nature methods

ARTICLES

https://doi.org/10.1038/s41592-021-01154-y

Check for updates

A pairwise distance distribution correction (DDC) algorithm to eliminate blinking-caused artifacts in SMLM

Christopher H. Bohrer[1,2], Xinxing Yang[1], Shreyasi Thakur[3], Xiaoli Weng[1], Brian Tenner[4], Ryan McQuillen[1], Brian Ross[4], Matthew Wooten[5], Xin Chen[5], Jin Zhang[4], Elijah Roberts[2], Melike Lakadamyali[3] and Jie Xiao[1]✉

Single-molecule localization microscopy (SMLM) relies on the blinking behavior of a fluorophore, which is the stochastic switching between fluorescent and dark states. Blinking creates multiple localizations belonging to the same fluorophore, confounding quantitative analyses and interpretations. Here we present a method, termed distance distribution correction (DDC), to eliminate blinking-caused repeat localizations without any additional calibrations. The approach relies on obtaining the true pairwise distance distribution of different fluorophores naturally from the imaging sequence by using distances between localizations separated by a time much longer than the average fluorescence survival time. We show that, using the true pairwise distribution, we can define and maximize the likelihood, obtaining a set of localizations void of blinking artifacts. DDC results in drastic improvements in obtaining the closest estimate of the true spatial organization and number of fluorescent emitters in a wide range of applications, enabling accurate reconstruction and quantification of SMLM images.

The development of super-resolution fluorescence microscopy has enabled the probing of macromolecular assemblies in cells with nanometer resolution. Among different super-resolution imaging techniques, SMLM has gained wide popularity due to its relatively simple implementation.

SMLM reconstructs a super-resolution image by stochastic photoactivation and subsequent post-imaging localization of single fluorophores[1–3]. A successful SMLM experiment requires the ability to localize and temporally separate individual fluorophores and obtain a sufficient number of localizations to meet the Nyquist criterion for spatial resolution[4]. Due to its nature allowing single-molecule detection, one major advantage of SMLM is the ability to determine the number of molecules in a macromolecular assembly quantitatively, allowing investigation of both the molecular composition and spatial arrangement at a level unmatched by other ensemble imaging-based super-resolution imaging techniques. SMLM has led to new discoveries and quantitative characterizations of numerous biological assemblies[5,6], such as those composed of RNA polymerase[7–9], membrane proteins[10–13], bacterial divisome proteins[14–17], synaptic proteins[18,19], the cytoskeleton[20], DNA-binding proteins[21,22], chromosomal DNA[23], viral proteins[24] and more.

One critical aspect in realizing the full quantitative potential of SMLM relies on the careful handling of the blinking behavior of fluorophores. A photoswitchable fluorophore can switch multiple times between activated and dark states before it is permanently photobleached, leading to 'repeat localizations' from the same molecule. These repeat localizations are often misidentified as multiple molecules, leading to the formation of false nanoclusters and errors in quantifying numbers of molecules (Fig. 1a)[25–30].

Multiple methods have been developed to correct for blinking-caused artifacts in SMLM (see Supplementary Table 1 for an overview of these methods). These methods can be coarsely divided into two categories depending on whether a method provides a corrected image void of repeat localizations or a statistical analysis summarizing the properties of the image at the ensemble level. Methods in the first category commonly use a variety of threshold values in time and space to group localizations that likely come from the same molecule[1,2,26,28,30]. The advantage of using thresholds is that it results in a corrected image, allowing one to observe the spatial distribution of fluorophores in cells and apply other quantitative analyses as needed. The disadvantage is that a constant threshold value is often insufficient in capturing the stochastic nature of fluorophore blinking and heterogeneous molecular assemblies. Furthermore, calibration experiments and/or a priori knowledge of the fluorophore's photokinetic properties are often needed to determine the appropriate threshold values[28,30–32]. Statistical analyses such as maximum likelihood or Bayesian approaches have been developed to take into account the stochastic behavior of blinking but have yet to produce corrected super-resolution images void of repeat localizations[33–35]. Additionally, many of these approaches are dependent on specific photokinetic models for the fluorophore, which can be complex and difficult to determine[31,32,36–39].

Methods in the second category use statistical methods to characterize the organization of molecules in uncorrected SMLM images at the ensemble level. Pair- or autocorrelation-based analyses have been used extensively[29,40], in which the shape of the correlation function can be fit to a model to extract quantitative parameters with specific assumptions[4]. New methods using experimentally varied labeling densities[27] and post-imaging computational analysis[41]

[1]Department of Biophysics and Biophysical Chemistry, Johns Hopkins School of Medicine, Baltimore, MD, USA. [2]Department of Biophysics, Johns Hopkins University, Baltimore, MD, USA. [3]Department of Physiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [4]Department of Pharmacology, University of California San Diego, San Diego, CA, USA. [5]Department of Biology, Johns Hopkins University, Baltimore, MD, USA. ✉e-mail: xiao@jhmi.edu

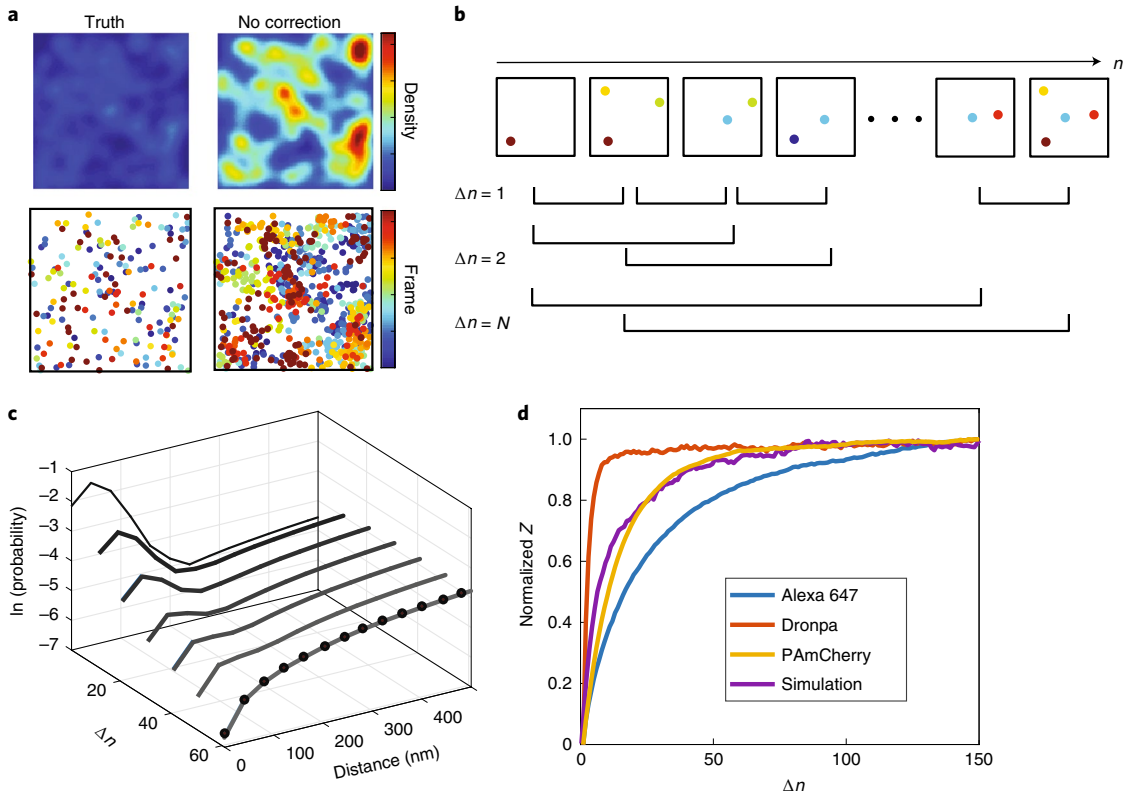


Fig. 1 | a, Simulated SMLM super-resolution images (top) of randomly distributed molecules without repeats (truth) and with repeats (no correction). The corresponding scatterplots (colored through time) are displayed in the bottom panel. **b**, Schematics of how the pairwise distance distributions at different frame differences (Δn) were calculated. **c**, Pairwise distance distributions at different Δn values (black to gray curves) converge to the true pairwise distribution (black dots) when Δn is large. **d**, Normalized Z values measured for three commonly used fluorophores and a simulated fluorophore (randomly distributed, Extended Data Fig. 1a). All Z values reach plateaus at large Δn values, indicating that, at large Δn values, the pairwise distance distributions converge to a steady state. The normalized Z values were calculated by taking the difference between the cumulative pairwise distance distribution at a Δn value and that at $\Delta n = 1$ ($Z(\Delta n) = \sum|cdf(P_d(\Delta r|\Delta n)) - cdf(P_d(\Delta r|\Delta n = 1))|$).

are powerful in determining whether a protein forms clusters or not, but they fail to provide a corrected image, which limits their use in analyzing heterogeneously distributed molecular assemblies and their spatial arrangement in cells.

Here we present an algorithm, termed Distance Distribution Correction (DDC), which is applied after the successful acquisition of an SMLM imaging stream and enables robust reconstruction and quantification of SMLM images with a near-complete reduction of blinking artifacts and without setting empirical thresholds or performing calibration experiments. We first validate our approach using a diverse set of simulated and experimental data and compare DDC to other existing methods. In each situation, DDC outperformed the existing methods in obtaining the closest representation of the 'true' image and in determining the number of fluorophores. We then applied DDC to experimentally collected SMLM images of membrane scaffolding proteins[42–44], dynein oligomers[45] and sister chromatin fibers[46]. Under all the conditions, DDC provided SMLM images substantially different from those obtained from other correction methods, allowing differential identification of membrane protein cluster properties, characterization of dynein assembly states and quantification of DNA content between sister chromatin fibers. These results demonstrate the broad application of DDC for SMLM imaging. Finally, we discuss critical considerations of how to apply DDC.

Results

Principle of DDC. DDC is based on the principle that the pairwise distance (Δr) distribution $P_d(\Delta r|\Delta n)$ of the localizations separated by a frame difference (Δn) much larger than the average number of frames that a molecule's fluorescence lasts (N), approximates the true pairwise distance distribution $P_T(\Delta r)$ (Supplementary Information). Note that N does not need to be precisely determined as long as it is in the regime where $P_d(\Delta r|\Delta n)$ approaches a steady state. This principle is also independent of the particular photokinetic property of the fluorophore (Fig. 1 and Supplementary Information). One intuitive way to understand this principle is that, if one collects an imaging stream that is long enough such that all the localizations in the first and last frames of the stream come from distinct sets of fluorophores, the pairwise distance distribution between the localizations of the two frames will then be devoid of repeat localizations and will reflect the true pairwise distance distribution $P_T(\Delta r)$. A mathematical justification of this principle is provided in the Supplementary Information with an in-depth discussion and illustration (Supplementary Fig. 1). Because of this principle, DDC is most applicable when a substantial proportion of the fluorophores bleach during the acquisition of an imaging stream, that is, $P_d(\Delta r|\Delta n)$ reaches a well-defined steady state when $\Delta n = N$ (Fig. 1 and Supplementary Figs. 2 and 3). Note that this condition is typically and easily met in direct stochastic optical reconstruction microscopy (dSTORM) and photoactivated localization microscopy (PALM) experiments when the acquisition time is sufficiently long.

To demonstrate the principle of DDC, we used simulated SMLM images of randomly distributed fluorophores that followed the photokinetic model shown in Extended Data Fig. 1b. One representative

super-resolution image and the corresponding scatterplot, colored through time, with and without repeat localizations, are shown in Fig. 1a. Using uncorrected images, we computed the pairwise distance distributions at all frame differences Δn (Fig. 1b). As shown in Fig. 1c and Extended Data Fig. 2, at small Δn values, there are large peaks at short distances, indicating that there were repeat localizations from the same fluorophores closely spaced in time and space. When Δn is large, the pairwise distance distributions approach a steady state converging upon the true pairwise distance distribution (Fig. 1c, dotted curve). This behavior supports the principle that, when Δn is sufficiently large, the pairwise distance distribution represents the true pairwise distance distribution. Using simulations, we also show that the pairwise distance distributions converge upon the true distributions at large Δn values, irrespective of the underlying photokinetics or molecular spatial distributions (Extended Data Fig. 2 and Supplementary Information).

Next, we used experimentally obtained SMLM images of three molecular assemblies labeled with different fluorophores in *Escherichia coli* cells: the bacterial transcription elongation factor NusA fused with the reversibly switching green fluorescent protein Dronpa[47], *E. coli* RNA polymerase fused with the photoactivatable red fluorescent protein PAmCherry[48] and precursor ribosomal RNA species (pre-rRNA) labeled with organic fluorophore Alexa 647-conjugated DNA probes[49] (Supplementary Fig. 4). We determined the pairwise distance distribution for each fluorophore and calculated the normalized, summed differences of the cumulative distributions for each Δn, relative to that of $\Delta n = 1$ ($Z(\Delta n) = \sum |cdf(P_d(\Delta r|\Delta n)) - cdf(P_d(\Delta r|\Delta n = 1))|$), where *cdf* is the cumulative distribution function. In all cases, the corresponding normalized Z values reach plateaus at large Δn values despite different photokinetics and spatial distributions (Fig. 1d). The rate at which each fluorophore reaches the plateau for the normalized Z value reflects the photokinetics of the fluorophore: the longer that a fluorophore blinks (such as Alexa 647 compared to Dronpa), the longer the time until Z plateaus. These experimental results further confirm the principle of DDC that the pairwise distance distributions converge upon a steady state distribution as Δn increases.

Once determined, $P_T(\Delta r)$ can then be used to calculate the likelihood of having a particular subset of true localizations (Supplementary Figs. 5–9) using the following equation:

$$\mathcal{L}(\{R, T\}|r, n) = \prod_{i,j \in \{T\}} P_T(\Delta r_{i,j}) \times \prod_{i \in \{R\}, j \in \{R,T\}} P_{R1}(\Delta r_{i,j}|\Delta n_{i,j}), \quad (1)$$

where $\{R, T\}$ are sets that contain the indices of the localizations considered repeats $\{R\}$ and the true localizations $\{T\}$, given coordinates r and associated frame numbers n obtained from experiment. The first term on the right of the equation is the probability of observing all distances Δr between every pair of true localizations ($i, j \in \{T\}$). Here the probability distribution $P_T(\Delta r_{i,j})$ is the true pairwise distance distribution. The second term is the probability of observing all distances between pairs of localizations with at least one being a repeat ($i \in \{R\}$ and $j \in \{R, T\}$). The probability distribution $P_{R1}(\Delta r_{i,j}|\Delta n_{i,j})$ gives the probability of observing a distance between a pair of localizations with a frame difference $\Delta n_{i,j}$ if at least one of the localizations is a repeat. This probability distribution can be easily determined once $P_T(\Delta r)$ is known (Supplementary Information). Maximizing the likelihood with respect to $\{R, T\}$ results in a subset of true localizations where the pairwise distance distributions $P_d(\Delta r|\Delta n)$ are equal to $P_T(\Delta r)$ (Extended Data Fig. 3). DDC maximizes the likelihood with respect to the two sets ($\{R, T\}$) using a Markov chain Monte Carlo method[50,51] to reconstruct the corrected image (Supplementary Figs. 8 and 9). In Extended Data Fig. 4 and the Supplementary Information, we provide a toy model to walk readers through the calculation. To validate equation (1), we performed six simulations of distinct spatial distributions with various fluorophore photokinetic models. We show that, when the likelihood reaches its maximum, more than 97% of the final localizations are true localizations (Extended Data Fig. 5).

DDC outperforms existing methods in image reconstruction and quantification. To compare the performance of DDC with those of commonly used blinking-artifact-eliminating methods, we simulated five systems: random distribution (no clustering), small clusters, dense clusters, parallel filamentous structures with low labeling density and intersecting filamentous structures with high labeling density (Fig. 2 and Supplementary Information). In these simulations, the fluorophore had two dark states and followed the photokinetic model shown in Extended Data Fig. 1a. The raw images without any repeat localizations for each simulation are shown in Fig. 2a. We applied DDC, three published thresholding methods (T1 (refs. [1,26]), T2 (ref. [30]) and T3 (ref. [28])) (Supplementary Figs. 8 and 9) and a customized thresholding method (T4, Supplementary Information) to all the images.

