Supplementary information

Nanoscale mapping of DNA dynamics reveals activity-driven genome organization in living human cells

Haitham A. Shaban[1,2] *, Roman Barth[1,3] and Kerstin Bystricky[1] *

1: Laboratoire de Biologie Moléculaire Eucaryote (LBME), Centre de Biologie Intégrative (CBI), CNRS; University of Toulouse, UPS; 31062 Toulouse; France; 2: Spectroscopy Department, Physics Division, National Research Centre, Dokki, Cairo, Egypt; 3: Present Address: Faculty of Applied Sciences, Delft University of Technology; 2628 CJ Delft; The Netherlands

* Correspondence: Email haitham.shaban@ibcg.biotoul.fr , kerstin.bystricky@ibcg.biotoul.fr

CONTENTS

SUPPLEMENTARY NOTE 1

Conversion of flow fields into trajectories

In high-density scenarios where Single Particle Tracking methods reach their limit, dense Optical Flow methods present a powerful tool to investigate local bulk motion of biological macromolecules. Here we determine flow fields of fluorescently labelled DNA and reconstruct virtual trajectories to extract motion at sub-pixel resolution and long-time intervals at the level of the whole nucleus (Shaban et al. 2018). Optical Flow algorithms estimate motion between frames as a field description (Eulerian description) of the underlying continuum motion, evaluated at fixed 'stations', i.e. the pixel positions in the Cartesian coordinate system, regardless of the actual position of particles, which is a powerful approach when the coordinates of single particles cannot be defined. In contrast, actual tracking of particles' coordinates over time is referred to as Lagrangian description. A continuum motion consisting of a finite number of particles can be described in both ways, according to continuum mechanics(Donea et al. 1999; Qiu et al. 2011; Onu et al. 2015).

We took advantage of the fact that every pixel carries information about the underlying continuum motion of particles, i.e. fluorophores, to develop an approach to reconstruct the trajectories of virtual particles. Let us assume a particle is unambiguously identified by its initial position $\vec{r}_0$. Let the coordinates of the particle in Cartesian space be $\vec{\xi}_{r_0}(t)$ at any time point t. Consider further that the Eulerian flow field is known only on fixed coordinates (e.g. a regular grid), but can be evaluated at any position by interpolation of the coordinates of interest. Then, the particle's (Lagrangian) velocity $\vec{q}_L(\vec{\xi}_{r_0}, t)$ at position $\vec{\xi}_{r_0}$ and time t is the same as the Eulerian velocity at $\vec{\xi}_{r_0}(t)$

$$\vec{q}_L\left(\vec{\xi}_{r_0}, t\right) = \frac{\partial \vec{\xi}_{r_0}}{\partial t} \tag{1}$$

Therefore, the trajectory, consisting of the consecutive positions of the particle can be obtained by integration of Equation (1) using the fact that Eulerian and Lagrangian velocities are equal when evaluated at the same position. In situations where particle detection is impossible (e.g. due to high

density of emitters), the Eulerian description of continuum motion can be translated to a Lagrangian description by considering *virtual* particles and using the flow field description to extract their hypothetical trajectories. We consider virtual particles with initial positions at the center of each image pixel, for which flow was estimated by Optical Flow (Supplementary Figure 1a, first flow field highlighted). Note that the flow fields describe the motion *between* frames, whereas particle coordinates are described at the imaging time of each frame. The time evolution, i.e. the trajectory of each virtual particle is reconstructed as follows: From the particle's initial position $\vec{\xi}_{r_0}(t_0) = \vec{r}_0$, the flow field dictates the displacement from frame 1 to frame 2 (dark blue trajectory segment in Supplementary Figure 1d, e), i.e. $\vec{\xi}_{r_0}(t_1) = \vec{r}_0 + \vec{q}_L(\vec{r}_0, t_0)\,\Delta t$, where $\Delta t = t_{i+1} - t_i$ denotes the time between consecutive frames. The current particle coordinates at t_1 do not necessarily coincidence with the regular grid on which the flow field is evaluated. We therefore interpolate the flow field at time t_1 to the particle coordinates $\vec{\xi}_{r_0}(t_1)$ (Supplementary Figure 1b), light blue flow field) and can evaluate the particle coordinates at t_2: $\vec{\xi}_{r_0}(t_2) = \vec{\xi}_{r_0}(t_1) + \vec{q}_L\big(\vec{\xi}_{r_0}(t_1), t_1\big)\,\Delta t$, where $\vec{q}_L\big(\vec{\xi}_{r_0}(t_1), t_1\big)$ denotes the interpolated flow field (i.e. displacement vectors) at time t_1 at the particle coordinates. This procedure is repeated until all flow fields are processed (Supplementary Figure 1c) and the resulting visual particle coordinates are connected to form trajectories (Supplementary Figure 1d, e). Note that extrapolation outside the nucleus and in nucleoli where no signal intensity and therefore no flow field is given, is not considered.

