Spot-On: robust model-based analysis of single-particle tracking experiments

Authors:

Anders S Hansen[1,2,6], Maxime Woringer[1,3,4,6], Jonathan B Grimm[5], Luke D Lavis[5], Robert Tjian[1,2], Xavier Darzacq[1]

Affiliation:

[1]Department of Molecular and Cell Biology, Li Ka Shing Center for Biomedical and Health Sciences, CIRM Center of Excellence, University of California, Berkeley, CA 94720, USA.
[2]Howard Hughes Medical Institute, Berkeley, CA 94720, USA
[3]Unité Imagerie et Modélisation, Institut Pasteur, 25 rue du Docteur Roux, 75015 Paris, France
[4]Sorbonne Universités, UPMC Univ Paris 06, IFD, 4 Place Jussieu, 75252 Paris cedex 05, France
[5]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA
[6]These authors contributed equally and are alphabetically listed.

Contact Information:

*Corresponding authors: anders.sejr.hansen@berkeley.edu (ASH); jmlim@berkeley.edu (RT); darzacq@berkeley.edu

ABSTRACT

Single-particle tracking (SPT) has become an important method to bridge biochemisty and cell biology since it allows direct observation of protein binding and diffusion dynamics in live cells. However, accurately inferring information from SPT approaches is challenging due to biases in both data analysis and experimental design. To address analysis biases, we introduce "Spot-On", an intuitive web-interface. Spot-On implements a kinetic modeling framework that accounts for known biases, including molecules moving out-of-focus, and robustly infers diffusion constants and subpopulations from pooled single-molecule trajectories. To minimize inherent experimental biases, we implement and validate stroboscopic photo-activation SPT (spaSPT), which minimizes motion-blur bias and tracking errors. We validate Spot-On using experimentally realistic simulations and show that Spot-On outperforms other methods. We then apply Spot-On to spaSPT data from live mammalian cells spanning a wide range of nuclear dynamics and demonstrate that Spot-On consistently and robustly infers subpopulation fractions and diffusion constants.

INTRODUCTION

Recent advances in imaging technologies, genetically encoded tags and fluorophore development have made single-particle tracking (SPT) an increasingly popular method for analyzing protein dynamics (Liu et al., 2015). Recent biological application of SPT have revealed that transcription factors (TFs) bind mitotic chromosomes (Teves et al., 2016), how Polycomb interacts with chromatin (Zhen et al., 2016), that "pioneer factor" TFs bind chromatin dynamically (Swinstead et al., 2016) and that different nuclear proteins adopt distinct target search mechanisms (Izeddin et al., 2014; Rhodes et al., 2017). Compared with indirect and bulk techniques such as Fluorescence Recovery After Photobleaching (FRAP) or Fluorescence Correlation Spectroscopy (FCS), SPT is often seen as less biased and less model-dependent (Mueller et al., 2013; Shen et al., 2017). In particular, SPT makes it possible to directly follow single molecules over time in live cells and has provided clear evidence that proteins often exist in several subpopulations that can be characterized by their distinct diffusion coefficients (Mueller et al., 2013; Shen et al., 2017). For example, nuclear proteins such as TFs and chromatin binding proteins typically show a quasi-immobile chromatin-bound fraction and a freely diffusing fraction inside the nucleus. However, while SPT of slow-diffusing membrane proteins is an established technology, 2D-SPT of proteins freely diffusing inside a 3D nucleus introduces several biases that must be corrected for in order to obtain accurate estimates of subpopulations. First, while a frame is acquired, fast-diffusing molecules move and spread out their emitted photons over multiple pixels causing a "*motion-blur*" artifact (Berglund, 2010; Izeddin et al., 2014), whereas immobile or slow-diffusing molecules resemble point spread functions (PSFs; Figure 1A). This

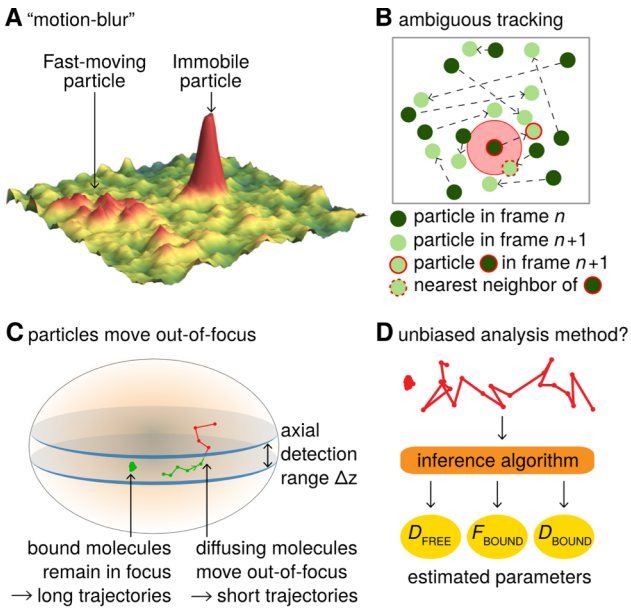

Figure 1. Bias in single-particle tracking (SPT) experiments and analysis methods. (**A**) "Motion-blur" bias. Constant excitation during acquisition of a frame will cause a fast-moving particle to spread out its emission photons over many pixels and thus appear as a motion-blur, which make detection much less likely with common PSF-fitting algorithms. In contrast, a slow-moving or immobile particle will appear as a well-shaped PSF and thus readily be detected. (**B**) Tracking ambiguities. Tracking at high particle densities prevents unambiguous connection of particles between frames and tracking errors will cause displacements to be misidentified. (**C**) Defocalization bias. During 2D-SPT, fast-moving particles will rapidly move out-of-focus resulting in short trajectories, whereas immobile particles will remain in-focus until they photobleach and thus exhibit very long trajectories. This results in a bias toward slow-moving particles, which must be corrected for. (**D**) Analysis method. Any analysis method should ideally avoid introducing biases and accurately correct for known biases in the estimation of subpopulation parameters such as D_{FREE}, F_{BOUND}, D_{BOUND}.

results in under-counting of the fast-diffusing subpopulation. Second, high particle densities tend to cause *tracking errors* when localized molecules are connected into trajectories. This can result in incorrect displacement estimates (Figure 1B). Third, since SPT generally employs 2D imaging of 3D motion, immobile or slow-diffusing molecules will generally remain in-focus until they photobleach and therefore exhibit long trajectories, whereas fast-diffusing molecules in 3D rapidly move out-of-focus, thus resulting in short trajectories (we refer to this as "*defocalization*"; Figure 1C). This results in a time-dependent under-counting of fast-diffusing molecules (Kues and Kubitscheck, 2002). Fourth, SPT *analysis methods* themselves may introduce biases; to avoid this an unbiased method is needed (Figure 1D).

Here, we introduce an integrated approach to overcome all four biases. The first two biases must be minimized at the data acquisition stage and we describe an experimental SPT method to do so (spaSPT), whereas the latter two can be overcome using a previously developed kinetic modeling framework (Hansen et al., 2017; Mazza et al., 2012) now extended and implemented in Spot-On. Spot-On is available as a web-interface (https://SpotOn.berkeley.edu) as well as Python and Matlab packages.