Method T1 links together localizations using a time threshold determined by an empirical estimation (similar to that in Betzig et al.)[1,26] (Supplementary Fig. 8). Method T2 uses experimentally quantified photokinetics of the fluorophore to set extreme thresholds so that the possibility of overcounting is extremely low[30]. Method T3 uses the experimentally determined number of repeats per fluorophore to choose thresholds that result in the correct number of localizations within each image[28] (Supplementary Fig. 9). T2 and T3 but not T1 require additional experiments to characterize fluorophore photo properties. Method T4 is an ideal thresholding method that scans all possible thresholds and uses the threshold that results in the least image error for each system (Supplementary Information). T4 cannot be applied in real experiments because the true image is unknown; we included it to illustrate the best scenario of what a thresholding method could achieve.

To quantitatively compare the ability of these methods, we calculated two metrics, the image error and the counting error (Fig. 2b and Supplementary Information). The image error was calculated by first summing the squared difference of each pixel's normalized intensity between the corrected and the true images and then dividing this squared difference by the error between the uncorrected and the true images (Supplementary Information). The image error quantifies the amount of error in determining the distribution of localizations without being penalized for the error in the number of localizations. The counting error was calculated as the difference between the true number of fluorophores and that determined from the corrected image divided by the actual number of fluorophores (Supplementary Information). Additionally, in Extended Data Fig. 6, we further show the degree of overcounting and undercounting of these different methods at the level of individual pixels[52].

As shown in Fig. 2b, DDC outperforms all methods by having the lowest image errors and the lowest (or close-to-lowest) counting errors. Interestingly, even with the best possible thresholds (T4), DDC still outperforms T4 for both metrics. This result suggests that thresholds cannot adequately account for the stochastic nature of blinking. Similar results are shown in Supplementary Fig. 10 for a fluorophore with one dark state (Extended Data Fig. 1b). When counting the number of localizations is the main concern, T3 performs equally to or slightly better than DDC because T3 was applied with an experimental calibration providing the average number of blinks per fluorophore (Fig. 2 and Supplementary Information). Nonetheless, DDC outperforms T3 by having lower image errors across all five simulation systems. In particular, for the dense cluster and the intersecting filament systems, two scenarios commonly encountered in biology, the average image errors of T3 were more than four times those of DDC (Fig. 2b). The advantage of DDC for these two systems highlights the unique superiority of

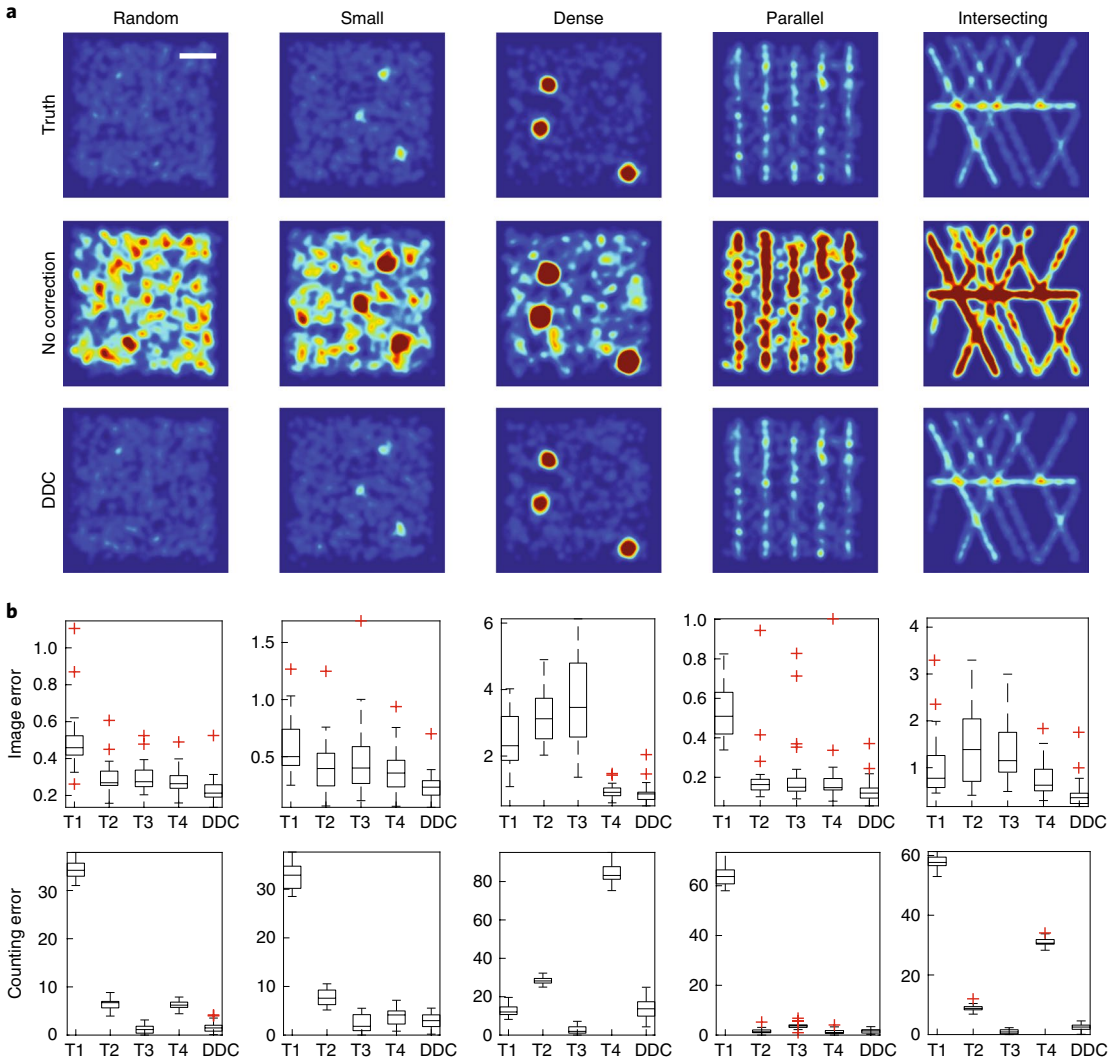


Fig. 2 | Comparison of four different thresholding methods with DDC on five spatial distributions (randomly distributed, small clusters, dense clusters and parallel filaments and intersecting filaments). Each simulation had 1,000 true localizations and was simulated for 30,000 frames (scale bar, 250 nm). **a**, True, uncorrected and DDC-corrected images for each spatial distribution. **b**, Image error and counting error calculated from T1–T4 and DDC for each spatial distribution. Lines are medians, box extends from 25% to 75%, whiskers extend to the most extreme data points not considered outliers, and the red plus signs are outliers beyond 1.5× interquartile range (number of images for each system, 24).

DDC for heterogeneously distributed proteins with uneven densities. These results demonstrate that DDC can be used to obtain the correct number of true localizations while producing the most accurate SMLM images.

Validating DDC using a labeled nuclear pore complex reference cell line. To validate the ability of DDC to produce SMLM images of real biological structures, we used a cell line that was recently developed as a reference standard for SMLM imaging[53]. In this cell line, the endogenous nucleoporin Nup96 of the nuclear pore complex (NPC) is tagged with monomeric enhanced GFP (mEGFP). The known arrangement and composition of the NPC (Fig. 3a) enabled us to quantitatively compare DDC and thresholding methods in their ability to produce images in which the number of localizations per NPC matched the copy number of Nup96 in NPCs.

We performed SMLM on the cell line using an anti-GFP nanobody labeled with Alexa 647. In Fig. 3b, we show a representative raw SMLM image of NPCs (top) and zoomed-in images of two individual NPCs with and without DDC (bottom). The image processed with DDC showed the expected ring-like arrangement with a more even intensity distributed along the circumference of the pore. Using the raw images and a previously published procedure[53], we determined the effective labeling efficacy (ELE) of the nanobody at 0.44 (Supplementary Information). The ELE is the proportion of target protein molecules labeled with a detectible fluorophore, and its calculated value is independent of the overcounting artifacts caused by fluorophore blinking[53]. Using the calculated ELE, we found the number of localizations per fluorophore obtained from DDC centered around unity (1.05 ± 0.3, $\mu \pm$ s.d., $n = 3{,}947$), whereas the number without DDC centered around 4.4 ± 2.5. Consequently, we found the mean number of localizations per NPC to be 33.6 ± 11 for DDC after taking into account the ELE. This number is indistinguishable from the known copy number of 32 for Nup96 within experimental noise (Fig. 3c). Note that the spread of the distributions in Fig. 3c is expected from the stochastic noise in the labeling efficiency of individual NPCs.

Next, we compared the performance of thresholding with three different distance thresholds (40 nm, 60 nm and 80 nm), each with time thresholds ranging from 1 to 5,000 frames (Fig. 3d). We found that all distance thresholds tended to overcount at short time

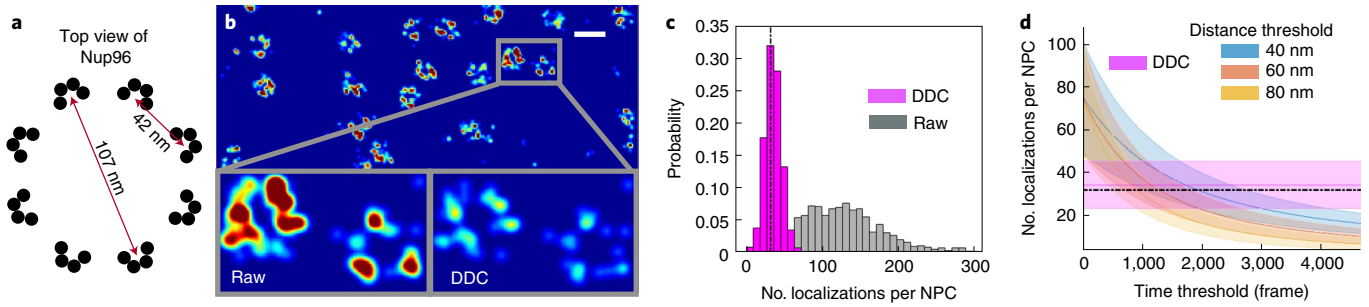


Fig. 3 | a, Schematic drawing of the top-down view arrangement of Nup96 (black dots) in the NPC. Each NPC contains 32 copies of Nup96. **b**, An SMLM image of Nup96–mEGFP labeled with an Alexa 647-tagged nanobody (top) and zoomed-in images of two representative NPCs (bottom) without (raw) or with the application of DDC. Scale bar, 200 nm; frame rate, 100 Hz. **c**, Number of localizations of Nup96 per NPC distribution (adjusted for mean ELE) obtained from DDC-corrected (magenta) or raw (gray) images. The dashed line shows the true copy number of Nup96 (32) per NPC. **d**, Number of localizations of Nup96 per NPC (adjusted for mean ELE) obtained from thresholding at different distances (40, 60 and 80 nm) and times (1–5,000 frames). The solid lines are the mean, and the shaded areas are the associated s.d. The mean and associated s.d. determined from the DDC algorithm were plotted as a magenta-shaded line for comparison.

thresholds and undercount at long time thresholds. Only when the time threshold was within a relatively small window, which varied for each of them, did they produce the correct number of localizations per NPC (adjusted for ELE). In realistic experiments in which the biological structure is unknown, it would be difficult to determine which time and distance threshold to use. In sum, these results validate the application of DDC to SMLM images as the method of choice to obtain the true representation of underlying cellular structures.

DDC identifies differential clustering properties of membrane microdomain proteins AKAP79 and AKAP150. Membrane microdomains have been observed in super-resolution imaging studies and have raised substantial interest in their biological functions[13]. However, concerns remain as to whether the characterizations of these microdomain protein clusters were impacted by blinking-caused artifacts[27]. Here we used DDC to investigate a membrane scaffolding protein, A-kinase anchoring protein (AKAP), which plays an important role in the formation of membrane microdomains[42–44]. The two orthologs AKAP79 (human) and AKAP150 (rodent) were previously shown to form dense membrane clusters, which are likely important for regulating anchored kinase signaling.

We performed SMLM imaging on AKAP150 in murine pancreatic β cells using an anti-AKAP150 antibody and analyzed the resulting SMLM data using DDC (Supplementary Information). For AKAP79, we applied DDC to previously acquired SMLM data from HeLa cells[42]. For comparison, we also applied the T1 method to both scaffolding proteins as in the previous study of AKAP79 (refs. [1,26,42]) (Extended Data Fig. 7 and Supplementary Fig. 11). We found that DDC-corrected images still showed substantial deviations from simulated random distributions, indicating the presence of clustering. However, the degree of clustering was significantly reduced when compared to those in uncorrected and T1-corrected images for both proteins (Fig. 4a). We further confirmed these results at the ensemble level by computationally varying the labeling density using a previously published method (Extended Data Fig. 8 and Supplementary Information)[41].

To quantitatively compare these images, we used a tree-clustering algorithm (Supplementary Information) to group localizations in individual clusters and plotted the corresponding cumulative distributions in Fig. 4b. Interestingly, the cumulative distributions showed that AKAP150 had a higher degree of clustering when compared to AKAP79, with more than 50% of the localizations within clusters containing more than 15 localizations, twice that of AKAP79. These results suggest that the clustering of the AKAP scaffolds are differentially regulated and that the context dependence is likely important in considering the microdomain-specific signaling functions of the clusters. These accurate, quantitative comparisons of cluster properties would be difficult to achieve by other threshold-based methods.

DDC identifies both subcellular locations and oligomeric states of dynein. Previously, using a well-defined DNA origami structure as a calibration standard, SMLM studies showed that dynein, a cytoskeletal motor protein responsible for retrograde transport on microtubules, can exist in monomeric, dimeric and multimeric states in different subcellular locations[45]. This system provides a previously quantified experimental system to investigate how blinking-caused artifacts can influence the assignment of individual assemblies.

We performed SMLM imaging on anti-GFP antibody-labeled HeLa IC74 cells that stably express a GFP-fused dynein intermediate chain (Fig. 5a and Supplementary Information)[45]. We applied the thresholding method (T1) and DDC to the resulting raw images, with zoomed-in sections shown (white box top) in Fig. 5b. We observed that both the threshold method and DDC yielded a lower amount of signal when compared to that from raw localizations (Fig. 5b, white box i), indicating that a substantial number of raw localizations were repeat localizations. Importantly, we also observed that the difference between threshold- and DDC-corrected images was not constant throughout the images (Fig. 5b, last row), suggesting different assignments of multimeric state for individual dynein assemblies between different methodologies.

To investigate further, we assigned oligomeric states to individual assemblies from each methodology such that the fractions of each oligomeric state matched those calibrated in the work of Zanacchi et al.[45] (Supplementary Information). We then compared the assignment of individual assemblies between the methodologies by calculating the probability of assigning the same oligomeric state to the same individual complex using two different methods. In Fig. 5c, we showed that, for single dynein monomers, both the raw and threshold methodologies were in relative agreement with the assignment of DDC (probability > 90%). However, we observed that the higher oligomeric states assigned by both the raw and threshold methods had considerable deviations from those of DDC, resulting in different spatial distribution of oligomeric dynein motors in cells. These results demonstrate the importance of using the correct method to obtain both subcellular locations and the quantitative properties of molecular assemblies.

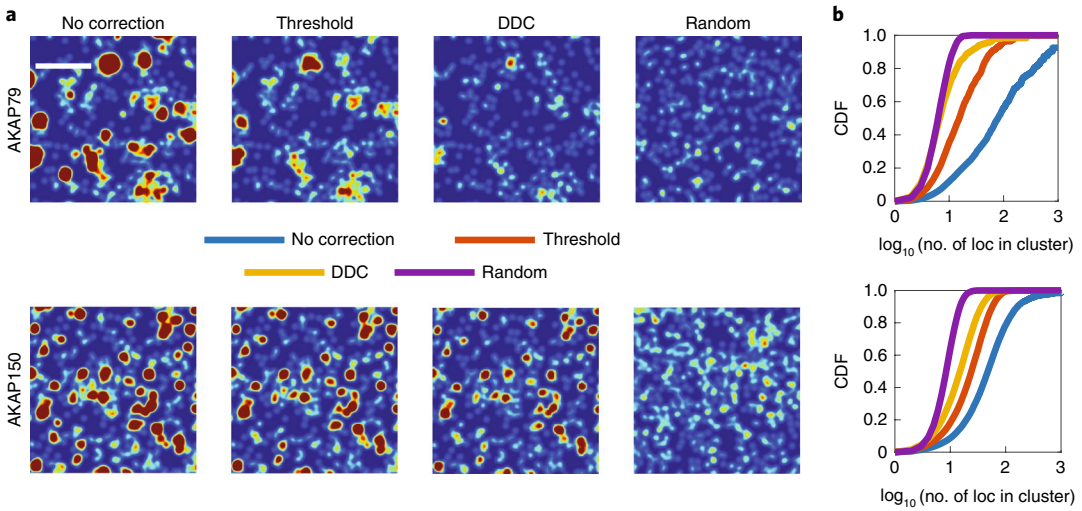


Fig. 4 | Application of DDC to experimentally measured spatial distributions of AKAP79 and AKAP150. a, SMLM images of the two scaffold proteins without correction, corrected using the thresholding methods T1 and DDC and that of a simulated random distribution using the same number of localizations as that for DDC-corrected images (scale bar, 1 µm). **b**, Cumulative distributions for the number of localizations (loc) within each cluster for each protein (number of AKAP79 clusters, 71,685; number of AKAP150 clusters, 141,694). CDF, cumulative distribution function.

