Genome-wide analysis of the biophysical properties of chromatin and nuclear proteins in living cells with Hi-D

Valades-Cruz, Cesar Augusto; Barth, Roman; Abdellah, Marwan; Shaban, Haitham A.

DOI
10.1038/s41596-024-01038-3

Publication date
2024

Document Version
Final published version

Published in
Nature Protocols

Citation (APA)
Valades-Cruz, C. A., Barth, R., Abdellah, M., & Shaban, H. A. (2024). Genome-wide analysis of the biophysical properties of chromatin and nuclear proteins in living cells with Hi-D. *Nature Protocols*, *20 (2025)*(1), 163-179. Article 228101. https://doi.org/10.1038/s41596-024-01038-3

nature protocols

https://doi.org/10.1038/s41596-024-01038-3

Protocol

Check for updates

Genome-wide analysis of the biophysical properties of chromatin and nuclear proteins in living cells with Hi-D

Cesar Augusto Valades-Cruz[1,2,8,10], Roman Barth[3,10], Marwan Abdellah[4,10] & Haitham A. Shaban[5,6,7,9]

Abstract

To understand the dynamic nature of the genome, the localization and rearrangement of DNA and DNA-binding proteins must be analyzed across the entire nucleus of single living cells. Recently, we developed a computational light microscopy technique, called high-resolution diffusion (Hi-D) mapping, which can accurately detect, classify and map diffusion dynamics and biophysical parameters such as the diffusion constant, the anomalous exponent, drift velocity and model physical diffusion from the data at a high spatial resolution across the genome in living cells. Hi-D combines dense optical flow to detect and track local chromatin and nuclear protein motion genome-wide and Bayesian inference to characterize this local movement at nanoscale resolution. Here we present the Python implementation of Hi-D, with an option for parallelizing the calculations to run on multicore central processing units (CPUs). The functionality of Hi-D is presented to the users via user-friendly documented Python notebooks. Hi-D reduces the analysis time to less than 1 h using a multicore CPU with a single compute node. We also present different applications of Hi-D for live-imaging of DNA, histone H2B and RNA polymerase II sequences acquired with spinning disk confocal and super-resolution structured illumination microscopy.

Key points

- This protocol covers the implementation in Python of a computational technique, termed high-resolution diffusion mapping, to detect, classify and map chromatin and protein dynamics via dense optical flow detection.
- High-resolution diffusion mapping relies on dense labeling of biomolecules. Alternative methods for the analysis of abundant and densely labeled molecules include image mean square displacement analysis (iMSD) and displacement correlation spectroscopy.

Key references

Shaban, H. A. et al. *Genome Biol.* **21**, 95 (2020): https://doi.org/10.1186/s13059-020-02002-6

Shaban, H. A. et al. *Nucleic Acids Res.* **46**, e77 (2018): https://doi.org/10.1093/nar/gky269

Shaban, H. A. et al. *Proc. Natl Acad. Sci. USA* **121**, e2311374121 (2024): https://doi.org/10.1073/pnas.2311374121

[1]SERPICO Project Team, Inria Centre Rennes–Bretagne Atlantique, Rennes, France. [2]SERPICO Project Team, UMR144 CNRS Institut Curie, PSL Research University, Paris, France. [3]Department of Bionanoscience, Delft University of Technology, Delft, the Netherlands. [4]Ecole polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland. [5]Spectroscopy Department, Institute of Physics Research National Research Centre, Cairo, Egypt. [6]Agora Cancer Research Center, Lausanne, Switzerland. [7]Precision Oncology Center, Department of Oncology Lausanne University Hospital, Lausanne, Switzerland. [8]Present address: Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China. [9]Present address: Faculty of Medicine, University of Geneva, Geneva, Switzerland. [10]These authors contributed equally: Cesar Augusto Valades-Cruz, Roman Barth, Marwan Abdellah. e-mail: haitham.ahmed@unige.ch

Introduction

The dynamic organization of chromatin structures is highly relevant to cell function, which may be reflected in the relative distribution of open euchromatin and dense heterochromatin regions[1,2]. According to structural models, the genome is folded into multiple hierarchical levels of chromatin structure, from domain folding and long-range looping to compartments and subcompartments[3]. Research points to the dynamic nature of nuclear compartments themselves being involved in the regulation of their organization and, thus, affecting the regulation of nuclear processes[4]. For example, the nucleosome density must be rearranged at transitions within and between euchromatin and heterochromatin to facilitate DNA processing[5] and coordinated for chromatin–protein interaction[6,7]. The dynamic activity of chromatin and nuclear proteins in living cells has been studied at high spatiotemporal resolution using specifically labeled molecules that can be identified and tracked in real time using various techniques[8,9]. Labeled loci and single genes can be tracked over various lengths and time scales using single particle tracking methods[10–15]. However, to fully understand the physical properties of a long-fiber structure, genomic chromatin and nuclear proteins must be comprehensively studied across the entire nucleus[16–18]. Nucleus-wide methods constitute a powerful approach because no locus-specific labeling is required[19,20]. Such methods are able to quantify the momentary displacement but yet fall short in quantifying the extracted motion spatially in terms of physical diffusion models[19,20] and at subpixel resolution[19]. Therefore, we recently developed a method, termed high-resolution diffusion (Hi-D), combining a dense optical flow algorithm to quantify the local motion of bulk structures with subpixel resolution[21]. Bayesian inference is used to precisely classify types of motion throughout and to assign biophysical parameters, in particular, the diffusion constant, anomalous exponent and drift velocity, to the motion modes throughout the nucleus. Using Hi-D, we investigate the dynamic roles of chromatin (both DNA and histones)[21–24], RNA polymerase II (RNA Pol II)[21,25] heterochromatin protein 1 (ref. 21) and transcription factors[26] in genome organization. Here, we provide instructions on how to use the Hi-D approach written in Python as an accessible and user-friendly tool, with an improvement in processing time. We present a step-by-step Hi-D analysis of DNA and nuclear proteins imaged with spinning disks confocal and super-resolution microscopy.

Development of the protocol

The Hi-D method was developed for analyzing densely labeled biomolecules in living cells and implemented in MATLAB. Here, we extend the approach by redesigning and implementing the workflow in Python, a high-level and noncommercial programming language that does not need any commercial licensing, contrary to MATLAB. Using Python, we reduced the execution time of the entire workflow by making use of parallel execution environments supported by Python including the multiprocessing and joblib packages by default. Even with a single-threaded execution, the Python implementation is faster than the MATLAB-based one. The new software version is more user-friendly and easier to run via documented Python Jupyter Notebooks, which guide the user in a step-by-step fashion to analyze the results on the fly. Documentation and code are released as part of the Hi-D project (https://github.com/haitham-shaban/hidpy).

Overview of the procedure

The full protocol of Hi-D is divided into two complementary procedures. The first procedure is performed in the laboratory to acquire video sequences of living cells showing the dynamics of chromatin and nuclear proteins. Further details on this procedure are available[21]. The second procedure analyzes these sequences using Hi-D. The presented protocol is focused on the Python workflow implemented for analyzing the video sequences. Nevertheless, we briefly describe a few details of the first procedure for users aiming to replicate the conditions and setups that were used to generate the presented sequences (refer to the video sequences provided within the source code repository and used as default input to the in the Jupyter Notebooks). The Python workflow is composed of three stages. In Stage 1, the input sequence

is preprocessed to compute the optical flow fields, estimate the trajectories and apply Bayesian inference. We then select the most suitable models for deconvolution in Stage 2. In Stage 3, we compare between average statistics of subpopulations from different biological conditions.

Stage 1: preprocessing, optical flow calculation, trajectories estimation, MSD calculations and Bayesian inference (Steps 1–12)

After the generation of the video sequence, we applied Stage 1 to preprocess the frames of the input sequence, compute the trajectories and their mean square displacements (MSDs) and use Bayesian inference to test the different models for any MSD curve (Stage 1 in Fig. 1). The preprocessing stage thresholding, denoising and edge detection operations are applied to match the specific requirements of the optical flow algorithm. For a sequence with N frames, $N-1$ optical flow maps are produced. Those maps are then interpolated and linked to compute trajectories on a per-pixel basis. Therefore, we allow the user to define a thresholding value based on image intensity to determine whether the value of the pixel is sufficiently high to compute a trajectory or not. Trajectory computations are implemented without taking into consideration the dimensions of the pixel. The pixel size is defined by the user in the input