SUPPLEMENTARY NOTE 2

Comparison of confined and anomalous diffusion

In order to test the ability of the Bayesian classification routine used in this study to resolve confined and anomalous diffusion (DR and DA respectively), we simulate particles undergoing either one of the two types of diffusion. Confined diffusion is characterized by a particle diffusing freely within a

sphere of radius R_c with diffusion constant D. The space outside the sphere has the form of an infinite potential impossible to overcome and hence resulting in confinement to the volume of the sphere. Exemplary simulated trajectories are shown in two dimensions in Supplementary Figure 4a-d (left column) for different values of diffusion constant and radius of confinement. Anomalous diffusion is characterized by an effective potential exerting a driving force towards the particle's origin, whose source may be to surround obstacles hindering free diffusion. We model the driving potential as a harmonic potential with the characteristic dimension L_{trap}. The particle feels a spring-like driving force with spring constant $\propto L_{trap}^{-2}$. Exemplary trajectories for anomalous diffusion are shown in Supplementary Figure 4a-d (middle column) for different values of L_{trap}. The potential strength is indicated by colour code. The theoretical MSD is given as in equations 1 and 3 for anomalous and confined diffusion respectively.

For confined diffusion, one can approximate the expression in Equation **Error! Reference source not found.** for short and long time scales:

$\boldsymbol{D\tau \ll R_c^2}$**:** The exponential can be expanded in a Taylor series yielding $MSD_{DR}(\tau) \approx R_c^2\left(1-\left(1-\frac{4D\tau}{R_c^2}+\cdots\right)\right) \approx 4D\tau$, i.e. free diffusion in first order. Effectively, the particle does not feel any confinement for short time lags as the explored volume is much smaller than the confinement volume.

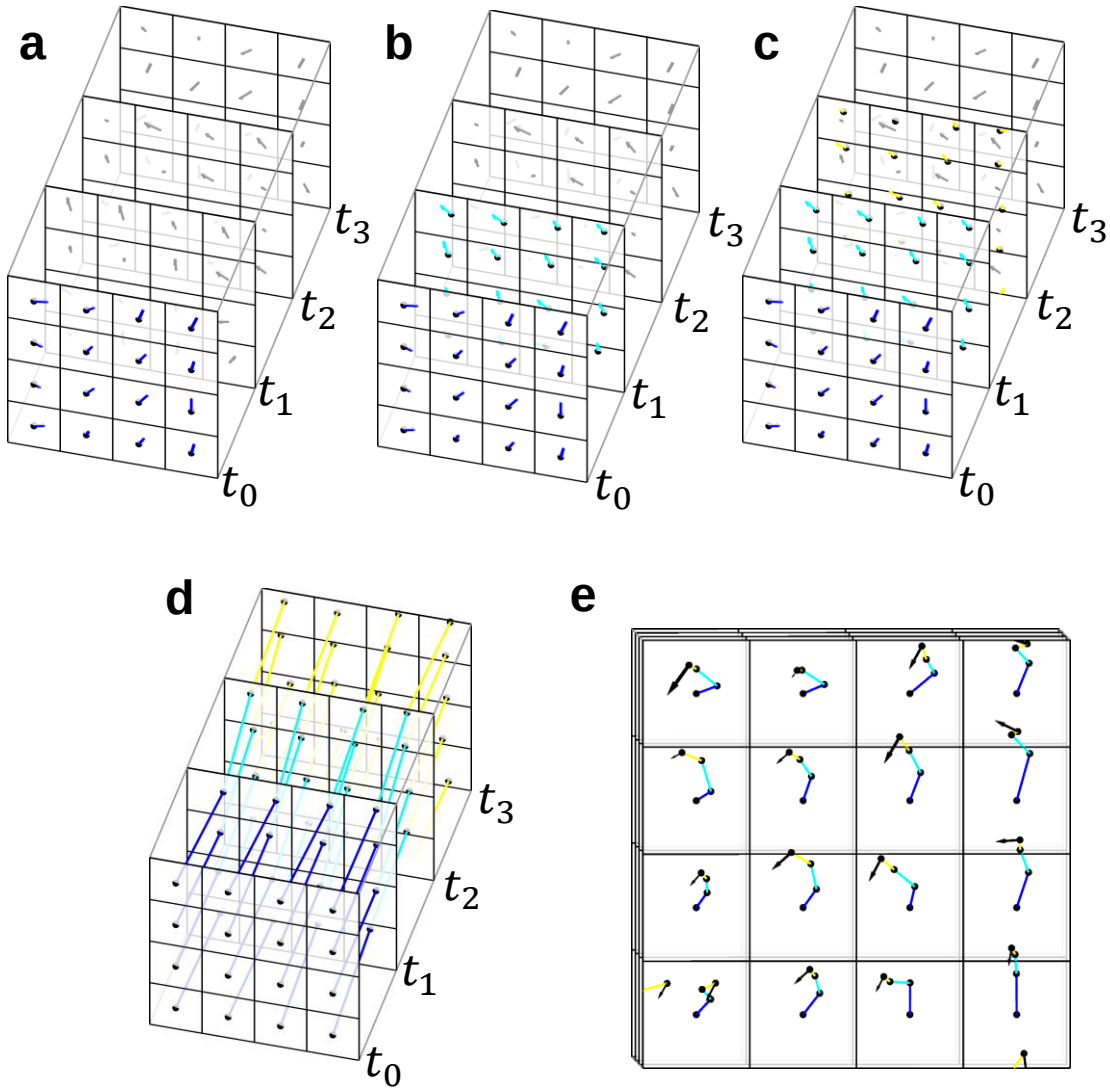
$\boldsymbol{D\tau \gg R_c^2}$**:** The exponential argument is small and therefore $MSD_{DR}(\tau) \approx R_c^2(1-1) = R_c^2$. The confinement is effectively a hard wall potential, impossible to overcome for the particle. For long time lags, the particle therefore explored the whole available volume, but cannot reach any further, resulting in a constant MSD (see Supplementary Figure 3c).