DDC minimizes measurement noise in labeled symmetric sister chromatin fibers. DDC can also be applied to minimize noise in the measurement of cellular structural features such as shape and symmetry. To demonstrate an application, we examined the symmetric structure of sister chromatin fibers. Previous studies have shown that, during stem cell differentiation, *Drosophila melanogaster* male germline stem cells undergo asymmetric division to produce a self-renewing stem cell and a differentiating daughter cell[54]. The asymmetric division is likely directed by biased replication fork movement and asymmetric histone incorporation between two sister chromatids[46,55].

To provide a quantitative comparison standard for analyzing DNA and protein contents in sister chromatids, we performed SMLM imaging on YOYO-1-stained chromatin fibers isolated from *D. melanogaster* embryos (Supplementary Information and Fig. 6a). Chromatin fibers isolated from embryonic, non-stem cells should exhibit homogeneous and symmetric labeling on both sisters. We then applied the threshold (T1) and DDC methods to the raw SMLM images (Fig. 6a). In many fibers, we could resolve two parallel sister chromatin fibers; the apparent width of each sister was ~140 nm, similar to that reported in Wang et al.[56] (full width at half maximum), and the separation between sisters was ~200 nm. These characteristics were measured from the projected localizations along the length of fibers.

Next, to determine whether the two sister chromatin fibers have a similar amount of DNA, we quantified the ratio of signal (number of localizations) between the two using segments of different lengths (~1 µm was used in the original work of Wooten et al.[46] (Supplementary Information)). Two sisters having identical replicated DNA content would have a ratio of 1, irrespective of the average length of segment used. As shown in Supplementary Fig. 12, while the ratios of signal between the two sisters for all three methodologies (raw, threshold (T1) and DDC) are approximately centered around 1.0, the spreads in the ratios vary considerably, suggesting that, while repeat localizations may not affect the accuracy of these measurements, they may instead affect the precision.

To investigate further, we calculated the s.e.m. for the different segment lengths (Fig. 6b). We observed that s.e.m. from DDC were consistently the lowest for segments greater than 300 nm. When the segment lengths became too short, the level of variation became indistinguishable between DDC and the thresholding method due to the intrinsic stochastic labeling density in the experiment. Nevertheless, the apparent s.e.m. in raw and threshold-corrected images at length scales of chromatin fibers (300 nm to 1 µm) could mask asymmetries in labeled sister chromatin fibers isolated from germline stem cells (previously quantified with this technique[46]), making it difficult to identify corresponding molecular mechanisms contributing to asymmetry. In summary, this example illustrates how the mishandling of repeat localizations lowers precision and demonstrates the need of DDC for measuring cellular structural features.

Considerations in the application of DDC. In this section, we evaluate the impact of localization density and activation rate on the performance of DDC using simulations. We also demonstrate that the practice of ramping the UV activation power in SMLM imaging should be avoided when applying DDC.

To quantify the influence of localization density on the performance of DDC, we simulated random distributions of fluorophores with densities ranging from 1,000 raw localizations to 15,000 localizations per 1 µm^2. Note that a density greater than 5,000 localizations per µm^2 corresponds to a Nyquist resolution of 30 nm or better. As shown in Extended Data Fig. 9a, the image error increases with localization density and reaches a plateau at ~0.35. We found that the increase in image error at high localization densities was mostly due to the decreased raw image error of the uncorrected images at high localization densities (Extended Data Fig. 10a). The decreasing improvement of DDC at increasing sampling rates suggests that a high sampling rate of the underlying structure reduces the image distortion caused by repeats, although very high labeling densities (>10,000 localizations per µm^2) are usually difficult to achieve for protein assemblies.

Next, to quantify the influence of the activation rate, we varied the activation probability of each simulated fluorophore from 0.025 to 0.15 per frame, with 1,000 fluorophores randomly distributed throughout a 1-µm^2 area. Extended Data Fig. 9b shows that the image error of DDC steadily increases with the activation rate. This increase was because, at high activation rates, the temporal overlaps of individual fluorophores that were spatially close increased, which made it difficult to distinguish the repeat localizations from different fluorophores. This trend holds true for all other

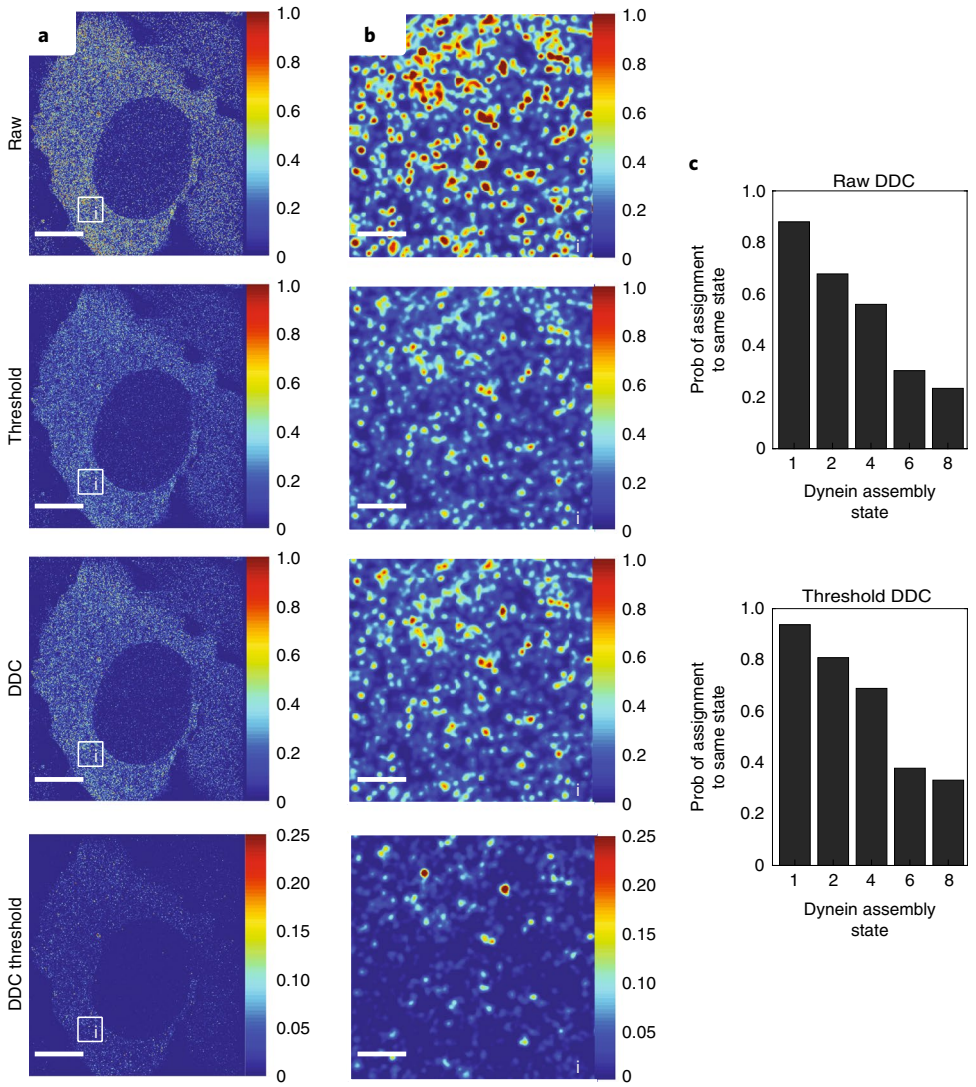


Fig. 5 | Application of DDC to experimentally measured spatial distributions of dynein. a, SMLM images of dynein for a whole cell with all three methods and the difference between the DDC- and threshold-corrected images (10-μm scale bar). **b**, Zoomed-in images showing the raw, threshold (T1)-corrected and DDC-corrected images and 'DDC minus threshold' images (1-μm scale bar). **c**, The probability (prob) of an individual assembly being assigned the same oligomerization state as that assigned with DDC for the raw (top) and threshold (T1, bottom) methodology (number of individual assemblies, 184,368; Supplementary Information). Note that, because a functional dynein motor is homodimeric, we only included even-number complexes and the monomeric state as previously carried out by Zanacchi et al.[45].

blinking-artifact-correction methodologies. Therefore, as with others, DDC obtains the best images when the activation rate is slow.

Finally, we illustrate that one critical requirement for the successful application of DDC, that is, the photokinetics (blinking behavior) of the fluorophore, must be kept constant throughout the acquisition of the SMLM imaging stream. One common practice in SMLM imaging is to ramp the activation power gradually throughout the imaging sequence to speed up the acquisition at later times. However, the activation power may not only change the activation rate of a fluorophore (that is, the probability of a fluorophore being activated per frame), such as that of Dendra[32], but may also change the photokinetics of a fluorophore's blinking behavior (that is, number of blinks, dark time and fluorescence-on time), such as that of mEos2 and PAmCherry[31,32]. If the fluorophore's blinking behavior varied during the acquisition, errors will be introduced into the calculation of its pairwise distance distributions at varying frame numbers, and they may not converge to the true pairwise distance distribution function (equation S1 in the Supplementary Information). Note that this requirement is also needed for all other blinking-artifact-correction methods[1,26,28,30].

To illustrate this critical point, we investigated the blinking behaviors of the photoactivatable fluorescent protein mEos3.2 and the organic fluorophore Alexa 647 at different activation (405-nm) intensities. We quantified three parameters: the number of blinks, off times (T_{off}) and on times (T_{on}) and reported the mean value for each parameter as a function of activation intensity (Extended Data Fig. 9c). We define one blink event as one continuous emission event that could span multiple fluorescence-on frames, the number of blinks as the number of repeated emissions separated by dark frames from the same fluorophore, T_{off} as the time between each blink and T_{on} as the time that the fluorophore remained fluorescent at each blink-on event (Extended Data Fig. 9c). We observed that both fluorophores had a similar dependence of T_{on} on UV intensity, where T_{on} initially increased and then decreased at higher UV

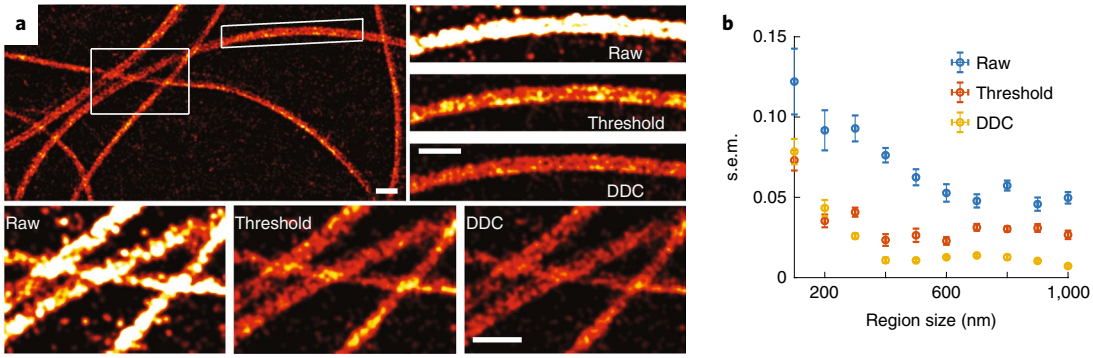


Fig. 6 | a, Representative sister chromatids analyzed with DDC and zoomed-in images showing the resulting images for each of the methodologies (scale bar, 1 μm; samples were prepared and imaged two separate times with similar image results). **b**, s.e.m. (determined from bootstrapping) versus region size for the different methodologies (error bars are s.e.m., determined from bootstrapping; number of 1-μm segments analyzed, 36).

intensities (Extended Data Fig. 9d, top), suggesting that UV light also participates in the fluorescence emission cycle of the fluorophores. Next, we found that T_{off} decreased nonlinearly as the UV intensity increased for both fluorophores (Extended Data Fig. 9d, middle). Finally, we observed that the average number of blinks for Alexa 647 increased dramatically with UV intensity while that of mEos3.2 remained largely constant (Extended Data Fig. 9d, bottom), suggesting a differential influence of UV light in changing the photokinetics of different fluorophores. Thus, varying the activation intensity during the acquisition of an SMLM image can indeed change the blinking characteristics of the fluorophores, which would affect the performance of DDC and all other algorithms that are based on a fluorophore's photokinetics. These results suggest that changing the activation intensity should only be permitted when a quantitative approach is not needed or the proper controls have been performed to show that the fluorophore is insensitive to variations in the activation intensity. We note that, if different activation intensities are needed to account for the labeling density, one could acquire a series of imaging streams; the activation intensity within each stream remains constant but gradually increases through the series. Each imaging stream can be analyzed independently using DDC and then combined together to obtain the final SMLM image.

Discussion

In this work, we provided a blinking-artifact-correction methodology, DDC, that does not depend upon exact thresholds or additional experiments to obtain accurate SMLM super-resolution images. DDC works by determining a 'ground truth' about the underlying organization of fluorophores, the true pairwise distance distribution. It also operates on the prerequisite that a successful SMLM experiment (no substantial spatial temporal overlaps between individual localizations) must be carried out before the application of DDC. DDC can also be applied to previously collected SMLM data, as long as, for each stream of imaging acquisition, the UV activation power stays constant, and the successful SMLM imaging condition is met. We verified, by simulations and experiments, that the true pairwise distance distribution is obtained naturally by taking the distances between localizations separated by a frame difference much longer than the average lifetime of the fluorophore. Using the true pairwise distribution, the likelihood can be calculated, where, upon maximization of the likelihood, one obtains an accurate representation of the true underlying structure. An advantage of DDC is that it does not require any prior knowledge or characterization of the fluorophore's photokinetics, as they can vary greatly depending on specific experimental systems and imaging conditions[53,57–59] (Supplementary Fig. 2), as long as a substantial portion of fluorophores is photobleached during imaging (Supplementary Fig. 3). Here we should note that DDC does not address the issue of limited labeling efficiency, which requires additional experimental considerations[60]. Additionally, DDC only counts the number of emitters, which does not necessarily equal the number of molecules that are labeled using dye-conjugated antibodies[45]. Lastly, DDC can be used to group localizations together to improve resolution and does not affect the localization precision of the individual localizations.

We compared the performance of DDC against those of thresholding methods using simulated data with various spatial distributions and photokinetic models. DDC outperformed these methods, providing the 'best' corrected images as well as excellent estimates of the number of molecules. We validated DDC using a reference standard cell line for which the spatial distribution pattern and copy number of the NPC subunit are known. We then experimentally demonstrated that blinking-caused repeat localizations can lead to artificial clustering of membrane scaffolding proteins, misassignment of dynein oligomeric state and misidentification of DNA content in chromatin fibers.

Finally, we demonstrated that the higher the activation rate and the density of fluorophores, the smaller the relative improvement DDC has. We also showed that, to use DDC, the common practice of ramping UV light should be avoided in certain cases. The complete MATLAB package (Supplementary Information) for DDC is available for download at https://github.com/XiaoLabJHU/DDC. Because of the simplicity and robustness of DDC, we expect that it will become a field standard in SMLM imaging for the most accurate reconstruction and quantification of SMLM images to date.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41592-021-01154-y.

Received: 12 September 2019; Accepted: 12 April 2021;
Published online: 31 May 2021

References

1. Betzig, E. et al. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).
2. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–796 (2006).
3. Hess, S. T., Girirajan, T. P. K. & Mason, M. D. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **91**, 4258–4272 (2006).
4. Coltharp, C., Yang, X. & Xiao, J. Quantitative analysis of single-molecule superresolution images. *Curr. Opin. Struct. Biol.* **28**, 112–121 (2014).

