scientific reports

OPEN Single molecule tracking and analysis framework including theory-predicted parameter settings

Timo Kuhn[1,3], Johannes Hettich[1,3], Rubina Davtyan[1,2] & J. Christof M. Gebhardt[1]✉

Imaging, tracking and analyzing individual biomolecules in living systems is a powerful technology to obtain quantitative kinetic and spatial information such as reaction rates, diffusion coefficients and localization maps. Common tracking tools often operate on single movies and require additional manual steps to analyze whole data sets or to compare different experimental conditions. We report a fast and comprehensive single molecule tracking and analysis framework (TrackIt) to simultaneously process several multi-movie data sets. A user-friendly GUI offers convenient tracking visualization, multiple state-of-the-art analysis procedures, display of results, and data im- and export at different levels to utilize external software tools. We applied our framework to quantify dissociation rates of a transcription factor in the nucleus and found that tracking errors, similar to fluorophore photobleaching, have to be considered for reliable analysis. Accordingly, we developed an algorithm, which accounts for both tracking losses and suggests optimized tracking parameters when evaluating reaction rates. Our versatile and extensible framework facilitates quantitative analysis of single molecule experiments at different experimental conditions.

Single-molecule experiments are gaining increasing importance when investigating dynamical and structural parameters such as binding kinetics, diffusion coefficients or spatial distributions of biomolecules in living systems[1–4]. In these experiments, the biomolecule of interest is typically fused to a fluorescent label, such that the signal of fluorescent photons in successive recordings reports on the position and movement of the biomolecule. Extracting quantitative information from such movies includes linking individual detections of biomolecules at consecutive time points to continuous tracks[5–10]. In recent years, several tracking algorithms have been adapted to tracking of biomolecules, including basic nearest neighbour and more complex algorithms such as Kalman filtering, combinatorial optimization, multiple hypothesis tracking or neural networks[6,11–15]. Still, the intuitive application of tracking algorithms is challenged by a significant dependence on empirical parameters, such as the tracking radius which is oftentimes determined on visual aspects[9]. For the extraction of diffusion coefficients, a tracking radius of three times the root mean squared displacement was shown to yield accurate results[12,16,17]. In addition, upper limits to minimize misconnections were estimated[18]. Similar rules for binding time analysis, however, are still missing. Furthermore, only few of the particle tracking and analysis software published up to now were made accessible to a broader audience by providing an intuitively operable graphical user interface[13,19–23]. In addition, the subsequent analysis steps to extract quantitative information from tracked molecules are mostly left to additional software such as SMTracker[24], Spot-ON[25] or vbSPT[26]. Thus, analyzing whole data sets consisting of multiple movies or comparing different experimental conditions is cumbersome as they require extensive manual intervention. Overall, there is high demand for a user-friendly, comprehensive program covering tracking, analysis, and data visualization of single molecule experiments.

Linking detections of biomolecules into tracks unavoidably comes with errors, amongst others due to missed detections, inappropriate tracking radius or mix-up at high molecule density[6,9,16]. The probability for errors can be reduced by low molecule densities, though[27]. Premature loss of a track also occurs if the fluorescent label photobleaches. Photobleaching can be conveniently corrected for, e.g. by comparison with immobile histone molecules[28], by ensemble measurements[29,30], or by time-lapse imaging[31]. In contrast, errors inherent to the linking process are more challenging to tackle and often only assessed qualitatively. Recently, the effect of allowing several

[1]Institute of Biophysics, Ulm University, Albert-Einstein-Allee 11, 89081 Ulm, Germany. [2]Present address: NanoLund and Solid State Physics, Lund University, Box 118, 22100 Lund, Sweden. [3]These authors contributed equally: Timo Kuhn and Johannes Hettich. ✉email: christof.gebhardt@uni-ulm.de

gaps in assembling tracks of immobile molecules has been considered[23]. Also the impact of the tracking radius on the diffusion coefficient has been discussed[9,16]. In diffusion analysis, molecules diffusing out of the focal plane need to be accounted for[25]. However, a theoretical description of tracking errors and how they can be corrected for is missing for the determination of the binding (or residence) time of fluorescently labelled biomolecules.

We introduce TrackIt, an integrative tracking and analysis software for fast and extensive analysis of single molecule data sets within a single framework. Our user-friendly graphical user interface (GUI) provides access to two different tracking algorithms and multiple analysis procedures for kinetic and spatial parameters. Moreover, it comes with several data visualization options. Additionally, tracking results can be exported to utilize external analysis algorithms such as Spot-ON and vbSPT[25,26]. Analysis parameters and associated data are organized in a batch structure such that the workflow starting from spot detection over tracking to quantitative analysis can be repeated in a single step. This enables convenient comparison of different tracking parameters as well as of data sets recorded using different experimental conditions. Furthermore, we introduce a formalism to estimate tracking errors when linking immobile molecules and quantify the decrease of tracking losses if a gap frame is allowed. Using this formalism, we calculate optimal parameter settings for tracking and subsequent residence time analysis. We apply the theory-suggested parameter set to extract residence times of the transcription factor CDX2 in the cell nucleus from a single molecule experiment[32].

Results

Single molecule tracking.
Our single molecule tracking and analysis framework is designed to simultaneously analyse and compare several multi-movie data sets corresponding to different experimental conditions such as movie acquisition schemes or biochemical treatments, thereby facilitating the workflow (Fig. 1 and Supplementary Material). The data-loading tool automatically scans selected folder structures for tiff-formatted movies. Specific regular expressions in filenames can be used to automatically determine frame cycle times or to select for experimental conditions. In addition, accompanying images or movies carrying information about regions of interests (ROIs) such as the cell nucleus can be loaded. We implemented common movie handling and visualization features such as brightness, contrast and z-projection. Movies within the same or a different data set can be conveniently accessed.

Our single molecule tracking approach includes four steps to detect individual molecules and link their motion through consecutive images. First, we apply a combination of two wavelet filters to enhance spots representing single molecules[33]. Wavelet filters performed well in the particle tracking challenge[6]. Second, we select spot candidates using a local maximum search approach and filtering candidates with a user-defined intensity threshold. Third, we refine the localization of the filtered spots using TrackNTrace's fast 2D Gaussian fit[19]. Finally, we link spot localizations into tracks using a simple model-free nearest neighbour algorithm, which is widely used at low or intermediate spot densities[6,9,10,16,27]. For high spot densities or movement models known a prior, we implemented u-track as an alternative tracking algorithm[11]. The nearest neighbour algorithm links spots that are nearest neighbours in two consecutive frames as long as their distance does not exceed a user-defined tracking radius. To account for fluorophore blinking and stochastic fluctuations in spot intensity, the tracking algorithm may bridge missing detections (gap frames) in a user-defined number of subsequent frames[23], as long as the first track segment contains a certain number of detections. The choice of tracking radius and concatenating track segments may introduce tracking errors, which we discuss below. Our GUI allows controlling all four steps of the tracking workflow. Importantly, we implemented the possibility to compare different choices of tracking parameters to enable assessing their influence on the tracking results.

Our framework simplifies the effort of analysing numerous movies of multiple data sets by applying the detection and linking steps to all loaded movies without further input by the user. The results and tracking parameters are stored in a single analysis batch structure. This unique data structure summarizes associated files, properties of detected spots and tracks and all tracking parameters in one single file. Thus, reproducing results and comparing multiple processed data sets is possible with minimal effort.

