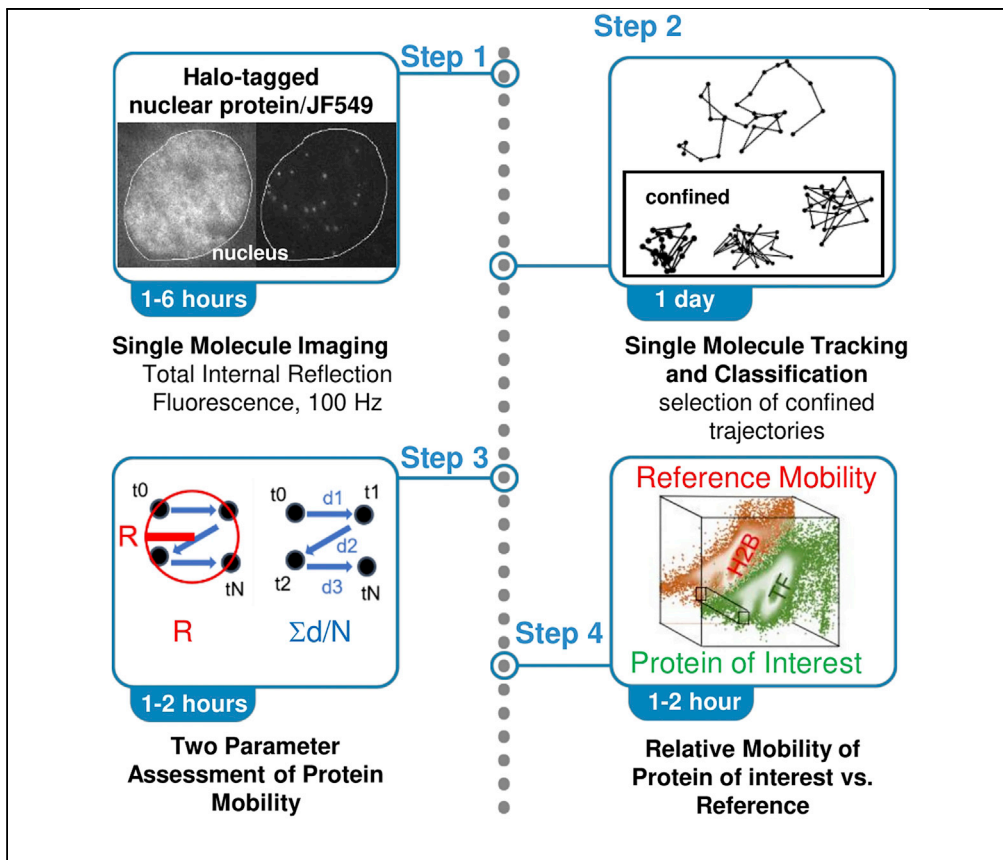


## Protocol

# Two-parameter single-molecule analysis for measurement of chromatin mobility



This protocol provides a two-parameter analysis of single-molecule tracking (SMT) trajectories of Halo-tagged histones in living adherent cell lines and unveils a chromatin mobility landscape composed of five chromatin types, ranging from low to high mobility. When the analysis is applied to Halo-tagged, chromatin-binding proteins, it associates chromatin interaction properties with known functions in a way that previously used SMT parameters did not.

Jonathan Lerner,  
Pablo Aurelio  
Gómez-García, Ryan  
L. McCarthy, Zhe  
Liu, Melike  
Lakadamyali,  
Kenneth S. Zaret

jler@penmedicine.  
upenn.edu

### HIGHLIGHTS

Single-molecule imaging to assay/compare motions of chromatin and nuclear factors

Protocol walks the user through imaging, detection, and tracking of single molecules

Two-parameter analysis allows assay of motions of chromatin-interacting molecules

Translates motions into interaction with functionally diverse chromatin domains

Lerner et al., STAR Protocols 1,  
100223

December 18, 2020 © 2020

The Authors.

[https://doi.org/10.1016/](https://doi.org/10.1016/j.xpro.2020.100223)

[j.xpro.2020.100223](https://doi.org/10.1016/j.xpro.2020.100223)



## Protocol

## Two-parameter single-molecule analysis for measurement of chromatin mobility

Jonathan Lerner,<sup>1,6,8,9,\*</sup> Pablo Aurelio Gómez-García,<sup>2,3,6</sup> Ryan L. McCarthy,<sup>1</sup> Zhe Liu,<sup>4</sup> Melike Lakadamyali,<sup>1,5</sup> and Kenneth S. Zaret<sup>1,7</sup>

<sup>1</sup>Institute for Regenerative Medicine and Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104-6058, USA

<sup>2</sup>Centre for Genomic Regulation (CRG), Barcelona Biomedical Research Park, 08003 Barcelona, Spain

<sup>3</sup>The Institute of Photonics Sciences (ICFO), Barcelona Institute of Science and Technology (BIST), 08003 Barcelona, Spain

<sup>4</sup>HHMI Janelia Research Campus, Ashburn, VA 20147, USA

<sup>5</sup>University of Pennsylvania, Department of Physiology, Philadelphia, PA 19104-6058, USA

<sup>6</sup>These authors contributed equally

<sup>7</sup>Senior author

<sup>8</sup>Technical contact

<sup>9</sup>Lead contact

\*Correspondence: [jlerner@penmedicine.upenn.edu](mailto:jlerner@penmedicine.upenn.edu)  
<https://doi.org/10.1016/j.xpro.2020.100223>

## SUMMARY

This protocol provides a two-parameter analysis of single-molecule tracking (SMT) trajectories of Halo-tagged histones in living adherent cell lines and unveils a chromatin mobility landscape composed of five chromatin types, ranging from low to high mobility. When the analysis is applied to Halo-tagged, chromatin-binding proteins, it associates chromatin interaction properties with known functions in a way that previously used SMT parameters did not.

For complete information on the use and execution of this protocol, please refer to Lerner et al. (2020).

## BEFORE YOU BEGIN

The wide range of chromatin compaction states visible in super-resolution and electron microscopy (Ou et al., 2017; Ricci et al., 2015) elicits a variety of mechanical and electrostatic constraints over the motions of chromatin domains in living cells (Dion and Gasser, 2013; Nozaki et al., 2017). Single-Molecule Tracking (SMT) allows the assessment of the displacements of individual Halo (derived from haloalkane dehalogenase) fusion proteins in living cells, with a high spatiotemporal resolution (Liu and Tjian, 2018).

The following protocol performs the characterization of Halo-tagged histone H2B trajectories in living cells with two parameters: the radius of confinement, estimating the area explored by the molecule, and the average displacement, representing the average frame-to-frame jump (Lerner et al., 2020). Though positively correlated, the two parameters present systematic deviations allowing the resolution of five chromatin mobility groups corresponding to functionally different chromatin domains (Lerner et al., 2020).

The protocol then uses the measured chromatin mobility landscape as a benchmark and scales the mobility of diverse chromatin-binding proteins, thus assaying the interactions with different chromatin domains while scanning the genome (Lerner et al., 2020).



### Construct Halo fusion expression vectors

⌚ Timing: 3–4 days

This protocol has been used and validated for various Halo-tagged nuclear proteins: histone H2B, transcription factors, heterochromatin regulators, lamins; and in a variety of somatic adherent cell lines, as well as pluripotent iPSCs (Lerner et al., 2020).