RESULTS

Overview of Spot-On

Spot-On is a user-friendly web-interface that pedagogically guides the user through a series of quality-checks of uploaded datasets consisting of pooled single-molecule trajectories. It then performs kinetic model-based analysis that leverages the histogram of molecular displacements over time to infer the fraction and diffusion constant of each subpopulation (Figure 2). Spot-On does not directly analyze raw microscopy images, since a large number of localization and tracking algorithms exist that convert microscopy images into single-molecule trajectories (for a comparison of particle tracking methods, see (Chenouard et al., 2014)).

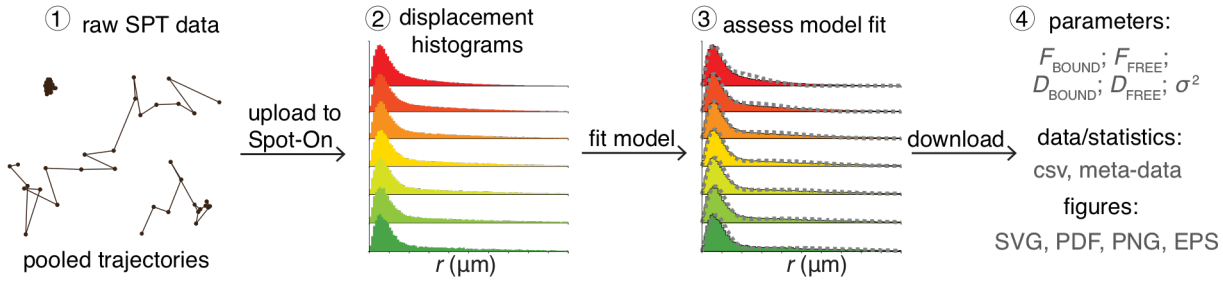

Figure 2. Overview of Spot-On interface. Overview of Spot-On. To use Spot-On, a user uploads raw SPT data in the form of pooled SPT trajectories to the Spot-On webinterface. Spot-On then calculates displacement histograms. The user inputs relevant experimental descriptors and chooses a model to fit. After model-fitting, the user can then download model-inferred parameters, meta-data and download publication-quality figures.

To use Spot-On, a user uploads their SPT trajectory data in one of several formats (Figure 2). Spot-On then generates useful meta-data for assessing the quality of the experiment (e.g. localization density, number of trajectories etc.). Spot-On also allows a user to upload multiple datasets (e.g. different replicates) and merge them. Spot-On then calculates and displays histograms of displacements over multiple time delays. The next step is model fitting. Spot-On models the distribution of displacements for each subpopulation using Brownian motion under steady-state conditions without state transitions (full model description in Experimental Procedures). Spot-On also accounts for localization errors (either user-defined or automatically inferred from the SPT data). Crucially, Spot-On corrects for defocalization bias (Figure 1C) by explicitly calculating the probability that molecules move out-of-focus as a function of time and their diffusion constant. In fact, Spot-On uses the gradual loss of freely diffusing molecules over time as additional information to infer the diffusion constant and size of each subpopulation.

Spot-On considers either 2 or 3 subpopulations. For instance, TFs in nuclei can generally exist in both a chromatin-bound state characterized by slow diffusion and a freely diffusing state associated with rapid diffusion. In this case, a 2-state model is generally appropriate ("bound" vs. "free"). Spot-On allows a user to choose their desired model and parameter ranges and then fits the model to the data. Using the previous example of TF dynamics, this allows the user to infer the bound fraction and the diffusion constants. Finally, once a user has finished fitting an appropriate model to their data, Spot-On allows easy download of publication-quality figures and relevant data (Figure 2).

Validation of Spot-On using simulated SPT data and comparison to other methods

We first evaluated whether Spot-On could robustly infer subpopulations (Figure 1D) and successfully account for known biases (Figure 1C) using simulated data. We compared Spot-On to a popular alternative approach of first fitting the mean square displacement (MSD) of individual trajectories of a minimum length and then fitting the distribution of estimated diffusion constants (we refer to this as "MSD_i") as well as a sophisticated Hidden-Markov Model-based Bayesian inference method (vbSPT) (Persson et al., 2013). Since most SPT data is collected using highly inclined illumination (Tokunaga et al., 2008) (HiLo), we simulated TF binding and diffusion dynamics (2-state model: "bound vs. free") confined inside a $4\ \mu m$ radius mammalian nucleus under realistic HiLo SPT experimental settings subject to a 25 nm localization error. We considered the effect of the exposure time (1 ms, 4 ms, 7 ms, 13 ms, 20 ms), the free diffusion constant (from