Data analysis.
We implemented a GUI-module to analyse multiple characteristics of single molecule tracks and to display the results, enabling direct comparison of different data sets (Fig. 1 and 2, "Materials and methods" and Supplementary Material). Tracked molecules can be visualized using a set of intuitive tools allowing both directly inspecting the effect of changes in tracking or analysis parameters and accessing their spatiotemporal dynamics. Besides conventional plotting of spots and tracks, heat maps of localizations[2,34,35] and jump distances[36] can be displayed. We distinguish mobile and immobile molecules based on the time spend within a certain area[28–30] ("Materials and methods"). Lifetimes of immobile molecules, imaged using different time-lapse conditions to allow for photobleaching correction[31], are collected in survival time distributions[29]. From these, the complete spectrum of dissociation rates is extracted by inverse Laplace transformation using GRID[32]. Also, the fractions of bound molecules can be assessed using interlaced time-lapse microscopy[37]. For mobile molecules imaged at sufficient acquisition speed, analysis of the jump distances within each track yields diffusion coefficients[38], the bound fraction as amplitude of apparent slow diffusion due to the localization error[29,39] and confinement radii as function of the mean displacement[40]. In addition, histograms of the angles between the jumps of a track can be displayed, informing on compact versus non-compact diffusion[18,41,42]. We further implemented the possibility to analyse the intensity profile of tracks[43]. The batch structure facilitates assessing differences between multiple experimental conditions.

To ensure transparency and flexibility, we included a data export option, with which track coordinates are stored in specific Matlab or csv file formats to enable utilizing external analysis environments, such as Spot-On or vbSPT for diffusion parameters[25,26].
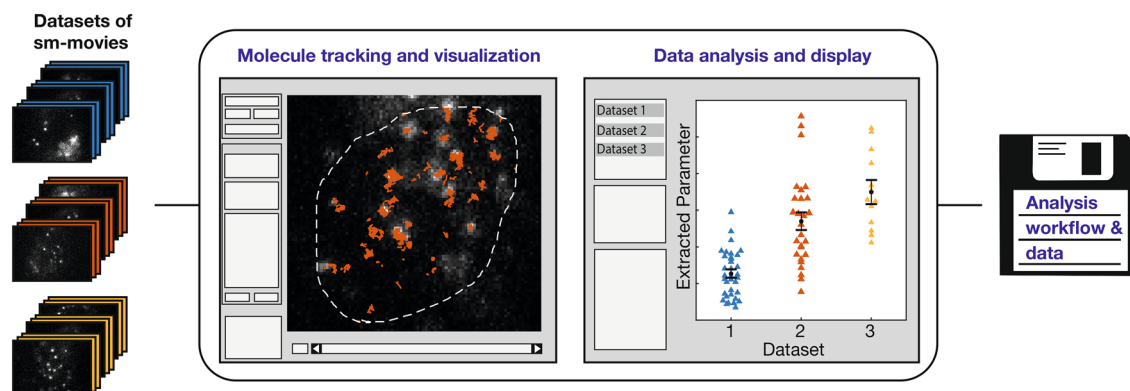
Figure 1. Workflow of TrackIt software. TrackIt enables comparative tracking and analysis of multiple single molecule data sets, each consisting of multiple movies. After loading and manual or automatic determination of tracking parameters, molecules are detected and tracked within a region of interest. Various display options of tracks provide visual feedback on the tracking process. Within the analysis tool, multiple quantitative parameters can be extracted, including bound fractions, residence times, diffusion coefficients, jump angles and confinement radii. Data is stored in a convenient batch structure for later reanalysis or comparison to other data sets.
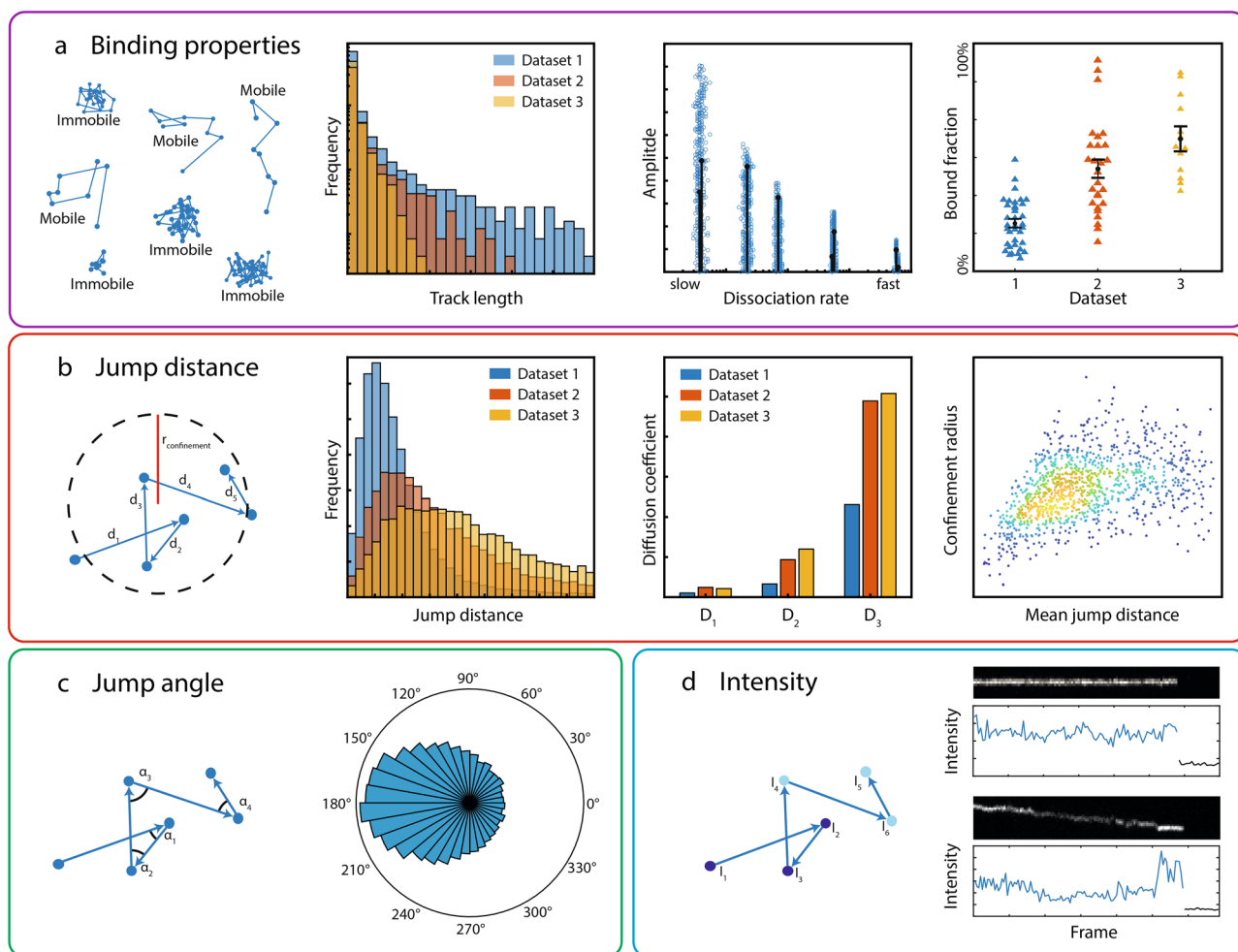