1. Clone the open reading frame (ORF) encoding the protein of interest in a vector containing a Halo cassette (see [Key resources table](#)), at the N- or C-terminal end.

#### Notes:

Fusion with Halo, a relatively large protein (33 kDa) might impair the biological function of the protein of interest. To circumvent possible functional interferences, we recommend using a flexible linker sequence between Halo-tag and the protein of interest. In the publication of reference (Lerner et al., 2020), we used a polyGlycine/Serine linker of the following sequence: SGGGGSGGGGSGGGGSGGGGS (Iwafuchi et al., 2020).

The Halo-tagged protein should present a proper nuclear and subnuclear localization (e.g., nucleoli, chromocenters, lamina, ...). Please note that some transcription factors display a normal partial cytoplasmic localization (Raccaud et al., 2019). We recommend validation of the biological activity of Halo fusion proteins through additional approaches: western blot, luciferase reporter assays, Electro Mobility Shift Assays, chromatin-immunoprecipitation, ...

Other labeling strategies for SMT in living cells such as SNAPtag or CLIPtag (Crivat and Tarska, 2012) are also compatible with this protocol.

For the recommended reagents and kits for this step, see [Key resources table](#).

### Deliver the vectors into living cells

⌚ Timing: 48 h

2. Plate cells in the imaging set up compatible with the single-molecule microscopy device (cover-slips, 8 well plates, ...). We recommend plating cells to reach a confluency of 70%–80% on the day of imaging. Too high confluency might result in cell overlays which can affect the imaging quality.
3. Using the appropriate transfection system, deliver the expression vectors in the cells.
4. Grow the cells in the adequate cell culture medium for 48 h.

#### Notes:

Using a Tetracycline-inducible lentiviral expression vectors (such as TETO-FUW) ensures the good delivery of the vector in a wide range of cell types, allows a controlled expression of the transgene, and can conveniently generate stable cell lines expressing the Halo-tagged protein of interest (see [Troubleshooting](#)). In the case of cells endogenously expressing the protein of interest, we recommend using expression levels of the Halo-tagged protein near those of the endogenous protein. This can be achieved through clonal selection, with expression levels assayed by western blot, FACS, or RT-qPCR. Alternatively, insertion of the Halo cassette at the endogenous locus through a CRISPR-Cas9 system is a recommended strategy when possible. Other strategies for moderating the transgene expression level are also available (Liu et al., 2018).

We recommend using Phenol red-free medium at all steps of cell culture, to reduce the background during imaging.

We also encourage running a western blot with an antibody specific of the protein of interest to verify the stability of the Halo fusion (see [Troubleshooting](#)).

### Computing setup

⌚ Timing: 1 h

Install MATLAB and upload MATLAB script package.zip (see [Key resources table](#)). Extract the .zip file in the folder of your choice. In the home section of MATLAB, hit "Set Path." Select "Add with subfolders" and browse to the folder containing the extracted files. Hit "Save." MATLAB can now find and run all the scripts necessary for this analysis, notably `scatplot.m`, `SLIMfast.m`, `SMT_classifier.m`, `Two_Parameter_SMT.m`, `densitydifferences.m`, `msdanalyzer.m`.

#### Notes:

A digest version of this protocol can also be found in the Scripts, Lerner-Zaret, as [STEP BY STEP simplified protocol.pdf](#). The protocol does not require an extensive knowledge of MATLAB; however, we recommend performing the tutorial available at [Matlab.com](#), particularly the section about setting up paths and running scripts. We ran this analysis successfully on a computer with ad minima a 2.70 GHz Processor and 6 GB of RAM. Higher specifications can lead to reduced times of analysis.

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant DNA		
pENTR4/HALO	Eric Campeau	Addgene W876-1
Chemicals, peptides, and recombinant proteins		
Janelia Fluor 549 Halo ligand	Promega	GA1110
Critical commercial assays		
MinElute PCR purification kit	QIAGEN	28004
MiniPrep	QIAGEN	27106
MAXIPREP	QIAGEN	12263
Gel extraction kit	QIAGEN	28706
Gibson Assembly kit	NEB	E5510
T4 DNA ligase high concentration	NEB	M0202T
Software and algorithms		
Fiji is Just ImageJ (Fiji) version 2.0.0-rc/1.51f	ImageJ	<a href="https://imagej.net/">https://imagej.net/</a>
MATLAB	MathWorks	<a href="https://www.mathworks.com/products/matlab.html">https://www.mathworks.com/products/matlab.html</a>
MATLAB script package.zip containing:	This paper	2D_tracking_matlab_package Mendeley <a href="https://doi.org/10.17632/hxnhttxpk.1">https://doi.org/10.17632/hxnhttxpk.1</a>
Step by step simplified protocol.pdf, a digest version of the protocol	This paper	STEP BY STEP simplified protocol.pdf
@msdanalyzer, containing scripts used for the analysis of T-MSD curves	( <a href="#">Tarantino et al., 2014</a> )	<a href="https://tinevez.github.io/msdanalyzer/">https://tinevez.github.io/msdanalyzer/</a>

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Confined diffusion model folder, containing the model equations for measuring the radius of confinement, used by the main scripts	This paper	Confined diffusion model
csv example folder, containing FOXA1 or H2B tracking .csv files that can be used to test the protocol	This paper	csv example
Geometry folder, containing scripts for assaying the geometry of trajectories, used by the main scripts	This paper	Geometry
Utils folder, containing various .m scripts used by the main scripts	This paper	Utils
Csv_converter.m, used to convert the SMT data .txt files into .csv files	This paper	Converter_final.m
SMT_Motion_Classifier.m, used for the classification of trajectories	This paper	SMT_Motion_Classifier.m
Two_Parameter_SMT.m, used to perform the two-parameter analysis described in this paper	This paper	Two_Parameter_SMT_V1.m
Density_Differences.m, used to compare the two-dimensional plots of different proteins	This paper	Density_Differences_V1.m
SLIMfast.m, used to perform SMT on the imaging data	(Sergé et al., 2008; Teves et al., 2016)	SLIMfast.m
scatplot.m, a graphical tool to assay the data density in the two-dimensional plots.	<a href="https://www.mathworks.com/matlabcentral/fileexchange/8577-scatplot">https://www.mathworks.com/matlabcentral/fileexchange/8577-scatplot</a>	scatplot.m
<b>Other</b>		
Oni Nanoimager S	ONI	n/a
Other TIRF/single-molecule imaging system	n/a	n/a

**MATERIALS AND EQUIPMENT**

The JF549 Halo ligand should be stored as a 5 μM stock solution in DMSO, at –20°C and protected from the dark, and aliquoted in small volumes to avoid repetitive freeze/thaw cycles. The ligand can be thawed the day of the experiment and diluted directly at 5 nM in the cell culture medium (see [Halo ligand](#) section).

Single-molecule imaging for this protocol is done with a microscope performing Highly Inclined and Laminated Optical sheet (HILO) illumination for optimal background-to-noise ratio, and with a camera compatible with a 100 Hz acquisition. Other imaging methods such as lattice light sheet (Chen et al., 2014a) have not been tested for this protocol. In (Lerner et al., 2020), we used a Nanoimager S from Oxford Nanoimaging Limited (ONI), comporting a temperature and humidity controlled chamber, a scientific Complementary metal–oxide–semiconductor (sCMOS) Hamamatsu Orca Flash 4 V3 camera with a 2.3 electrons rms read noise at standard scan, a 100×, 1.49 NA oil-immersion objective and a 561 nm green laser. Alternatively, we used a Nikon Eclipse Ti microscope equipped with a 100× oil-immersion objective lens (NA, 1.4), and EMCCD (iXon3, Andor) (Chen et al., 2014b).