For anomalous diffusion (Equation **Error! Reference source not found.**), the particles move freely for short times, but adapt sub-diffusive behaviour for longer times as the effect of the external potential becomes dominant. However, a particle undergoing anomalous diffusion is not confined in the sense

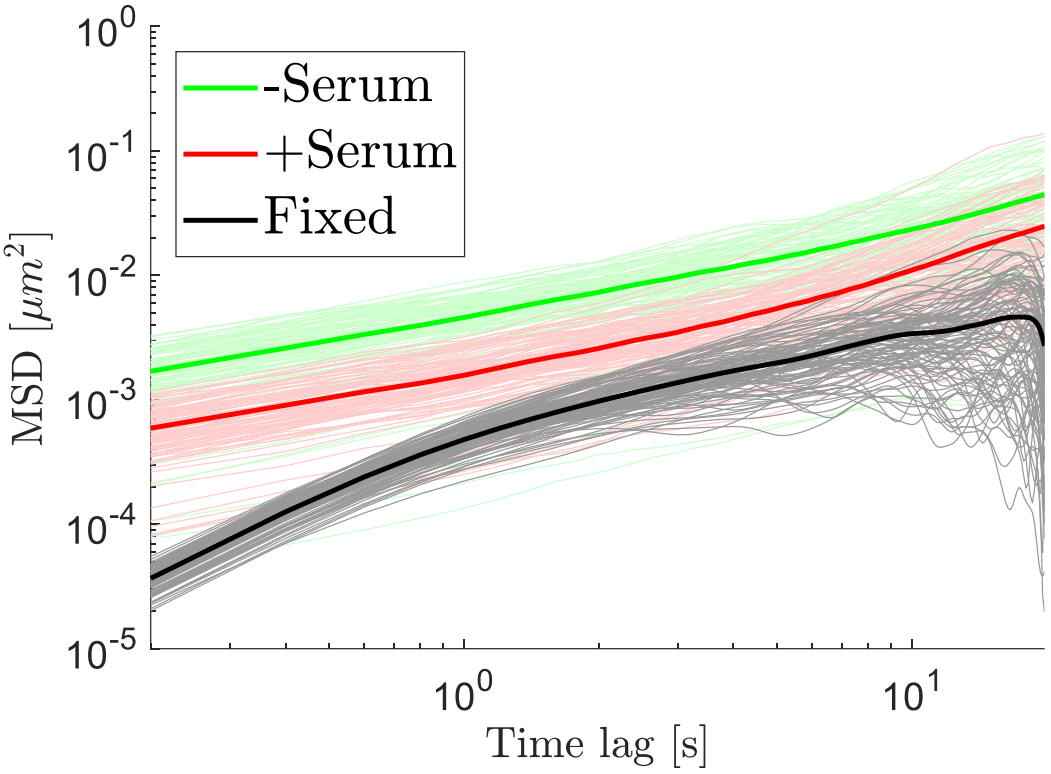
of a hard wall potential and can in principle diffuse in all space (Jin et al. 2007). This leads to a continuously rising MSD for $\tau \to \infty$. Despite the analytical form of the MSD for confined and anomalous diffusion, the behavior for long time lags is a main characteristic for the distinction of the two types of diffusion.

In order to illustrate the theoretical curve shape, exemplary scenarios are shown in Supplementary Figure 4a-d, where different levels of mobility and confinement / anomaly and the ability to resolve the correct type of motion by means of the MSD (right column) are explored. For the scenarios a-c) the limit $D\tau \ll R_c^2$ is not reached within a trajectory length of 150 steps and the shapes of mean MSD for confined (green) and anomalous diffusion (red) are similar. Consequently, the exact type of diffusion could not be resolved. However, for strong confinement (Supplementary Figure 4d), the curves are sufficiently dissimilar and allow extracting the correct MSD model. Experimentally, only a finite trajectory length can be recorded and it is questionable when the trajectory length is sufficient to observe the long time lag limit when no prior information about the particle environment is known. To test whether a reliable distinction can be made for experimentally observed trajectories, an example nucleus is analysed twice. First, the Bayesian classifier is given the choice between free, directed and confined diffusion (and combinations thereof: D, DR, V, DV, DRV). The model selection is shown in Supplementary Figure 4e (left). The majority of trajectories are classified as confined + directed diffusion (DRV), whereas only about 25% of trajectories are classified as purely confined diffusion. Next, the model for confined diffusion is replaced by anomalous diffusion and the analysis is carried out again. First of all, high agreement between the two modes of analysis is seen for trajectories classified as purely Brownian and directed as well as a combination thereof. However, the fraction of anomalous + directed diffusion (DAV) is small compared to purely anomalous diffusion. In particular, about 92% of trajectories classified as DAV were previously classified as DRV. From the remaining trajectories not classified DAV, about 72% are classified as purely anomalous diffusion. These results suggest that only a subset of trajectories classified as combination of confined and directed diffusion is consistent with the combination of anomalous and directed

