scientific **data**

Check for updates

OPEN

DATA DESCRIPTOR

Single-Molecule Tracking dataset for histone H3 (*hht1*) from live and fixed cells of *Schizosaccharomyces pombe*

Akriti Kumari [1], Nitesh Kumar Podh[1], Sucharita Sen[2], Kirti Kashyap[3], Sahil Islam[3], Anupam Gupta[3], Eerappa Rajakumara[4], Mridula Nambiar[2] & Gunjan Mehta [1] ✉

Single-molecule tracking provides direct real-time measurements of protein dynamics in live cells with high spatiotemporal resolution. In the last decade, this method has been adopted by several labs to understand the dynamics of DNA replication, transcription, telomerase, chromatin remodeling, etc. in a variety of model systems (bacteria, yeast *S. cerevisiae*, cell lines, embryo). However, it has not been employed for the yeast *Schizosaccharomyces pombe*, despite being a valuable model system. Here, we present single-molecule tracking datasets for chromatin-bound histone H3 (Hht1) in live and fixed cells and freely diffusing GFP in *S. pombe* nuclei to benchmark the diffusion parameters (diffusion coefficient, mean squared displacement, fraction of bound molecules, residence time). These parameters will be used to differentiate chromatin-bound molecules from unbound (free) molecules of any protein. It will be a valuable resource for any lab that starts employing this method for studying protein dynamics. This dataset can also be used to validate the performance of various tracking software and as a training dataset for machine learning-based automated tracking.

Background & Summary

Cellular processes such as replication, transcription, translation, DNA repair, cell division etc. require fast inter-actions between biomolecules (DNA, RNA, proteins, lipids). All these processes and biomolecular interactions are dynamic and highly heterogeneous in the cell population. Decades of research using ensemble averaging methods (such as chromatin immunoprecipitation (ChIP), immunofluorescence, western blotting etc.) have provided the static picture of all these processes and the biomolecular interactions[1]. However, the dynamic information was lost due to cell population-based average measurements. Advancements in protein/DNA labe-ling strategies, microscope optics, camera technologies, and computational image processing have made it pos-sible to visualize the dynamics of individual biomolecules in live cells and to track their movement to estimate their diffusion dynamics and target-search mechanism. This method, referred to as Single-molecule tracking in live cells, provides real-time visualization of biomolecules with high spatiotemporal resolution[1]. In the last decade, several labs have adopted this technology to understand the molecular mechanism of DNA replication, transcription, telomerase action, chromatin remodeling, and many more[2–12]. This method has revolutionized the way we understood biological processes before, by revealing dynamic information about the biomolecules in action. It has been used with several model systems, such as bacteria, yeast (*Saccharomyces cerevisiae*), cell lines, and embryos.

Here, we have optimized the single-molecule imaging and tracking method in the yeast *Schizosaccharomyces pombe* (fission yeast), which is a widely used model system in cell and molecular biology. The overall work-flow from genetic engineering to single-molecule imaging, tracking, and data analysis is presented in Fig. 1.

[1]Laboratory of Chromosome Dynamics and Gene Regulation, Department of Biotechnology, Indian Institute of Technology Hyderabad, Kandi, Sangareddy, Telangana, 502285, India. [2]Department of Biology, Indian Institute of Science Education and Research, Pune, India. [3]Department of Physics, Indian Institute of Technology Hyderabad, Kandi, Sangareddy, Telangana, 502285, India. [4]Macromolecular Structural Biology Laboratory, Department of Biotechnology, Indian Institute of Technology Hyderabad, Kandi, Sangareddy, Telangana, 502285, India. ✉e-mail: gunjanmehta@bt.iith.ac.in
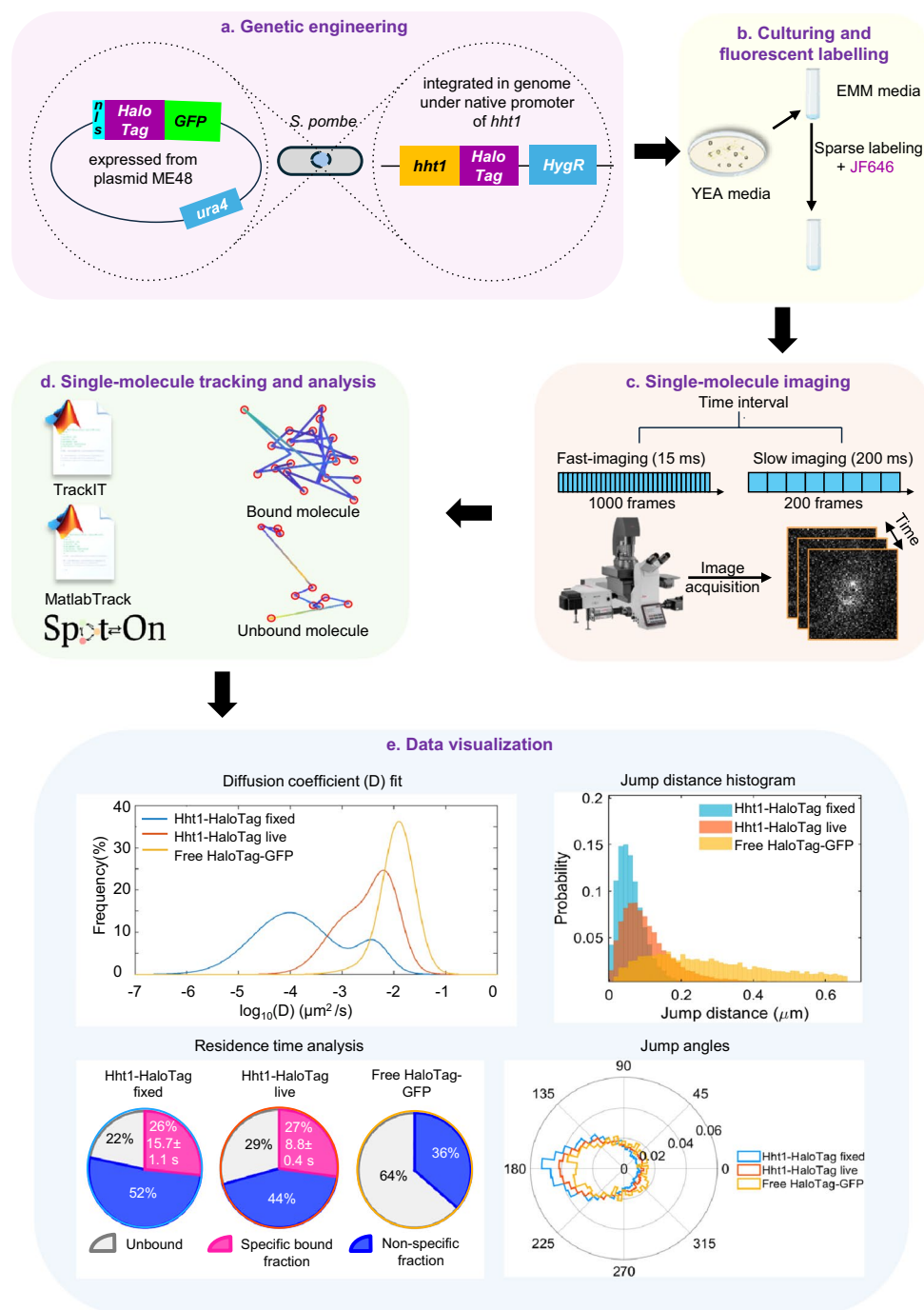
Fig. 1 Experimental workflow for single-molecule imaging and tracking. (**a**) Genetic engineering of yeast strains for visualizing the single-molecules of chromatin-bound histone H3 (Hht1) and free GFP. The *halotag* was fused at the C-terminus of the *hht1* gene endogenously by genetic engineering using the hygromycin resistance (HygR) gene as a selection marker. The recombinant gene *hht1-halotag* is expressed from its native promoter. Similarly, the *halotag* was fused at the N-terminus of the GFP gene in a cloned vector (with the *URA4* gene as a selection marker), and expressed from the *nmt1* promoter. *nls* = nuclear localization signal sequence. (**b**) Cells are taken out from the glycerol stocks (from −80°C freezer) and grown to the log phase before imaging. The Hht1-HaloTag protein was sparsely labeled with a JF646-HTL so that only 1 to 5 molecules per nucleus were labeled. (**c**) Single-molecule imaging is performed using fast and slow imaging regimes (time-lapse movies with 15 ms time intervals for 1000 frames and 200 ms time intervals for 200 frames, respectively). (**d**) Using tracking software (MatlabTrack and TrackIt), the single molecules are detected based on intensity thresholding, and the trajectories are made by connecting particles from one frame to the next frame. (**e**) Based on the X- and Y- positions of the molecules in their trajectories and how long the molecules survive, the biophysical parameters (such as jump distances, mean square displacement, diffusion coefficient, fraction of bound and free molecules, and residence time) can be quantified.