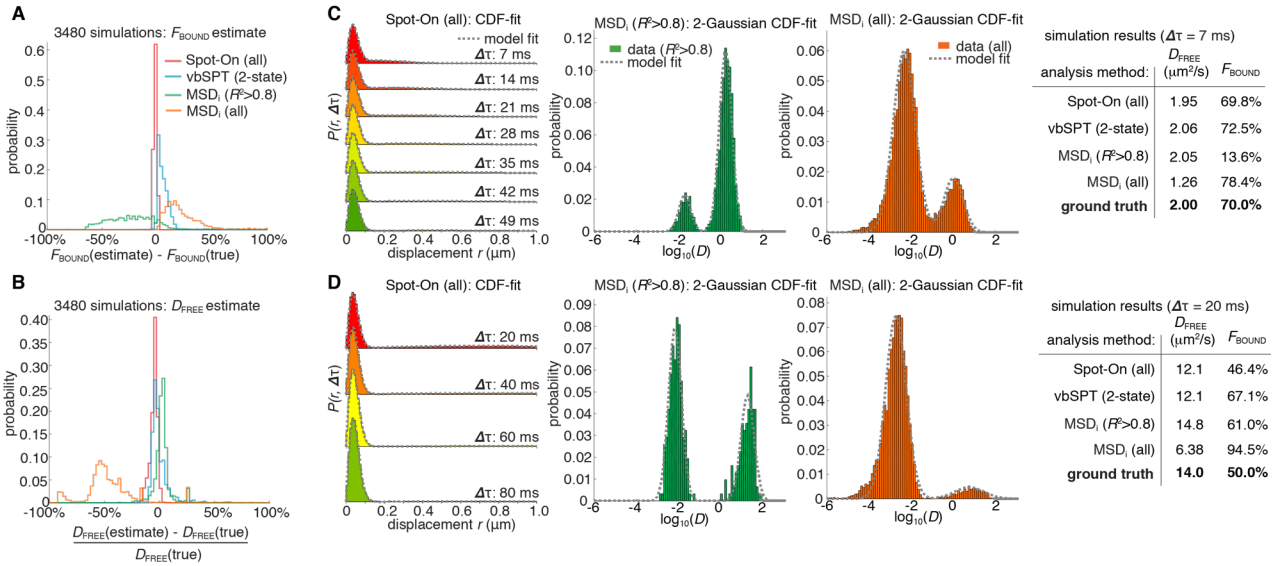

Figure 3. Validation of Spot-On using simulations and comparisons to other methods.
(**A-B**) Simulation results. Experimentally realistic SPT data was simulated inside a spherical mammalian nucleus with a radius of 4 μm subject to highly-inclined and laminated optical sheet illumination (Tokunaga et al., 2008) (HiLo) of thickness 4 μm illuminating the center of the nucleus. The axial detection window was 700 nm with Gaussian edges and particles were subject to a 25 nm localization error in all three dimensions. Photobleaching corresponded to a mean trajectory length of 4 frames inside the HiLo sheet and 40 outside. 3480 experiments were simulated with parameters of D_{FREE}=[0.5;14.5] in steps of 0.5 μm^2/s and F_{BOUND}=[0;95% in steps of 5% and the frame rate correspond to $\Delta\tau$=[1,4,7,10,13,20] ms. Each experiment was then fitted using Spot-On, using vbSPT (maximum of 2 states allowed) (Persson et al., 2013), MSD$_i$ using all trajectories of at least 5 frames (MSD$_i$ (all)) or MSD$_i$ using all trajectories of at least 5 frames where the MSD-curvefit showed at least R^2>0.8 (MSD$_i$ (R^2>0.8)). (**A**) shows the distribution of absolute errors in the F_{BOUND}–estimate and (**B**) shows the distribution of relative errors in the D_{FREE}–estimate. (**C**) Single simulation example with D_{FREE} = 2.0 μm^2/s; F_{BOUND} = 70%; 7 ms per frame. The table on the right uses numbers from CDF-fitting, but for simplicity the fits to the histograms (PDF) are shown in the three plots. (**D**) Single simulation example with D_{FREE} = 14.0 μm^2/s; F_{BOUND} = 50%; 20 ms per frame. Full details on how SPT data was simulated and analyzed with the different methods is given in Supplemental Information.

0.5 μm^2/s to 14.5 μm^2/s in 0.5 μm^2/s increments) and the bound fraction (from 0% to 95% in 5% increments) yielding a total of 3480 different conditions that span the full range of biologically plausible dynamics (Figure S1-6; Supplemental Information).

Spot-On robustly inferred subpopulation sizes with minimal error (Figure 3A-B, Table 1), but slightly underestimated the diffusion constant (-4.8%; Figure 3B; Table 1). However, this underestimate was due to particle confinement inside the nucleus: Spot-On correctly inferred the diffusion constant when the confinement was relaxed (Figure S3; 20 μm nuclear radius instead of 4 μm). This emphasizes that diffusion constants measured by SPT inside cells should be viewed as apparent diffusion constants. In contrast, the MSD$_i$ method failed under most conditions regardless of whether all trajectories were used (MSD$_i$ (all)) or a fitting filter applied (MSD$_i$ (R^2>0.8); Figure 3A-B; Table 1). vbSPT performed almost as well as Spot-On for slow-diffusing proteins, but showed larger deviations for fast-diffusing proteins (Figure S1-2).

To illustrate how the methods could give such divergent results when run on the same SPT data, we considered two example simulations (Figure 3C-D; more examples in Figure S2). First, we considered a mostly bound and relatively slow diffusion case (D_{FREE}: 2.0 μm^2/s; F_{BOUND}: 70%; $\Delta\tau$: 7

ms; Figure 3C). Spot-On and vbSPT robustly inferred both D_{FREE} and F_{BOUND}. In contrast, MSD_i (R^2>0.8) greatly underestimated F_{BOUND} (13.6% vs. 70%), whereas MSD_i (all) slightly overestimated F_{BOUND}. Since MSD_i-based methods apply two thresholds (first, minimum trajectory length: here 5 frames; second, filtering based on R^2) in many cases less than 5% of all trajectories passed these thresholds and this example illustrate how sensitive MSD_i-based methods are to these thresholds. Second, we considered an example with a slow frame rate and fast diffusion, such that the free population rapidly moves out-of-focus (D_{FREE}: 14.0 $\mu m^2/s$; F_{BOUND}: 50%; $\Delta\tau$: 20 ms; Figure 3D). Spot-On again robustly inferred F_{BOUND}, and slightly underestimated D_{FREE} due to high nuclear confinement (Figure S3). Although vbSPT generally performed well, because it does not correct for defocalization bias (vbSPT was developed for bacteria, where defocalization bias is minimal), vbSPT strongly overestimated F_{BOUND} in this case (Figure 3D). The MSD_i-based methods again gave divergent results despite seemingly fitting the data well. Thus, a good fit to a histogram of log(D) does not necessarily imply that the inferred D_{FREE} and F_{BOUND} are accurate. A full discussion of the simulations, methods comparison and Spot-On parameter sensitivity analysis is given in Supplemental Information.

Taken together, this analysis of simulated SPT data suggests that Spot-On successfully overcomes defocalization and analysis method biases (Figure 1C-D), accurately and consistently estimates subpopulations and diffusion constants across a wide range of dynamics and, finally, outperforms other methods.

	D_{FREE}			F_{BOUND}		
Analysis method	bias	std	iqr	bias	std	iqr
Spot-On (all)	-4.8%	3.3%	3.5%	-1.7%	1.2%	1.8%
vbSPT (2-state)	0.8%	12.5%	6.8%	5.0%	4.6%	6.1%
MSD_i (R^2>0.8)	8.0%	28.5%	4.9%	-20.6%	26.4%	32.1%
MSD_i (all)	-39.6%	41.8%	19.0%	22.0%	15.8%	17.8%

Table 1. Summary of simulation results and comparison of methods.
The table shows the bias (mean error), the standard deviation (std) and the interquartile range (iqr) for each method for all 3480 simulations. The left column shows the relative bias/std/iqr for the D_{FREE}-estimate and the right column shows the absolute bias/std/iqr for the F_{BOUND}-estimate.

spaSPT minimizes biases in experimental SPT acquisitions

Having validated Spot-On on simulated data, which is not subject to experimental biases (Figure 1A-B), we next sought to evaluate Spot-On on experimental data. To generate SPT data with minimal acquisition bias we performed stroboscopic photo-activation SPT (spaSPT; Figure 4A), which integrates previously and separately published ideas to minimize experimental biases.