**STEP-BY-STEP METHOD DETAILS**

**Halo ligand staining**

⌚ Timing: 30 min

At this step, cells are incubated with the fluorescent Halo ligand, which penetrates the cells to specifically and covalently bind to the Halo-tag. Janelia Fluorophore 549 (JF549, see [Key resources table](#)) has been broadly used for single-molecule applications and provides an excellent background/signal ratio. The Halo fusion protein is detected with a laser excitation optimal for JF549 (549 nm).

1. Incubate the cells for 15 min with 5 nM of fluorescent Halo ligand.
2. Remove the medium containing the fluorophore in excess and replace with fresh medium (free of Phenol Red). Repeat three times.
3. Leave the cells in the incubator for 15 min to release unbound fluorophore.
4. Remove the medium and replace with fresh medium (free of phenol red).

⚠ **CRITICAL:** We recommend concealing the sample from the light as much as possible to avoid photobleaching and leaving the light off in the cell culture hood.

### Single-molecule imaging

⌚ **Timing:** several hours

This is the step of detection of fluorescent single molecules of the Halo-tagged protein of interest, using HILO imaging.

5. Set the sample in the observation chamber, ideally allowing the control temperature and CO<sub>2</sub> levels. If no CO<sub>2</sub> control is possible, we recommend imaging for relatively short times (<1 h), alternating between samples in and out from a cell incubator to avoid impairing cell function.
6. Using a laser compatible with JF549 fluorophore excitation (peak at 549 nm) and starting at a moderate power coupled with a 100 ms exposure, detect the JF549 fluorescence in the sample. Adjust focus and illumination angle for optimal signal.
7. Take a single snapshot of the nucleus/nuclei to be acquired. It can be used for further detection of nuclear structures (e.g., nucleoli, condensates, ...), as well as for assaying the intensity of fluorescence prior to single-molecule acquisition.
8. Set the camera exposure time at 10 ms and increase progressively the laser power, until single molecules are visible. Photobleaching the nucleus might be necessary and take several seconds ([Methods Video S1](#)).
  - a. Adjust the illumination angle, z-axis and laser power to obtain an optimal molecule density and background-to-signal ratio (see [Troubleshooting](#) and [Methods Videos S1](#) and [S2](#)).
  - b. As the elevated laser power causes fast photobleaching, start recording the single molecules as soon as possible. Typically, 5,000 frames are recorded (total acquisition duration: 50 s).

#### Notes:

For step 6, we recommend starting with a low laser power. Typically, we use values ranging from 50 to 500 W/cm<sup>2</sup> for imaging. If you observe rapid photobleaching, decrease the laser power.

For step 8, the optimal range for acquisition lies between imaging as seen in [Methods Videos S1](#) and [S2](#). We recommend acquiring between 20–30 individual nuclei to ensure having enough data for further analysis. Our protocol was established using a 10 ms exposure for single-molecule imaging. Exposure times between 5 and 30 ms should be compatible with this protocol but might lead to differences in the accuracy of the measure of the average displacement and radius of confinement leading to two-dimensional plots different of the examples provided here.

## Single-molecule tracking analysis I: tracking

⌚ Timing: several hours

This is the first step of computational analysis for the single-molecule imaging data. The input is a stack of .tif files generated at the single-molecule imaging step (Figure 1A).

We preferentially use tracking software SLIMfast.m (Normanno et al., 2015; Teves et al., 2016), based on the tracking algorithm Multiple Target Tracing (MTT) (Sergé et al., 2008), both publicly available (please also cite these publications whenever referring to our protocol). Alternatively, this protocol is also compatible with tracking files generated by TrackMate (ImageJ), which can be directly used as inputs in the following section.

For each .tif stack, this step outputs a .txt tracking data file containing a series of successive spatio-temporal coordinates (x, y, t) for each single molecule trajectory (or motion track) over the time of acquisition. The tracking step here is using the MATLAB-based SLIMfast.m script (Teves et al., 2016), a modified version of MTT (Sergé et al., 2008), both publicly available. Please cite these publications whenever referring to this protocol.

9. Start MATLAB and run SLIMFast.m

10. In SLIMfast, hit Load > Batch, and select the .tif stacks to analyze. Once all the files loaded, adjust the Optics parameters: fluorophore (emission for JF549=580 nm), objective (Numerical Aperture, NA) and camera (Pixel size) used during single-molecule imaging (Figure 1B and Methods Video S3).

Methods video S3. Detection of molecules by SLIMfast.m, related to single-molecule tracking analysis I (step 10)

This video displays how to add the tracking script package to the path, and how to detect molecules in successive frames of the .tif single-molecule imaging stacks.

11. Hit START. A progress bar (“localization”) will pop up, representing the progressive detection of particles in the successive frames of each .tif stack (Methods Video S3). Once detection is finished, a signal (“Done!”) pops up. This step produces several output files sharing the name of the initial .tif stack, among which a .mat file.

12. In SLIMfast, Load > Particle data and select an individual .mat file. An overlay of all particles detected over the course of acquisition is displayed (Figure 1C and Methods Video S4).

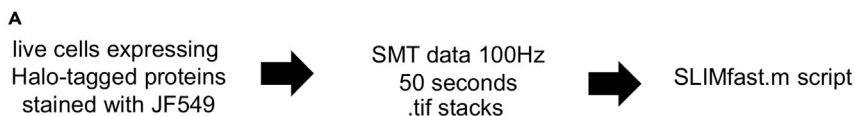
a. To track the molecules frame-by-frame, in the overlay menu, hit OPT > Tracking, and change the Max. expected Diffusion Coefficient to  $3 \mu\text{m}^2/\text{s}$ .

b. Hit “GEN TRAJ” in the overlay window. A “Processing Frame” progress bar pops up. Once completed, the output files can be saved in the directory of choice. This step generates a .txt file with the same name than the initial .mat file, which will be converted as .csv and used as an input for the two-parameter mobility analysis (Figure 1D).

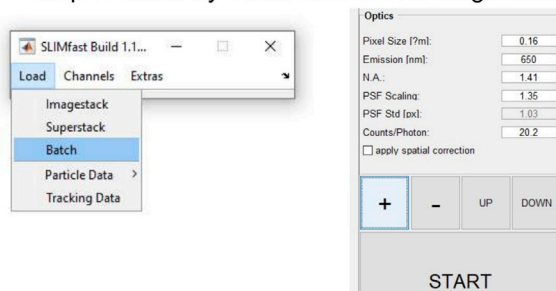
Methods video S4. Single-molecule tracking by SLIMfast.m, related to single-molecule tracking analysis I (step 12)

This video displays how to upload the file containing the successive position of molecules in the .tif stacks, and to run the MTT tracking algorithm to establish trajectories.

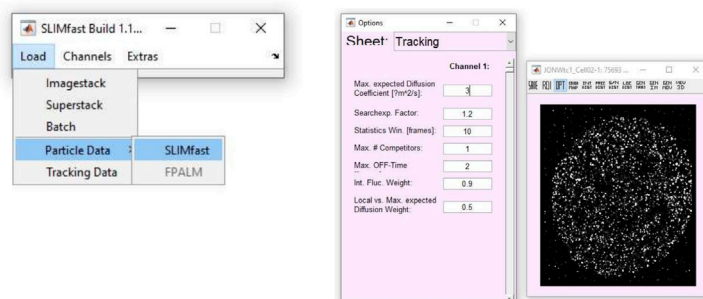
⏸ **Pause Point:** Step 11 can be timely and can take up to several hours to complete for large data sets. For example, for 50 cells, using the minimal configuration indicated in the Computing setup section, it can take up to 3–4 h.