To benchmark the diffusion parameters (mean-squared displacements (MSD), diffusion coefficient (D_{bound}), fraction of bound molecules (F_{bound})) and the residence time (τ) for chromatin-bound histones, we endogenously fused - *HaloTag* with histone H3 gene (*hht1*) using homologous recombination (Fig. 1a, Supplementary Fig. 1). To benchmark the diffusion parameters for free/unbound molecules in the nucleus, we expressed NLS (Nuclear Localization Signal)-HaloTag-GFP from a plasmid ME48 (Fig. 1a, Supplementary Fig. 2).

The HaloTag fusion proteins (Hht1 and GFP) were sparsely labeled with 5 nM and 50 pM concentrations of JF646-HaloTag Ligand (HTL), respectively (Fig. 1b). Time-lapse movies were acquired with a 15 ms time interval (fast imaging regime to quantify diffusion parameters MSD, D_{bound}, F_{bound}, jump angles) and with a 200 ms time interval (slow imaging regime to quantify residence time) (Fig. 1c). Single-molecule tracking was performed as described previously[13] using TrackIt[14], Spot-On[15], and MatlabTrack[13,16] (Fig. 1d). The workflow for single-molecule detection and tracking is depicted in Fig. 2a. For the quantification of the residence time using MatlabTrack, the values for R_{min}, N_{min}, and R_{max} are needed. We have defined these parameters in the methods section and Fig. 2b, along with the procedure to quantify them. The measured parameters (τ, D_{bound}, F_{bound}) for the chromatin-bound histone H3 and the free molecules of NLS-HaloTag-GFP are shown in Fig. 3a,d. For chromatin-bound histone H3 in live cells, $\tau = 8.8 \pm 0.4$ s (Fig. 3a), $D_{bound} = 0.016 \, \mu m^2/s$, $F_{bound} = 62\%$ (Fig. 3d). For free NLS-HaloTag-GFP, we couldn't observe any specific bound fraction (Fig. 3a), and the Spot-On analysis revealed $F_{free} = 86\%$, $D_{free} = 1.6 \, \mu m^2/s$ (Fig. 3d).

To check the sensitivity and specificity of this method and to validate the effect of system perturbations on the readouts provided, we fixed the cells of the *hht1-HaloTag* strain with formaldehyde for 10 min. Cells were washed twice with EMM media. The single-molecule imaging and tracking were performed for Hht1-HaloTag-JF646-HTL as above. We found stark differences in the diffusion parameters and residence time. For chromatin-bound histone H3 in fixed cells, $\tau = 15.7 \pm 1.1$ s (Fig. 3a), $D_{bound} = 0.002 \, \mu m^2/s$, $F_{bound} = 89\%$ (Fig. 3d). This result suggests that the single-molecule imaging and tracking method used here is very sensitive to detect system perturbations.

This data descriptor and the associated dataset contain raw time-lapse movies (acquired with fast and slow imaging regimes,.tif files) of single-molecule imaging of chromatin-bound histone H3 (Hht1-HaloTag-JF646-HTL) in live and fixed cells, and freely diffusing NLS-HaloTag-GFP. It also contains the tracking files (.mat files generated through tracking software MatlabTrack and TrackIt), and.csv files generated through TrackIt.

Single-molecule tracking of any other protein in *S. pombe* will require this dataset and the values presented for the chromatin-bound histones and free NLS-HaloTag-GFP (τ, MSD, D_{bound}, F_{bound}, R_{min}, R_{min}, N_{min}) to distinguish between chromatin-bound and free molecules and to quantify residence time. Also, the presented dataset can be used to validate the performance of various tracking software and as a training dataset for machine learning-based automated tracking in *S. pombe*.