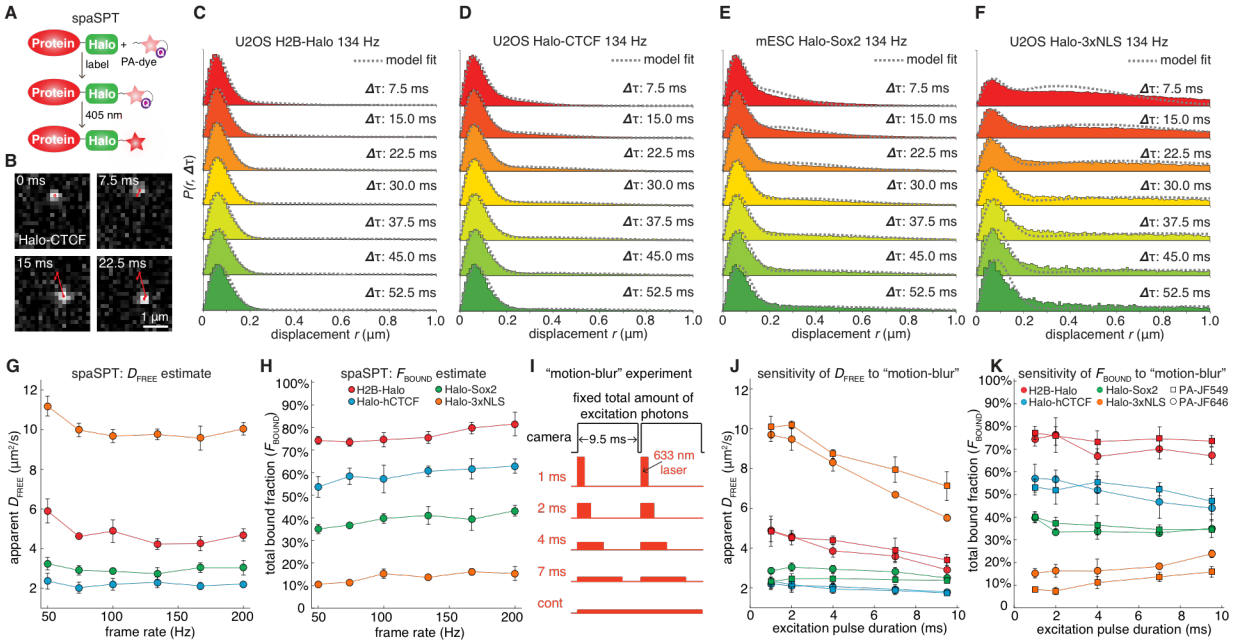

Figure 4. Overview of spaSPT and experimental results. (**A**) spaSPT. HaloTag-labeling with UV (405 nm) photo-activatable dyes enable spaSPT. spaSPT minimizes tracking errors through photo-activation which maintains low densities. (**B**) Example data. Raw spaSPT images for Halo-CTCF tracked in human U2OS cells at 134 Hz. (**C-F**) Histograms of displacements for multiple **Δτ** of histone H2B-Halo in U2OS cells (**C**), Halo-CTCF in U2OS cells (**d**), Halo-Sox2 in mES cells (**E**) and Halo-3xNLS in U2OS cells (**F**). (**G-H**) Effect of frame-rate on D_{FREE} and F_{BOUND}. spaSPT was performed at 200 Hz, 167 Hz, 134 Hz, 100 Hz, 74 Hz and 50 Hz using the 4 cell lines and the data fit using Spot-On and a 2-state model. Each experiment on each cell line was performed in 4 replicates on different days and ~5 cells imaged each day. Error bars show standard deviation between replicates. (**I**) Motion-blur experiment. To investigate the effect of "motion-blurring", the total number of excitation photons was kept constant, but delivered during pulses of duration 1, 2, 4, 7 ms or continuous (cont) illumination. (**J-K**) Effect of motion-blurring on D_{FREE} and F_{BOUND}. spaSPT data was recorded at 100Hz and 2-state model-fitting performed with Spot-On. The inferred D_{FREE} (**J**) and F_{BOUND} (**K**) were plotted as a function of excitation pulse duration. Each experiment on each cell line was performed in 4 replicates on different days and ~5 cells imaged each day. Error bars show standard deviation between replicates.

First, spaSPT minimizes motion-blurring, which is caused by particle movement during the camera exposure time (Figure 1A), by using stroboscopic excitation (Elf et al., 2007). We found that the bright and photo-stable dyes PA-JF_{549} and PA-JF_{646} (Grimm et al., 2016) in combination with the HaloTag ("Halo") labeling strategy made it possible to achieve a signal-to-background ratio greater than 5 with just 1 ms excitation pulses, thus providing a good compromise between minimal motion-blurring and high signal (Figure 4B). Second, spaSPT minimizes tracking errors (Figure 1B) by using photo-activation (Figure 4A)(Grimm et al., 2016; Manley et al., 2008). Tracking errors are generally caused by high particles densities. Photo-activation allows tracking at extremely low densities (≤1 molecule per nucleus per frame) and thereby minimizes tracking errors (Izeddin et al., 2014), whilst at the same time generating thousands of trajectories. To consider the full spectrum of nuclear protein dynamics, we studied histone H2B-Halo (overwhelmingly bound; fast diffusion; Figure 4C), Halo-CTCF (Hansen et al., 2017) (largely bound; slow diffusion; Figure 4D) and Halo-NLS (overwhelmingly free; very fast diffusion; Figure 4F) in human U2OS cells and Halo-Sox2

(Teves et al., 2016) (largely free; intermediate diffusion; Figure. 4E) in mouse embryonic stem cells (mESCs). We labeled Halo-tagged proteins in live cells with the HaloTag ligands of PA-JF_{549} or PA-JF_{646} (Grimm et al., 2016) and performed spaSPT using HiLo illumination. To generate a large dataset to comprehensively test Spot-On, we performed 1064 spaSPT experiments across 60 different conditions.

Validation of Spot-On using spaSPT data at different frame rates

First, we studied whether Spot-On could consistently infer subpopulations over a wide range of frame rates. Accordingly, we performed spaSPT at 200 Hz, 167 Hz, 134 Hz, 100 Hz, 74 Hz and 50 Hz using the four cell lines. Spot-On consistently inferred the diffusion constant (Figure 4G) and total bound fraction across the wide range of frame rates (Figure 4H). This is notable since all four proteins exhibit apparent anomalous diffusion (Figure S7). Moreover, Spot-On gave intuitively reasonable results (histone H2B was overwhelmingly bound; free Halo-3xNLS was overwhelmingly unbound). These results provide additional validation for the bias corrections implemented in Spot-On. We then sub-sampled the data sets and found that just ~3000 short trajectories (mean length ~3-4 frames) were sufficient for Spot-On to reliably infer the underlying dynamics (Figure S8). Thus, since many thousands of trajectories can normally easily be obtained with spaSPT for all but the most lowly expressed nuclear proteins, this now makes it possible to study biological cell-to-cell variability in TF dynamics.