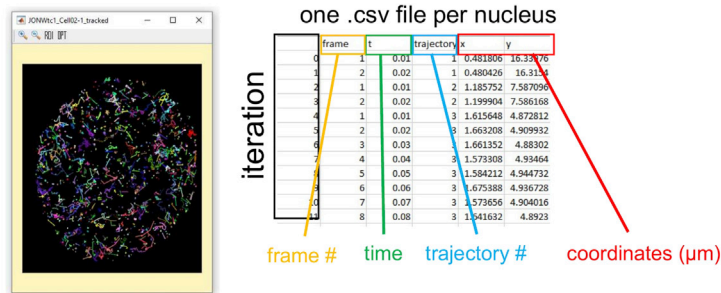
**B** Step 1: frame-by-frame detection of single molecules



**C** Step 2: motion tracking



**D** outputs  
one .csv file per nucleus



**Figure 1. Data pipeline for single-molecule tracking with SLIMfast.m**

(A) Single-molecule imaging .tif stacks are used as inputs for SLIMfast.m.

(B) Screenshots from SLIMfast.m indicating how to load several .tif stacks as a batch to detect molecules in each consecutive frame, as well as the parameters that can be modified by the user depending on the characteristic of the microscope in use (see Notes of the [Single-molecule tracking analysis I: tracking](#) section for details).

(C) Screenshot from SLIMfast.m indicating how to load the particle data from the precedent step, for the script to trace single molecule trajectories. An example of the overlay of all detected particles in a nucleus expressing FOXA1-HALO, over the course of 50 s is displayed.

(D) Example of trajectories established by the tracking algorithm in a single nucleus over the course of 50 s. A part of the corresponding .csv files is displayed, as provided after conversion by Csv\_converter.m. The consecutive columns indicate #1, iteration; #2, frame of detection of the trajectory first step; #3, time = frame × step duration(0.01 s); #4 and #5, x and y coordinates (µm).

**Notes:**

For step 10: tracking algorithm MTT performs the detection of single particles by fitting a two-dimensional Gaussian to the Point Spread Function (PSF), over sliding windows of 7 × 7 pixels



(Sergé et al., 2008). The shape and size of the PSF is function notably of the emission wavelength of the fluorophore in use (580 nm for JF549) and on the Numerical Aperture (NA) of the objective used for imaging (1.45 for the Oni Nanoimager S). In our experiments (Lerner et al., 2020), the PSF was approximately of 0.8. To discriminate false and true detected particle, a probability of false alarm (PFA) is set by default at  $10^{-6}$  by the tracking algorithm (Sergé et al., 2008). The Count/photons value is a technical feature intrinsic to the camera used to estimate the number of detected photons which does not affect the tracking parameters. Please note that all parameters can be freely modified by the user if desired.

For step 12, the Maximal expected Diffusion Coefficient value reflects the maximal frame-to-frame displacement expected for single particles, and was set as most accurately reflecting the motions of single particles seen in our experimental conditions. If desired, the user can input different values, which might be relevant if using exposure times different than 10 ms (see previous section).

### Single-molecule tracking analysis II: motion track classification

⌚ Timing: several hours

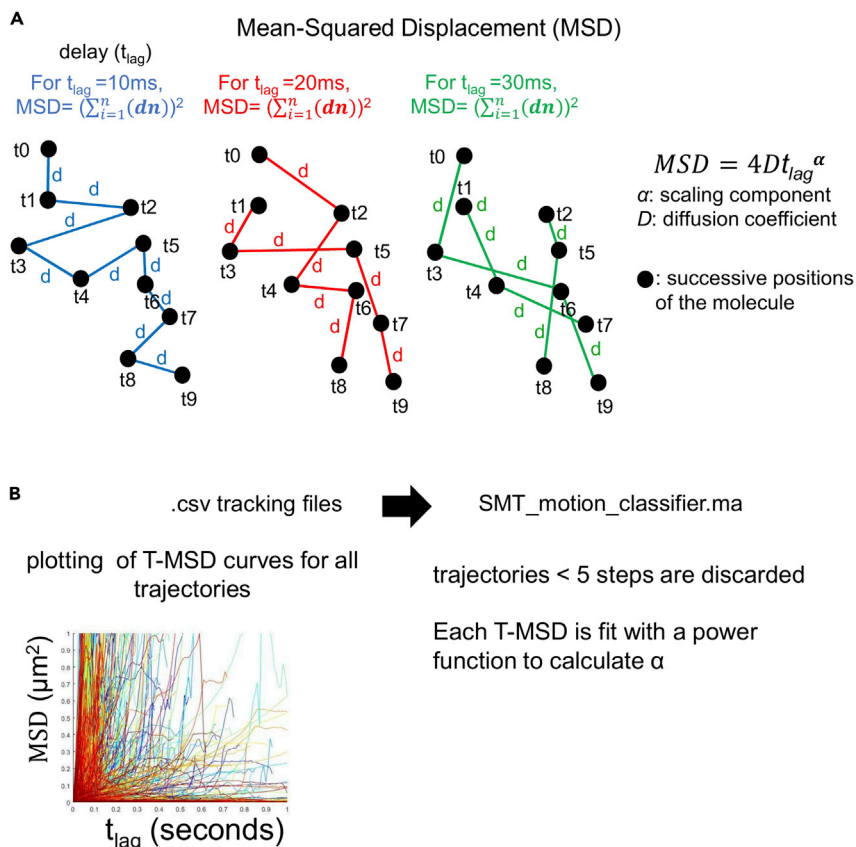
The two-dimensional mobility analysis is only performed on chromatin-interacting molecules. Thus, during the second step of data analysis, free-diffusing molecules are identified and filtered out by classifying trajectories generated by SLIMfast.m on the basis of their mean-squared displacement (or MSD, Figure 2A, Ernst and Köhler, 2012). The tracking .txt file generated at the previous step are converted to .csv and used as input.

The SMT\_Motion\_classifier.m script performs the following operations:

The SMT\_Motion\_classifier.m script performs the following operations: (1) discarding of trajectories shorter than 5 frames to optimize the classification; (2) computing the T-MSD curve (or  $MSD=f(t_{lag})$ , where  $t_{lag}$  is the delay in seconds) for each individual trajectory (Figure 2B); (3) fitting of each full T-MSD curve with a power law distribution (Michalet and Berglund, 2012) the scaling component  $\alpha$  (Figure 2B and see Ernst and Köhler, 2012); (4)  $\alpha$ -based classification of trajectories:  $\alpha \leq 0.7$ : confined;  $0.7 < \alpha < 1$ : Brownian;  $\alpha \geq 1$ : directed (Figure 3A); (5) in parallel, detection of “Butterfly” trajectories with mixed confined and directed/Brownian behaviors (“Butterfly”), on the basis of a displacement step (or jump) at least 1.5 times superior to the average displacement, and of a sum of displacements bigger than 8 times the average frame-by-frame displacement (Figure 3B); (6) splitting of Butterfly trajectories into confined and Directed/Brownian segments based on geometrical properties: Brownian/directed segments have at least 30% of its points outside of a polygon (convex hull) formed by the points of previous and posterior segments of the trajectory (Figure 3C); (7) classification of confined trajectories (including Butterfly confined segments) based on their average frame-to-frame displacement. Trajectories with average displacement below 100 nm are considered as confined, corresponding to molecules interacting with chromatin (Figure 3D).