5. Baddeley, D. & Bewersdorf, J. Biological insight from super-resolution microscopy: what we can learn from localization-based images. *Annu. Rev. Biochem.* **87**, 965–989 (2018).
6. Sauer, M. & Heilemann, M. Single-molecule localization microscopy in eukaryotes. *Chem. Rev.* **117**, 7478–7509 (2017).
7. Endesfelder, U. et al. Multiscale spatial organization of RNA polymerase in *Escherichia coli*. *Biophys. J.* **105**, 172–181 (2013).
8. Chen, X. et al. Study of RNA polymerase II clustering inside live-cell nuclei using Bayesian nanoscopy. *ACS Nano* **10**, 2447–2454 (2016).
9. Weng, X. & Xiao, J. Spatial organization of transcription in bacterial cells. *Trends Genet.* **30**, 287–297 (2014).
10. Lillemeier, B. F. et al. TCR and Lat are expressed on separate protein islands on T cell membranes and concatenate during activation. *Nat. Immunol.* **11**, 90–96 (2010).
11. Rossy, J., Owen, D. M., Williamson, D. J., Yang, Z. & Gaus, K. Conformational states of the kinase Lck regulate clustering in early T cell signaling. *Nat. Immunol.* **14**, 82–89 (2013).
12. Ehmann, N. et al. Quantitative super-resolution imaging of Bruchpilot distinguishes active zone states. *Nat. Commun.* **5**, 4650 (2014).
13. Garcia-Parajo, M. F., Cambi, A., Torreno-Pina, J. A., Thompson, N. & Jacobson, K. Nanoclustering as a dominant feature of plasma membrane organization. *J. Cell Sci.* **127**, 4995–5005 (2014).
14. Coltharp, C., Buss, J., Plumer, T. M. & Xiao, J. Defining the rate-limiting processes of bacterial cytokinesis. *Proc. Natl Acad. Sci. USA* **113**, E1044–E1053 (2016).
15. Buss, J. et al. In vivo organization of the FtsZ-ring by ZapA and ZapB revealed by quantitative super-resolution microscopy. *Mol. Microbiol.* **89**, 1099–1120 (2013).
16. Buss, J. et al. A multi-layered protein network stabilizes the *Escherichia coli* FtsZ-ring and modulates constriction dynamics. *PLoS Genet.* **11**, e1005128 (2015).
17. Fu, G. et al. In vivo structure of the *E. coli* FtsZ-ring revealed by photoactivated localization microscopy (PALM). *PLoS ONE* **5**, e12682 (2010).
18. Spühler, I. A., Conley, G. M., Scheffold, F. & Sprecher, S. G. Super resolution imaging of genetically labeled synapses in *Drosophila* brain tissue. *Front. Cell. Neurosci.* **10**, 142 (2016).
19. Bar-On, D. et al. Super-resolution imaging reveals the internal architecture of nano-sized syntaxin clusters. *J. Biol. Chem.* **287**, 27158–27167 (2012).
20. Xu, K., Zhong, G. & Zhuang, X. Actin, spectrin, and associated proteins form a periodic cytoskeletal structure in axons. *Science* **339**, 452–456 (2013).
21. Wang, W., Li, G.-W., Chen, C., Xie, X. S. & Zhuang, X. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science* **333**, 1445–1449 (2011).
22. Xie, X., Cosma, M. P. & Lakadamyali, M. Super resolution imaging of chromatin in pluripotency, differentiation, and reprogramming. *Curr. Opin. Genet. Dev.* **46**, 186–193 (2017).
23. Spahn, C., Endesfelder, U. & Heilemann, M. Super-resolution imaging of *Escherichia coli* nucleoids reveals highly structured and asymmetric segregation during fast growth. *J. Struct. Biol.* **185**, 243–249 (2014).
24. Lehmann, M. et al. Quantitative multicolor super-resolution microscopy reveals tetherin HIV-1 interaction. *PLoS Pathog.* **7**, e1002456 (2011).
25. Annibale, P., Scarselli, M., Kodiyan, A. & Radenovic, A. Photoactivatable fluorescent protein mEos2 displays repeated photoactivation after a long-lived dark state in the red photoconverted form. *J. Phys. Chem. Lett.* **1**, 1506–1510 (2010).
26. Annibale, P., Vanni, S., Scarselli, M., Rothlisberger, U. & Radenovic, A. Quantitative photo activated localization microscopy: unraveling the effects of photoblinking. *PLoS ONE* **6**, e22678 (2011).
27. Baumgart, F. et al. Varying label density allows artifact-free analysis of membrane-protein nanoclusters. *Nat. Methods* **13**, 661–664 (2016).
28. Coltharp, C., Kessler, R. P. & Xiao, J. Accurate construction of photoactivated localization microscopy (PALM) images for quantitative measurements. *PLoS ONE* **7**, e51725 (2012).
29. Sengupta, P. et al. Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis. *Nat. Methods* **8**, 969–975 (2011).
30. Puchner, E. M., Walter, J. M., Kasper, R., Huang, B. & Lim, W. A. Counting molecules in single organelles with superresolution microscopy allows tracking of the endosome maturation trajectory. *Proc. Natl Acad. Sci. USA* **110**, 16015–16020 (2013).
31. Hartwich, T. M. P., Subach, F. V., Cooley, L., Verkhusha, V. V. & Bewersdorf, J. Determination of two-photon photoactivation rates of fluorescent proteins. *Phys. Chem. Chem. Phys.* **15**, 14868–14872 (2013).
32. Lee, S.-H., Shin, J. Y., Lee, A. & Bustamante, C. Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM). *Proc. Natl Acad. Sci. USA* **109**, 17436–17441 (2012).
33. Rollins, G. C., Shin, J. Y., Bustamante, C. & Pressé, S. Stochastic approach to the molecular counting problem in superresolution microscopy. *Proc. Natl Acad. Sci. USA* **112**, E110–E118 (2015).
34. Hummer, G., Fricke, F. & Heilemann, M. Model-independent counting of molecules in single-molecule localization microscopy. *Mol. Biol. Cell* **27**, 3637–3644 (2016).
35. Nino, D., Rafiei, N., Wang, Y., Zilman, A. & Milstein, J. N. Molecular counting with localization microscopy: a Bayesian estimate based on fluorophore statistics. *Biophys. J.* **112**, 1777–1785 (2017).
36. Huang, Z. et al. Spectral identification of specific photophysics of Cy5 by means of ensemble and single molecule measurements. *J. Phys. Chem.* **110**, 45–50 (2005).
37. K L Yeow, E. et al. Characterizing the fluorescence intermittency and photobleaching kinetics of dye molecules immobilized on a glass surface. *J. Phys. Chem.* **110**, 1726–1734 (2006).
38. Widengren, J., Chmyrov, A., Eggeling, C., Löfdahl, P.-Å & Seidel, C. A. M. Strategies to improve photostabilities in ultrasensitive fluorescence spectroscopy. *J. Phys. Chem.* **111**, 429–440 (2007).
39. Vogelsang, J. et al. A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. *Angew. Chem. Int. Ed.* **47**, 5465–5469 (2008).
40. Veatch, S. L. et al. Correlation functions quantify super-resolution images and estimate apparent clustering due to over-counting. *PLoS ONE* **7**, e31457 (2012).
41. Spahn, C., Herrmannsdörfer, F., Kuner, T. & Heilemann, M. Temporal accumulation analysis provides simplified artifact-free analysis of membrane-protein nanoclusters. *Nat. Methods* **13**, 963–964 (2016).
42. Mo, G. C. H. et al. Genetically encoded biosensors for visualizing live-cell biochemical activity at super-resolution. *Nat. Methods* **14**, 427–434 (2017).
43. Zhang, J. & Shapiro, M. S. Mechanisms and dynamics of AKAP79/150-orchestrated multi-protein signalling complexes in brain and peripheral nerve. *J. Physiol.* **594**, 31–37 (2015).
44. Zhang, J., Carver, C. M., Choveau, F. S. & Shapiro, M. S. Clustering and functional coupling of diverse ion channels and signaling proteins revealed by super-resolution STORM microscopy in neurons. *Neuron* **92**, 461–478 (2016).
45. Cella Zanacchi, F., Manzo, C., Magrassi, R., Derr, N. D. & Lakadamyali, M. Quantifying protein copy number in super resolution using an imaging-invariant calibration. *Biophys. J.* **116**, 2195–2203 (2019).
46. Wooten, M. et al. Asymmetric histone inheritance via strand-specific incorporation and biased replication fork movement. *Nat. Struct. Mol. Biol.* **26**, 732–743 (2019).
47. Habuchi, S. et al. Reversible single-molecule photoswitching in the GFP-like fluorescent protein Dronpa. *Proc. Natl Acad. Sci. USA* **102**, 9511–9516 (2005).
48. Subach, F. V. et al. Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. *Nat. Methods* **6**, 153–159 (2009).
49. Malagon, F. RNase III is required for localization to the nucleoid of the 5′ pre-rRNA leader and for optimal induction of rRNA synthesis in *E. coli*. *RNA* **19**, 1200–1207 (2013).
50. Bohrer, C. H., Bettridge, K. & Xiao, J. Reduction of confinement error in single-molecule tracking in live bacterial cells using SPICER. *Biophys. J.* **112**, 568–574 (2017).
51. Das, R., Cairo, C. W. & Coombs, D. A hidden Markov model for single particle tracks quantifies dynamic interactions between LFA-1 and the actin cytoskeleton. *PLoS Comput. Biol.* **5**, e1000556 (2009).
52. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861–874 (2006).
53. Thevathasan, J. V. et al. Nuclear pores as versatile reference standards for quantitative superresolution microscopy. *Nat. Methods* **16**, 1045–1053 (2019).
54. Betschinger, J. & Knoblich, J. A. Dare to be different: asymmetric cell division in *Drosophila*, *C. elegans* and vertebrates. *Curr. Biol.* **14**, R674–R685 (2004).
55. Tran, V., Lim, C., Xie, J. & Chen, X. Asymmetric division of *Drosophila* male germline stem cell shows asymmetric histone distribution. *Science* **338**, 679–682 (2012).
56. Wang, Y., Maharana, S., Wang, M. D. & Shivashankar, G. V. Super-resolution microscopy reveals decondensed chromatin structure at transcription sites. *Sci. Rep.* **4**, 4477 (2014).
57. Dempsey, G. T., Vaughan, J. C., Chen, K. H., Bates, M. & Zhuang, X. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nat. Methods* **8**, 1027–1036 (2011).
58. Goossen-Schmidt, N. C., Schnieder, M., Hüve, J. & Klingauf, J. Switching behaviour of dSTORM dyes in glycerol-containing buffer. *Sci. Rep.* **10**, 13746 (2020).
59. Hirvonen, L. M. & Cox, S. STORM without enzymatic oxygen scavenging for correlative atomic force and fluorescence superresolution microscopy. *Methods Appl. Fluoresc.* **6**, 045002 (2018).
60. Shivanandan, A., Unnikrishnan, J. & Radenovic, A. Accounting for limited detection efficiency and localization precision in cluster analysis in single molecule localization microscopy. *PLoS ONE* **10**, e0118767 (2015).

Methods

Specifics for simulations. For non-filament simulations, six different sets of data were simulated: three different underlying structures and two different fluorophores. The two fluorophores followed the two models in Extended Data Fig. 1. In these simulations, the fluorophore only registered as a localization if it was in the active state. For the different simulations, the first condition contained no clusters (random), and all fluorophores were randomly distributed within a 1,000 nm-by-1,000 nm square and allowed to blink according to the kinetic models in Extended Data Fig. 1. The kinetic rates of the one-state system were chosen to be similar to those of mEos2 and Dendra2, as determined previously[33,61]. The parameters for the two-state kinetic scheme were also chosen to be relatable to these but with an additional dark state that was longer lived. For example, assuming a 50-ms exposure time, our one-dark-state system would have the following rates (in Hz): $k_{on}=5$, $k_{off}=10$, $k_{bleach}=5$. In Lee et al.[61], the rates of mEos2 were approximately $k_{on}=6$, $k_{off}=8$ and $k_{bleach}=5$. The second (small clusters) and third (dense) conditions had three clusters, each with 10% of the fluorophores distributed into the clusters for the small cluster system and 50% for the dense system. For each of the simulations with clusters, each cluster's central location was randomly defined and the localizations within each cluster followed a normal distribution around that center. For each of the six systems, 24 different images were generated and analyzed for each methodology.

For simulations involving filaments, we randomly distributed 50% of the true localizations along five lines and randomly deposited the rest. We simulated 24 images, with 1,000 true localizations each, with approximately 4,000 localizations in total, following the photokinetic model in Extended Data Fig. 1a. These simulations produced filaments that were clearly visible but not homogeneous along the filaments.

Third, to produce 'intersecting' continuous overlapping filaments, we simulated filaments with no varying label density and with a localization error of 20 nm. This was carried out by placing a fluorophore every 5 nm along a filament. These simulations also followed the model in Extended Data Fig. 1a and resulted in images similar to that in Fig. 2, far right.

For all simulations (except for when the activation rate was varied), the activation rate results in 0.5 active fluorophores per 1 μm^2 per frame. This photoactivation density per frame was chosen to minimize the detection of overlapping fluorophores similar to that in the experiment.

Methods for experiments used to calculate $Z(\Delta n)$. *Strains.* Strains with chromosomal fluorescent protein fusion tags were constructed using λ Red-mediated homologous recombination[62]. Some results used in this paper came from strains that also harbor a single chromosomal DNA site marker (*tetO6*); the DNA markers are positioned in various positions on the chromosome, and a portion of the results are not relevant and thus are not discussed in this publication. Details for the construction of these bacterial strains are described in detail in a previous publication[62].

Cell growth. For live-cell imaging, single colonies were picked from LB plates and cultured overnight in EZ Rich Defined Medium (EZRDM, Teknova) with 0.4% glucose at room temperature (RT) with shaking. The next morning, cells were reinoculated into fresh EZRDM with 0.4% glucose and grown at RT until they reached mid-log phase (OD_{600}, 0.3–0.4). For simultaneous visualization of DNA site markers (results are not reported here), cells were collected and resuspended in fresh EZRDM supplemented with 0.3% L-arabinose and 0.4% glycerol and allowed to grow for 2 additional hours; these cells were collected via centrifugation and imaged immediately. For fixed-cell experiments, cells were grown accordingly and fixed in 3.7% (vol/vol) paraformaldehyde (16% paraformaldehyde, EM grade, EMS) for 15 min at RT, washed with 1× PBS and imaged immediately.

Nascent rRNA labeling (smFISH). We performed smFISH using a previously published protocol[63,64]. Briefly, cells were grown in EZRDM with glucose as previously described; 5 ml mid-log-phase cells were fixed with 3.7% (vol/vol) paraformaldehyde (16% paraformaldehyde, EM grade, EMS) and placed on ice for 30 min. Next, cells were collected via centrifugation and subsequently washed two times with 1× PBS. Cells were then permeabilized by resuspending them in a mixture of 300 µl water and 700 µl 100% ethanol and incubating the suspension with rotation at RT for 30 min. Cells were stored at 4 °C until the next day. Wash buffer was freshly prepared with 40% formamide and 2× SSC and put on ice. Cells were pelleted in a bench-top centrifuge at 10,000 r.p.m. for 3 min, and the cell pellet was resuspended in 1 ml wash buffer. The sample was placed on a nutator to mix for 5 min at RT. Hybridization solution was prepared with 40% formamide and 2× SSC; subsequently, dye-labeled oligonucleotide probes were added to hybridization solution to a final concentration of 1 µM. Cell were pelleted again, and 50 µl hybridization solution with probe was added to the pellet. The hybridization sample was mixed well and placed overnight in an incubator at 30 °C. The next day, 10 µl hybridization sample was washed with 200 µl fresh wash buffer and incubated at 30 °C for 30 min; this step was repeated one more time. The washed sample was imaged immediately, without stochastic optical reconstruction microscopy (STORM) imaging buffer for ensemble fluorescence or with STORM buffer to induce dye blinking for super-resolution imaging. Glucose oxidase-supplemented thiol STORM buffer was used to image samples with only dye labeling (50 mM Tris (pH 8.0), 10 mM NaCl, 0.5 mg ml^{-1} glucose oxidase (Sigma-Aldrich), 40 g ml^{-1} catalase (Roche), 10% (wt/vol) glucose and 10 mM MEA (Fluka))[57]. Thiol-only STORM buffer (10 mM MEA, 50 mM Tris (pH 8.0), 10 mM NaCl) was used to image samples with both endogenously expressed fluorescent proteins and dye labeling. This was to preserve the fluorescent signal from fluorescent proteins, because the presence of glucose oxidase in the STORM buffer tended to quench the fluorescent protein signal. Pre-rRNA transcripts were detected with a single probe, L1, conjugated at the 5′ end with Alexa Fluor 647 (NHS ester) (IDT)[65]. Upon receiving commercial oligonucleotides, a working stock (50 µM) was made and aliquoted for storage at −20 °C.