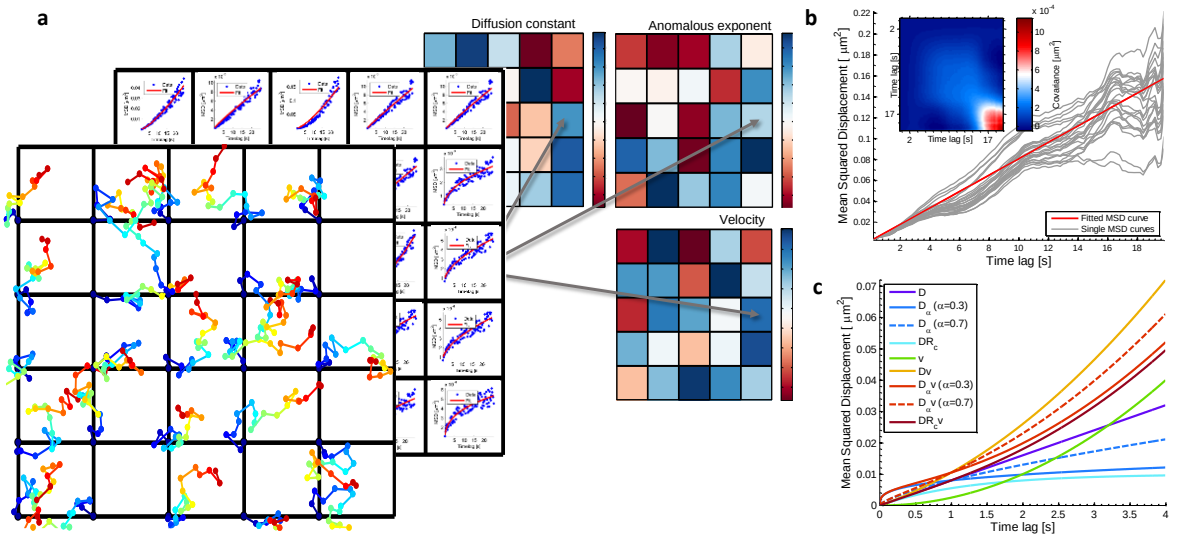
diffusion. The majority of trajectories classified as DRV is preferentially described by purely anomalous diffusion. A reason might be that in the case where only confined diffusion is allowed to describe experimental trajectories, an effective directed transport is needed to account for the continuous rise of the MSD even for large time lags. Finally, it remains unclear if confined or anomalous diffusion is present in experimental trajectories as long as the plateau in the MSD is not reached. Even though strong experimental evidence for confined diffusion of proteins and molecules exists(Saxton and Jacobson 1997) for large biomolecules such as chromatin, a confinement may have several forms such as anisotropic or temporally varying confinement radii with to date unknown sources. The idealized model of a hard wall potential defining the confinement volume may not be appropriate for most biological cases (except for membranes) resulting in a mixed state between confined and anomalous diffusion, which is hard or even impossible to resolve without high spatiotemporal resolution and long-time measurements. The exemplary results and reasoning above suggest that sub-diffusive behaviour observed for chromatin may be best described by anomalous diffusion rather than confined diffusion to prevent misclassification and misinterpretation.