For each imaged nucleus, seven “msd\_results”.mat files are generated as outputs. They correspond to each class of trajectories: All, Brownian, Directed, Confined with high (confined\_High\_Jump) or Low (confined\_Low\_Jump) average displacement, Butterfly Confined (segments\_confined) and Directed (segments\_directed) segments.

13. Gather the batch of .txt files generated by SLIMfast.m at the previous step (each .txt file corresponds to one initial .tif stack) in a unique folder, which will be made directory for MATLAB (Methods Video S5).
14. Run csv\_converter.m (Methods Video S5), which converts each .txt file into a .csv file containing the following information (column-by-column): iteration (1-n, with n number of rows of the csv



**Figure 2. Mean-squared displacements (MSD) of single-molecule trajectories**

(A) Measurement of the MSD for the same single-molecule trajectory, using different delays ( $t_{lag} = 10, 20$  or  $30$  ms), including the mathematical expression of MSD in function of the diffusion coefficient  $D$  and the scaling component  $\alpha$ . Each black dot represents the successive position of the molecule, with the associated timing of detection.

(B) .csv tracking files are used as inputs for the script SMT\_motion\_classifier. An example of T-MSD curves obtained for a single nucleus expressing FOXA1-HALO over 50 s of acquisition at 100 Hz is displayed.

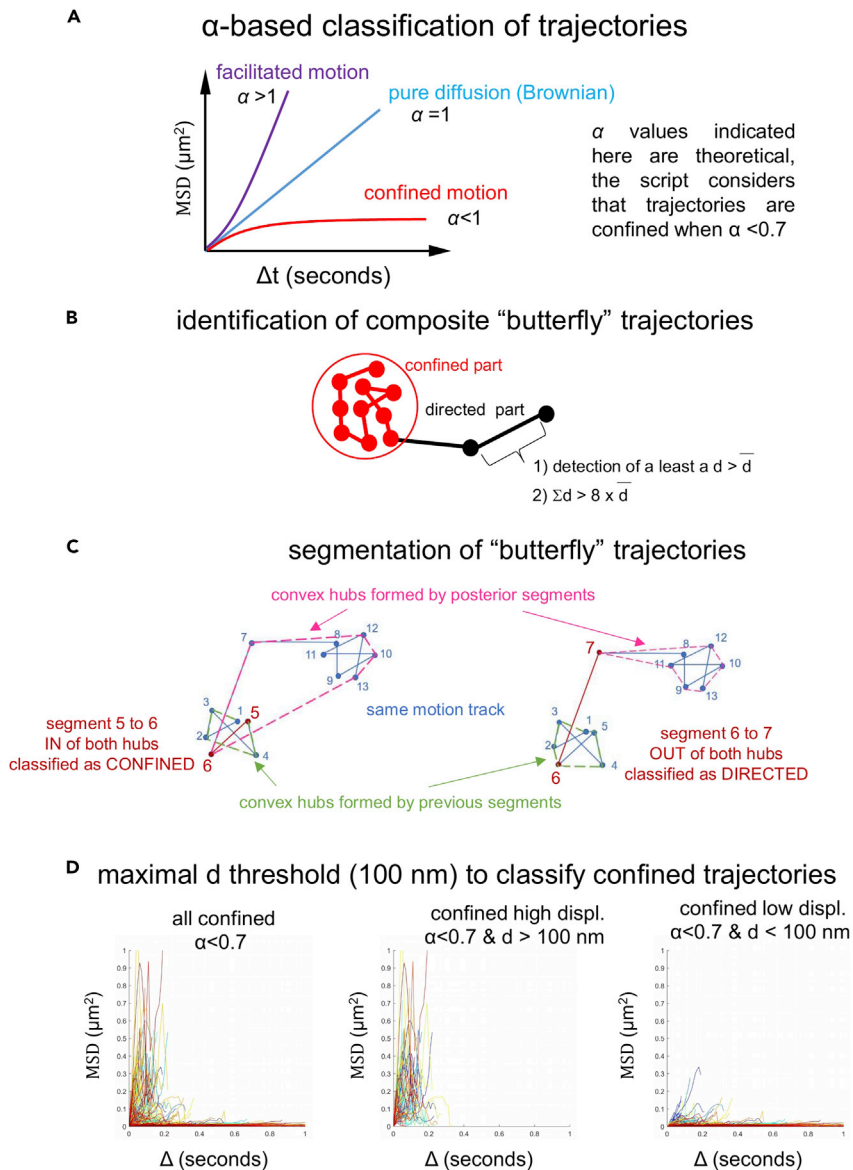
table), frame (the frame on which each single molecule was detected), time ( $t = \text{frame} \times \text{exposure}$ ), trajectory (Identification number of the trajectory), and x and y coordinates, in  $\mu\text{m}$ .

- Run SMT\_Motion\_Classifier.m. Select the batch of tracking .csv files to analyze. To display additional figures, or change parameters, please refer to the User Manual provided in the MATLAB script package. After execution, the script outputs a new folder starting with "MSD\_results" and containing 7 .mat files (starting with "msd\_results").

**Pause Point:** Step 15 can be timely for large datasets and depending on your computing power, this can be used as a Pause point for an hour or more (see [Troubleshooting](#)). For example, for 100,000 trajectories, using the minimal configuration indicated in the Computing setup section, the script can take up to 2–3 h.

**Notes:**

Thresholds for detection of butterfly trajectories were set empirically, optimizing the values of thresholds for an optimal classification, as assayed by visual inspection of the confined and directed parts. For a first-time analysis, we recommend using the defaults parameters. Nevertheless, advanced users can freely modify all parameters of the data analysis in lines 25–54 of the SMT\_Motion\_Classifier.m, to vary the conditions.



**Figure 3. Basis of trajectory classification by SMT\_motion\_classifier.m**

(A) The scaling component  $\alpha$ T-MSD curve is used to keep confined trajectories, corresponding to chromatin associated molecules; and discard facilitated and pure diffusion trajectories, corresponding to molecules non-associated with chromatin.

(B) Composite (“butterfly”) trajectories, presenting confined and diffusing components, are identified on the basis of two criteria: a displacement ( $d$ ) larger than the average frame-to-frame displacement ( $\bar{d}$ ), and a sum of displacements ( $\Sigma d$ ) larger than 8 times  $\bar{d}$ .

(C) “Butterfly” trajectories are segmented into their confined and Brownian/directed parts. Each number represents the successive time of detection of the single molecule. Example of a fragment considered as confined (from time 5 to 6) and directed (from time 6 to 7) are displayed. The Brownian/directed segment must fulfill the condition that a minimum percentage of its points lie outside the polygon (convex hull) formed by the points of previous and posterior segments of the trajectory.

(D) To discard confined tracks with high displacements, a threshold for a maximal frame-to-frame displacement of 100 nm is set. T-MSD curves are shown here before and after filtration for all confined tracks in a single nucleus expressing FOXA1-HALO. T-MSD curves after filtration (right panel) present a confined aspect only.

Users also have the option not to use the Confined segments of the butterfly tracks in the analysis, just by selecting the `Confined_low_jump.mat` file in the following part.