Figure 2. Examples of evaluation parameters and display options. (**a**) Duration of all tracks (track lengths) are collected in a histogram. Bound molecules can be segmented by the choice of tracking parameters, i.e. the tracking radius. For molecules classified as bound, the dissociation rate spectrum via GRID and bound fractions can be obtained. (**b**) Diffusion coefficients and confinement radii can be computed from jump distance distributions. (**c**) Angles between consecutive jumps are calculated and displayed in a polar histogram. (**d**) Each track can be individually explored by plotting the intensity over time together with a kymograph. The upper and lower graphs show examples of an immobile and mobile molecule, respectively.
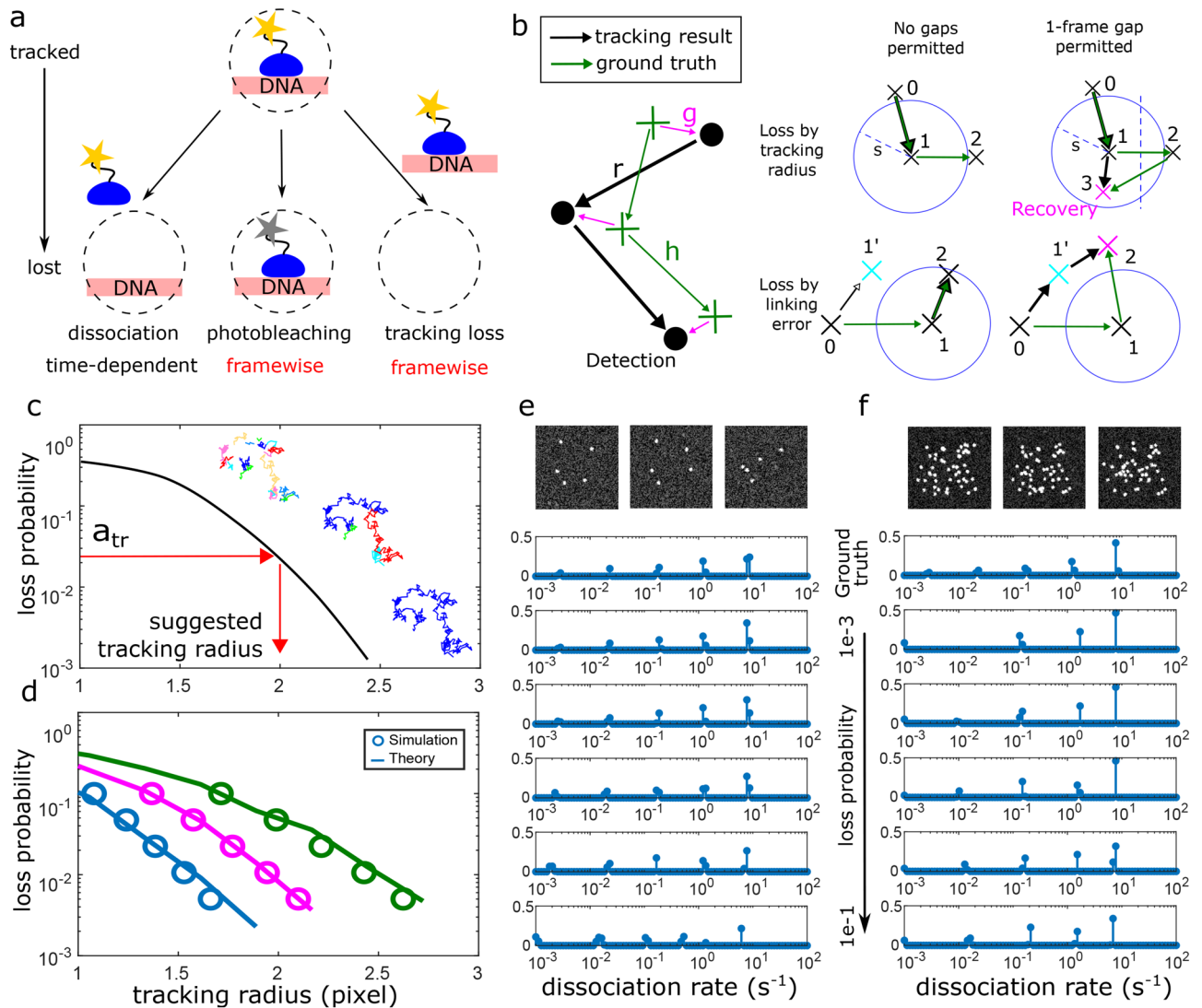
Figure 3. Prediction of loss probabilities in nearest neighbour tracking. (**a**) The track of a bound molecule stops after dissociation, which depends on the laboratory time. Additionally, photobleaching and tracking losses may falsely terminate a track. These losses depend on the frame-cycle time. (**b**) Sketches of tracking losses in the nearest neighbour algorithm. Tracking loss may occur due to jumps out of the tracking radius (s) or erroneous links to a different molecule in close proximity. If one gap frame is allowed, both types of losses can be partially recovered. Recovered tracks retain the correct temporal information, however the spatial properties of the track will be altered. (**c**) Simulated loss probability as function of tracking radius. Large tracking radii increase erroneous links to neighbouring detections, small tracking radii increase losses of the tracked molecule. After each loss a track is split into fragments. The colour coding indicates the different track fragments. (**d**) Comparison of loss probabilities (open circles) of simulated spots tracked with the nearest neighbour algorithm using different tracking radii with loss probability predicted by theory (lines). Three different time-lapse conditions where simulated, with frame-cycle times of 0.05 s (blue), 1 s (magenta) and 5 s (green) (Table 1 Scenario 1). (**e**) and (**f**) Dissociation rate spectra of molecules simulated with five different dissociation rate constants after tracking them using the tracking radii suggested by theory for a given loss probability (Table 1 Scenarios 2 and 3). (**e**) For low densities, dissociation rate spectra are correctly inferred if the loss probability was below 10% per frame. (**f**) For a density of 0.005 Spots per pixel, only fast dissociation rates were recovered in cases of low loss probabilities.

Description of tracking errors. The track of an immobile molecule stops when (i) the molecule exits the immobile state, (ii) the fluorescent label photobleaches (photobleaching loss) or (iii) a tracking error occurs (tracking loss) (Fig. 3a). Only case (i) depends on the properties of the molecule and carries information about its residence time in the immobile state. In contrast, both photobleaching and tracking errors falsely terminate a track and thus blur the survival time distribution if they are not corrected for. Although tracking loss may occur as often as or even exceed photobleaching loss, correction methods for tracking loss are less developed than for photobleaching loss.