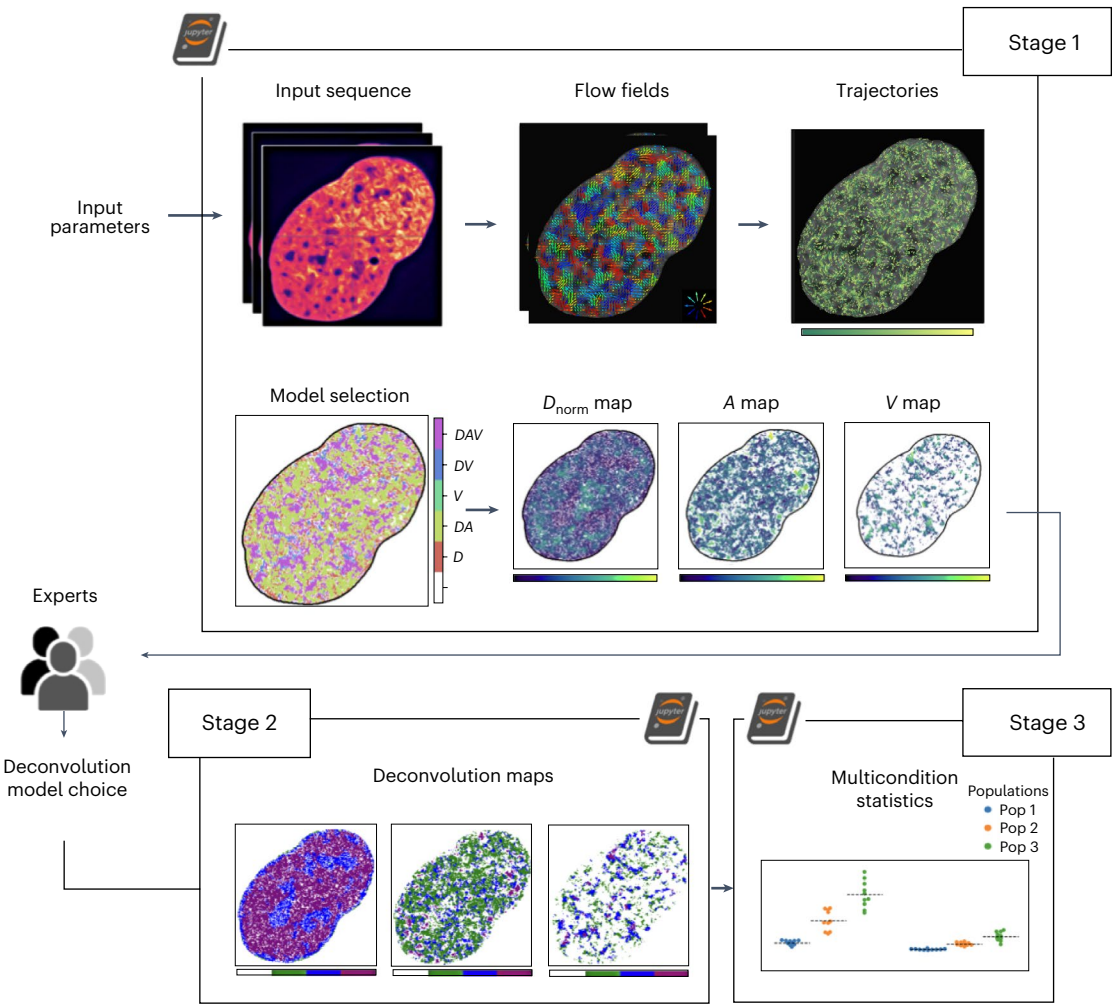

Fig. 1 | Overview of the Hi-D workflow. This workflow is composed of three stages. In stage 1, we preprocess the input sequence, compute the flow field and trajectories and then apply Bayesian inference to compute the parameters needed for the deconvolution. The normalized diffusion constant (D_{norm}) is defined as $D_{norm} = \log_{10}(D \cdot 1s^{\alpha}/\mu m^2)$, where D is the diffusion constant and α is the anomalous exponent (depicted in the 'A'). V corresponds to drift velocity. Users can read and select the best-fitting models for deconvolution in stage 2. Stage 3 compares the statistics of subpopulations from different biological conditions.

data before running the protocol based on the configuration of the microscope used to image the sequence. Therefore, we scale each estimated trajectory by multiplying the computed coordinates by the pixel size to convert from pixel units to micrometers. We then use the Scipy library to interpolate the trajectories. The resulting trajectories are then listed in a linear list and mapped to a three-dimensional array for visualization. After trajectory estimation, the MSD curves are calculated for every trajectory. The estimated MSD is a function of the initial position and the time lag (Supplementary Note 1). A nucleus mask is estimated during the preprocessing stage and used to select only nuclear-localized trajectories. Local MSD estimation is described in the original Hi-D publication[21].

Bayesian inference is used to test different models for any given MSD curve[27]. Given the data $Y=\{Y_1, \ldots, Y_n\}$, K model candidates $M=\{M_1, \ldots, M_K\}$ and parameter set $\theta=\{\theta_1, \ldots, \theta_K\}$, the goal is to find the model $M_k(Y, \theta_k)$ such that the probability of $M_k(Y, \theta_k)$ is maximal for the selected set of models; therefore, the optimal parameters for every model are calculated. A general multivariate Gaussian function[28] is used to represent the probability that the data Y is observed for a model M_k described by the function $M_k(x; \theta_k)$ for any parameter set θ_k as shown in Supplementary Note 2. After estimating the best parameter set for a model, the model and its chosen parameters are selected to maximize their probability of describing the data in which $\hat{\theta}_{k,MLE} = \arg\max_{\theta_K} P(Y|\theta_k, M_k)$. The Bayesian information criterion (BIC) is subsequently used to select the model M_k, which favors the model with fewer parameters that represent the data.

MSD analysis is performed locally[21], only the 3 × 3 neighborhood of a pixel is chosen to determine the MSD covariance matrix used for the Bayesian inference[29]. The choice of a window size of 3 × 3 is based on the selected filter size in the optical flow estimation. All the calculations are implemented in Python. Scipy[30] is used for numerical fitting. The type of diffusion characterizing each trajectory motion was chosen in an unbiased manner using the Bayesian inference described previously from a set of five common models to fit each trajectory's MSD. These methods were detailed earlier[21], and a summary of them is presented in Table 1. A constant offset o is added to consider experimental noise.

The user can select a subset or all the models listed in Table 1, where D is in units of μm²/s. D_α is in units of μm²/s$^\alpha$, α is its anomalous exponent, v (μm/s) is its velocity and R_C (μm) is the radius of a sphere within the particle that is confined. Free diffusion is expected for particle motion in an unrestrictive environment, directed motion may result from an active process or is expected for repelling particles at short time scales. Anomalous diffusion may arise from a variety of factors including diffusion in an environment with obstacles[31], transient binding events[32], crowding[33] or due to the polymeric nature of chromatin[34]. Confined diffusion applies specifically to the analysis of trapped particles, such as membrane proteins[35]. A free, anomalous and confined motion may co-occur with directed motion. These models are comprehensive, but the user might identify cases in which these models are not applicable. We encourage the user to implement their own, application-specific model into the pipeline. To implement custom diffusion models, we refer the experienced user to the source code file (hidpy/core/inference.py/msd_fitting()). Furthermore,

Table 1 | Overview of MSD models

Model	Equation
Free diffusion (D)	$\mathrm{MSD}_D(\tau) = 4D\tau + o$
Drift (V)	$\mathrm{MSD}_V(\tau) = v^2\tau^2 + o$
Free diffusion + drift (DV)	$\mathrm{MSD}_{DV}(\tau) = 4D\tau + v^2\tau^2 + o$
Anomalous diffusion (DA)	$\mathrm{MSD}_{DA}(\tau) = 4D_\alpha\tau^\alpha + o$
Anomalous diffusion + drift (DAV)	$\mathrm{MSD}_{DAV}(\tau) = 4D_\alpha\tau^\alpha + v^2\tau^2 + o$
Confinement + diffusion (DR)	$\mathrm{MSD}_{DR}(\tau) = R_C^2\left(1 - e^{\left(\frac{-4D\tau}{R_C^2}\right)}\right) + o$
Confinement + diffusion + drift (DRV)	$\mathrm{MSD}_{DRV}(\tau) = R_C^2\left(1 - e^{\left(\frac{-4D\tau}{R_C^2}\right)}\right) + v^2\tau^2 + o$

Table 2 | Example of a results table of a GMM deconvolution in Hi-D

No. population	Normal distribution			Log-normal distribution		
	1	**2**	**3**	**1**	**2**	**3**
$-\log_{10}(D \times 1\,s^{\alpha}/\mu m^2)$	0	10%	10%	0	30%	**50%**
A	0	0	**100%**	0	0	0
V	0	0	10%	20%	30%	**40%**

Table displaying the percentage of cells per distribution type and the number of subpopulations for each parameter (diffusion-normalized constant, anomalous exponent (*A*) and drift velocity (*V*)) of the data described in Fig. 2 after a GMM deconvolution in Hi-D. Optimal combination is highlighted in bold numbers for each parameter. $n = 10$ cells.

the diffusion constant D is transformed into the generalized diffusion constant (D_{norm}), denoted as $\log_{10}(D \cdot 1\,s^{\alpha}/\mu m^2)$ (ref. 36). D_{norm} is a dimensionless quantity.

Stage 2: deconvolution by a GMM (Steps 13–17)

The Bayesian inference of the trajectory MSD results in maps of the normalized D_{norm}, anomalous exponent (*A*) and drift velocity (*V*) for every cell (the best model parameters $\hat{\theta}_{k,MLE}$). These parameter values are represented in histograms and subsequently deconvolved by a general mixture model (GMM) approach[29]. Before starting the trajectory computation and Bayesian inference (Stage 2 in Fig. 1), the user selects the maximum number of subpopulations n_{ot} to be expected within the data set, as well as the type of distribution to be used by the GMM. Without prior knowledge, the number of subpopulations can be set to three, and, typically, both normal and log-normal distributions can be considered. The GMM is then evaluated on all types of distributions for a number of subpopulations from 1 to n_{pop}. The goodness of fit for each of those combinations is evaluated using the BIC[37] (Supplementary Note 3). The result of the deconvolution is a table representing which fractions of cells in the analyzed data set are best described by a particular combination of distribution type and the number of subpopulations for every parameter (D_{norm}, *A* and *V*) (Table 2). In principle, every cell might be best described by a different number of subpopulations of varying relative size and different shapes (that is, normal or log-normal). To enable a global GMM deconvolution across the entire data set, the user is advised to select the GMM settings from the table, which describes the largest fraction of cells within the dataset; this should be the default choice. Currently, Hi-D considers this automatically. However, if the user has reason to believe that another setting would be more appropriate to describe the data, potentially due to prior knowledge, the user may select another setting. The table is then used for the user to evaluate which fraction of cells within the dataset is well described by a particular choice of GMM settings. When a decision has been reached, the user could leave default choice or to supply the number of subpopulations and shapes and starts the final deconvolution stage with the chosen model(s) to obtain deconvolved histograms as well as a mapping of the respective distributions within the cell nucleus. If the Jupyter Notebook is used, the parameters are updated in the parameters panel directly (Fig. 2).