For step 13, we provide sample .csv files with our MATLAB script package

### Single-molecule tracking analysis III: measure of the radius of confinement and of the average displacement, and mapping of the data density

⌚ Timing: several hours

The third step of single-molecule data analysis consists of assessing the two parameters used for characterizing the mobility of individual molecules. The `Two_Parameter_SMT.m` script is run on trajectories of nuclear proteins in a chromatin-bound state, defined as the pool of confined trajectories and of the confined segments of the Butterfly trajectories (Figure 3 and Figure 4A).

To measure the radius of confinement (RC), the `Two_Parameter_SMT.m` script fits each T-MSD curves with a circle confined diffusion model (Wieser and Schütz, 2008) to provide (1) the corresponding radius of confinement RC; (2) the diffusion coefficient at short time scales D; (3) a constant offset corresponding to the localization precision limit (Figure 4B).

The average displacement (AVE) is computed by measuring the average Euclidean distance between two consecutive positions of the molecule in each motion track (Figure 4C).

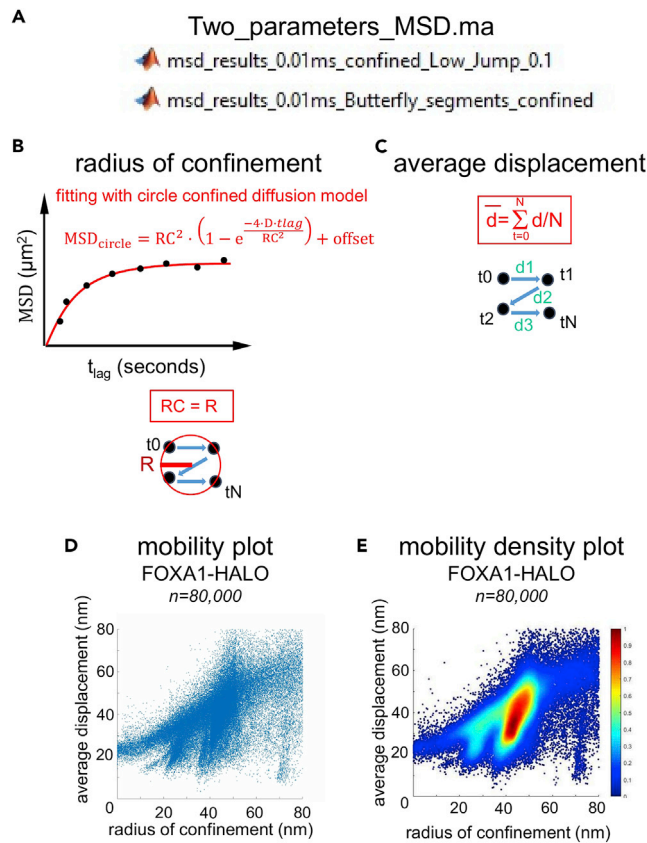
The different mobility chromatin domains are arbitrarily defined based on the correlation between the two parameters, and based on the data density: very low (vLMC) and low (LMC) mobility chromatin are seen as two marked deviations from the line of proportionality. Along the correlation line, Intermediate (IMC) and High (HMC) are distinguished based on the bimodality of the distribution of the radius of confinement: IMC presents lower RC than HMC (Figure 5A). The limits for radius of confinement (RC) and average displacement (AVE) are set as follows (Figure 5A): very Low Mobility Chromatin (vLMC):  $0.020 < RC < 0.030$  &  $0.015 < AVE < 0.029$ ; Low Mobility Chromatin (LMC):  $0.035 < RC < 0.05$  &  $0.020 < AVE < 0.035$ ; Intermediate Mobility Chromatin (IMC):  $0.015 < RC < 0.0375$  &  $0.029 < AVE < 0.036$ ; High Mobility Chromatin (HMC):  $0.0375 < RC < 0.055$  &  $0.0375 < AVE < 0.06$ ; very High Mobility Chromatin (vHMC)  $0.055 < RC < 0.3$  &  $0.06 < AVE < 0.3$ .

The outputs generated by the script are (1) `trackingdata.csv` files, in which each row represents a motion track, with its corresponding RC (column 1), average displacement (column 2) and track duration (column 3); (2) two representations of the radius of confinement versus average displacement assigned to each motion track: a scatter plot (Figure 4D), and a density scatter plot (Figure 4E) (see `Scatplot.m`). The scripts also generate a 3D mobility plot with x: radius of confinement, y: average displacement, z: data density as measured by `scatplot.m`.

16. Run `Two_Parameter_MSD` and hit "add." Go to the appropriate directory and open `Butterfly_segments_confined.mat` and `confined_Low_jump_0.1.mat` for the set of data to analyze. Hit "Done Selecting Files" and wait for the script to execute.
17. Save the figures generated by the script under the desired format.

⏸ **Pause Point:** Step 16 can be timely for large datasets and depending on your computing power, this can be used as a Pause point for an hour or more. For example, for 100,000 trajectories, using the minimal configuration indicated in the Computing setup section, the script can take up to 30 min to 1 h.

**Notes:**



**Figure 4. Two parameters of mobility are assessed for confined trajectories**

(A) Confined tracks, corresponding to chromatin-interacting molecules are defined as the pool of confined below the jump threshold (confined\_Low\_Jump\_0.1) and the confined portion of butterfly trajectories (Butterfly\_segments\_confined).

(B) The radius of confinement is measured by fitting individual T-MSD curves with a circle confined diffusion model, and estimates the circle encompassing the motion track.

(C) The average displacement corresponds to the average Euclidean distance between two consecutive positions of the molecule in the motion track.

(D) Scatter plot of the radius of confinement versus average displacement for  $n = 80,000$  individual trajectories of FOXA1-HALO.

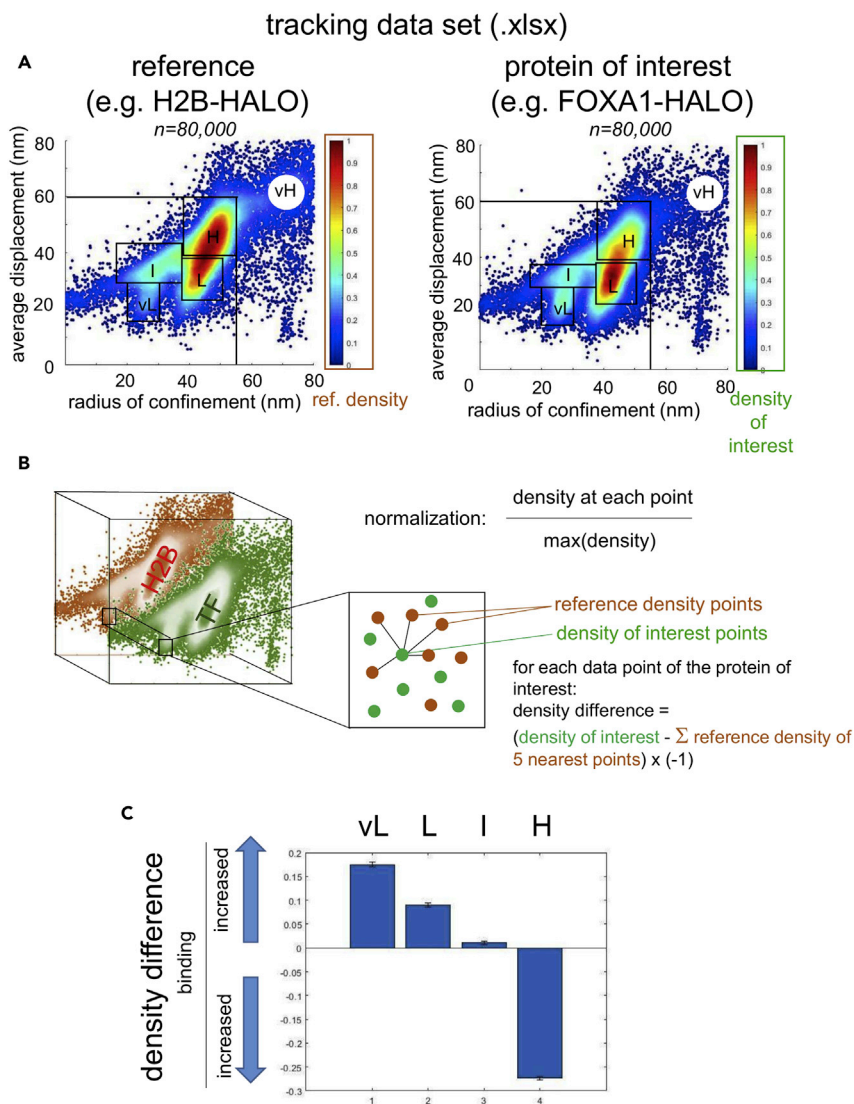
(E) Visualization of the same plot than D, but displaying data density using `scatplot.m`.