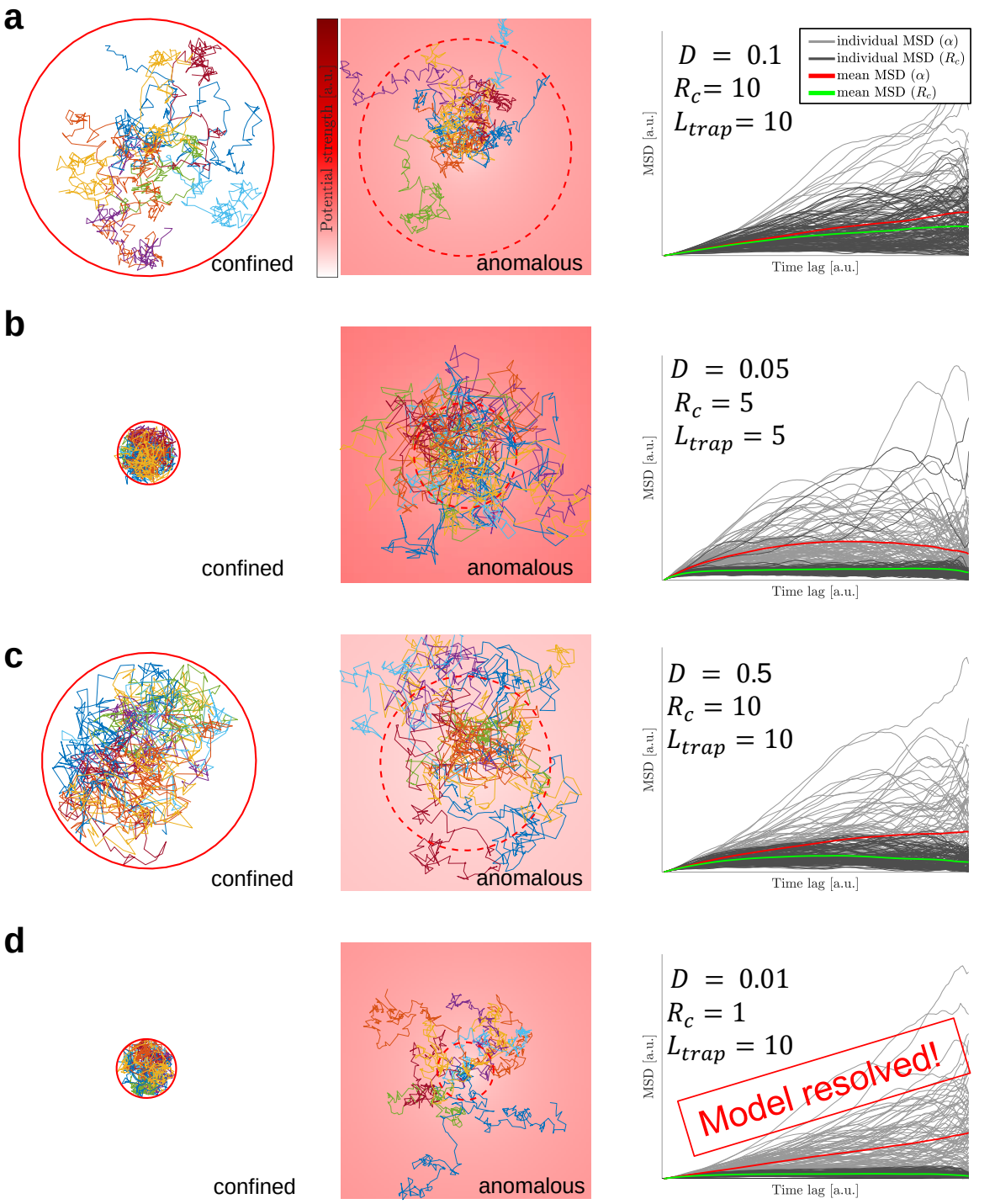
Supplementary Figure 1: Conversion from flow fields to trajectories. **a)** Flow fields are evaluated on a fixed grid given by the pixels of the input images to Optical Flow. The first flow field of the series is highlighted in dark blue. Virtual particles whose initial coordinates coincidence with the center of image pixels are displaced by the flow field at t_0. **b)** The flow field at t_1 is interpolated on the current positions of the virtual particles and displaced according to the interpolated flow field. **c)** The procedure is repeated for all flow fields in the series resulting in subsequent positions of virtual particles given by the displacement of flow fields. **d)** Particle positions are reconnected to form trajectories. The color of trajectory segments denote the flow field which was used to displace the particles from t_i to t_{i+1}. **e)** Quasi-2D representation of virtual particle trajectories over time as shown in d).