Methods

Genetic engineering of *S. pombe* strains for single-molecule imaging of chromatin-bound Hht1 and free GFP.
To generate the *hht1::hht1-HaloTag>>HygR* C-terminal fusion allele, we generated a DNA construct containing *hht1*-specific homology regions flanking a HaloTag>>HygR stitched product for integration in the genomic copy of *hht1* gene. The construct was designed to generate an in-frame fusion of HaloTag at the C-terminus of the *hht1* gene, followed by a transcription terminator sequence and an independently expressing *HygR* gene further downstream. The *hht1* left homology region from the ORF, 279 bp till the codon before the stop, was PCR amplified using primers MO255 and MO256 on *S. pombe* genomic DNA as a template. The 897 bp Halotag sequence was PCR amplified using primers MO186 and MO187 on the plasmid pTSK561 (addgene #190816). MO256 and MO186 add an 18 nt L1 linker sequence for facilitating the stitch between the *hht1* left homology and the HaloTag. The *hht1 left homology region-L1* was stitched to *L1-HaloTag-L2* by annealing the two products at the linker region L1 and allowing extension by PCR. This was followed by PCR amplification of the stitched product by using primers MO255 and MO187. A 2332 bp T*nmt1>>HygR* cassette was PCR amplified from plasmid ME13 (pFA6a derivative; Gift from Prof. Gerald Smith, USA) using primers MO192 and MO193. MO187 and MO192 add a 20 bp linker sequence L2 for facilitating the stitch between HaloTag and the HygR selection marker. The *hht1 left homology region-L1-HaloTag-L2* was then stitched with *L2-Tnmt1>>HygR-hht1 right homology* by annealing the two products at linker L2, followed by PCR amplification of the stitched product using primers MO255 and MO209. MO209 adds 80 bp post stop codon of *hht1* 3′UTR homology region. The 3627 bp final cassette constructed was *hht1L-L1-HaloTag-L2-Tnmt1>>HygR-hht1R*. The product was transformed into a Hygromycin-sensitive *S. pombe* strain using the lithium acetate method as described before[17]. The transformants were selected based on resistance to the antibiotic Hygromycin B (100 μg/ml). The genomic integrations were further verified using locus-specific primers (Supplementary Fig. 1) as described below in the 'Technical Validation' section.

For tracking free GFP in the *S. pombe* nucleus, we expressed NLS-HaloTag-GFP from an expression vector ME48 by transforming it into a wild-type strain MP561. To generate the vector ME48, NLS-HaloTag was amplified from plasmid pTSK561 (addgene #190816) using primers MO408 and MO409, containing XhoI and BamHI restriction enzyme sites, respectively. Primer MO408 additionally contained the sequence of the SV40 NLS (PKKKRKV). The double-digested products were ligated into pREP4x.2 (ME40) and the clones were confirmed using multiple restriction enzyme digestions as well as by PCR (Supplementary Fig. 2). The confirmed plasmid ME48 was then transformed into MP561 and *ura4* + transformants were selected on minimal media without uracil. Two transformants were selected and the presence of plasmid was confirmed by colony PCRs using locus-specific primers (Supplementary Fig. 2) as described in the 'Technical Validation' section.
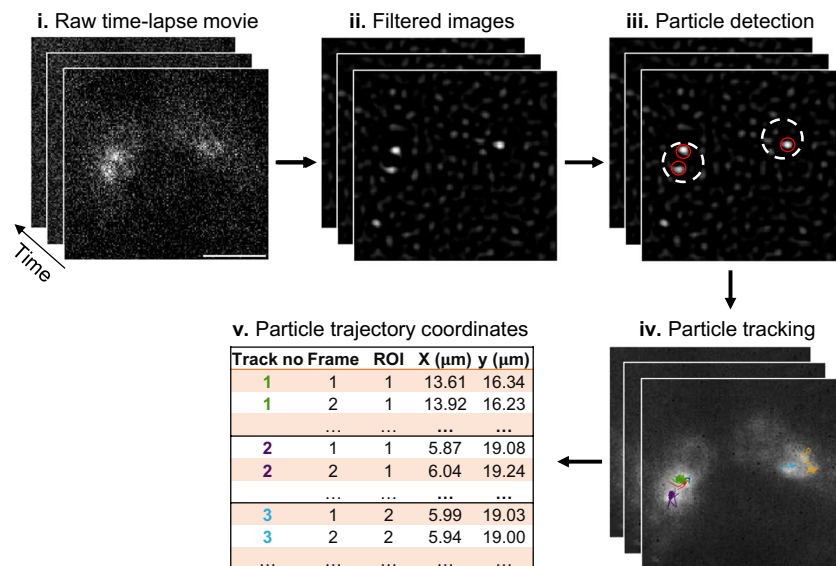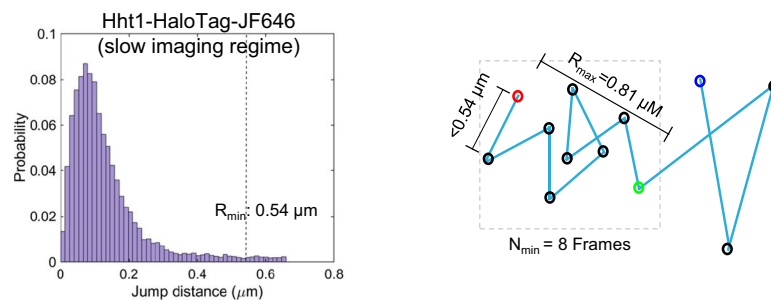
a. Workflow for single-molecule tracking

i. Raw time-lapse movie   **ii.** Filtered images   **iii.** Particle detection

v. Particle trajectory coordinates

Track no	Frame	ROI	X (µm)	y (µm)
1	1	1	13.61	16.34
1	2	1	13.92	16.23
…	…	…	…	…
2	1	1	5.87	19.08
2	2	1	6.04	19.24
…	…	…	…	…
3	1	2	5.99	19.03
3	2	2	5.94	19.00
…	…	…	…	…

iv. Particle tracking

b. Schematics to define R_{min}, N_{min}, and R_{max}

Hht1-HaloTag-JF646 (slow imaging regime)

R_{min}: 0.54 µm

R_{max} =0.81 µM

<0.54 µm

N_{min} = 8 Frames

Fig. 2 (**a**) Workflow for single-molecule tracking. **i.** raw time-lapse movies are acquired with low laser power to reduce photobleaching. So, the raw time-lapse images are noisy. Scale: 5 µm. **ii.** a computational filtering algorithm (bandpass filtering) was used to denoise the images. **iii.** The tracking software (MatlabTrack and TrackIt) detects the particles (red circles) from the filtered images using intensity thresholding within the defined region of interest (ROIs, white circles representing nuclei). **iv.** The tracking software connects the molecules from one frame to the next frame using the nearest neighbor algorithm and forms trajectories. **v.** The tracking software extracts the X- and Y- positions of the molecules throughout all the trajectories. (**b**) Schematic representation to define R_{min}, N_{min}, and R_{max}. These parameters are used in MatlabTrack to consider any molecule as a bound molecule. R_{min} is the frame-to-frame displacement (jump distance) traveled by >99% of histones (see histogram). N_{min} is the minimum number of frames for which the molecule should not travel more than R_{min} distance to be considered a bound molecule. Rmax is the maximum displacement that a particle can move after Nmin frames. This is to further discriminate diffusing molecules from bound. The software classifies segments as bound if they include at least Nmin time-points, and each position in the segment is less than Rmin away from the previous time-point, and less than Rmax away from the position measured in the time-point Nmin before. The red circle represents the start point of the trajectory, the green circle represents the end of the trajectory for which the molecule remained bound (as it traveled <0.54 µm distance for a minimum of 8 frames), and the blue circle represents the end of the trajectory. At this point, either the molecule was bleached or it left this site. Between green and blue circles, the molecule behaved as a free molecule as the jump distances were >0.54 µm.