In our model, IMC and LMC were distinguished on the basis of the mobility of heterochromatin constituents. In fact, heterochromatin regulators bind preferentially to vLMC and LMC (Lerner et al., 2020). IMC is bound by transcription factors and heterochromatin regulators and might be constituted by composite chromatin domains.

Boundaries in radius of confinement and average displacement chosen for each population were chosen as non-overlapping squares covering the best each mobility populations. The user can modify these limits as desired in the script.

As representing the circle best encompassing the motion track, rather than encompassing it strictly, the measurement of the radius of confinement is independent of the track duration (Lerner et al., 2020).

For point 14., the user might also use only confined\_Low\_jump.mat files, in the case no butterfly tracks have been detected.



**Figure 5. Comparing the mobility of a protein of interest with a mobility of reference**

(A) Example of a reference mobility plot (H2B-HALO, *n* = 80,000) and of the mobility plot for a protein of interest (FOXA1-HALO, *n* = 80,000). The squares represent the boundaries set for each chromatin mobility group. Differences in data density are measured by normalizing the data density in the reference plot by the maximal value in the plot of interest. vL, very low mobility chromatin; L, low mobility chromatin; I, intermediate mobility chromatin; H, high mobility chromatin; vH, very high mobility chromatin.

(B) Density values in the plot of reference and of interest are first normalized by the corresponding maximal density value in the plot (noted max(density)). Then, the Density\_differences.m script overlaps the two-dimensional density plots of reference (e.g., histone H2B in red) and of interest (in green), and subtracts the average density values of the 5 nearest points of reference to each density of interest. Density differences are then multiplied by -1 for intuitiveness. Then positive values for density differences indicate increased binding, while negative values indicate decreased binding.

(C) Data density differences between FOXA1 and H2B in the different mobility groups. Data are represented as means  $\pm$  SEM. vL, very low mobility chromatin; L, low mobility chromatin; I, intermediary mobility chromatin; H, high mobility chromatin.

### Single-molecule tracking analysis IV: comparison of data density respective to histone H2B

© Timing: several hours



This step consists in the quantitative measurement of differences between a reference two-dimensional mobility plot (e.g., histone H2B, wildtype protein, ...) and the two-dimensional mobility plot of a protein of interest (e.g., transcription factor, mutant) (Figure 5A). The trackingdata.csv files generated at the previous step are used as input for this step.

The Density\_Differences.m script performs the following operations: (1) measuring the local density in the different chromatin mobility populations of the two-parameter (RC and AVE) scatter plot with scatplot.m and assigning a density number to each trajectory (each individual point in the scatter plot). (2) Normalizing density values by the maximum density value for each condition (Figures 5A and 5B). (3) Computing the density differences between the protein of interest (any Chromatin-Binding Protein) and the reference (Figure 5B), by subtracting the average density of the five nearest points to the density of each trajectory of interest (Figure 5B). (4) Multiplying the result by  $-1$  for intuitiveness: in each chromatin mobility population, positive and negative values thus mean that the protein of interest is more or less dense than the reference in the considered chromatin mobility domain, respectively (Figure 5C).

The script generates 3 outputs plots: one individual density plot for the reference and the protein of interest, and a histogram plot displaying the average and standard error in 4 mobility population. The very high mobility population is excluded because the sparsity of the data in that region gives rise to high errors in the measurement.

18. On line 9 of the Density\_differences.m script, indicate the number of trajectories to analyze. We recommend running the script on the same number of randomly downsampled data, with  $n \geq 10,000$ .
19. Run the Density\_differences.m script, and first select the trackingdata.xlsx file corresponding to the reference plot (e.g., histone H2B). Then select the trackindata.xlsx file for the protein of interest.
20. Save the figures under the desired format.

## EXPECTED OUTCOMES

The two-parameter mobility plot for histone H2B was similar to Figure 4E in several differentiated cell lines (Lerner et al., 2020), but might be dependent on the cell context. For example, in the cited reference, ES cells presented a histone H2B mobility plot of higher mobility than other cell types, with mainly High Mobility Chromatin, and mitotic cells presented an overall decreased chromatin mobility (Lerner et al., 2020). These assessments from the two-parameter analysis agree with independent assessments of chromatin structure (Meshorer et al., 2006; Ou et al., 2017)

## QUANTIFICATION AND STATISTICAL ANALYSIS

For the radius of confinement measurement, to discard fitting errors, the trajectories with squared norm of the residual (or RSS) higher than  $10^{-5}$ . In addition, fitting of T-MSD curves can lead to aberrant values for radius of confinement (e.g., bigger than the radius of the nucleus). Thus, the threshold set for maximal radius of confinement is 300 nm, consistently with the maximal size of chromatin domains seen in super-resolution microscopy (Ricci et al., 2015) and in single-molecule imaging (Nozaki et al., 2017). Less than 5% of trajectories were consistently discarded by these filters.

The fitting of T-MSD curves with a circle diffusion model (Ernst and Köhler, 2012) also provides a constant offset due to the localization precision limit inherent to all the localization-based microscopy methods. In our case, we estimate a localization precision of 13 nm. Tracks shorter than 5 steps were arbitrarily excluded, as it seems a reasonable cutoff to reduce the standard error in the measurement of the average displacement.

When measuring density differences between a two-dimensional plot of reference and for a protein of interest, the plot displays the mean and the standard error of the mean in each chromatin mobility population. Statistical significance was assayed by one-way ANOVA in the original publication ([Lerner et al., 2020](#)).

### LIMITATIONS

Our protocol has been carried out only on adherent cells. Non-adherent cells pose a challenge for single-molecule imaging because of their constant movements in the medium, even when sedimented. That can be addressed by coating the support with components allowing immobilization of the cell.

### TROUBLESHOOTING

#### Problem

The protein of interest not compatible with a Halo fusion. In our experience, this is particularly true for very large protein (>100 kDa) and resulting in the detection of Halo signal both in the nucleus and the cytoplasm.