Effect of motion-blur bias on parameter estimates

Having validated Spot-On on experimental SPT data, we next applied Spot-On to estimate the effect of motion-blurring on the estimation of subpopulations. As mentioned, since most localization algorithms (Chenouard et al., 2014; Sergé et al., 2008) achieve super-resolution through PSF-fitting, this may cause motion-blurred molecules to be undersampled, resulting in a bias towards slow-moving molecules (Figure 1A). We estimated the extent of the bias by imaging the four cell lines at 100 Hz and keeping the total number of excitation photons constant, but varying the excitation pulse duration (1 ms, 2 ms, 4 ms, 7 ms, constant; Figure 4I). For generality, we performed these experiments using both PA-JF_{549} and PA-JF_{646} dyes (Grimm et al., 2016). We used Spot-On to fit the data and plotted the apparent free diffusion constant (Figure 4J) and apparent total bound fraction (Fig. 4K) as a function of the excitation pulse duration. For fast-diffusing proteins like Halo-3xNLS and H2B-Halo, motion-blurring resulted in a large underestimate of the free

diffusion constant, whereas the effect on slower proteins like CTCF and Sox2 was minor (Figure 4J). Regarding the total bound fraction, motion-blurring caused a ~2-fold overestimate for rapidly diffusing Halo-3xNLS (Figure 4K), but had a minor effect on slower proteins like H2B, CTCF and Sox2. Importantly, similar results were obtained for both dyes, though JF_{549} yielded a slightly lower bound fraction for Halo-3xNLS (Figure 4J-K). We note that the extent of the bias due to motion-blurring will likely be very sensitive to the localization algorithm. Here, using the MTT-algorithm (Sergé et al., 2008), motion-blurring caused up to a 2-fold error in both the D_{FREE} and F_{BOUND} estimates.

Taken together, these results suggest that Spot-On can reliably be used even for SPT data collected under constant illumination provided that protein diffusion is sufficiently slow and, moreover, provides a helpful guide for optimizing SPT imaging acquisitions (we include a full discussion of considerations for SPT acquisitions and a proposal for minimum reporting standards in SPT in Supplemental Information).

CONCLUSION

In summary, SPT is an increasingly popular technique and has been revealing important new biological insight. However, a clear consensus on how to perform and analyze SPT experiments is currently lacking. In particular, 2D SPT of fast-diffusing molecules inside 3D cells is subject to a number of inherent experimental (Figure 1A-B) and analysis (Figure 1C-D) biases, which can lead to inaccurate conclusions if not carefully corrected for.

Here, we introduce approaches for accounting for both experimental and analysis biases. Several methods are available for localization/tracking (Chenouard et al., 2014; Sergé et al., 2008) and for classification of individual trajectories (Monnier et al., 2015; Persson et al., 2013). Spot-On now complements these tools by providing a bias-corrected, comprehensive open-source framework for inferring subpopulations and diffusion constants from pooled SPT data. This platform can easily be extended to other diffusion regimes. Moreover, spaSPT provides an acquisition protocol for tracking fast-diffusing molecules with minimal bias. We hope that these validated tools will help make SPT more accessible to the community and contribute positively to the emergence of "gold-standard" acquisition and analysis procedures for SPT.

Acknowledgements

We are very grateful to Davide Mazza who inspired this work and provided invaluable comments on Spot-On, to Florian Mueller for suggestions on the web-application, Christophe Zimmer for insightful discussions, David McSwiggen and Sheila Teves for kindly providing cell lines, Carolyn Elya and Chiahao Tsui for the name “Spot-On”, and to members of the Tjian/Darzacq labs and Maxime Dahan for discussions. We also thank Astou Tangara and Anatalia Robles for microscope maintenance. ASH is a postdoctoral fellow of the Siebel Stem Cell Institute. This work was supported by NIH grants UO1-EB021236 and U54-DK107980 (XD), the California Institute of Regenerative Medicine grant LA1-08013 (XD), by the Howard Hughes Medical Institute (003061, RT) and used the computational and storage services (TARS cluster) provided by the IT department at Institut Pasteur, Paros.

Competing financial interests

JBG and LDL declare competing financial interests. The other authors declare no competing financial interests.

Author contributions

ASH, MW and XD conceived the project. ASH and MW developed Spot-On, performed simulations, analyzed experiments and drafted the manuscript. ASH performed experiments. JBG and LDL developed and contributed JF dyes. RT and XD supervised the project. All authors edited the manuscript. ASH and MW contributed equally and are listed in alphabetical order on the first page.

MATERIALS AND METHODS

Spot-On model. Spot-On implements and extends a kinetic modeling framework first described in Mazza *et al*. (Mazza et al., 2012) and later extended in Hansen *et al*. (Hansen et al., 2017) The evolution over time of a concentration of particles located at the origin as a Dirac delta function and which follows free diffusion in two dimensions with a diffusion constant D can be described by a propagator (also known as Green's function). Properly normalized, the probability of a particle starting at the origin ending up at a location $r = (x,y)$ after a time delay, $\Delta\tau$, is given by:

$$P(r,\Delta\tau) = N\frac{r}{2D\Delta\tau}e^{-\frac{r^2}{4D\Delta\tau}}$$

Here N is a normalization constant with units of length. Spot-On integrates this distribution over a small histogram bin window, Δr, to obtain a normalized distribution to compare to binned experimental data. For simplicity, we will therefore leave out N from subsequent expressions. Since experimental SPT data is subject to a significant localization error, σ, Spot-On also accounts for this (Matsuoka et al., 2009):

$$P(r,\Delta\tau) = \frac{r}{2(D\Delta\tau+\sigma^2)}e^{-\frac{r^2}{4(D\Delta\tau+\sigma^2)}}$$

Many proteins studied by SPT can generally exist in a quasi-immobile state (e.g. a chromatin-bound state in the case of transcription factors) and one or more mobile states. We will first consider the 2-state model. Under most conditions, state transitions can be ignored (Hansen et al., 2017). Thus, the steady-state 2-state model considered by Spot-On becomes:

$$P(r,\Delta\tau) = F_{\mathrm{BOUND}}\frac{r}{2(D_{\mathrm{BOUND}}\Delta\tau+\sigma^2)}e^{-\frac{r^2}{4(D_{\mathrm{BOUND}}\Delta\tau+\sigma^2)}} + (1-F_{\mathrm{BOUND}})\frac{r}{2(D_{\mathrm{FREE}}\Delta\tau+\sigma^2)}e^{-\frac{r^2}{4(D_{\mathrm{FREE}}\Delta\tau+\sigma^2)}}$$

Here, the quasi-immobile subpopulation has diffusion constant, D_{BOUND}, and makes up a fraction, F_{BOUND}, whereas the freely diffusing subpopulation has diffusion constant, D_{FREE}, and makes up a fraction, $F_{\mathrm{FREE}} = 1 - F_{\mathrm{BOUND}}$. To account for defocalization bias (Figure 1C), Spot-On explicitly considers the probability of the freely diffusing subpopulation moving out of the axial detection range, Δz, during each time delay, $\Delta\tau$. This is important. For example, only ~25% of freely-diffusing molecules will remain in focus for at least 5 frames (assuming $\Delta\tau$=10 ms; Δz=700 nm; 1 gap allowed; D=5 μm^2/s), resulting in a 4-fold undercounting if uncorrected for. If we assume absorbing boundaries such that any molecule that contacts the edges of the axial detection range located at $z_{\mathrm{MAX}} = \Delta z/2$ and $z_{\mathrm{MIN}} = -\Delta z/2$ is permanently lost, the fraction of freely diffusing molecules with diffusion constant, D_{FREE}, that remain at time delay, $\Delta\tau$, is given by (Carslow and Jaeger, 1959; Kues and Kubitscheck, 2002):

$$P_{\mathrm{remaining}}(\Delta\tau,\Delta z,D_{\mathrm{FREE}}) = \frac{1}{\Delta z}\int_{-\Delta z/2}^{\Delta z/2}\left\{1-\sum_{n=0}^{\infty}(-1)^n\left[\mathrm{erfc}\left(\frac{\frac{(2n+1)\Delta z}{2}-z}{\sqrt{4D_{\mathrm{FREE}}\Delta\tau}}\right)+\mathrm{erfc}\left(\frac{\frac{(2n+1)\Delta z}{2}+z}{\sqrt{4D_{\mathrm{FREE}}\Delta\tau}}\right)\right]\right\}\mathrm{d}z$$