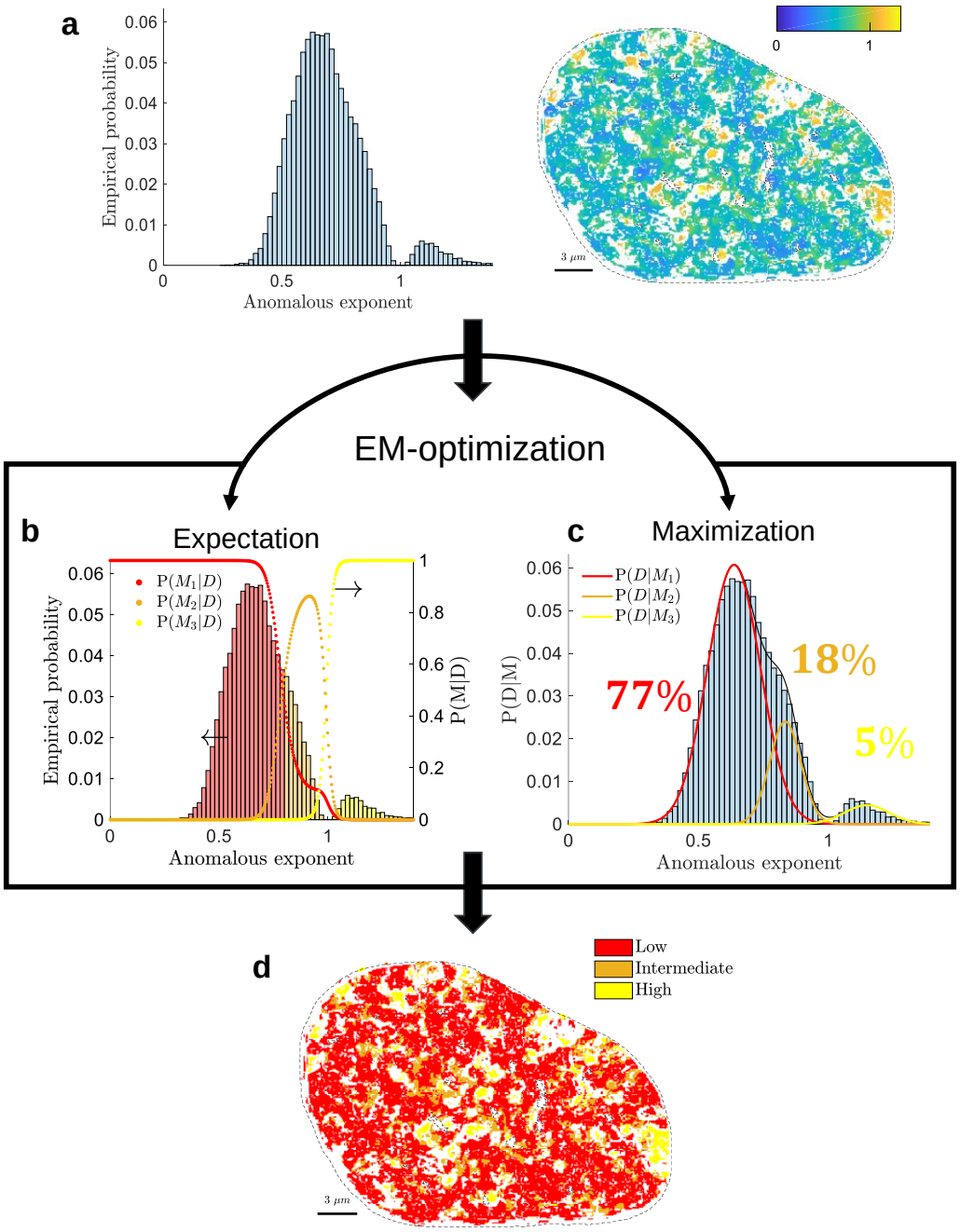
Supplementary Figure 2: Experimental method validation by cell fixation. Trajectories were reconstructed using Optical Flow and MSD curves were calculated. Exemplary individual and average MSD curves are shown for fixed and living U2OS cells labeled by using DNA-SiR kit with and without Serum. Diffusion coefficients for the three average curves were derived by regression yielding $D = (8.7 \pm 0.1) \cdot 10^{-4}\ \mu m^2/s$ for serum-starved cells, $D = (2.6 \pm 0.1) \cdot 10^{-4}\ \mu m^2/s$ for serum-stimulated and $D = (6.1 \pm 0.1) \cdot 10^{-6}\ \mu m^2/s$ for fixed cells. MSD curves show considerably higher MSD values for living cells and diffusion coefficients are one orders of magnitude higher for living cells thus confirming motion well above noise background.