All the primer sequences are provided in Supplementary Table 1, the genotypes of the *S. pombe* strains are provided in Supplementary Table 2, and the list of plasmids is provided in Supplementary Table 3. All strains and plasmids are available upon request.

Culturing of yeast cells for single-molecule imaging.   The cells were streaked on Yeast Extract Agar (YEA) media (Yeast Extract: 0.5%, Dextrose: 3%, Agar: 2%) from the glycerol stock and incubated at 30° C for 48 hr. A single colony was inoculated in 3 ml EMM media (Edinburgh's Minimal Media, MP Biomedicals, Cat no. 4110-012) and grown under shaking conditions (230 RPM, 30° C) for 24 hr. To bring cells to the log phase, 100 µl of this culture was inoculated in fresh 3 ml EMM broth and grown for 10 hr at 30° C under shaking conditions
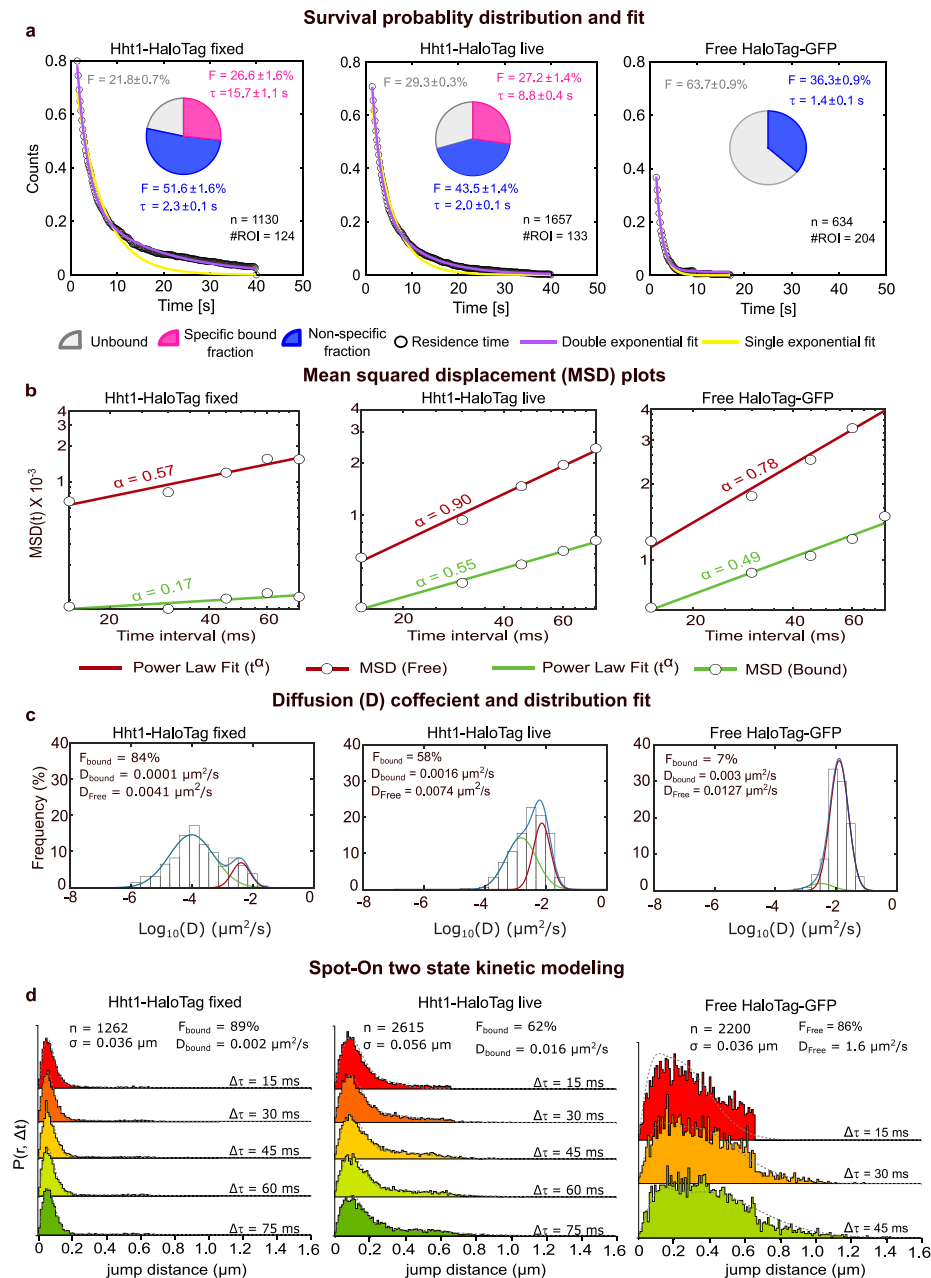
Fig. 3 Benchmarking parameters for histone H3 (Hht1-HaloTag-JF646-HTL) in live and fixed cells and for free GFP (NLS-HaloTag-GFP) in live cells of *S. pombe*. (**a**) Survival probability distributions generated through the MatlabTrack by tracking movies acquired with slow imaging regime. For Hht1-HaloTag, it fits well with the double exponential decay curve (purple line), suggesting two types of bound molecules: (1) bound with long residence time (specifically bound molecules, pink fraction in the pie chart), and (2) bound with short residence time (non-specifically bound molecules, blue fraction in the pie chart). For NLS-HaloTag-GFP, it fits well with the single-exponential decay curve (yellow line), suggesting only one type of bound molecules (bound with short residence time). The grey fraction represents unbound molecules. F = fraction, τ = mean residence time, n = number of tracks analyzed, #ROI = number of ROIs (cells) tracked. (**b**) Mean-squared displacement (MSD) for bound and free population, with the power-law fit ($MSD \propto t^{\alpha}$). α is an anomalous diffusion exponent that characterises the type of diffusion. To estimate α, log(MSD) and log(time) were fitted in a linear fit. $\alpha < 1$ indicates subdiffusive behaviour. (**c**) The movies acquired with fast imaging regime were tracked by TrackIt and the .csv files generated were used for making LogD plots using SPTAnalyzer. We used these plots only to identify the number of populations based on D, as the Spot-On analysis requires to select between the two-states or three-states kinetic models. The fitting of the logD plot estimates the values for the fraction of bound (F_{bound}) and free (F_{free}) molecules and their diffusion coefficients (D_{bound} and D_{free}), however, these values are not accurate as described in the methods section. (**d**) For robust estimation of F_{bound} and D_{bound}, the Spot-On analysis was performed. This analysis makes the histograms of jump distances for various time intervals (15 ms, 30 ms, 45 ms, 60 ms, and 75 ms). The two-state model fitting estimates the values for the F_{bound} and F_{free} and their D_{bound} and D_{free}, σ = localization error, n = number of tracks analyzed.