Stage 3: comparison between average statistics of subpopulations from different biological conditions (Steps 18–21)

In this stage, the users can compare the biophysical and diffusion parameters out of the Hi-D analysis from different biological conditions (Stage 3 in Fig. 1). The three stages are described in detail in the Procedure section, to assist the user in using the protocol in a step-by-step procedure. The timing of the three stages depends on the video length, the number of pixels per frame and the number of selected models. A full Hi-D workflow including the three stages are shown in Fig. 3.

Evaluation of the protocol implementation in Python

Dense optical flow fields are reconstructed by an optical flow algorithm based on the Horn–Schunck formulation[38]. The performance of the translation to Python was evaluated against the previously used MATLAB implementation[21,39] (Supplementary Fig. 1) on experimental data from the original work. The MATLAB implementation partially utilizes built-in functions for

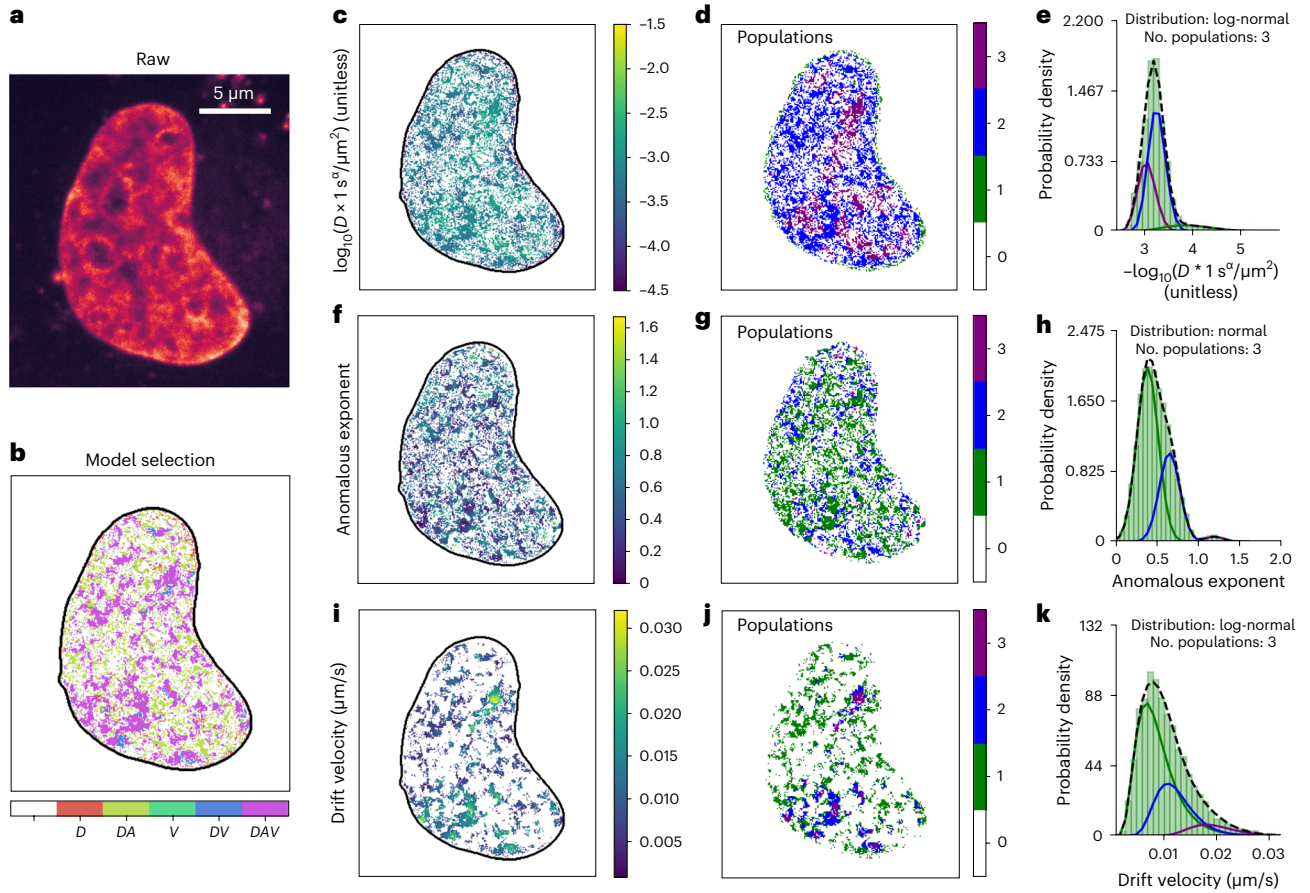

Fig. 2 | Deconvolution by a GMM approach in Hi-D. a,**b**, SiR-DNA-stained U2OS cells with serum were imaged at $\Delta t = 200$ ms. Optical flow estimation and Bayesian inference were done using Hi-D. Representative maps are shown for one cell: raw (**a**) and model selection (**b**). Optimal combinations of distribution type and the number of sub-populations for each biophysical parameter were estimated using deconvolution with a GMM approach applied to the dataset ($n = 10$ cells). **c**–**k**, Coefficient maps and GMM deconvolution results for a representative cell are presented for each biophysical parameter (diffusion-normalized constant (**c**–**e**), anomalous exponent (**f**–**h**) and drift velocity (**i**–**k**)). Estimated populations are shown as maps for **d**,**g** and **j** and deconvolved histograms for **e**,**h** and **k**.

which no source code is available. We carefully validated that the Python translation returns identical results to the MATLAB implementation for all available source code. Nevertheless, the translation of proprietary algorithms yields slightly different results, which ultimately result in small variations of the estimated flow fields (endpoint error median ± s.d. of 0.67 ± 0.39 pixels, ratio between flow magnitude of the Python and MATLAB implementation median ± s.d. of 0.996 ± 1.967; Supplementary Fig. 1). Supplementary Fig. 2 shows a comparison between the results of the Bayesian analysis[1] using MATLAB and that resulting from our Python implementation. Similar results are shown for the least-square fitting from the MATLAB curve fitting toolbox and Python Scipy library.

Alternative methods

Hi-D relies on dense labeling of biomolecules. Whenever this cannot be achieved and individual diffraction-limited spots originating from labeled biomolecules can be detected and spatially separated, conventional single-particle tracking methods are appropriate to use. Alternative methods for the analysis of abundant and densely labeled molecules include image mean square displacement analysis (iMSD)[17] and displacement correlation spectroscopy[19]. However, neither methods, per se, extract biophysical parameters describing the biomolecules' motion nor infer the diffusion constant at high resolution (subpixel resolution) and, thus, provide