Scenario	Molecules per frame	Relative frequency	Dissociation Rate (s⁻¹)	Photobleaching rate a (Probability per frame)	Time-lapse conditions (s)	Frames	Point spread function width	Conversion factor pixel to μm
1	1	1	1e−15	1e−15	0.05; 1; 5	500	1px	0.16
2	5	0.03,0.06,0.130,0.26,0.52	3e−3, 2e−2, 0.2, 1.3, 10	0.01	0.05,0.3,1.5,5	400	1px	0.16
3	50	0.03,0.06,0.130,0.26,0.52	3e−3, 2 e−2, 0.2, 1.3, 10	0.01	0.05,0.3,1.5,5	400	1px	0.16

Table 1. Simulation parameters for time-lapse data. Scenario 1 (Fig. 3 Panel d) corresponds to a single spot that has an infinite residence time and does not exhibit photobleaching. Scenario 2 (Fig. 3 panel e) and 3 (Fig. 3 panel f) correspond to a molecule that exhibits five dissociation rates from chromatin and is subject to photobleaching at different spot densities per frame.

To be able to account for tracking errors of immobile molecules, e.g. transcription factors bound to chromatin, we established a model for the tracking loss of the nearest neighbour algorithm ("Materials and methods"). We considered tracking loss due to jumps out of the tracking radius and erroneous links to a different molecule in close proximity (Fig. 3b). If one gap frame is allowed, these losses can be partially recovered and the lifetime of the immobile state stays unaltered ("Materials and methods"). Oversize jumps are recovered if the molecule returns back into the tracking radius after the gap frame and is closer to the centre of the tracking area than to its previous position. A falsely linked track can be correctly continued, if the subsequent detection is within the tracking radius of the falsely linked spot. We calculated the probabilities of recovery and final loss by considering the geometry of these situations and obtain an overall probability a_{tr} for the tracking loss of ("Materials and methods"):

$$a_{tr} = a_{NN}(\rho) + a_z(s)\big[1 - f\big]\big[1 - a_{NN}(\rho)\big] \qquad (1)$$

where a_{NN} is the loss due to erroneous linking, that depends on the spot density ρ and a_z is the loss due to jumps out of the tracking radius s. The factor $(1 - f)$ represents the reduction of the tracking loss by allowing one gap frame. The optimal loss probability balances obtaining fully linked tracks at large tracking radii and low erroneous linking of adjacent molecules (Fig. 3c).

To validate our approach, we simulated a single immobile spot without photobleaching and without dissociation from chromatin (Table 1 and "Materials and methods"). We linked it with our nearest neighbour algorithm, and obtained the loss probability from the average lifetime of the tracks. We compared the loss probability as function of tracking radius and at different time-lapse conditions with our theoretical prediction (Fig. 3d and Table 1). The theoretical expectation well described the in silico experiment.

Experimental correction of tracking errors. Tracking errors have a certain probability to occur after each frame of a movie, similar to photobleaching. When measuring the residence time of immobile molecules, tracking errors can be corrected for by applying a time-lapse imaging scheme with several time-lapse conditions, if the loss probability per frame is constant for all time-lapse conditions. In this case, correction using time-lapse imaging is similar to the correction of photobleaching[31]. The reason is that the residence time of the molecule does not depend on the time-lapse condition, while both tracking errors and photobleaching do. In effect, both photobleaching and tracking errors can be corrected for simultaneously by combining both losses in a single loss probability.

When analysing a time-lapse experiment, a different tracking radius has to be chosen for each time-lapse condition to yield an overall constant loss probability (Fig. 3d). We implemented an algorithm in our tracking and analysis framework, which, for a user-defined loss probability, calculates the corresponding tracking radius for each time-lapse condition.

We tested the performance of our correction approach in the analysis of residence times of immobile molecules. We simulated a time-lapse experiment for a scenario where immobile molecules at a low density of 5 spots in 100×100 px were subject to five different binding interactions with corresponding dissociation rates (Fig. 3e, Table 1 and "Materials and methods"). We tracked the molecules by specifying a loss probability and using the suggested tracking parameters for each time-lapse condition in the nearest neighbour algorithm. Subsequently, we used GRID to extract the spectrum of dissociation rates[32]. To determine to which extent tracking errors influenced the result we varied the loss probability over two orders of magnitude. We found that for high loss probabilities > 10%, small dissociation rates arising from long-lasting tracks could not be inferred correctly. In contrast, for low loss probabilities < 10%, the ground truth was well inferred with our analysis.

We further tested to which extent the density of molecules affected our approach. We simulated densities up to 50 spots in 100×100 px (Fig. 3f and Table 1) in steps of 0.0005 spots per pixel to provoke incorrect linking between individual spots. Molecules were again subject to five different binding interactions. The maximal density of molecules at which dissociation rate spectra and thus associated residence times could be well inferred was 0.0025 spots per pixel.

Analysis of the dissociation rate spectrum of the transcription factor CDX2. Finally, we applied our tracking and analysis framework and the algorithm which predicts optimal tracking parameters to re-ana-
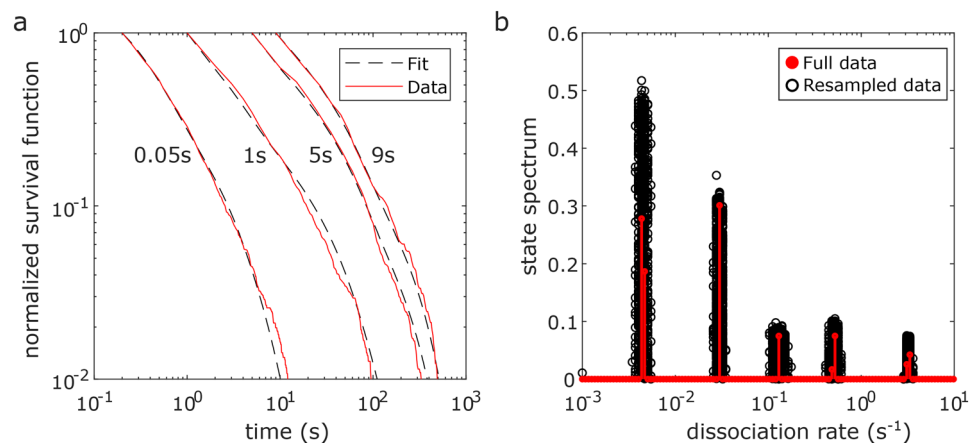
Figure 4. Re-analysis of CDX2 dissociation rate spectrum. Live-cell single molecule movies of SiR-Halo-CDX2[32] where analysed using the tracking radii suggested by theory for a tracking loss probability of 1e−2. (**a**) Fluorescence survival time distributions of tracked molecules (red lines) at different time-lapse conditions (indicated on top) and survival time functions obtained by GRID (black dashed lines). (**b**) State spectrum of SiR-Halo-CDX2 obtained by GRID using all data (red circles) and a superposition of 499 GRID results obtained by resampling 80% of data (black circles) as an error estimate of the state spectrum.