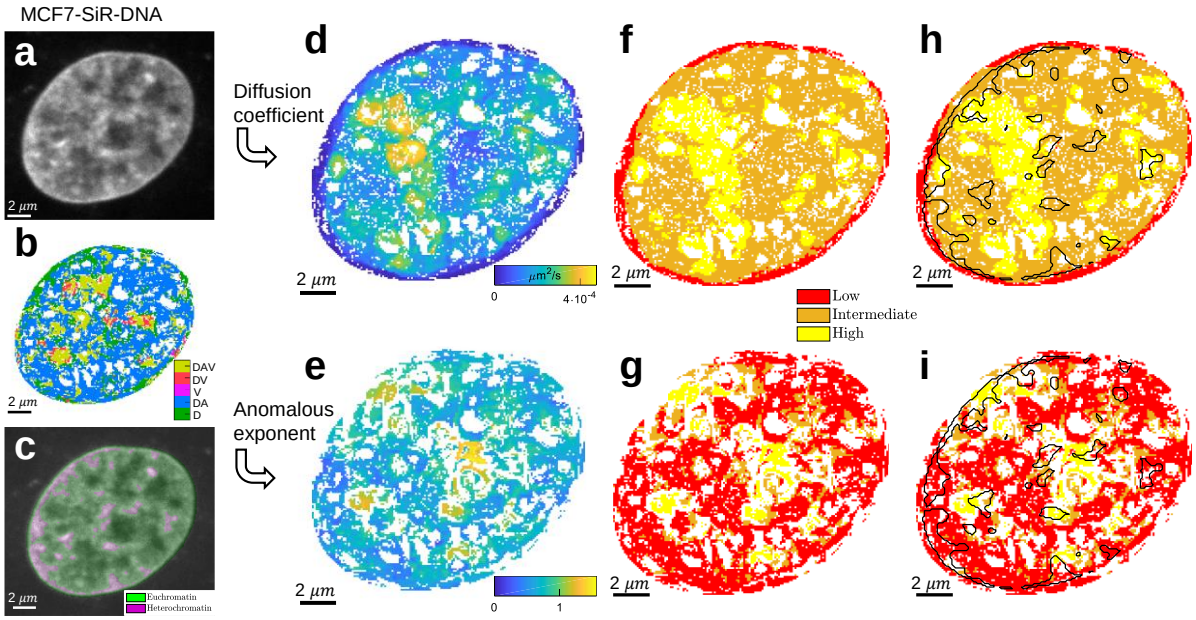
Supplementary Figure 3: Illustration of MSD and Bayesian selection illustration for parameter mapping. **a)** The trajectory of a virtual particle in every pixel inside the nucleus is calculated by using previously determined flow fields. For every pixel, the MSD is calculated. Fitting the MSD, taking into account its neighborhood, results in a set of parameters (diffusion coefficient, anomalous exponent, drift velocity) for each pixel. **b)** The MSD curves for one example pixel and its neighboring pixels. Because errors are correlated in MSD calculation, the covariance matrix is calculated for the adjacent MSD curves (inset). The Cholesky decomposition of the covariance matrix is used in a generalized least squares approach to transform the MSD model and experimental data into a coordinate system, in which the errors are no longer correlated and an ordinary least squares fit is appropriate to find a solution to the optimization problem. Note that the MSD calculation for large time lags has less statistical power than for small time lags due to the lack of pairs to average. Therefore, the covariance matrix shows a high variation for large time lags in turn resulting in small weights for large time lags. The fitted MSD curve is shown in red. **c)** Models to describe the empirical MSD for some exemplary parameter values.


a
confined
Potential strength [a.u.]
anomalous
$D = 0.1$
$R_c = 10$
$L_{trap} = 10$
individual MSD (α)
individual MSD (R_c)
mean MSD (α)
mean MSD (R_c)
MSD [a.u.]
Time lag [a.u.]
b
confined
anomalous
$D = 0.05$
$R_c = 5$
$L_{trap} = 5$
MSD [a.u.]
Time lag [a.u.]
c
confined
anomalous
$D = 0.5$
$R_c = 10$
$L_{trap} = 10$
MSD [a.u.]
Time lag [a.u.]
d
confined
anomalous
$D = 0.01$
$R_c = 1$
$L_{trap} = 10$
Model resolved!
MSD [a.u.]
Time lag [a.u.]
e
D
DR
V
DV
DRV
rel. population size
0
0.2
0.4
0.6
0.8
1
2 μm
D
DA
V
DV
DAV
rel. population size
0
0.2
0.4
0.6
0.8
1
2 μm

Supplementary Figure 4: Comparison of confined (DR) and anomalous diffusion (DA). Confined and anomalous diffusion was simulated in three dimension for 100 particles over 150 frames for different degrees of confinement / anomaly. The resulting MSD curves for anomalous and confined motion are calculated and were input to the Bayesian model classifier used in this study, where the model choice was only left between pure anomalous and confined motion. 10 example trajectories are shown as projection to two dimensions for confined motion (left column) and anomalous motion (middle column). The confinement radius is indicated by a circle (left column) and for comparison as dashed circle (middle column), respectively. The back-driving potential is indicated by color code. Individual and average MSD curves are shown for all trajectories. The following parameter pairs are shown: **a)** moderate diffusion and weak confinement / anomaly, **b)** slow diffusion and moderate confinement / anomaly, **c)** slow diffusion and weak confinement, **d)** very slow diffusion and strong confinement, but weak anomaly. All numerical values in a.u. **e)** An example nucleus stained with DNA-Sir analyzed in two modes: The Bayesian model selection is allowed to choose from free, directed, confined diffusion and corresponding combinations (left) or anomalous diffusion (right) as indicated by color and summarized as stacked bars.