However, this analytical expression overestimates the fraction lost since there is a significant probability that a molecule that briefly contacted or exceeded the boundary re-enters the axial detection range. The re-entry probability depends on the number of gaps allowed in the tracking (g), $\Delta\tau$, and Δz and can be approximately accounted for by considering a corrected axial detection range, Δz_{corr}, larger than Δz: $\Delta z_{\mathrm{corr}} > \Delta z$:

$$\Delta z_{\mathrm{corr}}(\Delta z,\Delta\tau,D_{\mathrm{FREE}},g) = \Delta z + a(\Delta z,\Delta\tau,g)\sqrt{D_{\mathrm{FREE}}} + b(\Delta z,\Delta\tau,g)$$

Although Δz_{corr} depend on the number of gaps (g) allowed in the tracking, we will leave it out for simplicity in the following. We determined the coefficients a and b from Monte Carlo simulations. For a given diffusion constant, D, 50,000 molecules were randomly placed one-dimensionally along

the z-axis drawn from a uniform distribution from $z_{\mathrm{MIN}} = -\Delta z/2$ to $z_{\mathrm{MAX}} = \Delta z/2$. Next, using a time-step $\Delta\tau$, one-dimensional Brownian diffusion was simulated along the z-axis using the Euler-Maruyama scheme. For time delays from $1\Delta\tau$ to $15\Delta\tau$, the fraction of molecules that were lost was calculated in the range of D=[1;12] μm^2/s. $a(\Delta z, \Delta\tau, g)$ and $b(\Delta z, \Delta\tau, g)$ were then estimated through least-squares fitting of $P_{\mathrm{remaining}}(\Delta\tau, \Delta z_{\mathrm{corr}}, D)$ to the simulated fraction remaining. The process was repeated over a grid of plausible values of $(\Delta z, \Delta\tau, g)$ to derive a grid of 134,865 (a,b) parameter pairs. This pre-calculated library of (a,b) parameters enables Spot-On to perform model fitting on nearly any SPT dataset with minimal overhead.

Thus, the 2-state model Spot-On uses for kinetic modeling of SPT data is given by:

$$P_2(r,\Delta\tau) = F_{\mathrm{BOUND}}\frac{r}{2(D_{\mathrm{BOUND}}\Delta\tau+\sigma^2)}e^{-\frac{r^2}{4(D_{\mathrm{BOUND}}\Delta\tau+\sigma^2)}} + Z_{\mathrm{CORR}}(\Delta\tau,\Delta z_{\mathrm{corr}},D_{\mathrm{FREE}})(1-F_{\mathrm{BOUND}})\frac{r}{2(D_{\mathrm{FREE}}\Delta\tau+\sigma^2)}e^{-\frac{r^2}{4(D_{\mathrm{FREE}}\Delta\tau+\sigma^2)}}$$

where:

$$Z_{\mathrm{CORR}}(\Delta\tau,\Delta z_{\mathrm{corr}},D_{\mathrm{FREE}}) = \frac{1}{\Delta z_{\mathrm{corr}}}\int_{-\frac{\Delta z_{\mathrm{corr}}}{2}}^{\frac{\Delta z_{\mathrm{corr}}}{2}}\left\{1-\sum_{n=0}^{\infty}(-1)^n\left[\mathrm{erfc}\left(\frac{\frac{(2n+1)\Delta z_{\mathrm{corr}}}{2}-z}{\sqrt{4D_{\mathrm{FREE}}\Delta\tau}}\right)+\mathrm{erfc}\left(\frac{\frac{(2n+1)\Delta z_{\mathrm{corr}}}{2}+z}{\sqrt{4D_{\mathrm{FREE}}\Delta\tau}}\right)\right]\right\}\mathrm{d}z$$

Having derived the 2-state model, generalization to a 3-state model with 1 bound and 2 diffusive states is straightforward. If the three subpopulations have diffusion constants D_{BOUND}, D_{SLOW}, D_{FAST}, and fractions F_{BOUND}, F_{SLOW}, F_{FAST}, such that $F_{\mathrm{BOUND}} + F_{\mathrm{SLOW}} + F_{\mathrm{FAST}}$=1, then the 3-state model considered by Spot-On becomes:

$$P_3(r,\Delta\tau) = F_{\mathrm{BOUND}}\frac{r}{2(D_{\mathrm{BOUND}}\Delta\tau+\sigma^2)}e^{-\frac{r^2}{4(D_{\mathrm{BOUND}}\Delta\tau+\sigma^2)}} + Z_{\mathrm{CORR}}(\Delta\tau,\Delta z_{\mathrm{corr}},D_{\mathrm{SLOW}})F_{\mathrm{SLOW}}\frac{r}{2(D_{\mathrm{SLOW}}\Delta\tau+\sigma^2)}e^{-\frac{r^2}{4(D_{\mathrm{SLOW}}\Delta\tau+\sigma^2)}} + Z_{\mathrm{CORR}}(\Delta\tau,\Delta z_{\mathrm{corr}},D_{\mathrm{FAST}})(1-F_{\mathrm{BOUND}}-F_{\mathrm{SLOW}})\frac{r}{2(D_{\mathrm{FAST}}\Delta\tau+\sigma^2)}e^{-\frac{r^2}{4(D_{\mathrm{FAST}}\Delta\tau+\sigma^2)}}$$

Where $Z_{\mathrm{CORR}}(\Delta\tau,\Delta z_{\mathrm{corr}},D)$ is as described above.

Numerical implementation of models in Spot-On. Spot-On calculates the empirical histogram of displacements based on a user-defined bin width. Spot-On allows the user to choose between PDF- and CDF-fitting of the kinetic model to the empirical displacement distributions; CDF-fitting is generally most accurate for smaller datasets and the two are similar for large datasets (Figure S6). The integral in $Z_{\mathrm{CORR}}(\Delta\tau,\Delta z_{\mathrm{corr}})$ was numerically evaluated using the midpoint method over 200 points and the terms of the series computed until the term falls below a threshold of 10^{-10}. Model fitting and parameter optimization was performed using a non-linear least squares algorithm (Levenberg-Marquardt). Random initial parameter guesses are drawn uniformly from the user-specified parameter range. The optimization is then repeated several times with different initialization parameters to avoid local minima. Spot-On constrains each fraction to be between 0 and 1 and for the sum of the fractions to equal 1.