(230 RPM). 1 ml of culture was harvested (centrifuge at 2000 RPM for 2 min) and resuspended in 1 ml of fresh EMM. JF646-HTL (Promega, Cat no. GA112A) was added at 5 nM concentration for tracking Hht1-HaloTag (unless specified in the figure) and 50 pM concentrations for tracking NLS-HaloTag-GFP and kept shaking for 30 min. Cells were pelleted down again by centrifugation (2000 RPM for 2 min) and washed with 1 ml of fresh EMM media twice to remove unbound JF646-HTL. Cells were finally resuspended in 20 μl of EMM media and 3 μl of this suspension was placed on the LabTekII imaging chamber, covered by an EMM agarose pad (EMM media + 2% Seakem Agarose (Lonza, Cat no. 50004), size: 8 × 8 mm))[15], and imaged under a Leica DMi8 infinity TIRF microscope using Highly Inclined and Laminated Optical Sheet (HILO) illumination. From each agarose pad in the LabTekII imaging chamber, cells were imaged for 60 min.

For fixing the cells with formaldehyde, after 30 min incubation with JF646-HTL, formaldehyde was added at 4% concentration to the sample and kept at room temperature for 10 min. Cells were washed twice with 1 ml of EMM media and taken for imaging as before.

Single-molecule imaging. Single-molecule imaging was performed using a Leica (Model: DMi8 automated) infinity TIRF microscope, equipped with a 100X/1.47 NA oil immersion TIRF objective lens (Leica, article no. 11506318), Prime 95B (Photometrics) sCMOS camera (camera pixel size: 11 × 11 μm, image pixel size: 110 nm/pixel). We used HILO (Highly Inclined and Laminated Optical) sheet illumination for single-molecule imaging. The HILO illumination was achieved on a TIRF microscope by adjusting the angle of the incident laser beam (~56° to 59°). The HILO illumination can achieve an imaging depth of up to ~10 μm, which makes the molecules in the nucleus visible. We used a 150 mW 638 nm laser for illumination and a filter cube with an excitation filter (640/20 nm), dichroic mirror (658 nm), and emission filter (705/90 nm). The time-lapse movies were acquired (in a single focal plane of 256 × 256 pixels) with two imaging regimes: (1) fast imaging regime (10 ms exposure, 5 ms camera processing time, 1000 frames, 100% laser power (0.25 mW at the objective under HILO illumination)) and (2) slow imaging regime (50 ms exposure, 150 ms time interval (includes 5 ms camera processing time), 200 frames, 30% laser power (0.1 mW at the objective under HILO illumination)). Refer[13,18] for detailed methodology. For imaging Hht1-HaloTag, we focussed the cells using bright field illumination and the region of interests (ROIs) were made against nuclei for tracking from the maximum intensity projection (MIP) images of the time-lapse movies. For imaging NLS-HaloTag-GFP, we used the FITC channel for focussing the cells and making nuclear ROIs.

Single-molecule tracking and data analysis. The movies acquired through a fast-imaging regime were used to calculate the diffusion parameters, such as diffusion coefficient (D), mean square displacement (MSD), fraction of bound/unbound(free) molecules (F_{bound}/F_{free}), and jump angles. We used a MATLAB-based TrackIt[14] which determines the precise position of single molecules by Gaussian intensity fitting, assembles particle trajectories over multiple frames, and provides features for data analysis and visualization. The raw time-lapse movies were loaded as.tif files. The following parameters were used for TrackIt analysis: spot detection parameters (threshold factor 0.7–1, frame range 1-inf), tracking parameters (tracking algorithm nearest neighbour; tracking radius: 6 pixels; minimum track length: 4 frames; gap frames: 4 frames; minimum track length before gap frame: 2 frames). Next, we used in-house code 'SPTAnalyzer' (details are in the next section) to classify further subpopulation (2 or 3) based on the logD histogram data, D being the diffusion coefficient. The.csv file generated from TrackIt was used with SPTanalyzer. Due to the strict data filtration based on trajectory length selection and linear fit of MSD-time data, this method is able classify the population in either two or three state kinetic models, but this also leads to underestimation of the bound fraction or diffusivity as it rejects significant number of populations. Having an idea of the resolvable kinetic modelling of the trajectories, we then used the SpotOn[15] web interface to have quantitative measurements of F_{bound} and D_{bound}. The.csv file generated through TrackIt was loaded for Spot-On analysis. The estimation by the Spot-On analysis is considered more robust because it considers the number of biases and uncertainties[15] (e.g tracking error, motion blurring, localization error, z-correction that take place during single-molecule tracking experiments.) For detailed information on Spot-On variables, please refer to the official documentation at https://spoton.berkeley.edu/SPTGUI/docs/latest. The following settings were used on the Spot-On web interface: bin width: 0.01 μm; number of time points: 6; jumps to consider: 4 pixels; use entire trajectory: No; max jump: 1.6 μm. For model fitting (two-state), the following parameters were selected: D_{bound} (μm²/s): [min, max] = [0.0005, 0.1]; D_{free} (μm²/s): [min, max] = [0.1, 5]; F_{bound}: [min, max] = [0, 1]; Localization error(μm): Fit from data: 'Yes', [min, max] = [0.01, 0.1]; use Z correction: 'No'; Model Fit: 'CDF'; Iterations: '3'. For free NLS-HaloTag-GFP, the fitting was done up to 4 time points keeping the localization error of 0.036 μm. We also used SPTAnalyzer to estimate the anomalous diffusion exponent (α) by linear fitting log(MSD) and log(time). α characterises the type of diffusion, where $\alpha < 1$ indicates subdiffusive behaviour.