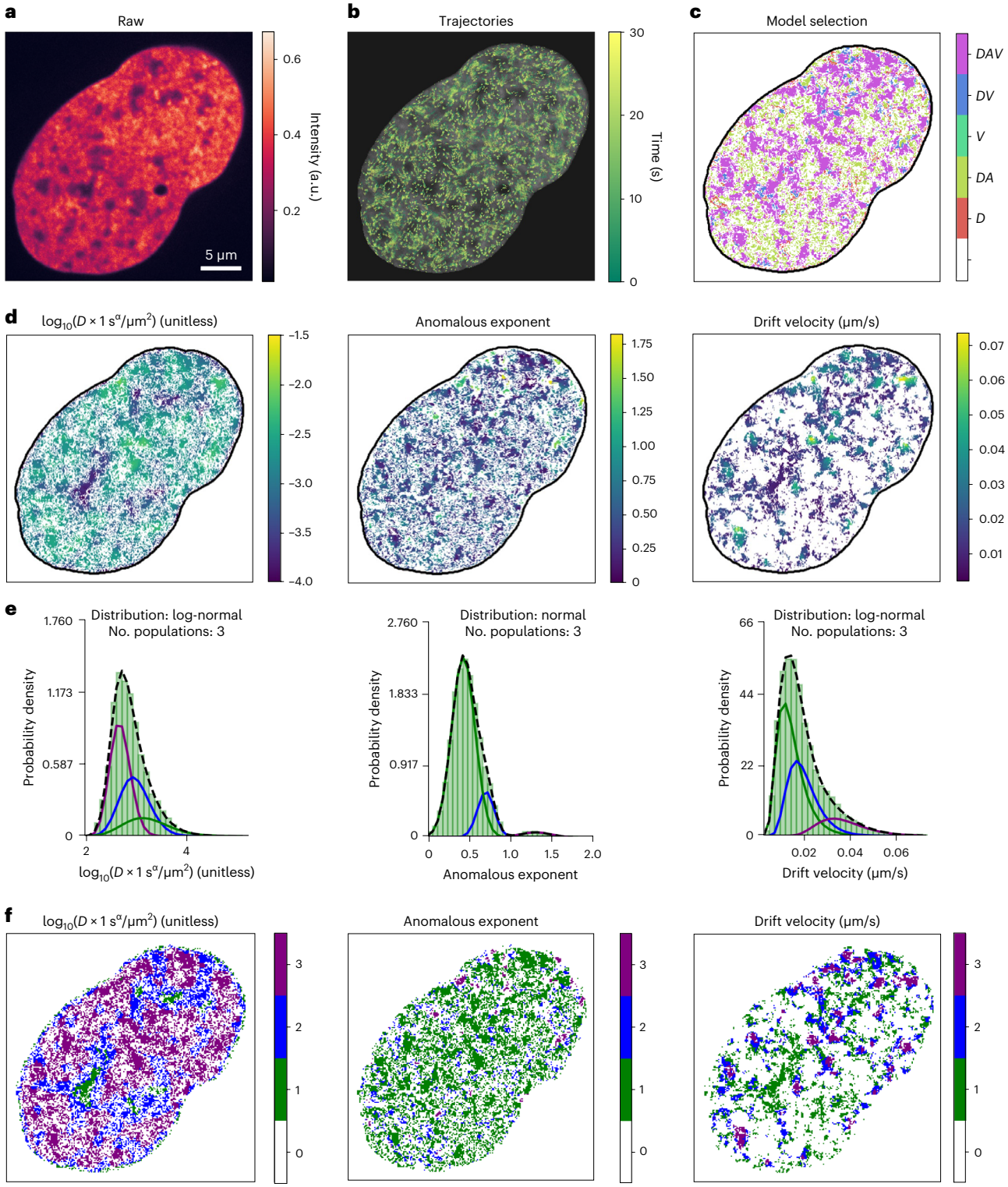

Fig. 3 | A complete analysis example using the Hi-D. a–d, The workflow was applied to a sequence of H2B (**a**) in U2OS with a time interval of $\Delta t = 200$ ms. a.u., arbitrary units. Trajectories were estimated (**b**) and Bayesian inference was used to map the results for the models (**c**) and biophysical parameters (diffusion-normalized constant (**d**, left), anomalous exponent (**d**, center) and drift velocity (**d**, right)). **e**,**f**, Results from deconvolution using GMM are shown, with histograms in **e** and the maps of populations in **f**.

limited information (even though an additional step to compute these could be added by users). Furthermore, iMSD may exhibit a bias toward particularly small and large values as noted in the comparison of the iMSD and Hi-D approaches (supplementary information in ref. 21). An optical flow algorithm has also been applied to boost the accuracy of the flow field calculation with the dense flow reconstruction and correlation (DFCC) method at subpixel resolution[19]. The particle image velocimetry algorithm underlying displacement correlation spectroscopy is less accurate than the optical flow algorithm employed in DFCC[19]. Hi-D extends the capabilities of DFCC by reconstructing trajectories and computing MSD curves for every pixel from the estimated flow fields, permitting the quantification of the extracted motion in terms of physical diffusion models at subpixel resolution[21].

Applications of Hi-D

Hi-D allows the monitoring of spatiotemporal dynamics within nuclei of living cells in action. Recently, we have shown the applicability of the Hi-D method in studying the dynamic properties and precisely mapping the diffusion parameters and physical diffusion models of chromatin, RNA Pol II and transcription factors, nuclear proteins at nanoscale resolution in single living cells[21,25,26]. Hi-D is useful for the dynamic analysis of chromatin and nuclear macromolecules across a variety of imaging scenarios[21–23] involving dense signals that are beyond the sensitivity of alternative methods. Recently, Hi-D was able to decipher the dynamic connection of gene expression to high-order chromatin structures genome-wide in eukaryotic cells[26]. Potentially, Hi-D can be applied to any biomolecule that is sufficiently abundant to densely stain the compartments within cells. Provided there is a high enough signal-to-noise ratio (≥20 dB, supplementary information in ref. 21), it should be possible to carry out Hi-D analysis of tissue imaging data. Hi-D can be combined with conventional[20,21] and super-resolution light microscopy techniques[22,23] and is amenable to screening different cell types, providing an approach for studying chromatin organization involved in loop formation[7,40] and other dynamic processes involved in gene expression in the future by using Hi-D[8,41].

Advantages and limitations

Commonly used methods to extract diffusion characteristics of nuclear loci such as single-particle tracking methods cannot be used with dense labeling or certain nuclear regions analysis and require special hardware, such as fluorescence correlation spectroscopy[42,43]. By contrast, Hi-D is effective in analyzing the dynamics of abundant molecules and dense structures directly in single living cells without sacrificing active fluorophore density. Therefore, there is no need for prior knowledge of sophisticated (sparse) labeling preparations or experience in using advanced microscopy techniques. However, Hi-D is adequate to analyze images acquired with both conventional and super-resolution microscopy techniques. Using Hi-D, meaningful diffusion characteristics are extracted at a single-pixel resolution, and the extracted diffusion is directly related to a physical description of various types of diffusion. By doing so, Hi-D overcomes the limitations of averaging the nucleus-wide analysis method[19]. Hi-D can also be applied separately to all kinds of different fluorophores, as long as the fluorophores can be spectrally unmixed. This property enables the processing of multichannel images to spatiotemporally correlate different nuclear factors, as shown in the original Hi-D article[21].

There are technical limitations of the original protocol of the Hi-D method. The original implementation of the protocol was done in MATLAB, a commercial software application that requires specific and hefty licensing. However, the GMM analysis module is implemented using the pomegranate package[44] in Python, and the interaction between MATLAB and Python requires the user to establish the connection between MATLAB and Python manually, which may limit inexperienced researchers to run the protocol in their research environments. The execution of the original protocol is time consuming; processing a single video sequence might take ~24 h or more on mid-range commodity hardware. This issue limits the analysis of lengthy time lapses with specific experimental conditions on personal computers and might require high-end computing clusters for large-scale processing. Finally, the original protocol does not have a flexible interface; it is merely a group of scripts that run independently. The user is required to visualize and compare the results of different cell lines/biological treatments manually.

Experimental design

Hi-D does not apply to biological scenarios in which only sparse labeling can be achieved, for example, due to low protein abundance or where the signal-to-noise ratio is below 20 dB (ref. 21). Furthermore, small, moving and particularly bright fluorescence spots might obscure Hi-D analysis as their contribution to the surrounding fluorescence intensity overwhelms the flow field quantification. Hi-D analysis should, thus, be performed on regions of images without labeling artifacts such as clusters of fluorophores. More extended regions of varying fluorescence intensity do not impose limitations on Hi-D[21]. As for all methods quantifying intracellular motion, the translational and rotational rigid-body motion of nuclei should be removed before Hi-D analysis. The user is encouraged to use a method suitable for the imaged system. Care should be taken to use methods that do not involve an interpolation step (that is, which shift or rotate images of cell nuclei by integer values). Interpolation nonlinearly distorts the pixel values, which might affect Hi-D performance. Nuclei that contract or expand during acquisition are in general unsuitable for Hi-D analysis because the nuclear area is assumed to remain constant for Hi-D. If the image series contains residual drift, consider limiting the acquisition time such that the nuclear morphology remains constant throughout the image series.

As a characteristic example, a full Hi-D workflow was applied to a spinning disk confocal sequence of histone H2B–green flourescent protein (GFP)-stained nuclei of a U2OS cell (Fig. 3a). Estimated trajectories are shown in Fig. 3b on which Bayesian inference is applied. Hi-D generates maps of diffusion models (Fig. 3c) and the normalized diffusion constant, anomalous exponent and drift velocity (and the radius of confinement, depending on the set of permitted MSD models; Fig. 3d). Finally, deconvolution by a GMM is applied to the biophysical parameters (D_{norm}, A and V). The result of this deconvolution stage is a particular combination of distribution type and the number of sub-populations for every parameter. Histograms and maps of populations are shown in Fig. 3e,f, respectively. The latter are useful to compare biological conditions with each other, including the appropriate controls for the biological question. However, since Hi-D is a noninvasive method, maps of biophysical parameters may also be compared across time, for example, before or after treatment of the same cell (for example, Fig. 3 in ref. 21). Preprocessing, optical flow computation and Bayesian fitting of the Hi-D analysis ('Procedure' section) can be executed in batch using the Python library papermill, as described in the Hi-D wiki (batch analysis). Stages 2 and 3 can be run in multiple files as described below.

This timing is computed based on the hardware described in the 'Materials' section. The processing time of the optical flow and Bayesian inference steps depends on the number of cores in the central processing unit (CPU) if the multicore functionality is enabled. This assumes that the running machine has sufficient memory if lengthy video sequences are supplied. The timing is estimated under the assumption that the user has basic coding experience in Python and understands how to run Python notebooks either via Jupyter Notebooks or relying on Visual Studio Code (Microsoft).

Materials

Reagents

- Dulbecco's modified Eagle medium, high glucose, pyruvate (DMEM; ThermoFisher, cat. no. 41966029)
- Fetal bovine serum (FBS; ThermoFisher, cat. no. 10270106)
- Sodium pyruvate solution (Sigma-Aldrich, cat. no. 113-24-6)
- Penicillin–streptomycin (Bioconcept, cat. no. 4-1F00H)
- Puromycin (ThermoFisher, cat. no. A1113803)
- Glutamax (ThermoFisher, cat. no. 35050061)
- Trypsin–EDTA solution (Sigma-Aldrich, cat. no. T4049)
- Doxycycline hyclate (Sigma-Aldrich, cat. no. D9891)

- SiR-DNA (5-610CP-Hoechst) (Grazvydas Lukinavicius, cat. no. NA)
- L-15 medium (Leibovitz) (ThermoFisher, cat. no. 11415064)
- HEPES buffer (Sigma-Aldrich, cat. no. 7365-45-9)

Biological materials

- U2OS cell line (ATCC, RRID: CVCL_0042)
- U2OS cell line stably expressing H2B–GFP (ATCC, a gift from Sébastien Huet, Rennes)
  ▲ **CRITICAL** The cell lines used should be regularly checked to ensure that they are authentic and are not infected with mycoplasma.