#### Potential solution

This can possibly be circumvented by tagging the other extremity (N- or -C-terminal) of the protein. Alternatively, trying several linkers between the protein and the Halo-tag might render the fusion possible.

#### Problem

Elevated fluorescence levels, resulting in high background, overlapping molecules, or artifacts (e.g., peri-nuclear inclusions of Halo fluorescence).

#### Potential solution

Use a low expression strategy for the Halo-tagged protein. For example, in the publication of reference we used Tetracyclin inducible promoters in the absence of activator and tetracycline. Other strategies for low expression of the protein can also be used ([Liu et al., 2018](#)). Alternatively, decreasing the concentration of Halo ligand, or incubation time is a good strategy for decreasing background levels.

#### Problem

No single-molecule are visible during the imaging step.

#### Potential solution

If no fluorescent signal can be seen at a 100 ms exposure, the problem is probably connected to the settings of the microscope. The 3 key parameters for signal optimization are : (1) laser power, (2) illumination angle, (3) z-position. These need to be finely tuned to find optimal imaging conditions. These parameters might have to be change when switching to another field.

#### Problem

The SMT\_Motion\_Classifier script is stuck on "Detecting Butterfly Motions."

#### Potential solution

This step can be timely as many operations are simultaneously performed and can sometimes take more than several hours to finish but will not crash. If the number of tracks analyzed is over 100,000, we recommend selecting less of the .csv files, which will make the analysis faster.



## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact and corresponding author, Dr. Jonathan Lerner at [jlerner@penncellpress.com](mailto:jlerner@penncellpress.com).  
[upenn.edu](http://www.upenn.edu).

### Materials availability

This study did not generate any new unique reagents.

### Data and code availability

Scripts used in this study are available on Mendeley <https://doi.org/10.17632/hxnhttxpk.1>

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.xpro.2020.100223>.

## ACKNOWLEDGMENTS

We thank Luke Lavis for kindly providing the Halo ligand coupled to JF549 fluorophore. J.L. was supported by a fellowship from Fondation pour la Recherche Medicale, code 40334. M.L. was supported by a Linda Pechenik Montague Investigator Award. The research was supported by NIH grant R01GM36477 to K.S.Z.

## AUTHOR CONTRIBUTIONS

Conceptualization, J.L., P.G.G., K.Z., and M.L.; methodology, J.L., P.G.G., K.Z., Z.L., and M.L.; software, P.G.G., Z.L., J.L., and R.M.; validation, J.L. and P.G.G.; formal analysis, J.L. and P.G.G.; investigation, J.L.; resources, K.Z., Z.L., and M.L.; data curation, J.L. and P.G.G.; writing – original draft, J.L. and P.G.G.; writing – review & editing, J.L., P.G.G., and K.Z.; visualization, J.L. and P.G.G.; supervision, K.Z., M.L., and Z.L.; funding acquisition, J.L., K.Z., and M.L.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## REFERENCES

- Chen, B.-C., Legant, W.R., Wang, K., Shao, L., Milkie, D.E., Davidson, M.W., Janetopoulos, C., Wu, X.S., Hammer, J.A., Liu, Z., et al. (2014a). Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. *Science* *346*, 1257998.
- Chen, J., Zhang, Z., Li, L., Chen, B.-C., Revyakin, A., Hajji, B., Legant, W., Dahan, M., Lionnet, T., Betzig, E., et al. (2014b). Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* *156*, 1274–1285.
- Crivat, G., and Taraska, J.W. (2012). Imaging proteins inside cells with fluorescent tags. *Trends Biotechnol.* *30*, 8–16.
- Dion, V., and Gasser, S.M. (2013). Chromatin movement in the maintenance of genome stability. *Cell* *152*, 1355–1364.
- Ernst, D., and Köhler, J. (2012). Measuring a diffusion coefficient by single-particle tracking: statistical analysis of experimental mean squared displacement curves. *Phys. Chem. Chem. Phys.* *15*, 845–849.
- Iwafuchi, M., Cuesta, I., Donahue, G., Takenaka, N., Osipovich, A.B., Magnuson, M.A., Roder, H., Seeholzer, S.H., Santisteban, P., and Zaret, K.S. (2020). Gene network transitions in embryos depend upon interactions between a pioneer transcription factor and core histones. *Nat. Genet.* *52*, 418–427.
- Lerner, J., Gomez-Garcia, P.A., McCarthy, R.L., Liu, Z., Lakadamyali, M., and Zaret, K.S. (2020). Two-parameter mobility assessments discriminate diverse regulatory factor behaviors in chromatin. *Mol. Cell* *79*, 677–688.e6.
- Liu, Z., and Tjian, R. (2018). Visualizing transcription factor dynamics in living cells. *J. Cell Biol.* *217*, 1181–1191.
- Liu, H., Dong, P., Ioannou, M.S., Li, L., Shea, J., Pasolli, H.A., Grimm, J.B., Rivlin, P.K., Lavis, L.D., Koyama, M., et al. (2018). Visualizing long-term single-molecule dynamics in vivo by stochastic protein labeling. *Proc. Natl. Acad. Sci. U S A* *115*, 343–348.
- Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* *10*, 105–116.
- Michalet, X., and Berglund, A.J. (2012). Optimal diffusion coefficient estimation in single-particle tracking. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* *85*, <https://doi.org/10.1103/PhysRevE.85.061916>.
- Normanno, D., Boudarène, L., Dugast-Darzacq, C., Chen, J., Richter, C., Proux, F., Bénichou, O., Voituriez, R., Darzacq, X., and Dahan, M. (2015). Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nat. Commun.* *6*, 7357.
- Nozaki, T., Imai, R., Tanbo, M., Nagashima, R., Tamura, S., Tani, T., Joti, Y., Tomita, M., Hibino, K., Kanemaki, M.T., et al. (2017). Dynamic organization of chromatin domains revealed by super-resolution live-cell imaging. *Mol. Cell* *67*, 282–293.e7.
- Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H., and O’Shea, C.C. (2017). ChromEMT: visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* *357*, eaag0025.
- Raccaud, M., Friman, E.T., Alber, A.B., Agarwal, H., Deluz, C., Kuhn, T., Gebhardt, J.C.M., and Suter, D.M. (2019). Mitotic chromosome binding predicts transcription factor properties in interphase. *Nat. Commun.* *10*, 487.

Ricci, M.A., Manzo, C., García-Parajo, M.F., Lakadamyali, M., and Cosma, M.P. (2015). Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* *160*, 1145–1158.

Sergé, A., Bertaux, N., Rigneault, H., and Marguet, D. (2008). Dynamic multiple-target tracing to probe

spatiotemporal cartography of cell membranes. *Nat. Methods* *5*, 687–694.

Tarantino, N., Tinevez, J.-Y., Crowell, E.F., Boisson, B., Henriques, R., Mhlanga, M., Agou, F., Israël, A., and Laplantine, E. (2014). TNF and IL-1 exhibit distinct ubiquitin requirements for inducing NEMO–IKK supramolecular structures. *J Cell Biol.* *204*, 231–245.

Teves, S.S., An, L., Hansen, A.S., Xie, L., Darzacq, X., and Tjian, R. (2016). A dynamic mode of mitotic bookmarking by transcription factors. *eLife* *5*, e22280.

Wieser, S., and Schütz, G.J. (2008). Tracking single molecules in the live cell plasma membrane—Do's and Don't's. *Methods* *46*, 131–140.