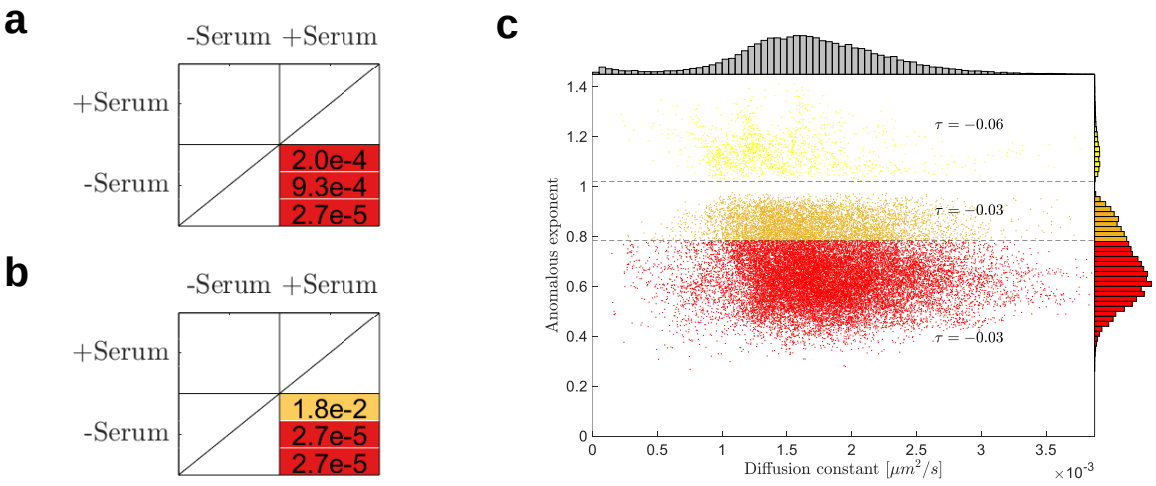


Supplementary Figure 5: General Mixture Model analysis. **a)** An example nucleus showing the spatial distribution of the anomalous exponent across the nucleus (right) and the corresponding histogram of the data with its empirical probability. The General Mixture Model analysis aims at estimating the parameters of underlying populations within the data, here given as three independent Gaussian distributions. Initial estimates are computed by a k-means clustering. The Expectation-Maximization (EM) algorithm iteratively refines data labels and population parameters to maximize the marginal likelihood of observed data until convergence. **b)** Calculating the probability of each data point to belong to one of the populations given the current parameter estimates of each population. The probability sums up to one for each data point and is shown for the three populations on the right axis. **c)** Given a labeling of data into populations, refine the population parameters by weighted maximum likelihood estimates. **d)** Each data point is labeled according to its maximal probability to belong to one of the populations, i.e. $\mathrm{M} = \mathrm{argmax_k} P(M_k|D)$ and is mapped back to the two-dimensional spatial distribution.



Supplementary Figure 6: Hi-D analysis for a different cell line (MCF-7 cells) confirmed that DNA dynamic properties and chromatin compaction are uncoupled. **a)** Fluorescence image of a MCF-7 cell nucleus stained by SiR-DNA to which Hi-D was applied. **b)** Model selection as indicated in the color bar. **c)** Spatial discrimination of high- and low compaction indicating hetero- and euchromatin respectively. **d)** The diffusion coefficient and **e)** anomalous exponent extracted from Hi-D. **f-g)** Mobility groups partitioning DNA dynamics within the nucleus can be classified into one of these populations and mapped back into two dimensions for f) diffusion coefficient and g) anomalous exponent, respectively. **h)** Overlay of diffusive populations, **i)** anomalous exponent and heterochromatin boundaries point to uncorrelated chromatin compaction and dynamics.

.



Supplementary Figure 7: Statistical significance of two-dimensional kernel density estimate and parameter independence. **a)** Statistical significance assessed by a multivariate extension of the Kolmogorov-Smirnov test(Peacock 1983) (gold: $p < 0.01$, red: $p < 0.001$). Columns and rows correspond to growing conditions and numerical values correspond to p-values indicating statistical significance. Shown for the diffusion coefficient (Figure 3c) and **b)** for the anomalous exponent (Figure 3d). **c)** Representation of histograms (Diffusion constant vs. anomalous exponent) from an exemplary nucleus as scatter plot revealing independence of parameters describing diffusion. For each population of the anomalous exponent, correlation quantification is assessed by Kendall's tau coefficient ($|\tau| \leq 1$, $\tau = \pm 1$ indicating perfect (anti-) correlation, $\tau = 0$ indicating parameter independence(Kendall 1938).

REFERENCES

Donea J, Huerta A, Ponthot J, Rodr A. 1999. Arbitrary Lagrangian – Eulerian Methods. *Encycl Comput Mech* 1–25.

Jin S, Haggie PM, Verkman AS. 2007. Single-particle tracking of membrane protein diffusion in a potential: Simulation, detection, and application to confined diffusion of CFTR Cl - channels. *Biophys J* **93**: 1079–1088.

Kendall MG. 1938. A New Measure of Rank Correlation. *Biometrika* **30**: 81. http://www.jstor.org/stable/2332226?origin=crossref.

Onu K, Huhn F, Haller G. 2015. LCS Tool: A computational platform for Lagrangian coherent structures. *J Comput Sci* **7**: 26–36.

Peacock J a. 1983. Two-dimensional goodness-of-fit testing in astronomy. *Mon Not R Astron Soc* **202**: 615.

Qiu G, Henke S, Grabe J. 2011. Application of a Coupled Eulerian-Lagrangian approach on geomechanical problems involving large deformations. *Comput Geotech* **38**: 30–39.

Saxton MJ, Jacobson K. 1997. Single-particle tracking: applications to membrane dynamics.

Annu Rev Biophys Biomol Struct **26**: 373–399.

Shaban HA, Barth R, Bystricky K. 2018. Formation of correlated chromatin domains at nanoscale dynamic resolution during transcription. *Nucleic Acids Res* **46**: e77. http://dx.doi.org/10.1093/nar/gky269.