For the quantification of the residence time, we used movies acquired with a slow imaging regime and tracked them using MatlabTrack[13,16]. The raw movies (.tif files) were loaded into the MatlabTrack. The following parameters were used for filtering and tracking: filtering (filter type: bandpass, lower limit: 1-pixel, upper limit: 5 pixels); find peaks: (threshold 2.5–4.5, window size 7, fit PSF to the particle: Yes); particle tracking (method: nearest neighbour; maximum jumps: 6 pixels; shortest track: 4 frames; gaps to close: 4 frames). The survival probability is defined as the probability that an event has survived past a time t. To estimate this, we have calculated the cumulative distribution function CDF(t) of dwell time that quantifies the probability of an event occurring before time t. Thus, the complement of CDF (i.e. 1-CDF) gives the survival probability. We used the following parameters to define any molecule as a bound molecule: maximum frame-to-frame jump (R_{min}): 0.54 μm, maximum end-to-end jump (R_{max}): 0.81 μm, minimum bound frames (N_{min}): 8 frames. These parameters were derived by tracking chromatin-bound histone Hht1-HaloTag in *S. pombe* from slow imaging regime movies (Fig. 2b[13]). The survival probability (1-CDF) of dwell times was fitted with the double-exponential decay function which estimates the residence time and fraction of slow- and fast-moving particles (specifically bound

and non-specifically bound molecules, respectively). This software allows us to apply photobleaching correction for residence time estimation and calculation of survival probability. It calculates the photobleaching curve by plotting the number of detected molecules as a function of time. This curve is then fitted by a double exponential, to obtain the photobleaching trend, and the survival distribution of bound molecules is divided by the estimated photobleaching trend for normalizing the photobleaching rate. Refer[13,16] for detailed methodology.

Details of SPTAnalyzer. The data file obtained from TrackIt in.csv format was subjected to analysis using SPTAnalyzer, an in-house MATLAB script. This script, accessible via (https://github.com/KashyapKirti/SPTAnalyzer), offers enhanced control over the dataset and fitting procedures. Notably, it enables data filtration based on user-defined minimum and maximum track lengths.

For every single particle tracking data set, after filtration, it fits the data linearly in the given temporal range (usually the first 5-6 frames). The linear fitting has a $R^2 \geq 0.8$ criterion by default that can be customized by the user. The diffusion coefficients (D) of each particle are then calculated by dividing the slope of the linear fit by 4. From the obtained D values it then calculates the probability distribution function (PDF) on a log_{10} scale, where the user can specify the bin width and normalisation parameter. The user can fit this PDF with either a single or sum of two/three Gaussians which finally gives the value of the diffusion coefficient and fractions for each of the Gaussian fits. The script uses a MATLAB built-in function called 'maximum likelihood estimate (mle)' for the Gaussian fit. Using SPTAnalyzer, the user can add correction due to localization error, if they have single particle tracking data for fixed cells. The localization error (σ) can be calculated from the y-intercept of the MSD vs time plot for the fixed cell data as, $\text{MSD}(t) = 4Dt^\alpha + 4\sigma^2$ [15].

Finally, the script can plot the individual Gaussian as well as the total PDF based on the user-defined plotting customizations.

Protein extraction and western blotting. Log phase culture was grown until OD_{600}: 0.8. Cells were harvested and resuspended with 775 µl autoclaved distilled water, 150 µl of 1.85 M NaOH and 50 µl of β-mercaptoethanol. The cells were resuspended and incubated on ice for 15 minutes. 165 µl of 50% TCA was added and maintained on ice for another 10 minutes. Cells were centrifuged at 16000 rpm, and the supernatant was discarded. Cell pellets were resuspended with 1X Laemmli sample buffer (2% SDS, 10% glycerol, 5% 2-mercaptoethanol, 0.002% bromophenol blue, and 60 mM Tris-HCl (pH 6.8)). 20–30 ul of Tris-Cl (pH 8) was added to turn solution alkaline (blue) and incubated at 99° C for 5 minutes. Then, lysates were centrifuged at 12000 rpm for 1 minute. Protein extracts were resolved on SDS–polyacrylamide gel, blotted onto PVDF membranes, and incubated with the following primary and secondary antibodies.

Antibody dilutions: Anti-Halo rabbit polyclonal antibody (Promega, Cat no. G928A, dilution: 1:2000), Anti-GAPDH mouse monoclonal antibody (Abcam, Cat no. Ab125247, dilution: 1:5000), HRP conjugated Goat anti-rabbit antibody (Jackson Immunoresearch, Cat no. 111-035-003, dilution: 1:10000), HRP conjugated Goat anti-mouse antibody (Jackson Immunoresearch, Cat no. 115-035-003, dilution: 1:10000).

Data Records

The dataset is available from BioImage Archive (BIA)[19]. The data is divided into folders and subfolders according to the software used for tracking and imaging regimes, as detailed in Supplementary Table 4.

File formats: The raw time-lapse movies of single-molecule imaging are saved as.tif, which can be opened using open-source freeware ImageJ/Fiji[20]. The tracking files are saved as.mat files that can be opened using MATLAB (MatlabTrack or TrackIt). The X- and Y- coordinates of the trajectories are saved as a.csv file, which can be opened using Microsoft Excel.

Folder naming: Folders are named based on the software used for tracking and data analysis.

The folder 'Histone_Live Cells' contains three subfolders: 'Hht1-HaloTag_slow imaging_(MatlabTrack)', 'Hht1-HaloTag_slow imaging_(TrackIt)' and 'Hht1-HaloTag_fast imaging_(TrackIt)'. The subfolder 'Hht1-HaloTag_slow imaging_(MatlabTrack)' contains all the individual tracking files (27.mat files) generated through MatlabTrack. These tracking files contain the trajectory information (X, Y coordinates of molecules in a particular trajectory). This folder also contains a merged file of all 27 tracking files for residence time analysis. This merged file contains survival probability distribution, pie chart, fit parameters, photobleaching correction, track statistics, total bound fraction C_{eq}, and k_{off}. The subfolder 'Hht1-HaloTag_slow imaging_(TrackIt)' contains all the raw time-lapse movies (27.tif files) acquired with slow imaging regime, their respective MIP images (27.tif files), and respective tracking files (27.mat files) generated through TrackIt. It also contains a merged file (.mat) of all 27 tracking files for plotting various graphs through TrackIt. The subfolder 'Hht1-HaloTag_fast imaging_(TrackIt)' contains all the raw time-lapse movies (27.tif files) acquired with fast imaging regime, and their respective MIP images (27.tif files), and the respective tracking files (27.mat files) generated through TrackIt. It also contains a merged file (.mat) of all 27 tracking files for plotting various graphs through TrackIt. It also contains a single.csv file generated through TrackIt for SPTAnalyzer and Spot-On analysis. This.csv file contains x- and y- coordinates of the trajectories.

The folder 'Histone_Fixed Cells' contains two subfolders: 'Hht1-HaloTag_fixed cells_slow imaging_(MatlabTrack)' and 'Hht1-Halotag_fixed cells_fast imaging_(TrackIt)'. The subfolder 'Hht1-Halotag_fixed cells_slow imaging_(MatlabTrack)' contains all the individual tracking files (33.mat files) generated through MatlabTrack by tracking 33 time-lapse movies and a merged file of all 33 tracking files for residence time analysis. The subfolder 'Hht1-Halotag_fixed cells_fast imaging_(TrackIt)' contains all the raw time-lapse movies (28.tif files) acquired with fast imaging regime, their respective MIP images (28.tif files) and their respective tracking files (28.mat files) generated through TrackIt. It also contains a merged file of all 28 tracking files and a single.csv file generated through TrackIt for SPTAnalyzer and Spot-On analysis.