Cluster No	1	2	3	4	5
Dissociation rate interval (1/s)[32] (Manual parameter settings)	0.004–0.007	0.018–0.026	0.08–0.09	0.50–0.55	4.6–5.2
State spectrum weight[32] (Manual parameter settings)	30.3%	38.5%	14.7%	10.5%	6.1%
Dissociation rate interval (1/s) (Theory predicted parameter settings)	0.003–0.006	0.026–0.035	0.1–0.15	0.44–0.56	2.6–3.9
State spectrum weight (Theory predicted parameter settings)	46.4%	30.1%	7.5%	9.2%	6.8%

Table 2. Comparison of the inferred CDX2 spectrum with[32]. Dissociation rate interval specifies the manually assigned dissociation rate intervals corresponding to a dissociation rate cluster. The spectral weight of each of the five distinct clusters of the CDX2-spectrum was obtained by integrating the GRID amplitudes resulting from 100% of the measured survival times.

lyse the dissociation rate spectrum of the transcription factor CDX2 using previously published single molecule imaging data (Fig. 4)[32]. The data consists of four time-lapse microscopy conditions of a HaloTag-CDX2 fusion protein labelled with a SiR dye. Time-lapse movies were obtained with 50 ms exposure time and an overall frame cycle time of 0.05 s, 1 s, 5 s and 9 s. In our tracking parameter prediction algorithm, we set the loss probability to 0.01 and the algorithm calculated a corresponding set of tracking parameters for each time-lapse condition. As a result, the algorithm revealed tracking radii and the corresponding time periods during which a HaloTag-CDX2 molecule should stay within this boundary to be identified as bound. The tracking radii where 100 nm, 240 nm, 410 nm and 430 nm for the frame cycle time condition of 0.05 s, 1 s, 5 s and 9 s, and the minimum track length was 5 frames for 0.05 s frame cycle time movies and 2 frames for the other conditions. For tracking, we used the nearest neighbour algorithm and allowed bridging detection gaps of one frame as long as a track already existed for at least 2 frames. Next, the resulting track durations were transferred to the GRID toolbox to extract the dissociation rate spectrum. Our approach with computationally determined tracking radii and the previous analysis using manually chosen tracking radii yielded comparable dissociation rate spectra (Table 2). The main deviations are in the amplitudes, not the values of dissociation rate clusters. For future experiments, computationally determined optimized tracking radii will ensure robust data analysis.

Discussion

We introduced the tracking and analysis framework TrackIt, which simplifies analysing and comparing multiple different large single molecule fluorescence data sets due to a comprehensive list of GUI modules and a batch structure. Thus, fast and reproducible analysis can be performed without the need for additional manual steps or programming. In particular, TrackIt is well suited to systematically compare different settings of tracking parameters and to scan data sets differing in experimental conditions for differences in kinetic or structural parameters.

We further introduced quantitative considerations of tracking losses of immobile molecules in a nearest neighbour tracking algorithm and implemented means to partially correct for them. The nearest neighbour algorithm is more prone to linking errors than more complex algorithms[6,11,12]. However, in contrast to complex

approaches that produce unpredictable tracking losses, the nearest neighbour algorithm allows for a theoretical prediction of tracking losses. This prediction allowed us to automatically determine consistent settings of tracking parameters for the analysis of immobile molecules. We note that tracking errors in every tracking algorithm can be minimized by measuring at low molecule densities, thereby trading optimized tracking for measurement throughput. Overall, we provided a pipeline to analyse transcription factor residence times including both photobleaching and tracking error corrections.

We used theory-suggested tracking parameters and GRID to determine the spectrum of dissociation rates of CDX2 proteins. We could verify our previous results, where manually assigned tracking parameters were carefully adjusted considering movies of all time-lapse conditions. Our theory-predicted tracking radii reach the same quality as manual parameters and thus enable reproducible, user-independent data analysis.

We only treated Brownian motion in our tracking formalism. Accordingly, we used the nearest neighbour algorithm that does not base the linking process on the molecule's past movement. Other modes of motion like the Ornstein–Uhlenbeck process[44], super- and anomalous diffusion[45,46] as well as Lévy flights[47] were discussed for the movement of biomolecules. In contrast to Brownian motion, these processes have a memory i.e. their current movement depends on their past. Thus, it may be advisable to use model based tracking algorithms e.g. Bayesian or Kalman filters that are able to predict the molecules movement based on the past track for these kinds of motion. Our idea of calculating empirical parameters from a given loss probability can be readily applied, however new calculations for the respective model are necessary.

Each research question comes with its own requirements for data analysis. Thus, while our framework provides state-of-the art analysis approaches applied in recent publications, care has to be taken whether an implemented approach optimally covers the analysis needs. For example, analysing bound fractions requires separating molecules into different kinetic classes. While we implemented a commonly used approach to identify bound molecules by their restricted mobility area[28–30], alternative classification approaches have been published[40,48–50]. Moreover, the bound fraction analysis we implemented is best suited for data sets captured using interlaced time-lapse microscopy[37]. Some important analysis approaches are not yet implemented in TrackIt, for example the possibility to analyse two-colour single molecule data. However, being implemented in commonly used Matlab format, our framework constitutes a broadly accessible platform to which novel analysis schemes can be added.

Materials and methods

Localization and jump distance mapping.
The spatial distribution of localizations and mobility parameters of tracked molecules contains valuable information about their function and environment[2,35,36]. We enabled creating a heat map of localizations by using all detected spots whose position is determined to sub-pixel precision with a 2D Gaussian fit. The positions of all spots were then accumulated in a 2D histogram. The pixel values therefore correspond to the amount of detections in each pixel. The image can be upscaled by dividing the original bin size (i.e. the original pixel size) into smaller units resulting in a super-resolved image, or downscaled by merging several bins. Similarly, we create a heat map of jump distances. We define a jump as a change in position between two consecutive frames of a track. Jumps involving gap frames are omitted. For all jumps within a track, a virtual line is drawn between the start and end positions of a jump. Each pixel touching this line is assigned with the corresponding jump distance. The resulting 2D histogram is then normalized by the amount of jump events in each pixel. Again, the image can be up- or downscaled by using an appropriate bin size of the histogram.

Jump distances and diffusion analysis.
To obtain a direct overview of the jump distance distribution of all tracks, a histogram is created using all single molecule jump distances between consecutive frames. The bin size Δr of the histogram can be set individually. In addition, the number of jumps of a tracked molecule to be considered in the histogram can be chosen. If only one jump per track is chosen, a biased weight of immobile over mobile molecules is minimized.

To extract diffusion coefficients from mobile molecules, we created a cumulative histogram of the squared jump distances, normalized to the total number of jump events. We fitted the resulting cumulative density distribution of squared displacements with a Brownian diffusion model including either two or three different diffusion components[29,31,39]. For the two-component model we used

$$F(X) = A_1 \left(1 - \exp\left(-\frac{X}{D_1}\right)\right) + \frac{(1 - A_1)\left(\exp\left(-\frac{C_1}{D_2}\right) - \exp\left(-\frac{X}{D_2}\right)\right)}{\exp\left(-\frac{C_1}{D_2}\right) - \exp\left(-\frac{C_2}{D_2}\right)} \tag{2}$$

while for the three-component model we used

$$F(X) = A_1 \left(1 - \exp\left(-\frac{X}{D_1}\right)\right) + A_2 \left(1 - \exp\left(-\frac{X}{D_2}\right)\right) + \frac{(1 - A_1 - A_2)\left(\exp\left(-\frac{C_1}{D_3}\right) - \exp\left(-\frac{X}{D_3}\right)\right)}{\exp\left(-\frac{C_1}{D_3}\right) - \exp\left(-\frac{C_2}{D_3}\right)} \tag{3}$$

here, D_i denotes the diffusion constant of the i-th component, $X = (x^2 + y^2)/(4\tau)$ with the camera frame cycle time τ and A_i denotes amplitude. The last term is normalized by $\exp\left(-\frac{C_1}{D_{2/3}}\right) - \exp\left(-\frac{C_2}{D_{2/3}}\right)$ to account for the cut off due to the lower limit of jump distances, which is $C_1 = 0$, and the upper limit of jump distances, C_2, which is given by the tracking radius, respectively.