Cell imaging and SMLM analysis. A 3% gel pad made with low-melting agarose (SeaPlaque, Lonza) in EZRDM was prepared. Live cells of an optimal imaging density were deposited onto the gel pad and immobilized with a coverslip for imaging as previously described[63]. An Olympus IX81 inverted microscope with a ×100 oil objective (UPlanApo, numerical aperture (NA) of 1.4) was used, with ×1.6 additional amplification. Images were captured with an iXon DU 895 (Andor) EMCCD with a 13-µm pixel size using MetaMorph (Molecular Devices). Illumination (405 nm, 488 nm, 561 nm, 647 nm) was provided by solid-state lasers Coherent OBIS 405, Coherent OBIS 488, Coherent Sapphire 561 and Coherent OBIS 647, respectively. Fluorescence was split using a multi dichroic filter (ZT 405/488/561/647rpc, Chroma), and the far-red, red and green channels were further selected using HQ705/55, HQ600/50 and ET525/50 bandpass filters (Chroma). Gold fiducial beads (50 nm, Microspheres-Nanospheres) were used to correct for any sample drift during imaging. All super-resolution images were acquired with a 10-ms exposure time with 3,000–9,000 frames. Activation of fluorescent proteins was carried out simultaneously with fluorophore excitation, and the activation laser (405 nm) was kept at a constant power throughout the imaging session (1 mW at the laser output). For two-color imaging, simultaneous, multicolor acquisition was achieved using the Optosplit II or Optosplit III device (Cairn Research); colored channels were overlaid using calibration images from TetraSpeck beads (Life Technologies, T-7279) as previously described[66]. Initial fitting of raw imaging data was performed via the ThunderSTORM plugin[67]. Later analysis of localizations with DDC was processed using custom MATLAB scripts, which will be made available upon request.

Methods for nuclear pore complex experiments. *Cell line.* Human U-2 OS genome-edited Nup96–mEGFP cells (clone 195, 300174) were obtained from CLS Cell Lines Service. Cells were grown at 37 °C with 5% CO_2 in DMEM (Life Technologies) supplemented with MEM NEAA, GlutaMAX and 10% FBS. Cells were plated on 8-well Lab-Tek #1 coverglass chambers (Nunc) for immunostaining and imaging purposes.

Immunostaining. Cells were fixed with 4% (vol/vol) paraformaldehyde in PBS for 25 min and blocked in blocking buffer (3% BSA, 0.2% Triton X-100 in PBS) for 1 h. Cells were labeled for 2 h on a rocker with the FluoTag-Q anti-GFP Alexa Fluor 647 nanobody (1:100 dilution, NanoTag Biotechnologies). Cells were then washed with washing buffer (0.2% blocking buffer, 0.05% Triton X-100 in PBS) four times, 10 min each.

Single-molecule localization microscopy imaging. Single-molecule imaging was performed using imaging buffer comprising 50 mM Tris, pH 7.5, 10 mM NaCl, 0.5 mg ml^{-1} glucose oxidase (Sigma, G2133), 40 µg ml^{-1} catalase (Roche Applied Science, 106810), 10% (wt/vol) glucose and 30 mM Ciseamine (stock concentration, 77 mg ml^{-1} of 360 mM HCl). Images were acquired on the Oxford Nanoimager-S microscope, which has the following configuration: 405-, 488-, 561- and 640-nm lasers, 498–551- and 576–620-nm bandpass filters in channel 1 and 665–705-nm bandpass filters in channel 2, ×100 1.4-NA objective (Olympus) and a Hamamatsu Flash 4 version 3 sCMOS camera. Localizations were acquired with a 10-ms exposure time over 40,000 frames with constant laser power for each acquisition. Images were processed, and localizations were obtained using the NimOS localization software (Oxford Nanoimaging).

Quantifying the effective labeling efficacy. To determine the ELE, we used the raw localizations from the Nup96 SMLM images and applied a simple tree-clustering algorithm (centroid, threshold = 100 nm) to segment individual NPCs. We then followed the methodology of Thevathasan et al.[53] to further filter the segmented NPCs to well-defined structures; we only used the well-defined structures in our analyses.

First, we fit the segmented complexes to a circle (with the radius as a free parameter) and eliminated complexes with a radius <40 nm or >70 nm, as these structures did not resemble single NPCs. Second, we refit these complexes with a circle of fixed radius and eliminated complexes when more than 25% of the localizations were within 40 nm of the center or 40% of the localizations were greater than 70 nm. Finally, we eliminated background localizations (those that were more than 70 nm away from the center of the refitted circle or those that were within 30 nm of the center of the circle).

Next, (again) as in Thevathasan et al.[53], we then quantified the number of detected corners of each NPC. To do this, we rotated the filtered localizations of the complexes so that the following function was minimized for each complex:

$$\theta_{\mathrm{rot}} = \arg \quad \min_{\theta_{\mathrm{rot}}}(\theta_{\mathrm{rot}} - \theta_i \mathrm{mod}\pi/4),$$

where θ_i is the angle of the ith localization within the complex under consideration. We then rotated the localizations by θ_{rot}. We then constructed a histogram of each complex's new θ_i values with bins starting at $-\pi/8$ of width $\pi/4$. The number of detected corners (for each complex) was equal to the number of bins with at least one localization detected inside.

With the number of detected corners for each complex on hand (again, as in Thevathasan et al.[53]), we then used the following equation to determine the ELE (see Thevathasan et al.[53] for further details):

$$B(k|8, (1 - B(0|4, \mathrm{ELE}))),$$

where k is the number of detected corners, and $B(k|n, p) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$ is the binomial probability density function.

Methods used for AKAP150. For fixed-cell STORM imaging, cells were fixed with 4% paraformaldehyde for 20 min and then washed with 100 mM glycine in Hanks' balanced salt solution (HBSS) to quench the free paraformaldehyde. Cells were permeabilized and blocked in a permeabilization solution with 0.1% Triton X-100, 0.2% BSA, 5% goat serum and 0.01% sodium azide in HBSS. Cells were then incubated overnight at 4 °C with an anti-AKAP150 (MilliporeSigma, 07-210; EMD Millipore, 07-210) antibody at a 1:500 dilution, followed by a 1–2-h incubation with goat anti-rabbit Alexa 647-conjugated antibodies at a 1:1,000 dilution. Cells were then post-fixed again in 4% paraformaldehyde, quenched with 100 mM glycine in HBSS and washed with HBSS to prepare for imaging. Immediately before imaging, the medium was changed to STORM-compatible buffer (50 mM Tris-HCl (pH 8.0), 10 mM NaCl and 10% glucose) with glucose oxidase (560 mg ml^{-1}), catalase (170 mg ml^{-1}) and mercaptoethylamide (7.7 mg ml^{-1}). STORM images were obtained using a Nikon Ti total internal reflection fluorescence (TIRF) microscope with N-STORM, an Andor iXon3 Ultra DU 897 EMCCD and a ×100 oil-immersion TIRF objective. Photoactivation was driven by a Coherent 405-nm laser, while excitation was driven with a Coherent 647-nm laser. Puncta localization was performed using both Nikon Elements analysis software.

Methods used for sister chromatid experiments. Chromatin-fiber preparation from *D. melanogaster* embryos with YOYO-1 staining. Young embryos (<2 h old, 15–20 embryos per experiment) were collected and washed three times with lysis buffer at RT (100 mM NaCl, 25 mM Tris base, 0.2% Joy detergent, pH 10; adapted from McKnight and Miller[68]). Embryos were transferred to the center of a clean glass slide (Fisherbrand Superfrost Plus Microscope Slides) and subsequently drained of residual lysis buffer. Following removal of residual lysis buffer, 20 µl fresh lysis buffer was then added to the surface of the glass slide to immerse embryos. Embryos were then manually broken apart with dissecting forceps to release embryonic nuclei from the intact embryo. After breaking open the embryo, the protective outer layers of the embryo (chorion layers, waxy layer and vitelline membrane) were removed, and the nuclei were allowed to sit in lysis buffer until fully lysed (2 min). A sucrose–formalin solution (10 µl, 1 M sucrose, 10% formaldehyde) was then added on top of the lysed nuclei, after which a large coverslip (22 × 50 mm, Thermo Scientific Rectangular Cover Slips) was placed on top of the lysed chromatin solution and incubated for 2 min at RT. Following incubation, slides containing chromatin fibers derived from lysed embryonic nuclei were transferred to liquid nitrogen and allowed to sit for 2 min. Slides were then removed from liquid nitrogen, after which the coverslip was removed with a razor blade. Slides were then transferred to cold (−20 °C) 95% ethanol and incubated for 10 min. Slides were removed from ethanol and placed at a 45° angle for 2 min (or until almost all ethanol had evaporated from the slide, but it was not completely dry). A fixative solution (500 µl, 0.5% formaldehyde in 1× PBS with 0.1% Triton) was then slowly added to the surface of the slide, after which the slide was incubated for 2 min. Slides were then drained of fixative solution and transferred to a coplin jar containing 50 ml 1× PBS. To fully wash chromatin-fiber samples, slides were then removed from the coplin jar and drained of remaining 1× PBS. Used PBS in the coplin jar was then discarded, and the coplin jar was refilled with 50 ml fresh PBS. Slides were then placed back inside the coplin jar and incubated at RT for 2 min. Slides were removed from the coplin jar and placed in fresh PBS two additional times to complete the wash process. Following washing, slides were transferred to a humid, dark place and preblocked with 500 µl blocking solution (2% MilliporeSigma BSA in 1× PBS) for 30 min. Blocking solution was then drained, and 500 µl DNA-labeling solution containing 1 µM YOYO-1 DNA dye (Thermo Fisher Scientific Invitrogen, YOYO-1) was then slowly added to the surface of the slide. Slides were then incubated for 120 min in a humid, dark place. Following incubation, slides were drained of DNA-labeling solution and transferred to a coplin jar containing 50 ml 1× PBS. Slides were removed from the coplin jar and placed in fresh PBS two additional times to complete the wash process. Following washing, slides were removed from the coplin jar and drained of residual 1× PBS. Slides were then mounted in preparation for STORM imaging.

Single-molecule localization microscopy imaging. SMLM imaging of DNA fibers is based on the DNA-intercalating dye method[75]. The fibers on cover slides were labeled with 1 µM YOYO-1 for 120 min. dSTORM buffer (8–10 µl)[76] was added on the top of the fibers and sandwiched with a clean coverglass (#1, Fisher Scientific). The coverglass was then sealed with nail polish. The sample can be imaged within 4–5 h with reasonable localizations. Image acquisitions were performed on an Olympus IX71 inverted microscope with a 1.49-NA ×100 TIRF objective, a ZT405/488/561 dichroic mirror (Chroma), an ET525/50 emission filter (Chroma) and an Andor iXon Ultra 897 EMCCD camera. Ten to 30 3,000-frame acquisitions of YOYO-1 signal were then obtained at a rate of 30 frames per second with 488-nm laser power at 1 kW cm^{-2}. During imaging, the activation 405-nm laser was ramped up stepwise (images were analyzed individually and then recombined) by 1 W cm^{-2} per movie (3,000 frames) to obtain more localizations. Specifically, different activation intensities were needed to account for the labeling density; therefore, we acquired an image with different set activation intensities and used DDC to analyze them separately and then recombined them at the end (as discussed in the last section of the main text). dSTORM data were first localized using 2D Gaussian fitting in an ImageJ plugin, ThunderSTORM. A bandpass filter (70 500 nm) for sigma was applied to remove the single pixel noise and out-of-focus molecules. The cross-correlation method in ThunderSTORM was applied to correct the long timescale drift.

Analysis. To quantify the number of localizations between sister chromatids, we first fit a spline function to cropped-out regions that showed single filaments. We then projected the localizations along this new axis, such that there was no curvature within the filaments and they were centered. We then split the filament into as many specifically sized segments as possible (as varied in the corresponding figures) and quantified the number of localizations in the upper sister relative to that in the lower sister for the different blinking-artifact methods.

Methods used for dynein experiments. *Cell line.* Stably transfected HeLa IC74-mfGFP cells (the dynein intermediate chain is GFP labeled, from the T. Murayama lab, Juntendo University School of Medicine) were plated on an 8-well Lab-Tek #1 coverglass chamber (Nunc). Cells were cultured under standard conditions (DMEM, high glucose, pyruvate, 10% FBS and 2 mM glutamine).

Immunostaining. Cells were fixed with paraformaldehyde (4% in PBS) at RT for 20 min and incubated with blocking buffer (3% (wt/vol) BSA (Sigma) in PBS and 0.2% Triton X-100 (Thermo Fisher Scientific) for 1 h. Dynein intermediate chain–GFP was immunostained with primary antibody (chicken polyclonal anti-GFP, Abcam, 13970) diluted 1:500 in blocking buffer for 45 min at RT. Cells were rinsed three times with blocking buffer and incubated for 45 min with secondary antibodies (donkey-anti chicken labeled with photoactivatable dye pairs) for STORM (Alexa Fluor 405–Alexa Fluor 647).

Imaging. Imaging was performed using the Nanoimager-S microscope (Oxford Nanoimaging) with the following specifications: 405-, 488-, 561- and 640-nm lasers and 665–705-nm bandpass filters, ×100 1.4-NA objective (Olympus) and a Hamamatsu Flash4 version 3 sCMOS camera. Localization microscopy images were acquired with 16-ms exposure and 50,000 frames. Activation (405 nm, ~0.5 W cm^{-2}) was kept constant and then processed using NimOS localization software (Oxford Nanoimaging).

Analysis. To quantify each 'cluster' as a particular oligomerization state, we first quantified the number of localizations within each individual cluster using the hierarchical tree-clustering algorithm built in MATLAB. We then assigned the oligomer state of dynein (for each method) such that the fractions of each state were the same as those in ref. [45]. We then compared the assigned state for each individual 'cluster' as described in the main text.

Methods used for characterizing blinking. *Sample preparation.* The Plac::mEos3.2 plasmid (pXY329) was constructed based on pJL005 (Plac::FtsZwt-mEos3.2)[69] using In-Fusion cloning (Takara) to remove the *ftsZ* gene. MG1655 cells were transformed with pXY329 and grown in M9+ medium. Cells were collected at log phase and fixed with 3.8% paraformaldehyde in 1× PBS buffer. Fixed cells were washed with 1× PBS three times and stored at 4 °C for no longer than 1 week.

Streptavidin conjugated with Alexa Fluor 647 (SA–AF647) was purchased from Thermo Fisher Scientific. The SA–AF647 working solution was made fresh every time by diluting the original stock (36 µM) to 10 pM in 1× PBS with 0.5% Tween-20.

Imaging. Photoactivated localization microscopy. Fixed MG1655-Plac::mEos3.2 cells were sandwiched between a 3% PBS agar pad and a coverglass as previously described[70]. PALM imaging was performed as in a previous study[69] on an Olympus IX71 inverted microscope with a ×100, 1.49-NA oil-immersion objective. The 561-nm excitation laser power was tuned to 1,500 W cm^{-2}, while the 405-nm laser power varied from 0 to 3.5 W cm^{-2}. For the 0 W cm^{-2} condition, a short laser pulse (1 s) at 3.5 W cm^{-2} and 405 nm was applied to activate some mEos3.2 molecules to a red fluorescent state. At each 405-power condition, six movies of 3,000-frame

images with an exposure time of 10 ms were collected continuously. Three repeats of all the 405 conditions were performed to obtain the average blinking behavior.

