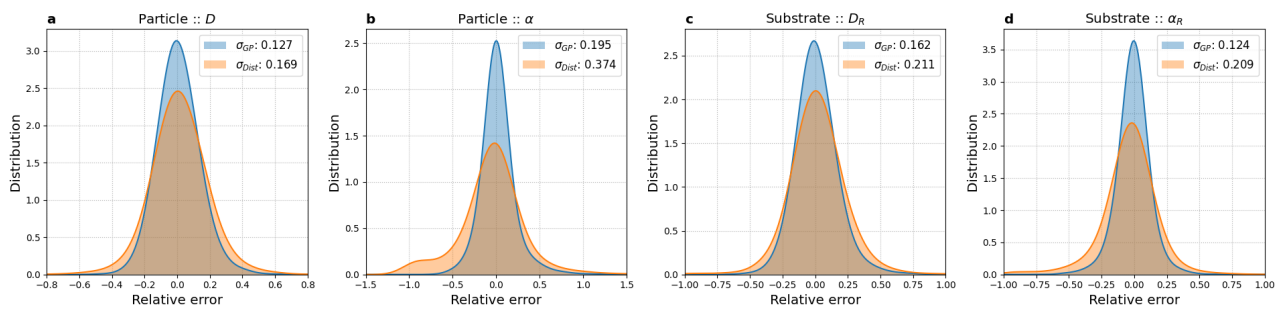


Figure S3: Bench-marking GP-FBM and DDB approaches for inference of diffusion properties with substrate movement A set of 10,000 trajectories of two particles with substrate movement were generated. Simulated values were inferred using GP-FBM (in blue) and DDB (in orange) and relative errors on D and α were plotted for particles and substrate. σ -values correspond to one standard deviation.

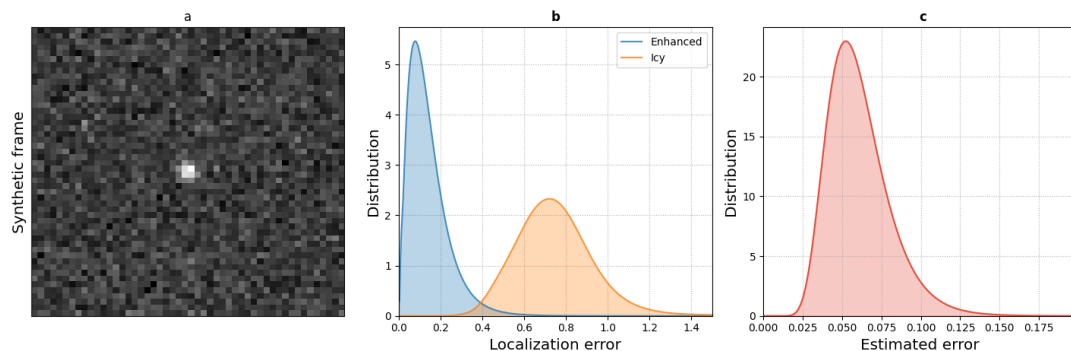


Figure S4: Verifying localization enhancement algorithm via a synthetic movie. To test this method, we generated a synthetic movie with 500 frames for a single particle trajectory. The spot was generated using a 2D symmetric Gaussian shape with deviation 1, background signal was set to 100 and spot intensity to 200. Signal noise was generated in the form of a Poisson distribution by taking original signal as mean. In (a) we have a frame of such a synthetic movie. (b) Spots were tracked with ICY and further enhanced as described in Methods. Distribution of positional errors were calculated using simulated trajectories as reference. (c) Distribution of positional error standard deviations, computed per spot with MCMC.