Using the GUI, the original cumulative histogram is overlaid with the fitted function for visual inspection. Moreover, the resulting diffusion constants and amplitudes are displayed and compared between different data sets. We give 95% confidence intervals as an estimate of the error of fit-parameters. Adjusted R^2 values are used to estimate how well the data is represented by one of the models.

Confinement radius analysis. We implemented a two-parameter representation of tracks in which tracks are sorted according to their confinement radius and their mean jump distance[40]. This representation gives insights into different mobility classes of single molecules. We calculated the confinement radii as previously described[40]. In brief, the mean squared displacement as function of time of each track is fitted with a confined diffusion model[16,40]

$$MSD_{circle} = R^2 \cdot \left(1 - e^{-4 \cdot D^* \cdot \frac{t_{lag}}{R^2}} \right) + \textit{offset} \tag{4}$$

where R is the radius of confinement, D^* is the local diffusion coefficient, and the offset is introduced to account for the finite localization precision. In order to select only confined tracks for analysis, the mean squared displacement as function of time of each track is fitted with a power law $MSD = 4 \cdot D \cdot t^\alpha$, where D is the diffusion coefficient and α an exponent that indicates the motion type[40]. A threshold can be chosen for the maximum value of α that should be considered for further analysis.

Analysis of the dissociation rate spectrum and corresponding residence times. For immobile molecules, we implemented the possibility to analyse their dissociation rate spectrum and corresponding residence times. To obtain the dissociation rate spectrum, the survival time distribution of track durations of all tracks in continuous video or a time-lapse data set were calculated in the GRID toolbox. In time-lapse microscopy, the acquired frames are separated by a time period without illumination of the sample. Next, dissociation rate spectra were extracted in a global analysis of all time-lapse data sets using GRID[32]. In brief, solving the inverse Laplace transformation of each survival time distribution is translated into a single minimization problem that can be handled by a gradient method.

Bound fraction analysis. The fraction of molecules bound to a stable structure such as chromatin can be determined by interpreting the amplitudes of diffusion components of tracked molecules[29,39]. In another approach, the fractions of molecules belonging to two different binding time classes can be approximated using the interlaced time-lapse microscopy (ITM) illumination scheme[37,51]. In ITM, two subsequent frame acquisitions are followed by a longer dark time. Detected molecules are sorted into different binding time classes. Tracks, which survive at least one dark period, are classified as long bound, tracks which persist for two consecutive frames are classified as short bound and single detections are classified as unbound/diffusing. To obtain accurate fractions, they have to be corrected for photobleaching[37]. Continuous movies may contain information comparable to ITM, however will be highly affected by photobleaching.

Once classified, we calculate the overall bound fraction using

$$BF = \frac{N_{bound}}{N_{bound} + N_{unbound}} \tag{5}$$

where $N_{unbound}$ denotes the number of bound molecules and $N_{unbound}$ the amount of unbound molecules.

In order to distinguish transient and stable bound molecules we calculate the fraction of long bound molecules

$$BF_{long} = \frac{N_{long}}{N_{long} + N_{short}} \tag{6}$$

where N_{long} denotes the number of long bound molecules and N_{short} the amount of short bound molecules. A determination of bound fractions for each movie leads to a variation in the bound fraction from movie-to-movie and the final bound fraction is calculated from the mean value over all movies. To avoid an over-representation of movies with low molecule counts, events for each binding time class are additionally summed over all movies resulting in a single "pooled" bound fraction of all movies.

Intensity and kymograph analysis. We implemented the possibility to select individual trajectories and plot them in a separate window together with its intensity and the associated kymograph[43]. To visualize the intensity over time, the mean intensity in a 3×3 pixel window around the spot centre is calculated and plotted until the track is lost. Kymographs are visualized in separate plots for each spatial dimension. One spatial dimension is displayed versus time, while the other spatial dimension is maximum projected. After the end of track, both intensity plot and kymographs are continued for another 20 frames where the projection window remains centred on the last position of the track. Thus kymographs indicate whether the molecule was lost due to tracking errors or diffusion out of the focal plane. The extracted intensity trajectory furthermore informs on photobleaching steps.

Jump angle analysis. To analyse the angles between consecutive jumps[18,42], we calculated the scalar product between the two normalized vectors representing the directions of the two consecutive jumps. Angles involving jumps over gap frames are omitted. We then took the inverse cosine to calculate the angle between the two vectors. The resulting angles are visualized in an angular histogram[18].

Simulation of single molecule time lapse data. We generated single molecule movies using a custom simulator implemented in Matlab 2019a. We simulated diffusion of a protein and photobleaching of an attached fluorescent label. Diffusion is altered upon association to or dissociation from chromatin by the protein. The protein can enter different binding states that are characterized by their on- and off-rates.

In favour of fast simulations, we renounced simulating the molecule position at fixed small time steps, but used the Gillespie direct method. With this approach, motion blur is not included in the simulation. We first determined the times until a photobleaching event or until a transition event from a diffusing to a bound state or vice versa. We then determined the position of the molecule for each frame. In case a state transition occurred outside a frame interval, we also determined the position of the molecule at the corresponding time. Jump distances were drawn from the 2D diffusion probability density corresponding to the diffusion coefficient D of the current state (free diffusion or apparent diffusion due to the localization error if bound).

We then used the simulated trajectories to generate images. For each spot, we simulated a point spread function with intensities corresponding to a lognormal photon count distribution. To add background, we used uniform random numbers for background noise. To approximate non-uniform background, we applied a band-pass filter, which additionally enhanced features of the background.

Tracking loss in nearest neighbour tracking. The nearest neighbour algorithm links detected spots into tracks by comparing their positions in two consecutive frames. It links spots that are closest to each other if their distance does not exceed a certain tracking radius. Spots that cannot be linked to an existing track start a new track. Errors occur if a jump distance is larger than the tracking radius or if the track is erroneously linked to a different molecule in close proximity.

To estimate the probability of losing a track we assumed that the detected position of a spot depends on diffusion and a localization error. The effective squared jump distance between two consecutive frames is given by

$$\sigma^2(\tau_{tl}) = \sigma_0^2 + 4D\tau_{tl} \tag{7}$$

which results in the corresponding probability density for the detected position

$$p(r) = \frac{1}{\pi\sigma^2} \exp\left(-\frac{r^2}{\sigma^2}\right) \tag{8}$$

The probability $P(|\vec{r}| > s)$ of the spot position to exceed the tracking radius s is obtained by integrating the above equation over the interval $s \leq |\vec{r}| \leq \infty$

$$P(|\vec{r}| > s) = \int\limits_{|\vec{r}|>s} p(r)d\vec{r} = \exp\left(-\frac{s^2}{\sigma^2}\right) = \exp\left(-z\right) \tag{9}$$

where we introduced the dimensionless variable

$$z = \frac{s^2}{\sigma^2} \tag{10}$$

We further considered an interruption of the track due to an erroneously linked molecule. The position of the disturbing molecule is denoted by $\vec{r}_\#$. In order to disrupt the track, the molecule has to be closer to the second detected position. The probability to lose a track due to linking errors in dense environments is obtained by considering all possible configurations of this scenario:

$$a_{NN} = \iiint\limits_{\substack{|\vec{r}_1|>|\vec{r}_\#| \\ |\vec{r}_1+\vec{r}_2-\vec{r}_\#|>|\vec{r}_2|}} \rho p(\vec{r}_1)p(\vec{r}_2)d\vec{r}_\# d\vec{r}_1 d\vec{r}_2 = 0.76 \cdot \pi\rho\sigma^2 \tag{11}$$

where ρ is the density of spots/frame.