Equipment

- Spinning disk confocal microscopy system (both CSU-X1-M1N and CSU22, Yokogawa)
  ▲ **CRITICAL** The CSU-X1-M1N system was used to image live DNA 150-frame sequences with an exposure time of 200 ms per frame. Other imaging modalities like widefield microscopy can be used, but users should be aware that the resulting dynamics might be slightly altered at low signal-to-noise levels below 20 dB (ref. 21). While a Nipkow-disk confocal microscopy (CSU22) system was used to image live H2B and RNA Pol II 150-frame sequences with an exposure time of 200 ms per frame.
- Super-resolution structured illumination microscopy (DeltaVision OMX V3 Blaze)
  ▲ **CRITICAL** This system is used to image 20 frames of live H2B sequences with a frame duration of 2 s. The system is equipped with a PlanApo 60/1.42 oil immersion objective, a 488 nm laser and sCMOS cameras.
- Standard personal computer
  ▲ **CRITICAL** All protocols have been tested on three machines with different operating systems to ensure that our implementation is cross platform. The reference machine runs Linux (Ubuntu 20.04) and has an Intel Corei9 CPU with 2.8 GHz and 128 GB of DDR4 memory. Hi-D can be downloaded and run locally as a stand-alone application on Windows, Linux and macOS machines.

Software

▲ **CRITICAL** Software is installed by cloning the open-source GitHub repository at https://github.com/haitham-shaban/hidpy. If you do not have a GitHub account, the project can be downloaded as a.zip file. The code relies on the following open-source packages:

- Scientific computing package: OpenCV (opencv.org) and its Python bindings
- Python's scientific computing package: Numpy (numpy.org)
- Python's scientific computing package: Scipy (scipy.org)
- Python's plotting package: Matplotlib (matplotlib.org)
- Python's plotting package: Seaborn (seaborn.pydata.org)
- Python's probabilistic models' package: Pomegranate (pomegranate.readthedocs.io)
- Python's pipelining package: joblib (joblib.readthedocs.io)
- Python's concurrency package: multiprocessing (docs.python.org/3/library/multiprocessing.html)
- Jupyter Notebook with Visual Studio Code (https://code.visualstudio.com)
  ▲ **CRITICAL** All the Python packages can be directly installed by relying on the pip command. The code has been tested on Unix-based operating systems (Ubuntu 20.04 and macOSX 11) and Windows 10. It is recommended to use Visual Studio Code (code.visualstudio.com) to run the Python notebooks. The running machine must have at least 16 gigabytes of memory in a multicore CPU. There are no special requirements regarding the graphics card.
  ▲ **CRITICAL** Hi-D can be easily executed from a list of Python notebooks, allowing the user to run every stage interactively. All the notebooks are documented and labeled, making the connections between all the stages seamless. The code consists of three principal notebooks 01-hidpy-stage-1.ipynb, 02-hidpy-stage-2.ipynb and 03-hidpy-stage-3.ipynb. The 00-install-dependencies.ipynb contains a single panel that allows the user to verify if the installed Python environment where the notebooks are executed has all the needed

dependencies to successfully run all the code. Note that the code supports parallel execution using all available CPU cores in your computing node as executed from the Jupyter Notebooks. We also added detailed documentation of how to use the code in a step-by-step fashion on the Hi-D wiki page of the online repository.

Experimental setup

Cell culturing and staining for H2B imaging using spinning disk confocal microscopy

- A human osteosarcoma U2OS cell line stably expressing GFP tagged histones (H2B) was used. The same cell line U2OS that expresses RPB1 (subunit of RNA Pol II) fused with Dendra2 is used for RNA Pol II imaging, as previously described in ref. 45
- Grow the cells grown in DMEM containing phenol red-free Glutamax with 50 g/ml gentamicin, 1 mM sodium pyruvate, 10% FBS and G418 with 0.5 mg/ml at 37 °C in an incubator with 5% CO_2
- Grow cells in 100-mm cell culture dishes and split every other day
- One day before the imaging, plate the cell at a density of ~1×10^5 cells per Petri dish (no. 1.5 coverslip-like bottom)
- For starvation mode, incubate the cells for 24 h at 37 °C before imaging with serum-free medium (DMEM, Glutamax containing 50 µg/ml gentamicin, 1 mM sodium pyruvate and G418 0.5 mg/ml)
- Just before cell imaging, mount the cells in L-15 medium
- For cell stimulation, add 10% FBS to the L-15 medium for 10–15 min

Cell culturing and staining for DNA staining using spinning disk confocal microscopy

- On the day of imaging add the SiR-DNA (5-610CP-Hoechst) dye to the medium of the cultured cells at 37 °C for 30 min to 1 h after at a final concentration of 1 M
- Gently wash the cells three times with prewarmed PBS
- Before the nuclei imaging, change the medium to Leibovitz's L-15 medium adding all the medium supplements needed for live cell imaging

Cell culturing and staining for H2B imaging using super-resolution structured illumination microscopy

- Seed HeLa H2B–GFP in phenol red-free DMEM supplemented with 10% FBS and 1% penicillin–streptomycin
- Incubate the cells at 37 °C, 5% CO_2 in a humidified incubator
- At 90% confluence, trypsinize cells with 0.05 to 0.25% trypsin in PBS or 1× trypsin replacement solution (TrypLE Express, Gibco) and passaged to a new culture dish at appropriate dilutions (1:5 to 1:12 every 2–3 d)
- For live-cell structured illumination microscopy (SIM) imaging, seed cells in a 35-mm microdish, high glass bottom (Ibidi), using phenol red-free DMEM supplemented with 10 µM Hepes

Live-cell imaging and sequence generation

Live-H2B and RNA Pol II imaging using spinning disk confocal microscopy. The spinning disk confocal microscope CSU22 (Yokogawa) was used to record a 150-frame sequence with an exposure time per frame of 200 ms. A diode-pumped solid-state laser with a single wavelength of 488 nm (25 mW) was used to image the H2B–GFP. For RNA Pol II (Rbp1-Dendra2) imaging, a laser source (5–10% light intensity) with a wavelength of 488 nm was applied. Imaging was performed with a Plan Apo (100×/1.42 NA) oil immersion objective from Nikon and detected using a cooled electron-multiplying charge-coupled device camera (iXon Ultra 888), with a sample pixel size of 88 nm. The imaging was performed in a humid chamber at 37 °C.

Live-DNA imaging using spinning disk confocal microscopy. The microscope CSU-X1-M1N (Yokogawa) was used to record a 150-frame sequence with an exposure time per frame of 200 ms of DNA-SiR. For DNA-SiR molecules imaging, an excitation laser source with 140 mW power and a wavelength of 647 nm was used. The emitted light was collected with a Leica HCX-PL-APO (100×/1.4 NA) oil immersion objective lens. A CMOS camera (ORCA-Flash4.0 V2)

with (1 × 1 binning) was used to record the imaged videos with a sample pixel size of 65 nm. The imaging was performed in a humid chamber at 37 °C.

Live-H2B imaging using super-resolution SIM. A human (HeLa) somatic cell line stably expressing GFP-tagged H2B was imaged using SIM. The microscope DeltaVision OMX V3 Blaze was used to record a 20-frame series with an interval between frames of 2 s with a sample pixel size of 41 nm. At 37 °C and 5% CO_2 supply, live-cell imaging was carried out utilizing an objective heater and stage-top incubator.

Procedure

Preprocessing, optical flow calculation, trajectories construction, MSD estimation and Bayesian inference

● **TIMING 1–5 h per image sequence**

1. Open the Jupyter Notebook 01-hidpy-stage-1.ipynb, and then, execute it either by using 'shift-Enter' or by clicking the 'Play' button in the cell.
2. Users can use their own data or data from online available sources. TIFF image stacks or AVI videos are suitable input formats. Users should supply crops of single cells/nuclei (care should be taken that adjacent nuclei do not touch each other).
   ◆ **TROUBLESHOOTING**
3. In the Jupyter Notebook 01-hidpy-stage-1.ipynb, the user should set the input parameters. The input parameters are 'pixel_threshold', 'pixel_size', 'dt' and 'R square_threshold'. The pixel threshold is used to exclude pixels under a given intensity value. The pixel size depends on the user's experimental setup and camera. R_square_threshold is used to select pixels based on the quality of the fitting performed in the Bayesian step. Default values are suggested for the demo dataset.
   ◆ **TROUBLESHOOTING**
4. Denoising: denoise the entire sequence by using noniterative bilateral filtering to verify the difference in intensity values to preserve the edges of the structures. Further details on the denoising setting parameters are provided in the code documentation page (https://github.com/haitham-shaban/hidpy/wiki).
   ◆ **TROUBLESHOOTING**
5. Optical flow computations: run the defined cell for optical flow computations, then optical flow fields of the entire sequence of N frames, a list of $N-1$ are automatically computed.
6. Trajectory estimation: execute the Jupyter Notebook cell of trajectories estimation, then the trajectories of every pixel whose value is above a given intensity threshold (Step 3) will be generated.
7. Visualize the resulting trajectories: run the Jupyter Notebook cell for trajectory visualization. This step shows plots for the estimated trajectories overlayed onto the first frame of the sequence. In this step, the user needs to ensure that the trajectories are located within the extent of the nucleus. Common problems are shown in Supplementary Fig. 3.
   ◆ **TROUBLESHOOTING**
8. Trajectory maps reconstruction: once the estimated trajectories have been calculated and verified in the previous step, the user can execute the trajectory map reconstruction Jupyter Notebook cell. This process converts the trajectories into flow field maps (in microns).
9. MSD estimation: run the MSD estimation Jupyter Notebook cell. In this step, Hi-D automatically computes the MSD for each trajectory within the nucleus mask.
10. Bayesian inference: execute the Bayesian inference Jupyter Notebook cell. Automatically, Hi-D will perform a Bayesian inference[21,27] on the previously estimated MSD to choose the best fitting model and thereby the corresponding parameters (D, A and V).
    ▲ **CAUTION** The time step (Δt) between the frames in the input sequence depends on the imaging procedure and the configuration of the microscope.
11. Model selection and parameters are mapped: create maps of the diffusion parameters that were chosen by the Bayesian selection by clicking on the Jupyter Notebook cell

after the Bayesian inference cell (examples of these mappings are shown in Fig. 3c,d). Additionally, a map of the coefficient of determination (R^2) from least-square fitting is shown (Supplementary Fig. 4). Hi-D implements a default threshold value of $R^2 > 0.9$. A higher R^2 threshold implies higher statistical significance, but it could also lead to a reduction in the number of analyzed pixels (Supplementary Fig. 4). Advanced users could rerun this cell, increasing or decreasing the R^2 threshold.