The folder 'GFP_Live Cells' contains two subfolders: 'NLS-HaloTag-GFP_slow imaging_(MatlabTrack)' and 'NLS-HaloTag-GFP_fast imaging_(TrackIt)'. The subfolder 'NLS-HaloTag-GFP_slow imaging_(MatlabTrack)' contains all the individual tracking files (43.mat files) generated through MatlabTrack by tracking 43 time-lapse movies and a merged file of all 43 tracking files for residence time analysis. The subfolder 'NLS-HaloTag-GFP_fast imaging_(TrackIt)' contains all the raw time-lapse movies (40.tif files) acquired with fast imaging regime, their respective MIP images (40.tif files) and their respective tracking files (40.mat files) generated through TrackIt. It also contains a merged file of all 40 tracking files and a single.csv file generated through TrackIt for SPTAnalyzer and Spot-On analysis.

The folder 'Optimization of JF646-HTL concentration' contains 4 subfolders: '0 nM JF646-HTL', '1 nM JF646-HTL', '5 nM JF646-HTL', and '10 nM JF646-HTL'. These four subfolders contain a raw time-lapse movie (.tif files) acquired by labeling Hht1-HaloTag with four different concentrations of JF646-HTL (0 nM, 1 nM, 5 nM, 10 nM) respectively.

The folder 'Comparison of filtered images with raw images' contains 3 raw time-lapse movies (.tif files) of Hht1-HaloTag-JF646-HTL in live cells acquired with slow imaging regime and their respective filtered movies (.tif files) obtained by MatlabTrack using bandpass filter.

Technical Validation

Genetic engineering of yeast strains for single-molecule imaging. The hygromycin-resistant transformants were verified for *hht1*-specific integration by colony PCR. The integration on the left was verified using primers MO275 and MO276 which gave an amplicon of 1.32 kb, whereas on the right was done by using primers MO8 and MO277 which gave an amplicon of 950 bp upon correct integration (Supplementary Fig. 1). The *S. pombe* strain confirmed to contain *hht1::hht1-HaloTag>>HygR* in a wild type background was named MP792.

To further check if the fusion protein Hht1-HaloTag is functional and the cell viability is not compromised, we performed a spot assay (Supplementary Fig. 3). 10-fold-serial dilutions of log phase cells with O.D.$_{600}$ 0.6 of MP1 and MP792 were made and 10 μl were spotted on YEA supplemented media. The cells were incubated at 30°C for 3-4 days. We couldn't observe any significant growth retardation in the *hht1-HaloTag* strain as compared to the wild type, suggesting that the fusion protein was largely functional and the cells were healthy, since *hht1* is an essential gene.

To further confirm the expression of the Hht1-HaloTag protein, we performed western blotting using an anti-HaloTag antibody (Supplementary Fig. 4). We could detect the presence of Hht1-HaloTag protein at the expected molecular weight of 48 kDa, suggesting a successful in-frame c-terminal fusion of HaloTag with Hht1. As a loading control, the expression level of GAPDH was probed using an anti-GAPDH antibody.

To confirm the presence of the NLS-HaloTag insert in ME48 by PCR, primers MO198 and MO276 were used which gave an amplification of 655 bp (Supplementary Fig. 2, Lane 4). To confirm the presence of the plasmid in the transformants, colony PCRs were done using MO198 and MO276, and two negative control strains MP789 and MP561 were also used (Supplementary Fig. 2, Lanes 5,6).

Validating and benchmarking the in-house code. To validate our in-house MATLAB script, we have reproduced the log(D) histogram plots provided by Nguyen *et al.*[9,10] and compared the PDF (on a log_{10} scale) to match the D$_{bound}$ values. We downloaded the data sets[10] for wild type H2B, Med14, and SUA7 and analyzed them with the code written. The D$_{bound}$ values and PDF match well with the data reported in the paper[9] (Supplementary Fig. 5). After validating the script, we benchmarked the values of bin width, minimum track length, and fitting range, we have used uniform values for different proteins throughout the paper.

From single particle fast-tracking data (frame rate 15 ms), we have discarded any track that contains <6 displacements. From that filtered data to calculate diffusion coefficient D, we have linearly fitted the data in the 30 ms to 75 ms range. The slope of the linear fitting that has $R^2 \geq 0.8$ criterion divided by 4 was defined as the diffusion coefficient (D) for that track. We have fixed the bin-validation of SPTAnalyzer using published dataset[9,10]. Datasets were downloaded for H2B, Med14, and Sau7 and analyzed using the SPTAnalyzer. We have fixed a bin width to a value of 0.02 μm throughout the analysis to maintain consistency. From the fitting, we get an estimate of D$_{bound}$ and D$_{free}$ for the sum of two Gaussian fits (for three Gaussian, we have an intermediate D$_{intermediate}$) and it also gives the fractions of bound and free molecules (F$_{bound}$ and F$_{free}$). After separating the population into two distinct populations, bound and free, based on D values, we did the power law fit ($MSD \propto t^\alpha$) for each subpopulation within the 30 ms to 75 ms range.

Optimization of the concentration of JF646-HTL for sparse labeling of Hht1-HaloTag. For sparse labeling of the Hht1-HaloTag, the log phase cells of the strain (MP792) were incubated with different concentrations of JF646-HTL (0 nM, 1 nM, 5 nM, and 10 nM) for 30 min. Cells were washed twice with the EMM media and observed under the single-molecule imaging microscope under the HILO illumination mode. The optimum concentration of JF646-HTL was determined by counting the percentage of cells that show a countable number of single molecules (1 to 5 molecules) per nucleus (Fig. 4, blue arrowheads). We found that a 5 nM concentration of JF646-HTL is optimum as it shows ~66% cells with a countable number of single molecules in the nucleus (Fig. 4).

Usage Notes

In this dataset, we have used two different software for tracking single-molecules: (1) MatlabTrack, (2) TrackIt. Both these software packages are user friendly, however, they provide different options for quantifying various parameters and data visualization. MatlabTrack is best suited for residence time analysis, as it offers correction for photobleaching. TrackIt offers a variety of data visualization options for individual tracks and all the tracks together (batch analysis). TrackIt also offers jump angle analysis.