Partial recovery of tracking loss by introducing one gap frame. Tracking loss can be partially recovered if the tracking algorithm allows bridging one frame with missed detection.

We first calculated the probability I that two consecutive jumps including localization error lie within a given area Ω:

$$I = \int\int\int\int\int_{\vec{r}_1,\vec{r}_2 \in \Omega} p\left(\vec{g}_1, \frac{\sigma_0^2}{2}\right) p\left(\vec{g}_2, \frac{\sigma_0^2}{2}\right) p\left(\vec{g}_3, \frac{\sigma_0^2}{2}\right) p\left(\vec{h}_1, 4Dt\right) p\left(\vec{h}_2, 4Dt\right) d\vec{g}_1 d\vec{g}_2 d\vec{g}_3 d\vec{h}_1 d\vec{h}_2 \tag{12}$$

with the integration borders

$$\vec{r}_1 = \vec{h}_1 - \vec{g}_1 + \vec{g}_2 \quad \vec{r}_2 = \vec{h}_2 - \vec{g}_2 + \vec{g}_3; \tag{13}$$

where h is the jump of the molecule and g is the detected position. By inserting the probabilities (8) and detected spot positions (13) in (12) and integrating with respect to $\vec{g}_1, \vec{g}_2, \vec{g}_3$ we obtain

9

$$I = \frac{1}{\pi^2 \sigma^4} \frac{1}{1-c^2} \iint_{\vec{r}_1, \vec{r}_2 \in \Omega} \exp\left(-\frac{1}{1-c^2}\frac{1}{\sigma^2}\left(\vec{r}_1^2 + \vec{r}_2^2 - 2c\vec{r}_1\vec{r}_2\right)\right) d\vec{r}_1 d\vec{r}_2 \tag{14}$$

We next used I to estimate the effect of a gap frame. We calculated the probability $I(\Omega)$ of a molecule to return inside the tracking radius after having left it in the previous frame. For this case, we need to consider that the molecule outside the tracking radius was detected but not linked to the existing track and therefore started a new track. The detection in the current frame has to be closer to the starting position than to the molecule outside the tracking radius otherwise the track will be cut. The area $\Omega = \Omega_{Gap}$ corresponding to this situation is given by:

$$\Omega_{Gap} = |\vec{r}_1| > s \text{ and } |\vec{r}_1 + \vec{r}_2| > \min(s, |\vec{r}_2|) \tag{15}$$

Solving the Integral (14) with area Ω_{Gap} yields the probability that the track is recovered. The integral $I(\Omega)$ was calculated numerically. The corresponding probability to lose a track after a gap frame is given by $1 - I(\Omega_{Gap}) = a_{Gap} = P(|\vec{r}| > s)$. Compared with the tracking loss without gap frame, the probability to lose a track if tracked with a gap frame is upmost a factor of 0.5 smaller.

The overall probability a_{tr} for losing a track in presence of a gap frame and in presence of erroneous linking is given by:

$$a_{tr} = 1 - \left(1 - a_{Gap}\right)\left(1 - a_{NN}\right) = a_{NN} + a_z\left(1 - f\right)\left(1 - a_{NN}\right) \tag{16}$$

Calculation of tracking radius to ensure a certain loss probability. To obtain the tracking radius for a given loss probability we need to solve equation (16) for s. Since no closed-form equation can be given for $s(a_{tr})$, we employed an iteration scheme. We started the iteration at a starting point s_0 and iterated until the change between the i-th and i+1-th iteration is smaller than 1e−2. The equations for the step $i \rightarrow i+1$ are given by

$$\tilde{z} = -\log\left(\frac{a_{tr} - a_{NN}(s_i)}{(1 - a_{NN}(s_i))\left(1 - a_{Gap}(s_i)\right)}\right)$$
$$z(i+1) = \tilde{z}\frac{1 - \exp(-\tilde{z})}{1 - (1+\tilde{z})\exp(-\tilde{z})} \tag{17}$$
$$s(i+1) = \sqrt{\sigma^2(s(i)) \cdot z(i+1)}$$

In our iteration, we accounted for the fact that jump distance distributions are cut by the tracking radius[16]. In each iteration, we determined the mean of squared jump distances $\sigma^2(s_i)$ by tracking with the nearest neighbour algorithm with tracking radius s_i.

Simulation parameters. We simulated videos with frame sizes of 100×100 pixels. The SNR of spots was chosen as 25. The diffusion constant of chromatin bound molecules was set to D = 1e−3 μm^2 s^{-1}, the free diffusion constant was set to D = 10 μm^2 s^{-1}.

Data availability
Data supporting the findings of this manuscript will be available from the corresponding author after publication upon reasonable request. Single particle movies of CDX2, described in[32], and tracking results are freely available at https://doi.org/10.5061/dryad.0zpc866wh.

Code availability
The TrackIt software is freely available. TrackIt is written in Matlab and available at https://gitlab.com/GebhardtLab/TrackIt.

References
1. Lionnet, T. & Wu, C. Single-molecule tracking of transcription protein dynamics in living cells: Seeing is believing, but what are we seeing?. *Curr. Opin. Genet. Dev.* **67**, 94–102 (2021).
2. Appelhans, T. *et al.* Nanoscale organization of mitochondrial microcompartments revealed by combining tracking and localization microscopy. *Nano Lett.* **12**, 610–616 (2012).
3. Persson, F., Barkefors, I. & Elf, J. Single molecule methods with applications in living cells. *Curr. Opin. Biotechnol.* **24**, 737–744 (2013).
4. Liu, Z., Lavis, L. D. & Betzig, E. Imaging live-cell dynamics and structure at the single-molecule level. *Mol. Cell* **58**, 644–659 (2015).
5. Aaron, J., Wait, E., DeSantis, M. & Chew, T. L. Practical considerations in particle and object tracking and analysis. *Curr. Protoc. Cell Biol.* **83**, e88 (2019).
6. Chenouard, N. *et al.* Objective comparison of particle tracking methods. *Nat. Methods* **11**, 281–289 (2014).
7. Shen, H. *et al.* Single particle tracking: From theory to biophysical applications. *Chem. Rev.* **117**, 7331–7376 (2017).
8. Meijering, E., Dzyubachyk, O. & Smal, I. Methods for cell and particle tracking. *Method. Enzym.* **504**, 183–200 (2012).
9. Kerkhoff, Y. & Block, S. Analysis and refinement of 2D single-particle tracking experiments. *Biointerphases* **15**, 021201 (2020).
10. Manzo, C. & Garcia-Parajo, M. F. A review of progress in single particle tracking: From methods to biophysical insights. *Rep. Prog. Phys.* **78**, 124601 (2015).
11. Jaqaman, K. *et al.* Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods* **5**, 695–702 (2008).