Direct stochastic optical reconstruction microscopy. SA–AF647 (10 pM) was flown into a preassembled chamber with biotin–PEG-coated coverglasses for 5 min and washed three times with 1× PBS. STORM buffer was made fresh using the recipe described in ref. [71] and injected into the chamber to replace the PBS buffer before imaging. All STORM images were taken after 60 min, as the oxygen level in the buffer was shown to be stable after 1 h. dSTORM imaging was performed on an Olympus IX81 inverted microscope with a ×100, 1.45-NA oil-immersion objective. The 647-nm excitation laser power was tuned to 1,800 W cm^{-2}, while the 405-nm laser power varied from 0 to 13.9 W cm^{-2}. At each 405 condition, two to three 5,000-frame movies at different regions on the coverglass were taken with an exposure time of 30 ms. Two repeats of all the 405 conditions were performed.

Data processing. The single fluorophore spots in both PALM and dSTORM movies were localized with an ImageJ[72] plugin, ThunderSTORM[73]. All spots with irregular properties (abnormal sigma, intensity too high or low or multiple spots within a 500-nm range) were removed. A customized MATLAB code was used to link the same spots within three- to fourfold of localization limitation (100 nm) throughout the whole movie using a nearest-neighbor algorithm. Continuous frames with localization from the same linked fluorophore were counted as on frames. Other frames before the last on frame were counted as off frames. Blinking number was calculated as the sum of on-frame number. The mEos3.2 sample size was 540, 1,634, 2,571, 3,548, 5,062; the AF647 sample size was 1,833, 3,919, 6,341, 8,795, 11,400, 13,848.

Bleaching of Alexa 647. To provide the bleaching timescale of AF647 under our imaging conditions, we used the nuclear pore (NUP) experiment (Fig. 3) and quantified the bleaching time of single nanobody-conjugated AF647 fluorophores sparsely distributed in the cytoplasm of cells.

As shown in Supplementary Fig. 2a, Nup96 labeled with nanobody-conjugated AF647 fluorophores predominantly localized on the nuclear membrane (dense area in the cell). Single nanobody–AF647 fluorophores were also found sparsely distributed in the cytoplasm, similar to previously observed results[53] (Supplementary Fig. 2b). We used the sparse distribution of cytoplasmic fluorophores to analyze the photobleaching behavior of AF647.

To identify individual AF647 fluorophores in cropped regions of the cytoplasm, we determined the best protocol to link localizations into individual trajectories using the following procedure. We first used the number of repeats per AF647 fluorophore (determined from the analysis in the main text based on the known true number of Nup96 molecules in the NPC) and divided the total number of localizations in a cropped cytoplasmic region by this number to obtain the true number of AF647 fluorophores in this region. We then simulated a random distribution of the same number of fluorophores within the same region with the same number of repeats per fluorophore that bleached over a timescale of 10,000 frames. These simulated fluorophores also had the same localization precision as our imaging condition. Using this simulated image, we determined that the following protocol resulted in the best estimation of photobleaching times: (1) we used a distance threshold of 40 nm to link any two localizations that were <40 nm into the same trajectory, and (2) we eliminated trajectories in which the maximal distance between any two localizations within the trajectory was >40 nm (data not shown). Additionally, we only used trajectories for which the first localization appeared within the first 1,000 frames to avoid potential incomplete photobleaching toward the end of the imaging. Note that we did not set any time threshold to link localizations. Using these criteria, we obtained a total number of 2,813 single AF647 fluorophores and plotted the CDF of the total number of frames that a fluorophore lasts (that is, the photobleaching time, calculated as maximum frame of trajectory − minimum frame of trajectory, Supplementary Fig. 2c). The CDF shows that, under our imaging conditions, 90% of AF647 fluorophores bleached within 6,100 frames (61 s, frame time = 10 ms). The average blinking number per AF647 fluorophore is 4.4.

Additionally, we performed the same analysis (with one modification) upon the publicly available data of Thevathasan et al.[53] (Supplementary Methods), who also imaged Nup96 labeled by nanobodies conjugated with AF647. Here we used trajectories for which the first localization appeared within the first 10,000 frames instead of 1,000 frames due to the smaller amount of data available and the longer acquisition time of these images (number of single AF647 molecules, 173; total acquisition time, 65,000 frames). As shown in Supplementary Fig. 2d, in this experiment, 90% of AF647 molecules bleached within 22,000 frames (330 s, frame time = 15 ms). Note here that the 90% bleaching time has a large error range because the CDF curve was compiled from only 173 molecules; the original paper only contained a limited amount data of the sparsely distributed AF647 fluorophores in the cytoplasm due to the fact that the work focused on NUP in the nucleus. The average blinking number per AF647 fluorophore in this work is 3.9.

Finally, we performed the same analysis on precursor ribosomal RNA species (pre-rRNA) labeled with AF647-conjugated DNA probes in *E. coli* cells using the same data that contributed to Fig. 1d. We selected cells with less than 500 total localizations to enrich for more sparse samples and used the same parameters described above (40-nm linkage distance, maximal distance of 40 nm, first localization appeared within the first 1,000 frames) to analyze the photobleaching timescale of DNA-conjugated AF647 fluorophores. The resulting CDF of bleaching times showed that 90% of fluorophores bleached within 1,200 frames (12 s, Supplementary Fig. 2e). Note that this photobleaching time could be an overestimate, as pre-rRNA species tend to cluster together[74].

In sum, these results illustrate that, in our experiments, the AF647 fluorophore can be substantially photobleached, satisfying the criterion for the application of DDC. In all data acquired in this work to which DDC was applied, the acquisition time was ~6 min or greater. Furthermore, these results demonstrate that the specific photokinetics of a fluorophore can vary depending on experimental systems and imaging conditions. The great variability of AF647's photokinetics under similar buffer conditions but in different experimental systems has also been observed before and is especially apparent in the variability of the switching cycles[53,57–59]. Finally, we note that, if the dark time of AF647 within an imaging condition is much longer than the imaging time, the reappearance of these molecules would be limited. Therefore, if this scenario were indeed the case, the lack of reappearance of AF647 would be approximately equivalent to AF647 bleaching over the experimental timescale, and DDC still applies.

Data specifics of Thevathasan et al.. The particular publicly available data we used for analyzing AF647 are available in Thevathasan et al.[53]. The particular condition that we analyzed in the Bleaching of Alexa 647 section was the GLOX + 35 mM MEA condition (mEGFP-NB-Q-AF647), using localizations fit with at least a localization precision of 15 nm to exclude uncharacteristic spots. See Thevathasan et al. for further details[53].

Methodology of Spahn et al.. The implementation of Spahn et al. was carried out by randomly selecting subsets of localizations (with replacement) and then using the threshold of 2.5 (just as in ref. [41]) as the definition of a cluster to create the cluster masks. The normalized average density within the clusters (P/P_o) versus the relative area of the image that the clusters covered (η) was plotted for all subsets of localizations to determine whether clustering was significant for the system of interest. For this methodology, clustering is deemed significant if P/P_o rises above 1 and stays above 1.

We tested this method on three different simulation systems (random, small clusters, dense clusters) with the two-state fluorophore, and we show these results in Extended Data Fig. 8a. We observed that the randomly distributed fluorophores maintained *a* P/P_o equal to 1, while the dense cluster system rose substantially well above 1, demonstrating that the methodology could adequately recognize that there were clusters in the dense cluster system and that there were not clusters in the random system. As expected, an intermediate value for the small cluster system was also observed.

Next, to investigate the clustering of AKAP79 and AKAP150 with a method orthogonal to DDC, we applied the methodology of Spahn et al.[41] to the super-resolution data of each of the two orthologs. The results of this analysis are shown in Extended Data Fig. 8b, where P/P_o values for both orthologs rose slightly above $P/P_o = 1$. These results support the previous findings that the two orthologs are significantly clustered, supporting the analysis as quantified by DDC. Although, we should note that P/P_o did not reach high values (like those for the dense cluster system), suggesting that, just as with DDC, the clustering of the two orthologs is not 'extreme.'

Statistics. Means were calculated with the following equation.

$$\mu = \frac{1}{N}\sum_{i=1}^{N}(x_i)$$

Standard deviation and s.e.m. (through bootstrapping) were calculated with the following equation.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

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

Data availability

The data in this paper are shown in the main figures and Extended Data figures. All raw data for each of the simulation systems (Figs. 1 and 2) are also included at https://github.com/XiaoLabJHU/DDC. Source data are provided with this paper. All other data are available upon request. The complete package of DDC (data, code, user guide) is available for download at https://github.com/XiaoLabJHU/DDC (there are no access restrictions).

References

61. Lee, S. H., Shin, J. Y., Lee, A. & Bustamante, C. Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM). *Proc. Natl Acad. Sci. USA* **109**, 17436–17441 (2012).
62. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
63. Hensel, Z., Fang, X. & Xiao, J. Single-molecule imaging of gene regulation in vivo using cotranslational activation by cleavage (CoTrAC). *J. Vis. Exp.* e50042 https://doi.org/10.3791/50042 (2013).
64. Skinner, S. O., Sepúlveda, L. A., Xu, H. & Golding, I. Measuring mRNA copy number in individual *Escherichia coli* cells using single-molecule fluorescent in situ hybridization. *Nat. Protoc.* **8**, 1100–1113 (2013).
65. Malagon, F. RNase III is required for localization to the nucleoid of the 5′ pre-rRNA leader and for optimal induction of rRNA synthesis in *E. coli*. *RNA* **19**, 1200–1207 (2013).
66. Hensel, Z., Weng, X., Lagda, A. C. & Xiao, J. Transcription-factor-mediated DNA looping probed by high-resolution, single-molecule imaging in live *E. coli* cells. *PLoS Biol.* **11**, e1001591 (2013).
67. Sage, D. et al. Quantitative evaluation of software packages for single-molecule localization microscopy. *Nat. Methods* **12**, 717–724 (2015).
68. McKnight, S. L. & Miller Jr., O. L. Electron microscopic analysis of chromatin replication in the cellular blastoderm *Drosophila melanogaster* embryo. *Cell* **12**, 795–804 (1977).
69. Lyu, Z., Coltharp, C., Yang, X. & Xiao, J. Influence of FtsZ GTPase activity and concentration on nanoscale Z-ring structure in vivo revealed by three-dimensional superresolution imaging. *Biopolymers* **105**, 725–734 (2016).
70. Yang, X. et al. GTPase activity-coupled treadmilling of the bacterial tubulin FtsZ organizes septal cell wall synthesis. *Science* **355**, 744–747 (2017).
71. Nahidiazar, L., Agronskaia, A. V., Broertjes, J., van den Broek, B. & Jalink, K. Optimizing imaging conditions for demanding multi-color super resolution localization microscopy. *PLoS ONE* **11**, e0158884 (2016).
72. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
73. Ovesný, M., Křížek, P., Borkovec, J., Svindrych, Z. & Hagen, G. M. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics* **30**, 2389–2390 (2014).
74. Weng, X. et al. Spatial organization of RNA polymerase and its relationship with transcription in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **116**, 20115–20123 (2019).
75. Flors, C., Ravarani, C. N. J. & Dryden, D. T. F. Super-resolution imaging of DNA labelled with intercalating dyes. *Chemphyschem* **10**, 2201–2204 (2009).
76. Nahidiazar, L., Agronskaia, A. V., Broertjes, B., van den Broek, B. & Jalink, K. Optimizing imaging conditions for demanding multi-color super resolution localization microscopy. *PLoS ONE* **11**, e0158884 (2016).

Acknowledgements

This work was supported by the NIH (5T32GM007231, C.H.B.; F31GM115149-01A1, M.W.; R01 GM086447; and R01 DK073368 and R35CA197622, J.Z.) and the NSF (MCB1817551), a Johns Hopkins Discovery Award, a Hamilton Innovation Research Award (J.X.), NIGMS/NIH (R01GM112008, J.X. and X.C.; R35GM127075, X.C.; and R01GM133842, M.L.) and the Howard Hughes Medical Institute (55108512, X.C.).

Author contributions

C.H.B., conceptualization, discovery and development of theory, software, data collection, data analysis, writing (original draft, review and editing); X.Y., experimental data collection, data analysis, writing (review and editing), project guidance; S.T., experimental data collection, data analysis; X.W., experimental data collection, data analysis; B.T., experimental data collection, data analysis, writing (review and editing); R.M., development of the user guide, software, writing (review and editing); B.R., experimental data collection, data analysis; M.W., experimental data collection, writing (review and editing); X.C., project guidance, funding acquisition, writing (review); J.Z., project guidance, funding acquisition, writing (review); E.R., project guidance, writing (review); M.L., experimental conceptualization, funding acquisition, writing (review and editing); J.X., conceptualization, project guidance, writing (original draft), supervision, funding acquisition, project administration, writing (review and editing).

Competing interests

The authors declare no competing interests.

Additional information

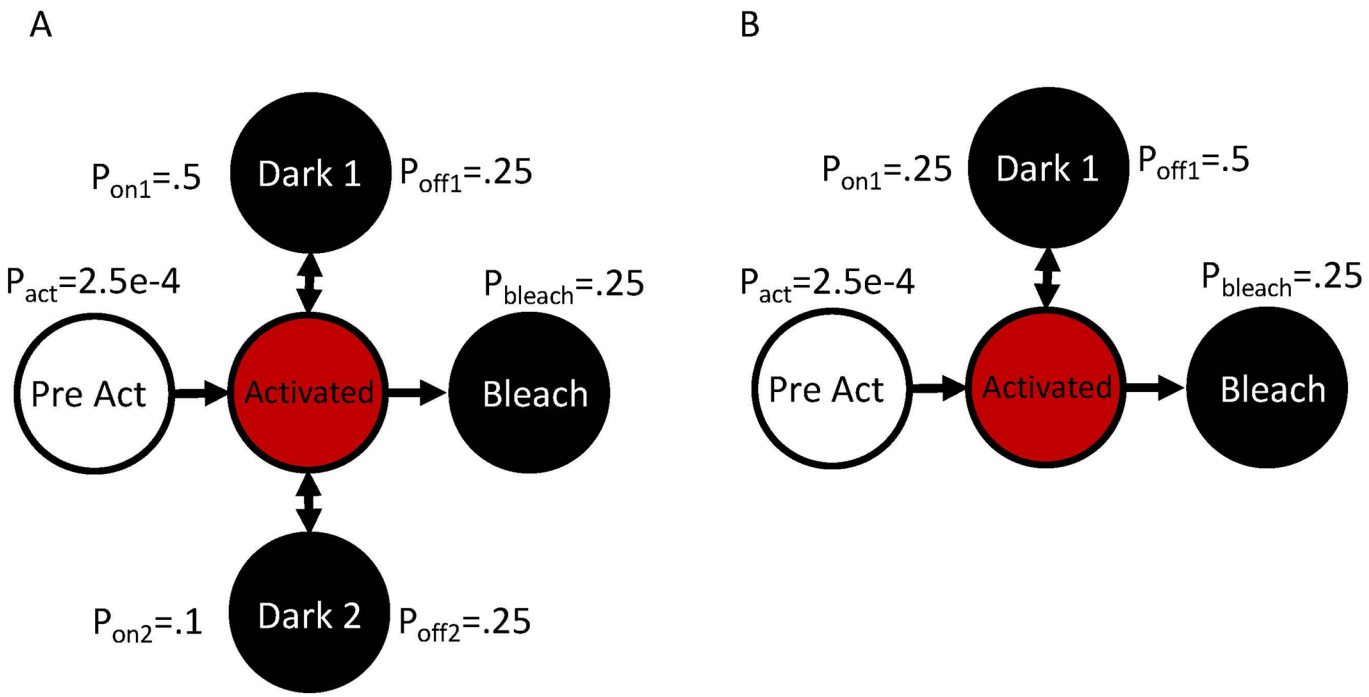
Extended data is available for this paper at https://doi.org/10.1038/s41592-021-01154-y.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41592-021-01154-y.