◆ **TROUBLESHOOTING**

12. Storing the Bayesian results: in the last step of the Jupyter Notebook 01-hidpy-stage-1.ipynb. *Hi*-D, save the Bayesian results for all the parameters (*D*, *A* and *V*) in a pickle file[*.pickle].

Deconvolution by a GMM

● **TIMING a few seconds to a minute**

13. Setting deconvolution parameters: open the notebook 02-hidpy-stage-2.ipynb and set the input parameters of the pickle files directory and models in the input parameters panel. In addition, before GMM deconvolution is performed, users must choose the maximum number of distributions (n_{pop}), the model parameter to deconvolve and the type of distribution (normal or log-normal) to be used by the GMM.

    ▲ **CAUTION** The number of histogram bins is selected by the user for visualization purposes After the user agrees with the input parameters, the user can run this Jupyter Notebook cell.

    ◆ **TROUBLESHOOTING**

14. GMM evaluation: run GMM evaluation Jupyter Notebook cell to perform GMM on all types of distributions (normal and log-normal) and for subpopulations (1 to n_{pop}). Evaluation is done using the BIC. The deconvolution results are summarized in Table 2 representing the fraction of cells from the analyzed dataset.
15. GMM table evaluation: users should inspect the GMM table and select the optimal combination of distribution type and the number of subpopulations for each parameter within the analyzed data set. Unless prior knowledge regarding the number of populations and/or the distribution is available, users are advised to select the optimal combination. This means the chosen number of populations and distribution type encompasses the largest fraction of cells in the dataset.
16. Deconvolution: after the user executes the Jupyter Notebook cell, Hi-D will generate two visualizations: histograms for each parameter to identify subpopulations and subpopulation maps within the cell nucleus for each chosen model parameter. Examples of results from GMM deconvolution are shown as histograms (Fig. 3e) and maps (Fig. 3f) for each biophysical parameter.

    ◆ **TROUBLESHOOTING**

17. By the end of this analysis, GMM deconvolution generates a pickle file named Statistics. pickle for each analyzed video.

Comparison between average statistics of subpopulations from different biological conditions

● **TIMING a few seconds to a minute**

18. Create a folder for each condition that includes all Statistics.pickle files that are created by GMM of the analyzed nuclei.
19. Open the Jupyter Notebook 03-hidpy-stage-3.ipynb.
20. Set comparison parameters: before performing the comparison, users should create a list containing the paths of each Statistics.pickle file for each condition in the input panel of notebook 03-hidpy-stage-3.ipynb. Users should write the list of paths in the input panel as follows: Pickles_Cond[0] = [path0]. Pickles_Cond[i] = [path i], where *I* represent the number of conditions; an example is provided in the notebook. Additionally, the names of each condition and the list of deconvolved parameters should be defined inside the Jupyter cell notebook. The name of each condition will be used for Hi-D as a label for the plots.
21. Comparison: once the input panel is set, execute 03-hidpy-stage-3.ipynb. This will generate a swarmplot that compares the different conditions and populations. An example output from this stage is shown in Fig. 4. In this evaluation, SiR-DNA-stained U2OS cells, both serum

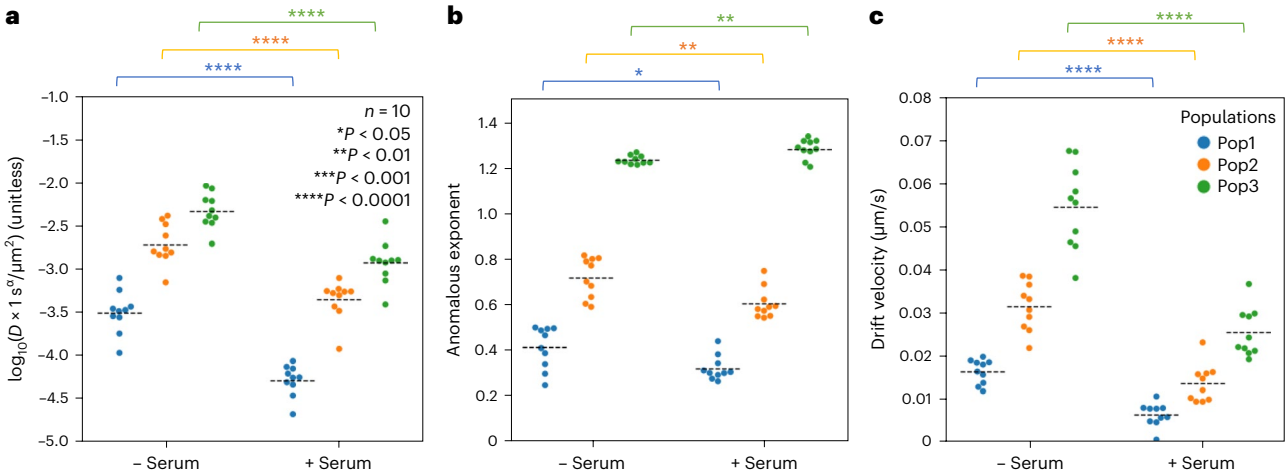

Fig. 4 | Comparison between two experimental conditions using Hi-D. **a**–**c**, SiR-DNA-stained U2OS cells, both serum-starved and serum-stimulated (n =10 cells per condition) were analyzed. Comparisons of average values for diffusion-normalized constants (**a**), anomalous exponents (**b**) and drift velocities (**c**) per population and cell are shown. Statistically, a t-test was performed using the Python library SciPy. The dashed lines represent average values.

starved and serum stimulated (n = 10 cells per condition), were used, as described in the original Hi-D publication[21]. Identically to the original Hi-D results[21], the normalized diffusion constant (D_{norm}) and anomalous exponent (A) values of DNA in serum-starved U2OS nuclei showed the same changing patterns in all three populations upon the addition of serum (Fig. 4a,b). Similarly, the drift velocity of DNA in serum-starved U2OS nuclei decreased in all three populations upon the addition of serum (Fig. 4c).

Troubleshooting

Troubleshooting advice can be found in Table 3.

Table 3 | Troubleshooting table

Step	Problem	Possible reason	Solution
2	Input video sequences are not loaded	Unsupported file format	Convert the sequence into any supported format, for example, tiff stack
3	Results are not generated to the specified output directory	The directory is not valid	Verify the path of the output directory
4	Noisy video sequences after bilateral denoising	Imaging artifacts such as noise	Use any image processing software, for example, Fiji, to denoise the sequence using median filters. The window size of the bilateral denoising might be increased by two pixels
7	Trajectories are appearing outside the locus of the nuclei	Imaging artifacts such as noise	Readjust the pixel threshold value and repeat Steps 1–10 until a convenient mapping is reached. The acceptable range of the threshold is between 0 and 1, and the default value is 0.05. The value can only be adjusted by trial and error due to several factors, including the imaging conditions and the nature of noise in the images
	Trajectories all move collectively in a certain direction	Nucleus contains drift	Drift may be eliminated using Fiji or another image processing program. Care should be taken to correct lateral drift smoothly and only use integer pixel values to avoid interpolation of pixel's intensity, which alters pixel intensity values and often blurs images
11	Parameters are higher than expected	Error in units	Verify pixel size and time step in input configuration. The units for the pixel size are micrometers
	Bayesian maps show a region outside of the cell	Error in pixel threshold	Adjust the pixel intensity threshold value
	Execution failure in case of long or high-resolution videos	Memory limitations	Use a single core for processing instead of multiple cores
	Fitting failure in Bayesian inference	Fitting parameters are not adapted	Correct fitting parameters in the Python file core/inference.py

Table 3 (continued) | Troubleshooting table

Step	Problem	Possible reason	Solution
13	Error empty sequence in deconvolution stage	Empty pickle folder	Verify that at least one pickle file is in the folder given to the deconvolution stage
16	Execution failure during GMM constrained	The model fitting did not converge	Try different numbers of populations and/or distribution type

Timing

The timing reported in the protocol is registered based on several video sequences with 150 frames. However, we expect a total running time, on average, of 1–5 h to run the entire pipeline. Average timing estimates for the entire protocol based on the exemplar datasets:
Cell culturing, staining, and imaging: 2–4 d
Steps 1–4, data preprocessing: several seconds to a few minutes
Step 5, optical flow computations: 10–100 s per frame
Steps 6–8, Trajectory computations: a few minutes
Steps 9–12, MSD-Bayes Analysis: a few to several minutes
Steps 13–17, GMM deconvolution: a few seconds
Steps 18–21, Average statistics comparison for subpopulations from different biological conditions: a few seconds