12. Sergé, A. *et al.* Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat. Methods* **5**, 687–694 (2008).
13. Tinevez, J. Y. *et al.* TrackMate: an open and extensible platform for single-particle tracking. *Methods* **115**, 80–90 (2017).
14. Winter, M. R., Fang, C., Banker, G., Roysam, B. & Cohen, A. R. Axonal transport analysis using multitemporal association tracking. *Int. J. Comput. Biol. Drug Des.* **5**, 35–48 (2012).
15. Spilger, R. *et al.* Deep particle tracker: Automatic tracking of particles in fluorescence microscopy images using deep learning. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* 128–136 (Springer, 2018).
16. Wieser, S. & Schütz, G. J. Tracking single molecules in the live cell plasma membrane-Do's and Don't's. *Methods* **5**, 687–694 (2008).
17. Wöll, D., Kölbl, C., Stempfle, B. & Karrenbauer, A. A novel method for automatic single molecule tracking of blinking molecules at low intensities. *Phys. Chem. Chem. Phys.* **15**, 6196–6205 (2013).
18. Izeddin, I. *et al.* Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *Elife* **3**, e02230 (2014).
19. Stein, S. C. & Thiart, J. TrackNTrace: a simple and extendable open-source framework for developing single-molecule localization and tracking algorithms. *Sci. Rep.* **6**, 37947 (2016).
20. Niewidok, B. *et al.* Single-molecule imaging reveals dynamic biphasic partition of RNA-binding proteins in stress granules. *J. Cell Biol.* **217**, 1303–1318 (2018).
21. Vallotton, P. *et al.* Diatrack particle tracking software: review of applications and performance evaluation. *Traffic* **18**, 840–852 (2017).
22. Lee, B. H. & Park, H. Y. HybTrack: a hybrid single particle tracking software using manual and automatic detection of dim signals. *Sci. Rep.* **8**, 212 (2018).
23. Ball, D. A. *et al.* Single molecule tracking of Ace1p in Saccharomyces cerevisiae defines a characteristic residence time for non-specific interactions of transcription factors with chromatin. *Nucleic Acids Res.* **44**, 1–12 (2016).
24. Rösch, T. C. *et al.* SMTracker: a tool for quantitative analysis, exploration and visualization of single-molecule tracking data reveals highly dynamic binding of B. subtilis global repressor AbrB throughout the genome OPEN. *Sci. Rep.* **8**, 1–12 (2018).
25. Hansen, A. S. *et al.* Robust model-based analysis of single-particle tracking experiments with spot-on. *Elife* **7**, e33125 (2018).
26. Persson, F., Lindén, M., Unoson, C. & Elf, J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nat. Methods* **10**, 265–269 (2013).
27. Jaqaman, K. & Danuser, G. Computational image analysis of cellular dynamics: a case study based on particle tracking. *Cold Spring Harb. Protoc.* **4**, 1–10 (2009).
28. Chen, J. *et al.* Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274–1285 (2014).
29. Mazza, D., Abernathy, A., Golob, N., Morisaki, T. & McNally, J. G. A benchmark for chromatin binding measurements in live cells. *Nucleic Acids Res.* **40**, 1–13 (2012).
30. Normanno, D. *et al.* Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nat. Commun.* **6**, 1–10 (2015).
31. Gebhardt, J. C. M. *et al.* Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nat. Methods* **10**, 421–426 (2013).
32. Reisser, M. *et al.* Inferring quantity and qualities of superimposed reaction rates from single molecule survival time distributions. *Sci. Rep.* **10**, 1758 (2020).
33. Izeddin, I. *et al.* Wavelet analysis for single molecule localization microscopy. *Opt. Express* **20**, 2081–2095 (2012).
34. Walder, R., Nelson, N. & Schwartz, D. K. Super-resolution surface mapping using the trajectories of molecular probes. *Nat. Commun.* **2**, 515 (2011).
35. Manley, S. *et al.* High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nat. Methods* **5**, 155–157 (2008).
36. Xiang, L., Chen, K., Yan, R., Li, W. & Xu, K. Single-molecule displacement mapping unveils nanoscale heterogeneities in intracellular diffusivity. *Nat. Methods* **17**, 524–530 (2020).
37. Reisser, M. *et al.* Single-molecule imaging correlates decreasing nuclear volume with increasing TF-chromatin associations during zebrafish development. *Nat. Commun.* **9**, 5218 (2018).
38. Schütz, G. J., Schindler, H. & Schmidt, T. Single-molecule microscopy on model membranes reveals anomalous diffusion. *Biophys. J.* **73**, 1073–1080 (1997).
39. Speil, J. *et al.* Activated STAT1 transcription factors conduct distinct saltatory movements in the cell nucleus. *Biophys. J.* **101**, 2592–2600 (2011).
40. Lerner, J. *et al.* Two-parameter mobility assessments discriminate diverse regulatory factor behaviors in chromatin. *Mol. Cell* **79**, 677–688 (2020).
41. Condamin, S., Bénichou, O., Tejedor, V., Voituriez, R. & Klafter, J. First-passage times in complex scale-invariant media. *Nature* **450**, 77–80 (2007).
42. Burov, S. *et al.* Distribution of directional change as a signature of complex dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19689–19694 (2013).
43. Grosse-Berkenbusch, A. *et al.* Myosin VI moves on nuclear actin filaments and supports long-range chromatin rearrangements. *bioRxiv Cell Biol* https://doi.org/10.1101/2020.04.03.023614 (2020).
44. Wiggins, P. A. An information-based approach to change-point analysis with applications to biophysics and cell biology. *Biophys. J.* **109**, 346–354 (2015).
45. De La Díaz Rosa, M. A. *et al.* Dynamic strategies for target-site search by DNA-binding proteins. *Biophys. J.* **98**, 2943–2953 (2010).
46. Saxton, M. J. Anomalous diffusion due to binding: a Monte Carlo study. *Biophys. J.* **70**, 1250–1262 (1996).
47. Viswanathan, G. M., Raposo, E. P. & da Luz, M. G. E. Lévy flights and superdiffusion in the context of biological encounters and random searches. *Phys. Life Rev.* **5**, 133–150 (2008).
48. Hansen, A. S., Amitai, A., Cattoglio, C., Tjian, R. & Darzacq, X. Guided nuclear exploration increases CTCF target search efficiency. *Nat. Chem. Biol.* **16**, 257–266 (2020).
49. Caccianini, L., Normanno, D., Izeddin, I. & Dahan, M. Single molecule study of non-specific binding kinetics of LacI in mammalian cells. *Faraday Discuss.* **184**, 393–400 (2015).
50. Kapadia, N., El-Hajj, Z. W. & Reyes-Lamothe, R. Bound2Learn: a machine learning approach for classification of DNA-bound proteins from single-molecule tracking experiments. *bioRxiv* https://doi.org/10.1101/2020.02.20.958512 (2020).
51. Hipp, L. *et al.* Single-molecule imaging of the transcription factor SRF reveals prolonged chromatin-binding kinetics upon cell stimulation. *Proc. Natl. Sci. U.S.A.* **116**, 880–889 (2019).

Acknowledgements

Author contributions

Funding

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/s41598-021-88802-7.

Correspondence and requests for materials should be addressed to J.C.M.G.

Reprints and permissions information is available at www.nature.com/reprints.

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