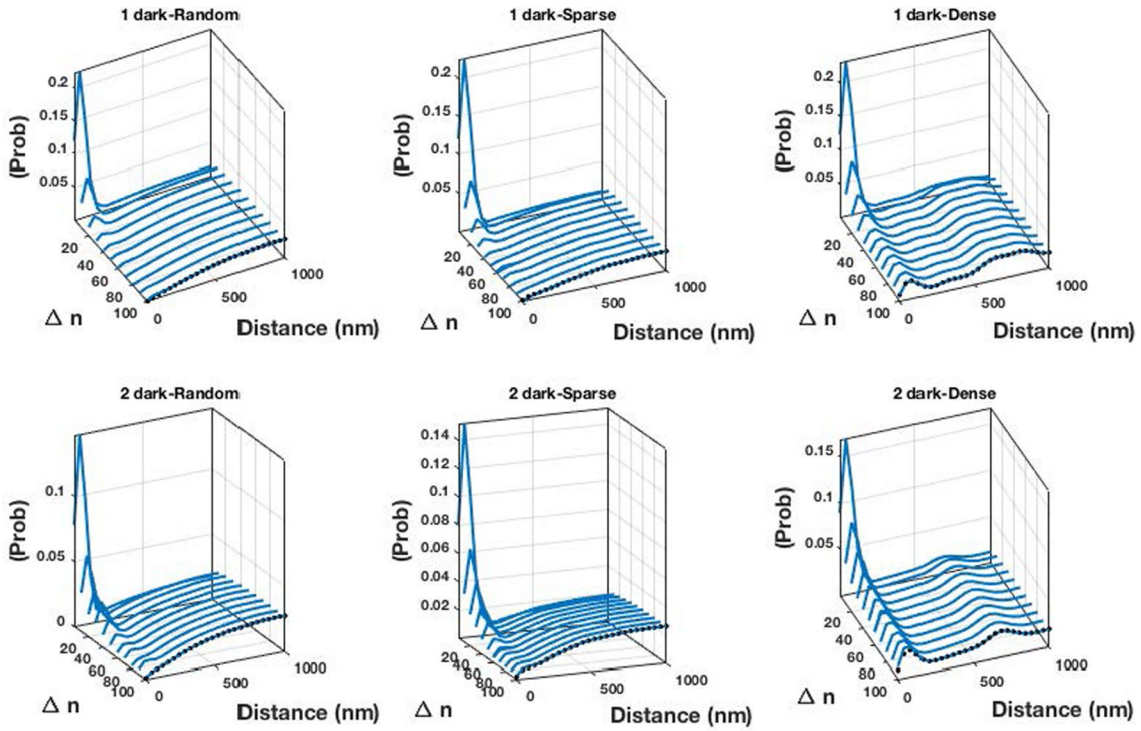
Correspondence and requests for materials should be addressed to J.X.

Peer review information Rita Strack was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. *Nature Methods* thanks Paolo Annibale, Sebastian Malkusch, and the other, anonymous reviewer(s) for their contribution to the peer review of this work.

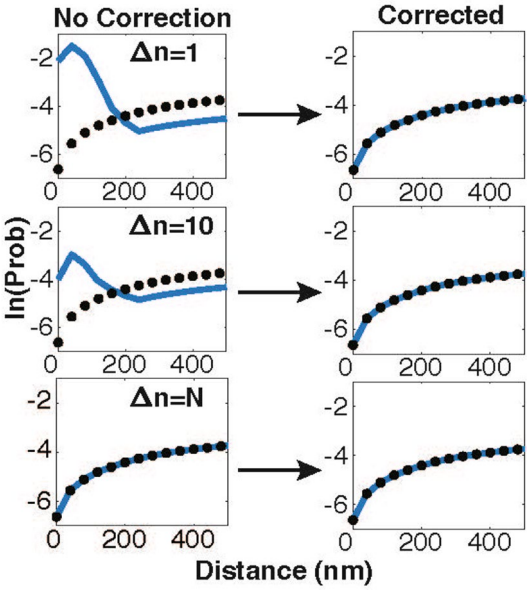
Reprints and permissions information is available at www.nature.com/reprints.



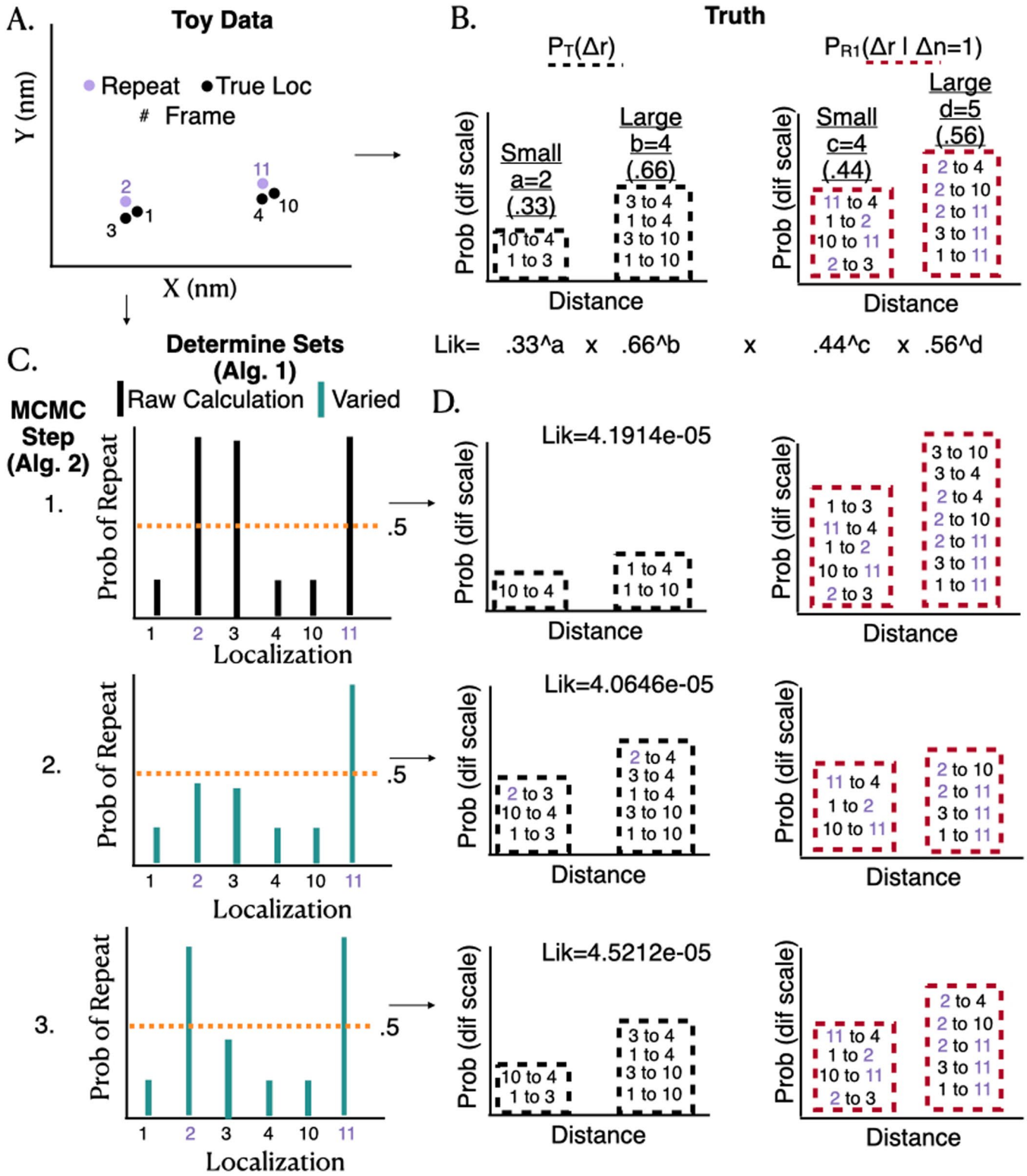
Extended Data Fig. 1 | Photokinetic Models. The two kinetic models used to simulate blinking, (**a**) 2 dark state and (**b**) 1 dark state. The transition probabilities per frame are shown in the figure.



Extended Data Fig. 2 | Converging of Pairwise distance distributions. The pairwise distance distributions for both photo-kinetic models shown in Extended Data Figure 1 and 6 molecular assemblies. Note here that the axis is no longer log scale as in the main text and the true pairwise distance distribution is shown as black dots.

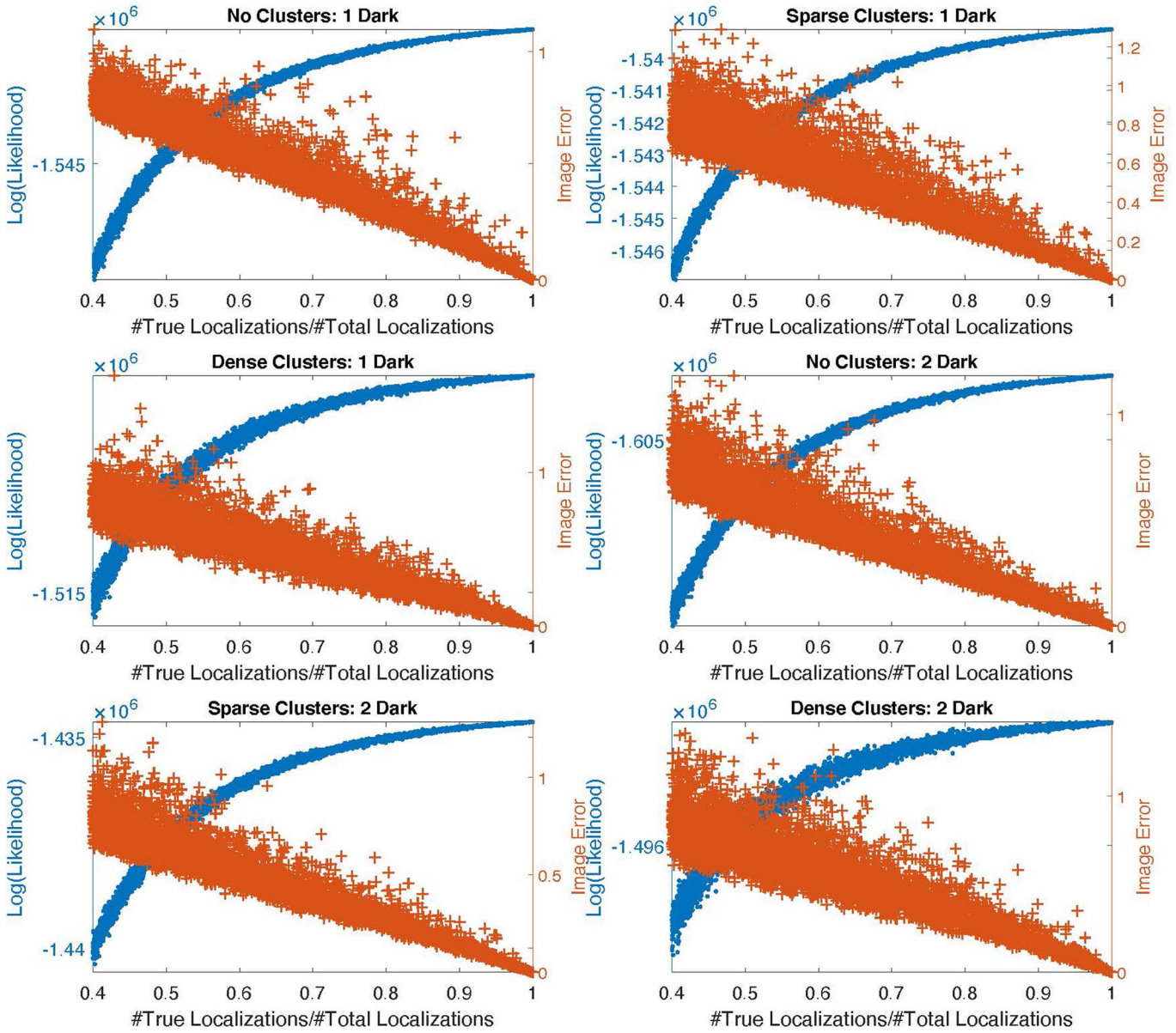


Extended Data Fig. 3 | Matching the true pairwise distance distribution. An illustration of the pairwise distance distributions at a certain frame difference, Δn, before and after being corrected with DDC. When the likelihood is maximized all of the pairwise distance distributions will match the true pairwise distance distribution. [The true pairwise distance distribution is shown as black dots.].

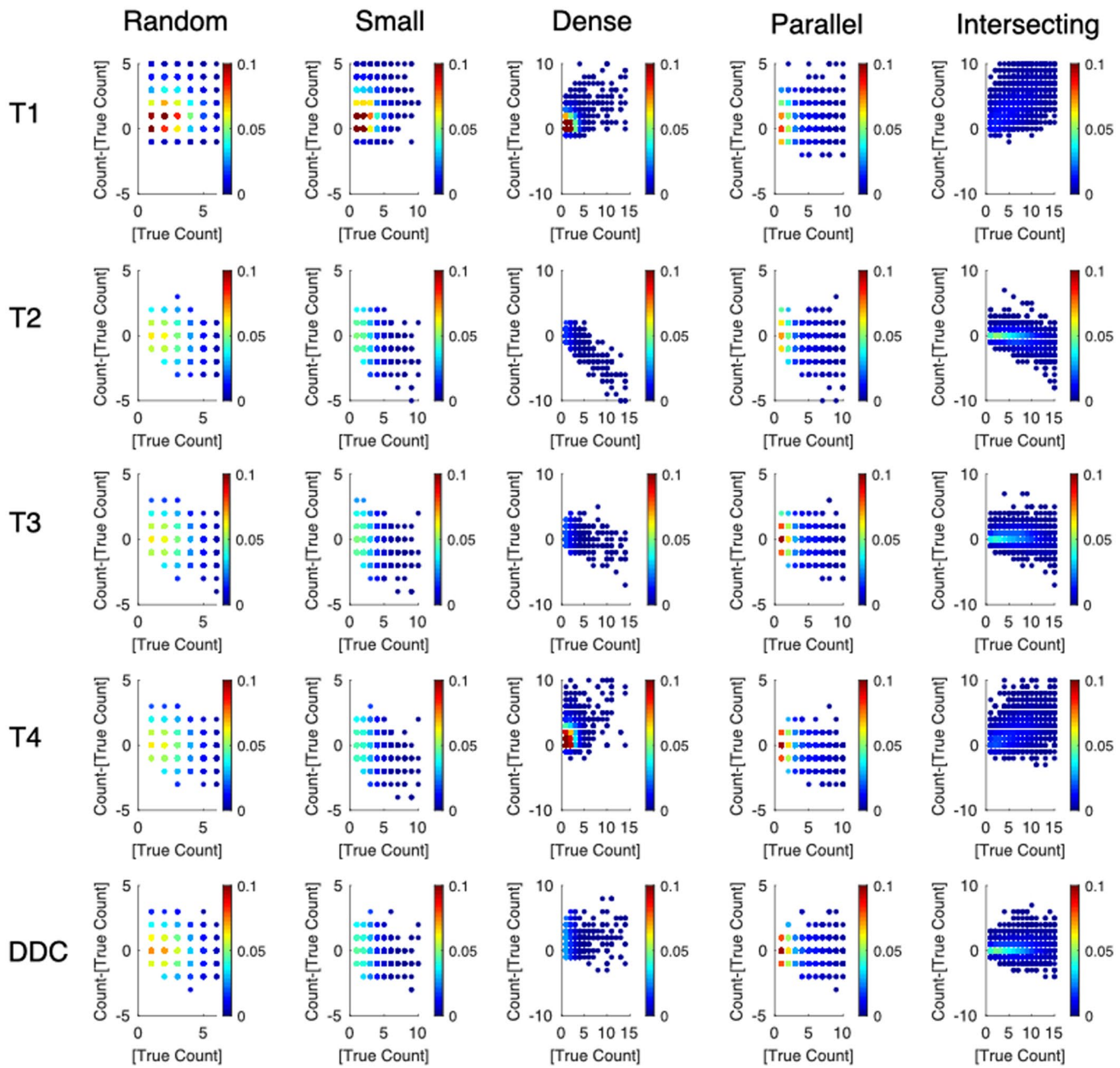


Extended Data Fig. 4 | See next page for caption.