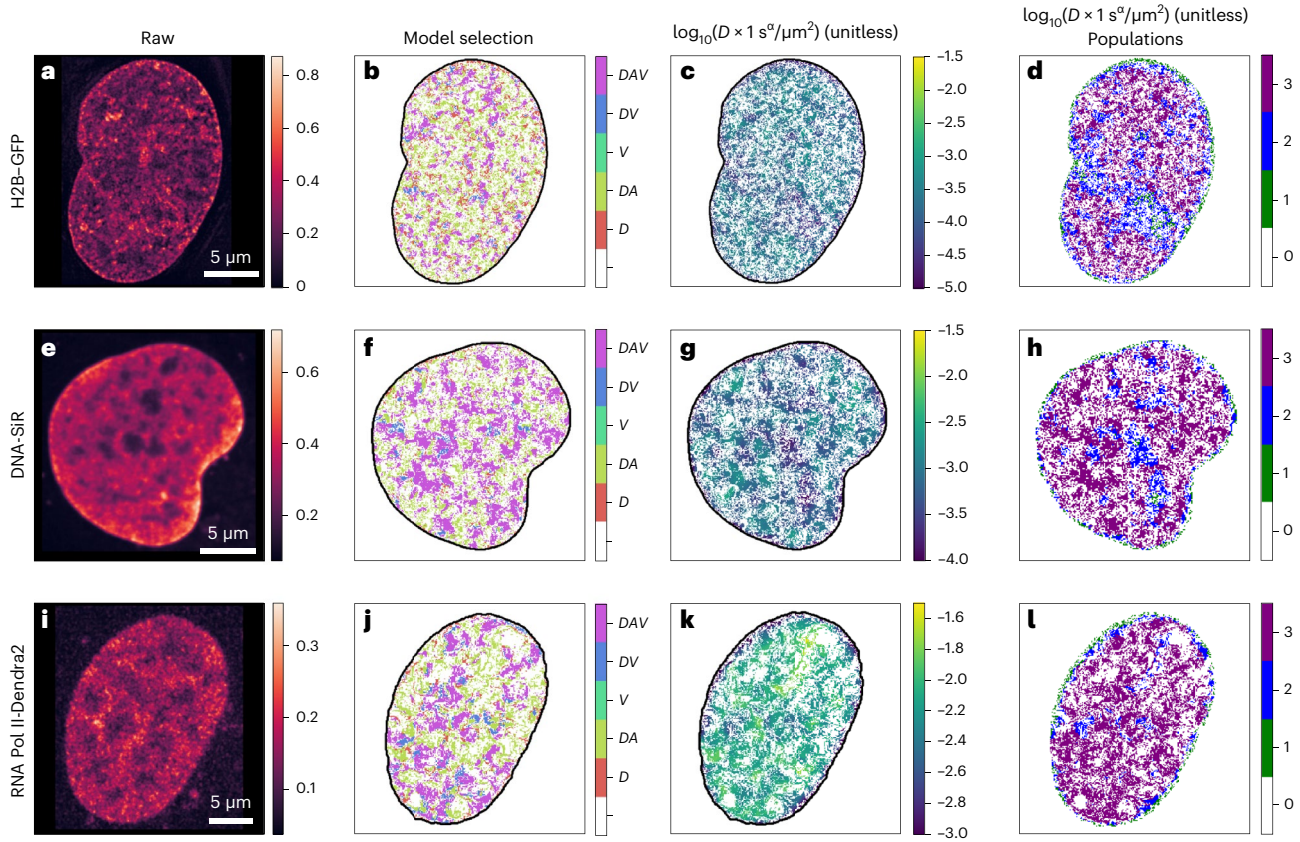

Fig. 5 | Hi-D applied to analyze live cell imaging of chromatin and nuclear proteins using two microscopy modalities. a–**d**, Live-nucleus imaging of Hela cell expressing H2B–GFP imaged using SIM with $\Delta t = 2$ s for raw (**a**), model selection (**b**), $\log_{10}$ calculation (**c**) and $-\log_{10}$ calculation (**d**). **e**–**l**, Live cell imaging of DNA-SiR stained (**e**–**h**) (for raw (**e**), model selection (**f**), $\log_{10}$ calculation (**g**) and $-\log_{10}$ calculation (**h**)) and RNA Pol II (Rbp1-Dendra2) expression in U2Os cells (**i**–**l**) (raw (**i**), model selection (**j**), $\log_{10}$ calculation (**k**) and $-\log_{10}$ calculation (**l**)) imaged using a spinning disk microscope with $\Delta t = 0.2$ s. Representative maps are displayed for each sample type: model selection for **b**,**f** and **j**; diffusion-normalized constant for **c**,**g** and **k**; and the estimated population map for **d**,**h** and **l** from deconvolution using GMM are shown for the analyzed cells H2B, DNA and RNA Pol II, respectively.

Anticipated results

The Hi-D software permits the quantification, classification, and mapping of diffusion parameters and physical diffusion models over the entire genome in living cells. Hi-D takes advantage of the Python programming environment to improve speed and user-friendliness through Jupyter Notebooks. Hi-D has so far been used with data obtained from two different microscopy techniques (Fig. 5): structured illumination microscopy for the analysis of H2B (Fig. 5a–d) and spinning disk confocal microscopy for the analysis of DNA (Fig. 5e–h) and RNA (Fig. 5i–l).

The different mobility classes of the diffusion constant, anomalous exponent and drift velocity reflect the cell's dynamic landscape. Outputs generated by Hi-D may be used to determine the impact of a certain biological condition (a drug interfering with nuclear processes or external stimuli) on a cell to determine mobility differences between cell lines. The estimated number of populations by GMM is informative about, for example, the extent of chromatin-bound, chromatin-processing and unbound fractions of a nuclear molecule. Similarly, chromatin (re)arrangements and different complications that are associated with different cell states and pathogenic conditions can be correlated to chromatin mobility[41]. The mean/median values of D_{norm}, A and V per subpopulation allow us to express mobility changes quantitatively. We, thus, encourage the chromatin organization community to complement the presented Hi-D method with external stimuli and additional markers of nuclear processes of interest, both locus-specific as well as genome-wide. The incorporation of Hi-D as a simple, fast and user-friendly analysis package for nuclear mobility greatly enhances the aspects of nuclear biology research beyond static and population-average methods as well as restrictive sparse loci methods.

Reporting summary

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

Data availability

The main data discussed in this protocol are available in the supporting primary research papers[21,22].

Code availability

All source codes of Hi-D are publicly available under the GPL-3.0 license at https://github.com/haitham-shaban/hidpy. Documentation is available on GitHub at https://github.com/haitham-shaban/hidpy/wiki.

Received: 29 October 2022; Accepted: 22 April 2024;
Published online: 28 August 2024

References

1. Misteli, T. The self-organizing genome: principles of genome architecture and function. *Cell* https://doi.org/10.1016/j.cell.2020.09.014 (2020).
2. Agbleke, A. A. et al. Advances in chromatin and chromosome research: perspectives from multiple fields. *Mol. Cell.* https://doi.org/10.1016/j.molcel.2020.07.003 (2020).
3. Dekker, J. & Mirny, L. The 3D genome as moderator of chromosomal communication. *Cell* **164**, 1110–1121 (2016).
4. Bhat, P., Honson, D. & Guttman, M. Nuclear compartmentalization as a mechanism of quantitative control of gene expression. *Nat. Rev. Mol. Cell Biol.* **22**, 653–670 (2021).
5. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* https://doi.org/10.1038/s41576-018-0089-8 (2019).
6. Shaban, H. A., Barth, R. & Bystricky, K. Navigating the crowd: visualizing coordination between genome dynamics, structure, and transcription. *Genome Biol.* https://doi.org/10.1186/s13059-020-02185-y (2020).
7. Gabriele, M. et al. Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* **376**, 496–501 (2022).
8. Shaban, H. A. & Seeber, A. Monitoring the spatio-temporal organization and dynamics of the genome. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkaa135 (2020).
9. Shaban, H. A. & Seeber, A. Monitoring global chromatin dynamics in response to DNA damage. *Mutat. Res.* https://doi.org/10.1016/j.mrfmmm.2020.111707 (2020).
10. Marshall, W. F. et al. Interphase chromosomes undergo constrained diffusional motion in living cells. *Curr. Biol.* https://doi.org/10.1016/S0960-9822(06)00412-X (1997).