Cell Culture. Halo-Sox2 (Teves et al., 2016) knock-in JM8.N4 mouse embryonic stem cells (Pettitt et al., 2009) were grown on plates pre-coated with a 0.1% autoclaved gelatin solution (Sigma-Aldrich, G9391) under feeder free conditions in knock-out DMEM with 15% FBS and LIF (full recipe: 500 mL knockout DMEM (ThermoFisher #10829018), 6 mL MEM NEAA (ThermoFisher #11140050), 6 mL GlutaMax (ThermoFisher #35050061), 5 mL Penicillin-streptomycin (ThermoFisher #15140122), 4.6 μL 2-mercapoethanol (Sigma-Aldrich M3148), 90 mL fetal bovine serum (HyClone FBS SH30910.03 lot #AXJ47554)) and LIF. mES cells were fed by replacing half

the medium with fresh medium daily and passaged every two days by trypsinization. Halo-3xNLS, H2B-Halo-SNAP and knock-in C32 Halo-CTCF(Hansen et al., 2017) Human U2OS osteosarcoma cells were grown in low glucose DMEM with 10% FBS (full recipe: 500 mL DMEM (ThermoFisher #10567014), 50 mL fetal bovine serum (HyClone FBS SH30910.03 lot #AXJ47554) and 5 mL Penicillin-streptomycin (ThermoFisher #15140122)) and were passaged every 2-4 days before reaching confluency. For live-cell imaging, the medium was identical except DMEM without phenol red was used (ThermoFisher #31053028). Both mouse ES and human U2OS cells were grown in a Sanyo copper alloy IncuSafe humidified incubator (MCO-18AIC(UV)) at 37°C/5.5% CO_2. Cell lines were pathogen tested and authenticated through STR profiling (U2OS) as described previously (Hansen et al., 2017; Teves et al., 2016).

Single-molecule imaging. The indicated cell line was grown overnight on plasma-cleaned 25 mm circular no 1.5H cover glasses (Marienfeld High-Precision 0117650) either directly (U2OS) or MatriGel coated (mESCs; Fisher Scientific #08-774-552 according to manufacturer's instructions just prior to cell plating). After overnight growth, cells were labeled with 5-50 nM PA-JF_{549} or PA-JF_{646} (Grimm et al., 2016) for ~15-30 min and washed twice (one wash: medium removed; PBS wash; replenished with fresh medium). At the end of the final wash, the medium was changed to phenol red-free medium keeping all other aspects of the medium the same. Single-molecule imaging was performed on a custom-built Nikon TI microscope equipped with a 100x/NA 1.49 oil-immersion TIRF objective (Nikon apochromat CFI Apo TIRF 100x Oil), EM-CCD camera (Andor iXon Ultra 897; frame-transfer mode; vertical shift speed: 0.9 µs; -70°C), a perfect focusing system to correct for axial drift and motorized laser illumination (Ti-TIRF, Nikon), which allows an incident angle adjustment to achieve highly inclined and laminated optical sheet illumination (Tokunaga et al., 2008). The incubation chamber maintained a humidified 37°C atmosphere with 5% CO_2 and the objective was also heated to 37°C. Excitation was achieved using the following laser lines: 561 nm (1 W, Genesis Coherent) for PA-JF_{549}; 633 nm (1 W, Genesis Coherent) for PA-JF_{646}; 405 nm (140 mW, OBIS, Coherent) for all photo-activation experiments. The excitation lasers were modulated by an acousto-optic Tunable Filter (AA Opto-Electronic, AOTFnC-VIS-TN) and triggered with the camera TTL exposure output signal. The laser light is coupled into the microscope by an optical fiber and then reflected using a multi-band dichroic (405 nm/488 nm/561 nm/633 nm quad-band, Semrock) and then focused in the back focal plane of the objective. Fluorescence emission light was filtered using a single band-pass filter placed in front of the camera using the following filters: PA-JF549: Semrock 593/40 nm bandpass filter; PA-JF_{646}: Semrock 676/37 nm bandpass filter. The microscope, cameras, and hardware were controlled through NIS-Elements software (Nikon).

spaSPT experiments and analysis. The spaSPT experimental settings for Figure 4G-H were as follows: 1 ms 633 nm excitation (100% AOTF) of PA-JF_{646} was delivered at the beginning of the frame; 405 nm photo-activation pulses were delivered during the camera integration time (~447 µs) to minimize background and their intensity optimized to achieve a mean density of ≤1 molecule per frame per nucleus. 30,000 frames were recorded per cell per experiment. The camera exposure times were: 4.5 ms, 5.5 ms, 7 ms, 9.5 ms, 13 ms and 19.5 ms.

For the motion-blur spaSPT experiments (Figure 4I-K), the camera exposure was fixed to 9.5 ms and photo-activation performed as above. To keep the total number of delivered photons constant, we generated an AOTF-laser intensity calibration curve using a power meter and adjusted the AOTF transmission accordingly for each excitation pulse duration. The excitation settings were as follows: 1 ms, 561 nm 100% AOTF, 633 nm 100% AOTF; 2 ms, 561 nm 43% AOTF, 633 nm 40% AOTF; 4 ms, 561 nm 28% AOTF, 633 nm 27% AOTF; 7 ms, 561 nm 20% AOTF, 633 nm 19%

AOTF; constant illumination, 561 nm 17% AOTF, 633 nm 16% AOTF.
spaSPT data was analyzed (localization and tracking) and converted into trajectories using a custom-written Matlab implementation of the MTT-algorithm (Sergé et al., 2008) and the following settings: Localization error: $10^{-6.25}$; deflation loops: 0; Blinking (frames): 1; max competitors: 3; max D ($\mu m^2/s$): 20. The spaSPT trajectory data was then analyzed using the Matlab version of Spot-On (v1.0; GitLab tag 1f9f782b) and the following parameters: dZ=0.7 μm; GapsAllowed=1; TimePoints: 4 (50 Hz), 6 (74 Hz), 7 (100 Hz), 8 (134 Hz), 9 (167 and 200 Hz); JumpsToConsider=4; ModelFit=2; NumberOfStates=2; FitLocError=0; LocError=0.035 μm; D_Free_2State=[0.4;25]; D_Bound_2State=[0.00001;0.08];