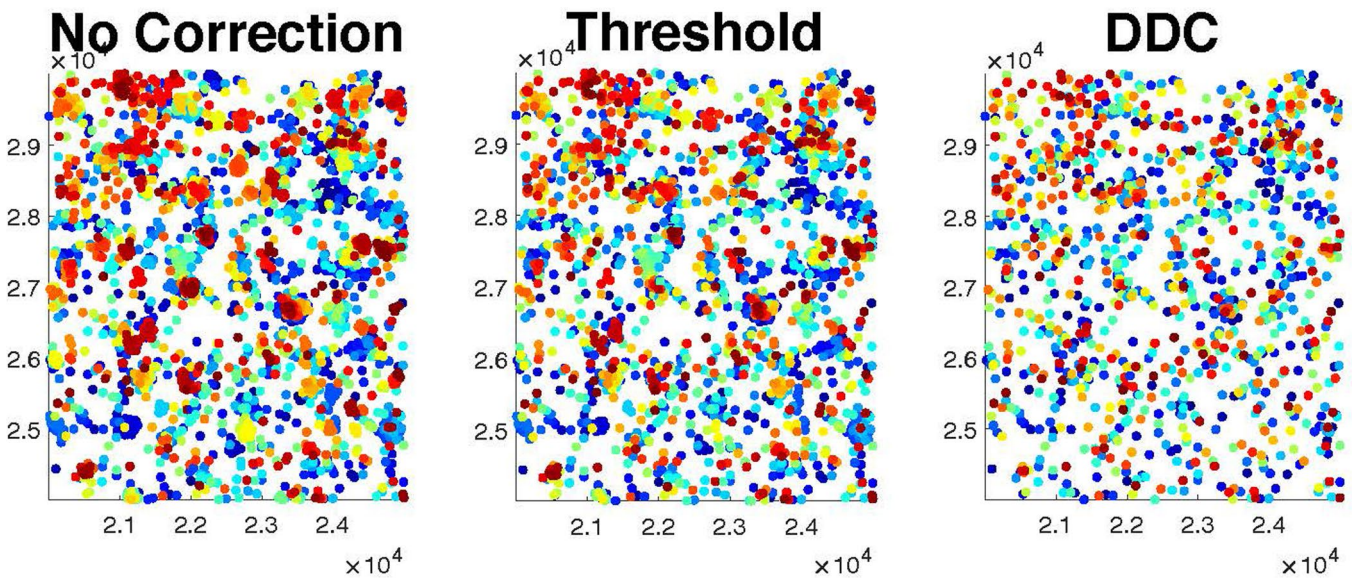
Extended Data Fig. 4 | Toy model illustration for inner workings of DDC. Toy model illustration for inner workings of DDC (See text within SI for in depth description): **a**, Simple toy model with 4 true localizations and 2 repeats (color coded), with the number showing the frame of each localization (can also be used to identify each localization for this example). **b**, The true pairwise distance distribution ($P_T(\Delta r)$) and the distribution of distances between loci given that at least one is a repeat ($P_{R1}(\Delta r|\Delta n=1)$) for the localizations within (**a**) The number (and probability) for 'small' distances and 'large' distances for each distribution is above each bar, with an assigned variable (**a**, **b**, **c**, **d**) used in the calculation of the Likelihood (Lik). We also show the specific pairs of loci under the bars to illustrate how assigning a particular loci to a certain set influences the likelihood calculation. Note: for this specific example blinks only appear with $\Delta n = 1$, and hence we ignore the distributions with $\Delta n > 1$ (See text). **c**, Simplified illustration of how Alg. 1 and Alg. 2 work together and assign localizations as a true localization or repeat localization. Multiple steps of the MCMC are shown with different rows (1 to 3) (See Text). Alg. 1 essentially calculates the probability that a localization is a repeat (green bars), if this value is above .5 it is assigned to that set. Alg. 2 varies this calculation by a small amount each step, generating new sets **d**, The sets assigned in (**c**) lead to different likelihoods (due to the particular distribution the distance between each pair is assigned (changing (**a**,**b**,**c**,**d**), note how the specific distances between each pair change with each assigned set), when the distributions of the assigned sets match the correct distributions (those in (B)) Lik is maximized. (See text for further details).



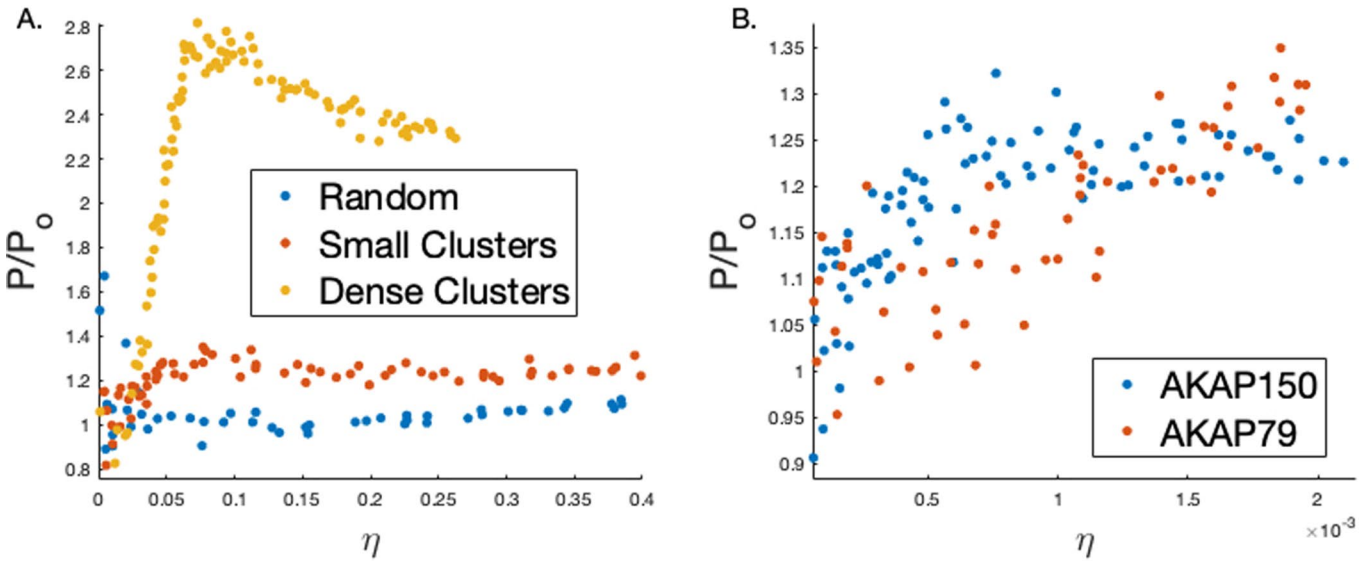
Extended Data Fig. 5 | Maximization of Likelihood Results in Correct Conformation of Localizations. Maximization of Likelihood Results in Correct Conformation of Localizations: For 6 systems investigated within this work, we randomly varied the percentage of true localizations and calculated the log(Lik) and the image error for each conformation (See Text).



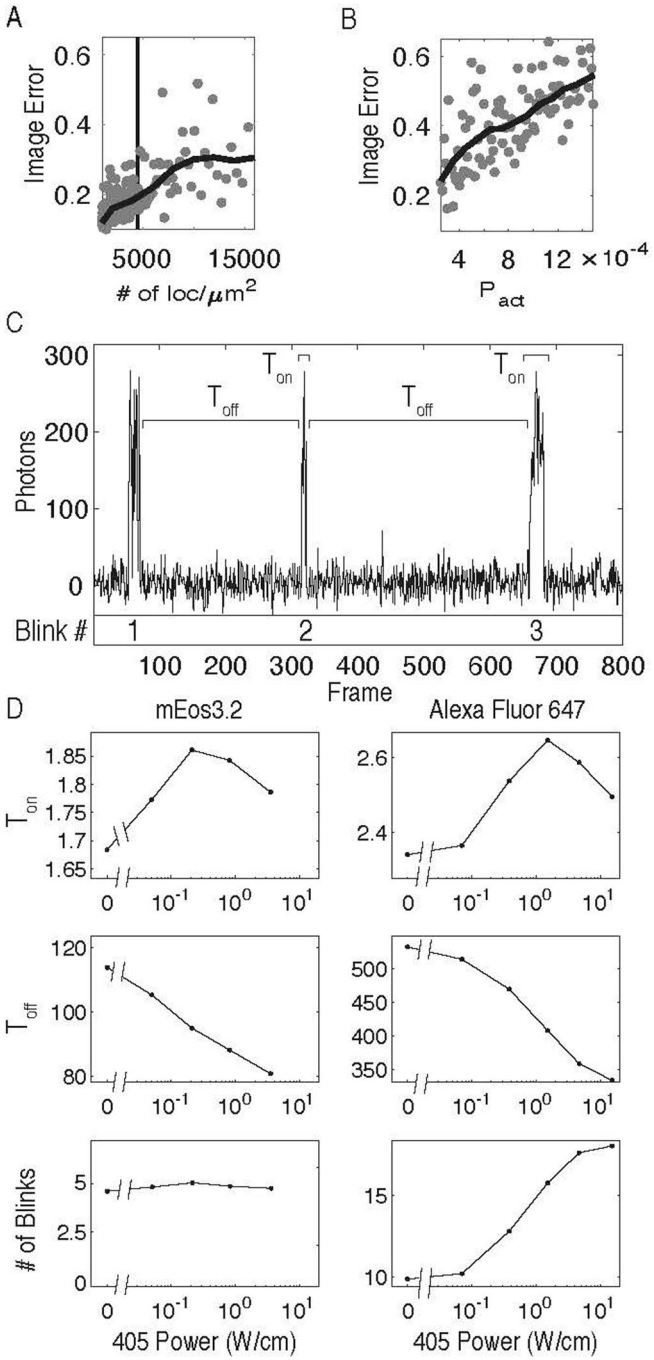
Extended Data Fig. 6 | Overcounting and undercounting in individual pixels. Overcounting and undercounting in individual pixels: Comparison of four different thresholding methods with DDC in counting the number of true localizations in individual pixels on five spatial distributions as depicted and simulated in main text Fig. 2. The y axis is the difference between the true count and the method-identified count expressed as Count-[True Count], with positive values indicating the degree of over-counting and negative values the degree of under-counting. The x-axis is the number of true counts in individual pixels. The pixel size was set to 50 nm. Note that only DDC shows consistent distributions of y values near zero at different true count values and across all five spatial patterns. [Each scatter point is colored to illustrate the estimated probability density - allowing one to visualize the regions of high density (red) and regions of low density (blue).].



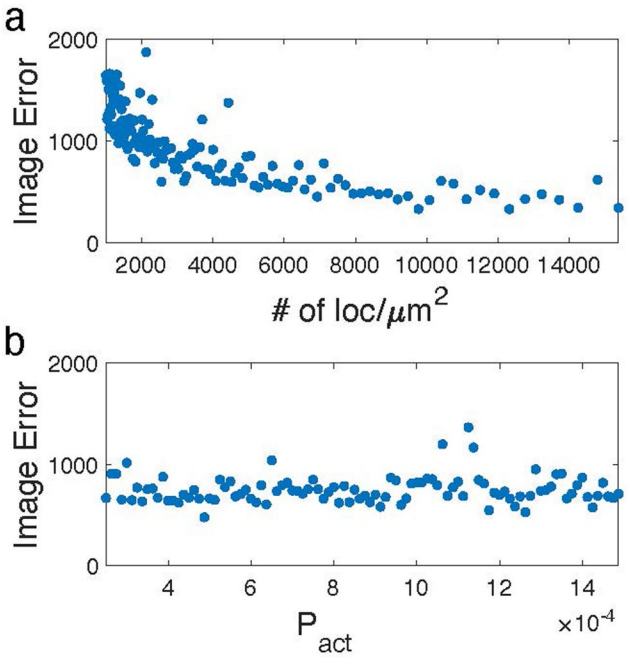
Extended Data Fig. 7 | AKAP scatter plots through time. Scatter plots for a section of a cell with the localizations from AKAP79 with the color indicating the frame of the localization (Blue is early and Red is late) for the three different methodologies.



Extended Data Fig. 8 | Computationally varying the label density. a, The results of computationally varying the label density on some of the simulation systems. **b**, The results of computationally varying the label density on AKAP79 and AKA150. (Values greater than 1 indicate significant clustering.).



Extended Data Fig. 9 | Experimental Concerns. Image Error at different densities of localizations (**a**) and activation probability per frame (**b**). The raw data points are shown as gray points and the moving average is shown in black (Supporting Material). **c**, An intensity trajectory of a single mEos3.2 molecule with labels showing the definitions of T_{on} and T_{off}. **d**, The average Ton, Toff (per frame, frame rate 33Hz), and number of blinks for Alexa647 and mEos3.2 at different UV activation intensities (405 Power).



Extended Data Fig. 10 | Varying Raw Image Error. The raw Image Error (Not Normalized) for the uncorrected SMLM images for varying the density of the localizations and the activation energy.

nature research

Corresponding author(s): Jie Xiao

Last updated by author(s): 2021-3-15

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	All custom code written was in MATLAB [https://github.com/XiaoLabJHU/DDC] (Version 2018), NIS-Elements AR software (v3.22.00), NimOS localization software (Version: 1.1.6115-2b51109), Metamorph 7.8.13.0,
Data analysis	Custom code written in MATLAB [https://github.com/XiaoLabJHU/DDC] (Version 2018), ImageJ (Version 1.53c), ThunderSTORM(dev-2016-09-10-b1)

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data of the manuscript are shown in the main figures and Extended Data figures (Source Data). All raw data for each of the simulation systems (Figure 1 and Figure 2) are also included at https://github.com/XiaoLabJHU/DDC. All other data are available upon request.

Field-specific reporting

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

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	Sample size was determined by the availability of data from experiment and was not limited in other cases. For each experimental system images were pooled from different regions of the same sample and on multiple samples on different cover slips.
Data exclusions	No data was excluded within this work except for initial experiments that obviously failed (different concentrations or wrong imaging setup). The need for adequate images was predetermined.
Replication	Fig. 3: Images of 3947 nuclear pore complexes were gathered from from multiple cells and showed similar results with the characteristic NPC structure. Final data was combined and analyzed together in the end. Fig. 4: Images of 141694 AKAP150 Clusters were gathered from 7 different cells of which all methodologies showed similar differences across the different cells. Data was combined in the end to generate the final distributions. Fig. 5: Images of 184368 individual dynein motor assemblies were acquired from 5 different cells of which all methodologies showed similar differences across the different cells. Data was combined in the end to show the bar plots. Fig. 6: Samples were prepared and imaged two separate times with similar image results. Extended Data Fig. 9: Blinking characteristics were quantified using two biological replicates which showed similar trends.
Randomization	Samples were not randomized. Our experimental workflow did not allow/need randomization. Randomization for bootstrapping was performed with the bootstrap function built into Matlab. ROIs for cells were selected at random for analysis.
Blinding	Blinding was not done. Our experimental workflow did not allow/need blinding

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used	1. anti-AKAP150: Millipore Sigma, Catalog No: 07-210, Clonality: Polyclonal 2. chicken polyclonal anti GFP, Abcam, Catalog No: 13970, Clonality: Polyclonal 3. donkey-anti chicken, Jackson ImmunoResearch, RRID (Research Resource Identifiers): AB_2340346, Catalog No: 703-005-155, Clonality: Polyclonal. 4. goat anti-rabbit, invitrogen, Catalog No: A21245, Clonality: Polyclonal
Validation	l. Routinely evaluated by immunoblot on a rat brain cytosolic preparation, further info can be found at (https://www.emdmillipore.com/US/en/product/AntiAKAP-150-Antibody, M M_N F-07-2 l0?cid= Bl OS-S-EPDF-1148-1107-RC) 2. Tested Applications: IHC-P, WB, IHC - Wholemount, IHC-FrFI, ICC/IF, IHC-Fr, IHC-FoFr, with positive control: Transgenic mouse spinal cords expressing GFP. ICC/IF: GFP-transfected NIH/3T3 cells. (https://www.abcam.com/gfp-antibody-ab13970.html) 3. The antibody has been tested by ELISA and/or solid-phase adsorbed to ensure minimal cross-reaction with bovine, goat, guinea pig, syrian hamster, horse, human, mouse, rabbit, rat and sheep serum proteins (https://www.jacksonimmuno.com/catalog/products/703-005-155) 4. This antibody has been tested with Immunofluorescence and Flow cytometry, (https://www.thermofisher.com/antibody/product/Goat-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21245)

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	1. human U-2 OS genome edited NUP96-mEGFP cells, clone 195 (cat. 300174) was obtained from CLS Cell Lines Service GmbH (Eppelheim, Germany) 2. Hela IC74-mfGFP cells (The dynein intermediate chain is GFP labeled, from Takashi Murayama lab, Juntendo University School of Medicine, Tokyo, Japan) 3. MIN6 (Miyazaki laboratory, Osaka University)
Authentication	1. Confirmed with Super-resolution imaging 2. Hela IC74-mfGFP was verified by western blotting and superresolution imaging. 3. MIN6 line was not authenticated
Mycoplasma contamination	1. U2OS-Nup96-mEGFP were not tested 2. Hela IC74-mfGFP cells were not tested 3. MIN6 cell lines were determined to be free of mycoplasma contamination based on weekly DNA staining.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used within this study.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Species: Drosophila melanogaster Strain: yellow white (yw) Sex: Male and Female Age: <2hours
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

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