11. Levi, V., Ruan, Q., Plutz, M., Belmont, A. S. & Gratton, E. Chromatin dynamics in interphase cells revealed by tracking in a two-photon excitation microscope. *Biophys. J.* **89**, 4275–4285 (2005).
12. Germier, T. et al. Real-time imaging of a single gene reveals transcription-initiated local confinement. *Biophys. J.* **113**, 1383–1394 (2017).
13. Gu, B. et al. Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science* https://doi.org/10.1126/science.aao3136 (2018).
14. Hansen, A. S. et al. Robust model-based analysis of single-particle tracking experiments with spot-on. *eLife* https://doi.org/10.7554/eLife.33125 (2018).
15. Bronstein, I. et al. Transient anomalous diffusion of telomeres in the nucleus of mammalian cells. *Phys. Rev. Lett.* **103**, 18102 (2009).
16. Eshghi, I., Eaton, J. A. & Zidovska, A. Interphase chromatin undergoes a local sol-gel transition upon cell differentiation. *Phys. Rev. Lett.* **126**, 228101 (2021).
17. Di Rienzo, C., Gratton, E., Beltram, F. & Cardarelli, F. Fast spatiotemporal correlation spectroscopy to determine protein lateral diffusion laws in live cell membranes. *Proc. Natl Acad. Sci. USA* https://doi.org/10.1073/pnas.1222097110 (2013).
18. Hebert, B., Costantino, S. & Wiseman, P. W. Spatiotemporal image correlation spectroscopy (STICS) theory, verification, and application to protein velocity mapping in living CHO cells. *Biophys. J.* https://doi.org/10.1529/biophysj.104.054874 (2005).
19. Zidovska, A., Weitz, D. A. & Mitchison, T. J. Micron-scale coherence in interphase chromatin dynamics. *Proc. Natl Acad. Sci. USA* **110**, 15555–15560 (2013).
20. Shaban, H. A., Barth, R. & Bystricky, K. Formation of correlated chromatin domains at nanoscale dynamic resolution during transcription. *Nucleic Acids Res.* **46**, e77 (2018).
21. Shaban, H. A., Barth, R., Recoules, L. & Bystricky, K. Hi-D: nanoscale mapping of nuclear dynamics in single living cells. *Genome Biol.* **21**, 95 (2020).
22. Miron, E. et al. Chromatin arranges in chains of mesoscale domains with nanoscale functional topography independent of cohesin. *Sci. Adv.* https://doi.org/10.1126/sciadv.aba8811 (2020).
23. Barth R, Bystricky K, Shaban HA. Coupling chromatin structure and dynamics by live super-resolution imaging. *Sci. Adv*. https://doi.org/10.1126/sciadv.aaz2196 (2020).
24. Barth, R., Fourel, G. & Shaban, H. A. Dynamics as a cause for the nanoscale organization of the genome. *Nucleus* **11**, 83–98 (2020).
25. Barth, R. & Shaban, H. A. Spatially coherent diffusion of human RNA Pol II depends on transcriptional state rather than chromatin motion. *Nucleus* **13**, 194–202 (2022).
26. Shaban, H. A. et al. Individual transcription factors modulate both the micromovement of chromatin and its long-range structure. *Proc. Natl Acad. Sci. USA* **121**, e2311374121 (2024).
27. Monnier, N. et al. Bayesian approach to MSD-based analysis of particle motion in live cells. *Biophys. J.* **103**, 616–626 (2012).
28. Seber, G. & Wild, C. *Nonlinear Regression* (John Wiley & Sons, 2003).
29. Schreiber, J. M. & Noble, W. S. Finding the optimal Bayesian network given a constraint graph. *Peer J. Comput. Sci.* **2017**, 1–16 (2017).
30. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
31. Saxton, M. J. Anomalous diffusion due to obstacles: a Monte Carlo study. *Biophys. J.* **66**, 394–401 (1994).
32. Saxton, M. J. Diffusion of DNA-binding species in the nucleus: a transient anomalous subdiffusion model. *Biophys. J.* **118**, 2151–2167 (2020).
33. Banks, D. S., Tressler, C., Peters, R. D., Höfling, F. & Fradin, C. Characterizing anomalous diffusion in crowded polymer solutions and gels over five decades in time with variable-lengthscale fluorescence correlation spectroscopy. *Soft Matter* **12**, 4190–4203 (2016).
34. Nieman, G. C. & Robinson, G. W. Rapid triplet excitation migration in organic crystals. https://doi.org/10.1063/1.1733439 (1962).
35. Saxton, M. J. & Jacobson, K. Single-particle tracking: applications to membrane dynamics. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 373–399 (1997).
36. Benelli, R. & Weiss, M. Probing local chromatin dynamics by tracking telomeres. *Biophys. J.* **121**, 2684–2692 (2022).
37. Wit, E., van den Heuvel, E. & Romeijn, J. W. 'All models are wrong. ': an introduction to model uncertainty. *Stat. Neerl.* **66**, 217–236 (2012).
38. Sun, D., Roth, S. & Black, M. J. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *Int. J. Comput. Vis.* **106**, 115–137 (2014).
39. Farnebäck, G. Two-frame motion estimation based on polynomial expansion. *Lect. Notes Comput. Sci.* https://doi.org/10.1007/3-540-45103-x_50 (2003).
40. Mach, P. et al. Cohesin and CTCF control the dynamics of chromosome folding. *Nat. Genet.* **54**, 1907–1918 (2022).
41. Shaban, H. A. & Gasser, S. M. Dynamic 3D genome reorganization during senescence: defining cell states through chromatin. *Cell Death Differ.* https://doi.org/10.1038/s41418-023-01197-y (2023).
42. Hinde, E., Cardarelli, F., Digman, M. A. & Gratton, E. In vivo pair correlation analysis of EGFP intranuclear diffusion reveals DNA-dependent molecular flow. *Proc. Natl Acad. Sci. USA* **107**, 16560–16565 (2010).
43. Di Bona, M. et al. Measuring mobility in chromatin by intensity-sorted FCS. *Biophys. J.* **116**, 987–999 (2019).
44. Schreiber, J. pomegranate: fast and flexible probabilistic modeling in python. *J. Mach. Learn. Res.* **18**, 1–6 (2018).
45. Cisse, I. I. et al. Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* **341**, 664–667 (2013).

Acknowledgements

H.A.S. acknowledges the financial support of the ISREC foundation. C.A.V.-C. acknowledges the funding by Labex Cell(n)Scale (ANR-11-LABX-0038) as part of the Idex PSL (ANR-10-IDEX-0001-02). M.A. and R.B. contribute to this work independently.

Author contributions

H.A.S. conceived the study. C.A.V.-C. and M.A. implemented the core framework. R.B. implemented the optical flow and validation modules. C.A.V.-C. implemented the Bayesian and deconvolution modules. M.A. implemented the trajectory computations. H.A.S. performed the live-cell experiments and imaging protocols. C.A.V.-C., R.B., M.A. and H.A.S wrote the manuscript. C.A.V.-C., R.B. and M.A. contributed equally to this work. H.A.S. supervised the project. All authors approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41596-024-01038-3.

Correspondence and requests for materials should be addressed to Haitham A. Shaban.

Peer review information *Nature Protocols* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Related links

Key references using this protocol:
Shaban, H. A. et al. *Genome Biol.* **21**, 95 (2020): https://doi.org/10.1186/s13059-020-02002-6
Shaban, H. A. et al. *Nucleic Acids Res.* **46**, e77 (2018): https://doi.org/10.1093/nar/gky269
Shaban, H. A. et al. *Proc. Natl Acad. Sci. USA* **121**, e2311374121 (2024): https://doi.org/10.1073/pnas.2311374121

nature portfolio

Corresponding author(s): Haitham A. Shaban

Last updated by author(s): 2024-04-01

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
	A description of all covariates tested
	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Spinning disk confocal microscopy system (Both CSU-X1-M1N and CSU22, Yokogawa). Super-resolution structured illumination microscopy
Data analysis	Hi-D (matlab) can be found in https://github.com/romanbarth/Hi-D.Hi-D(python) can be found in https://github.com/haitham-shaban/hidpy.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The authors declare that the main data discussed in this protocol are available in the supporting primary research papers (https://doi.org/10.1186/s13059-020-02002-6 and https://doi.org/10.1126/sciadv.aba8811).

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

◉ Life sciences ○ Behavioural & social sciences ○ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For this manuscript, cell number was chosen to illustrate the difference capabilities of the software.
Data exclusions	No data were excluded from the analyses.
Replication	All data were found to be reproducible over different cells measured.
Randomization	There was not specific allocation to experimental groups in this manuscript.
Blinding	There was not need for data allocation or blinding in this manuscript.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing	
Data exclusions	
Non-participation	
Randomization	

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing and spatial scale	
Data exclusions	
Reproducibility	
Randomization	

Blinding

Did the study involve field work? ○ Yes ○ No

Field work, collection and transport

Field conditions

Location

Access & import/export

Disturbance

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
○	Antibodies
●	Eukaryotic cell lines
○	Palaeontology and archaeology
○	Animals and other organisms
○	Clinical data
○	Dual use research of concern

Methods

n/a	Involved in the study
○	ChIP-seq
○	Flow cytometry
○	MRI-based neuroimaging

Antibodies

Antibodies used

Validation

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s): U2OS cell line (RRID:CVCL_0042) (ATCC). U2OS cell line stably expressing H2B-GFP (ATCC. A gift from S. Huet (Rennes)).

Authentication

Mycoplasma contamination: U2OS cells are tested regularly for mycoplasma contamination.

Commonly misidentified lines (See ICLAC register)

Palaeontology and Archaeology

Specimen provenance

Specimen deposition

Dating methods

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	
Wild animals	
Reporting on sex	
Field-collected samples	
Ethics oversight	

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	
Study protocol	
Data collection	
Outcomes	

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
○	○	Public health
○	○	National security
○	○	Crops and/or livestock
○	○	Ecosystems
○	○	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
○	○	Demonstrate how to render a vaccine ineffective
○	○	Confer resistance to therapeutically useful antibiotics or antiviral agents
○	○	Enhance the virulence of a pathogen or render a nonpathogen virulent
○	○	Increase transmissibility of a pathogen
○	○	Alter the host range of a pathogen
○	○	Enable evasion of diagnostic/detection modalities
○	○	Enable the weaponization of a biological agent or toxin
○	○	Any other potentially harmful combination of experiments and agents

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication*	
Files in database submission	
Genome browser session (e.g. UCSC)	

Methodology

Replicates

Sequencing depth

Antibodies

Peak calling parameters

Data quality

Software

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Instrument

Software

Cell population abundance

Gating strategy

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type

Design specifications

Behavioral performance measures

Acquisition

Imaging type(s)

Field strength

Sequence & imaging parameters

Area of acquisition

Diffusion MRI ○ Used ○ Not used

Preprocessing

Preprocessing software

Normalization

Normalization template

Noise and artifact removal

Volume censoring

Statistical modeling & inference

Model type and settings

Effect(s) tested

Specify type of analysis: ○ Whole brain ○ ROI-based ○ Both

Statistic type for inference (See Eklund et al. 2016)

Correction

Models & analysis

n/a | Involved in the study

- Functional and/or effective connectivity
- Graph analysis
- Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Graph analysis

Multivariate modeling and predictive analysis