Labeling Hht1-HaloTag with different concentrations of JF646-HTL
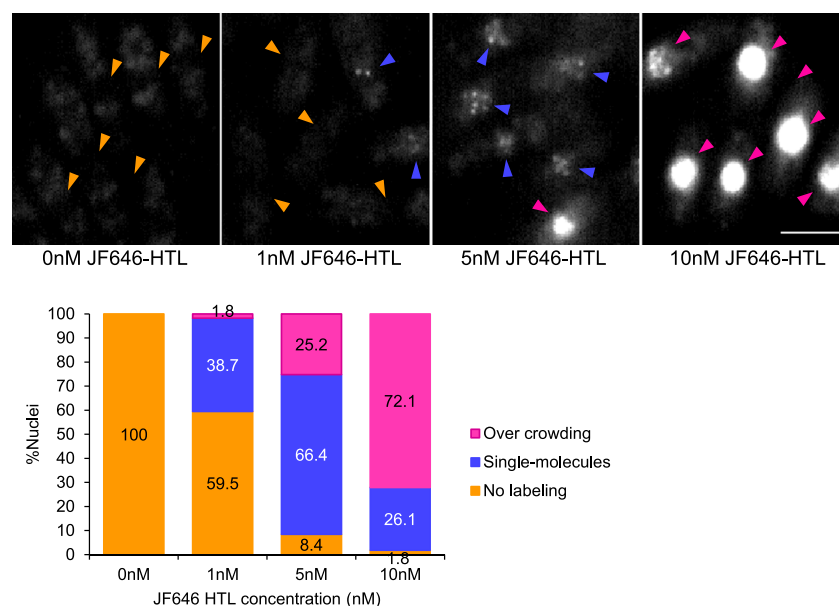

Fig. 4 Optimization of the JF646-HTL concentration for visualization of single-molecules of Hht1-HaloTag. Log phase cells of MP792 were treated for 30 min with four different concentrations of JF646-HTL (0 nM, 1 nM, 5 nM, 10 nM). Cells were washed twice with EMM media to remove excess dye and imaged using HILO illumination mode. Cells were classified according to the labeling density: No labeling (orange arrowheads), single-molecule labeling (blue arrowheads), and overcrowding (pink arrowheads). The counting data is presented by a bar graph. Scale: 5 μm.

Code availability

The following versions of the software/web tools were used:

    LAS_X_3.8.1
    MATLAB R2020a
    MatlabTrack_v6.0
    TrackIt_v1_5_1
    Spot-On_v0.11
    The script for the D plot can be accessed from GitHub (https://github.com/KashyapKirti/SPTAnalyzer).

References

1. Podh, N. K. *et al. In-vivo* Single-Molecule Imaging in Yeast: Applications and Challenges. *Journal of Molecular Biology* **433**, 167250 (2021).
2. Schmidt, J. C., Zaug, A. J. & Cech, T. R. Live Cell Imaging Reveals the Dynamics of Telomerase Recruitment to Telomeres. *Cell* **166**, 1188–1197.e9 (2016).
3. Teves, S. S. *et al*. A stable mode of bookmarking by TBP recruits RNA polymerase II to mitotic chromosomes. https://doi.org/10.7554/eLife.35621.001.
4. Ranjan, A. *et al*. Live-cell single particle imaging reveals the role of RNA polymerase II in histone H2A.Z eviction. *Elife* **9** (2020).
5. Kapadia, N. *et al*. Processive Activity of Replicative DNA Polymerases in the Replisome of Live Eukaryotic Cells. *Mol Cell* **80**, 114–126.e8 (2020).
6. Izeddin, I. *et al*. Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *Elife* **2014** (2014).
7. Presman, D. M. *et al*. DNA binding triggers tetramerization of the glucocorticoid receptor in live cells. *Proc Natl Acad Sci USA* **113**, 8236–8241 (2016).
8. Morisaki, T., Müller, W. G., Golob, N., Mazza, D. & McNally, J. G. Single-molecule analysis of transcription factor binding at transcription sites in live cells. *Nat Commun* **5** (2014).
9. Nguyen, V. Q. *et al*. Spatiotemporal coordination of transcription preinitiation complex assembly in live cells. *Mol Cell* **81**, 3560–3575.e6 (2021).
10. Nguyen, V. Q. Spatiotemporal coordination of transcription preinitiation complex assembly in live cells. Mendeley Data, V1, https://doi.org/10.17632/3xktk72wbd.1 (2021).
11. Kim, J. M. *et al*. Single-molecule imaging of chromatin remodelers reveals role of atpase in promoting fast kinetics of target search and dissociation from chromatin. *Elife* **10** (2021).
12. Price, R. M. *et al*. Heat shock transcription factors demonstrate a distinct mode of interaction with mitotic chromosomes. *Nucleic Acids Res* **51**, 5040–5055 (2023).
13. Podh, N. K., Das, A., Dey, P., Paliwal, S. & Mehta, G. Single-molecule tracking for studying protein dynamics and target-search mechanism in live cells of S. cerevisiae. *STAR Protoc* **3** (2022).

14. Kuhn, T., Hettich, J., Davtyan, R. & Gebhardt, J. C. M. Single molecule tracking and analysis framework including theory-predicted parameter settings. *Sci Rep* **11** (2021).
15. Hansen, A. S. *et al.* Robust model-based analysis of single-particle tracking experiments with Spot-On. *eLife* **7**, e33125 (2018).
16. Mazza, D., Ganguly, S. & McNally, J. G. Monitoring dynamic binding of chromatin proteins *in vivo* by single-molecule tracking. *Methods in Molecular Biology* **1042**, 117–137 (2013).
17. Bahler, J. R., *et al.* Heterologous Modules for Efficient and Versatile PCR-Based Gene Targeting in Schizosaccharomyces Pombe. *Yeast* **14** (1998).
18. Podh, N. K., Das, A., Dey, P., Paliwal, S. & Mehta, G. Single-molecule tracking dataset of histone H3 (Hht1) in Saccharomyces cerevisiae. *Data in brief* **47** (2023).
19. Kumari, A., Podh, N.K., Sen, S., Nambiar, M. & Mehta, G. BioStudies. https://doi.org/10.6019/S-BIAD1137 (2024).
20. Schindelin, J., *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9** (2012).

Acknowledgements

Author contributions

Akriti Kumari: Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing-Original Draft, Writing-Review and Editing, Visualization. Nitesh Kumar Podh: Methodology, Software, Writing-Review and Editing. Sucharita Sen: Methodology, Investigation, Writing-Review and Editing. Kirti Kashyap: Formal Analysis, Software, Writing-Review and Editing. Sahil Islam: Formal Analysis, Software, Writing-Review and Editing. Anupam Gupta: Formal Analysis, Software, Resources, Writing-Review and Editing. Eerappa Rajakumara: Resources, Writing-Review and Editing, Project Administration, Funding Acquisition. Mridula Nambiar: Conceptualization, Resources, Writing-Original Draft, Writing-Review and Editing, Supervision, Project Administration, Funding Acquisition. Gunjan Mehta: Conceptualization, Resources, Writing-Original Draft, Writing-Review and Editing, Visualization, Supervision, Project Administration, Funding Acquisition.

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/s41597-024-04258-0.

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

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.