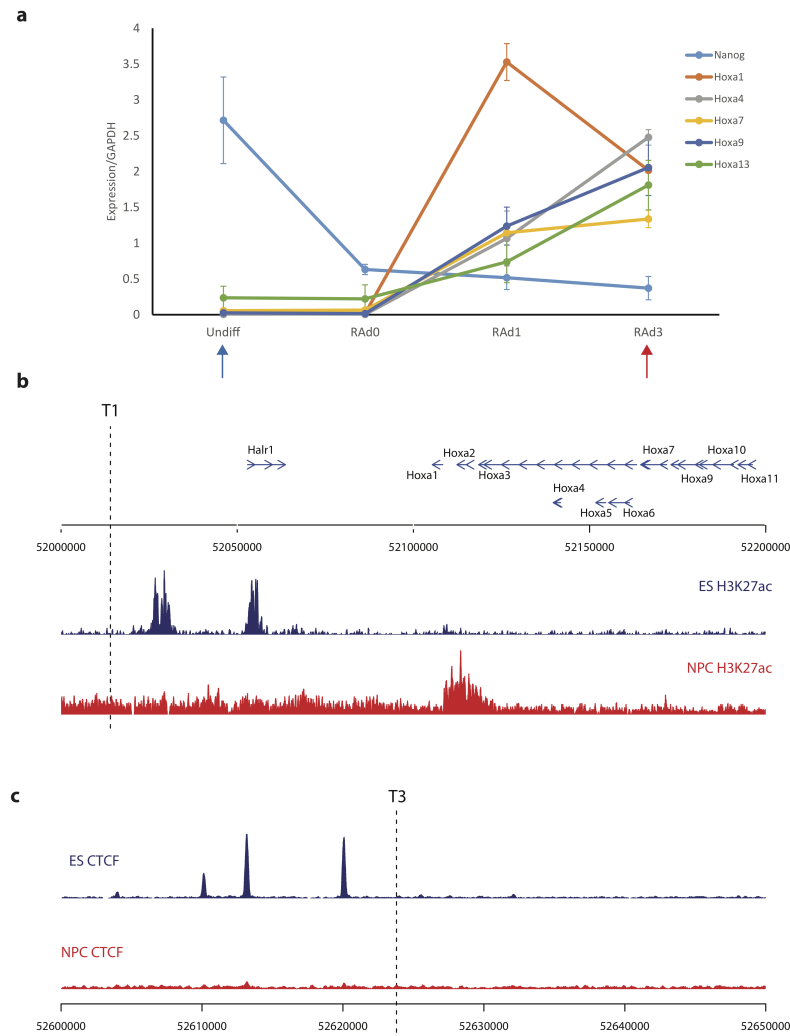


Figure S5: (a) RT-qPCR results, showing expression of Nanog and HoxA genes, normalized to GAPDH, in ES ANCHOR cells before and at different stages of induction by retinoic acid. The graph shows the average results over triplicate experiments. The blue and red arrows indicate the timepoints at which microscopy is performed, as in Fig 4. (b) Gene positions and H3K27ac ChIP-seq tracks in ES (blue) and neuronal precursor cells (red) are plotted for a 200 kb window containing the HoxA cluster and the position of ANCHOR label T1. T1 is <15 kb from an ES-specific putative enhancer. (c) CTCF ChIP-seq tracks in ES (blue) and neuronal precursor cells (red) are plotted for a 50 kb window flanking the ANCHOR label T3. T3 is very close to ES-specific CTCF sites.

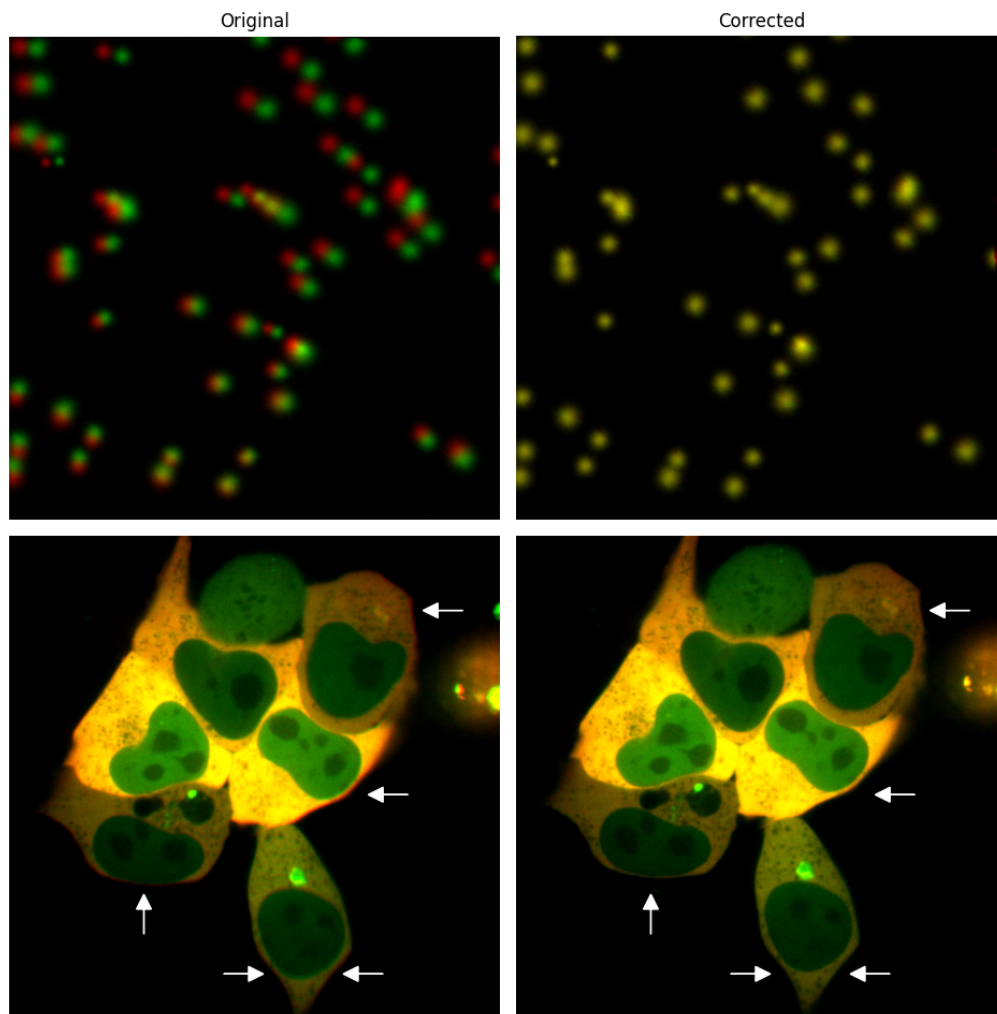


Figure S6: Top images display simulated spots in which we exaggerate effects of camera misalignment. On the bottom, white arrows point to regions with pronounced effects caused by chromatic aberrations. After alignment algorithm is deployed, both channels are properly superposed.