SPT simulations. We developed a utility to simulate diffusing proteins in a confined geometry (simSPT). Briefly, simSPT simulates the diffusion of an arbitrary number of populations of molecules characterized by their diffusion coefficient, under a steady state assumption. Particles are drawn at random between the populations and their location in the 3D nucleus is initialized following a uniform law within the confinement volume. The lifetime of the particle (in frames) is also drawn following an exponential law of mean lifetime β. Then, the particle diffuses in 3D until it bleaches. Diffusion is simulated by drawing jumps following a normal law of parameters $\sim N\left(0, \sqrt{2D\Delta\tau}\right)$, where D is the diffusion coefficient and $\Delta\tau$ the exposure time. Finally, a localization error ($\sim N(0,\sigma)$) is added to each (x,y,z) localization in the simulated trajectories. For this work, we parameterized simSPT to consider that two subpopulations of particles diffuse in a sphere (the nucleus) of 8 μm diameter illuminated using HiLo illumination (assuming a HiLo beam width of 4 μm), with an axial detection range of ~700 nm, centered at the middle of the HiLo beam. Molecules are assumed to have a mean lifetime of 4 frames (when inside the HiLo beam) and of 40 frames when outside the HiLo beam. The localization error was set to 25 nm and the simulation was run until 100,000 in-focus trajectories were recorded. More specifically, the effect of the exposure time (1 ms, 4 ms, 7 ms, 13 ms, 20 ms), the free diffusion constant (from 0.5 $\mu m^2/s$ to 14.5 $\mu m^2/s$ in 0.5 $\mu m^2/s$ increments) and the fraction bound (from 0 % to 95 % in 5 % increments) were investigated, yielding a dataset consisting of 3480 simulations. More details on the simulations, including scripts to reproduce the dataset, are available on GitLab as detailed in the "Computer code" section. Full details on how the simulations were analyzed by Spot-On, vbSPT and MSD_i are given in Supplemental Information.

Data availability. All raw 1064 spaSPT experiments as well as the 3480 simulations (Figure 3) are freely available in Spot-On readable Matlab and CSV file formats in the form of SPT trajectories at Zenodo. The experimental data is available at Zenodo after publication and upon request before publication; The simulations are available in Matlab format at: https://zenodo.org/record/835541; The simulations are available in CSV format at: https://zenodo.org/record/834787; And supplementary software used for MSD_i and vbSPT analysis as well as for generating the simulated data at: https://zenodo.org/record/835171

Computer code. Spot-On is fully open-source. The web-interface can be found at: https://SpotOn.berkeley.edu. All raw code is available at GitLab: https://gitlab.com/tjian-darzacq-lab. The web-interface code can be found at https://gitlab.com/tjian-darzacq-lab/Spot-On; the Matlab command-line version of Spot-On can be found at: https://gitlab.com/tjian-darzacq-lab/spot-on-matlab; the Python command-line version of Spot-On can be found at https://gitlab.com/tjian-darzacq-lab/Spot-On-cli; finally, the SPT simulation code (simSPT) can be found at: https://gitlab.com/tjian-darzacq-lab/simSPT.

REFERENCES

Berglund, A.J. (2010). Statistics of camera-based single-particle tracking. Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. *82*.

Carslow, H.S., and Jaeger, J.C. (1959). Conduction of Heat in Solids.

Chenouard, N., Smal, I., de Chaumont, F., Maska, M., Sbalzarini, I.F., Gong, Y., Cardinale, J., Carthel, C., Coraluppi, S., Winter, M., et al. (2014). Objective comparison of particle tracking methods. Nat Meth *11*, 281–289.

Elf, J., Li, G.-W., and Xie, X.S. (2007). Probing transcription factor dynamics at the single-molecule level in a living cell. Science *316*, 1191–1194.

Grimm, J.B., English, B.P., Choi, H., Muthusamy, A.K., Mehl, B.P., Dong, P., Brown, T.A., Lippincott-Schwartz, J., Liu, Z., Lionnet, T., et al. (2016). Bright photoactivatable fluorophores for single-molecule imaging. Nat. Methods 66779.

Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R., and Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. Elife *6*.

Izeddin, I., Recamier, V., Bosanac, L., Cisse, I.I., Boudarene, L., Dugast-Darzacq, C., Proux, F., Benichou, O., Voituriez, R., Bensaude, O., et al. (2014). Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. Elife *2014*, 1–27.

Kues, T., and Kubitscheck, U. (2002). Single molecule motion perpendicular to the focal plane of a microscope: Application to splicing factor dynamics within the cell nucleus. Single Mol. *3*, 218–224.

Liu, Z., Lavis, L.D., and Betzig, E. (2015). Imaging Live-Cell Dynamics and Structure at the Single-Molecule Level. Mol. Cell *58*, 644.

Manley, S., Gillette, J.M., Patterson, G.H., Shroff, H., Hess, H.F., Betzig, E., and Lippincott-Schwartz, J. (2008). High-density mapping of single-molecule trajectories with photoactivated localization microscopy. Nat. Methods *5*, 155–157.

Matsuoka, S., Shibata, T., and Ueda, M. (2009). Statistical analysis of lateral diffusion and multistate kinetics in single-molecule imaging. Biophys. J. *97*, 1115–1124.

Mazza, D., Abernathy, A., Golob, N., Morisaki, T., and McNally, J.G. (2012). A benchmark for chromatin binding measurements in live cells. Nucleic Acids Res. *40*.

Monnier, N., Barry, Z., Park, H.Y., Su, K.-C., Katz, Z., English, B.P., Dey, A., Pan, K., Cheeseman, I.M., Singer, R.H., et al. (2015). Inferring transient particle transport dynamics in live cells. Nat Meth *12*, 838–840.

Mueller, F., Stasevich, T.J., Mazza, D., and McNally, J.G. (2013). Quantifying transcription factor kinetics: at work or at play? Crit Rev Biochem Mol Biol *48*, 492–514.

Persson, F., Lindén, M., Unoson, C., and Elf, J. (2013). Extracting intracellular diffusive states and transition rates from single-molecule tracking data. Nat. Methods *10*, 265–269.

Pettitt, S.J., Liang, Q., Rairdan, X.Y., Moran, J.L., Prosser, H.M., Beier, D.R., Lloyd, K.C., Bradley, A., and Skarnes, W.C. (2009). Agouti C57BL/6N embryonic stem cells for mouse genetic resources. Nat. Methods *6*, 493–495.

Rhodes, J., Mazza, D., Nasmyth, K., and Uphoff, S. (2017). Scc2/Nipbl Hops Between Chromosomal Cohesin Rings After Loading. bioRxiv.

Sergé, A., Bertaux, N., Rigneault, H., and Marguet, D. (2008). Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. Nat. Methods *5*, 687–694.

Shen, H., Tauzin, L.J., Baiyasi, R., Wang, W., Moringo, N., Shuang, B., and Landes, C.F. (2017). Single Particle Tracking: From Theory to Biophysical Applications. Chem. Rev. *117*, 7331–7376.

Swinstead, E.E., Miranda, T.B., Paakinaho, V., Baek, S., Goldstein, I., Hawkins, M., Karpova, T.S., Ball, D., Mazza, D., Lavis, L.D., et al. (2016). Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. Cell *165*, 593–605.

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

Tokunaga, M., Imamoto, N., and Sakata-Sogawa, K. (2008). Highly inclined thin illumination enables clear single-molecule imaging in cells. Nat. Methods *5*, 159–161.

Zhen, C.Y., Tatavosian, R., Huynh, T.N., Duc, H.N., Das, R., Kokotovic, M., Grimm, J.B., Lavis, L.D., Lee, J., Mejia, F.J., et al. (2016). Live-cell single-molecule tracking reveals co-recognition of H3K27me3 and DNA targets polycomb Cbx7-PRC1 to chromatin. Elife *5*.