ARTICLES
https://doi.org/10.1038/s41592-022-01463-w

nature methods

Check for updates

Correction of multiple-blinking artifacts in photoactivated localization microscopy

Louis G. Jensen[1]✉, Tjun Yee Hoh[2], David J. Williamson[3], Juliette Griffié[4], Daniel Sage[5], Patrick Rubin-Delanchy[2,7]✉ and Dylan M. Owen[6,7]✉

Photoactivated localization microscopy (PALM) produces an array of localization coordinates by means of photoactivatable fluorescent proteins. However, observations are subject to fluorophore multiple blinking and each protein is included in the dataset an unknown number of times at different positions, due to localization error. This causes artificial clustering to be observed in the data. We present a 'model-based correction' (MBC) workflow using calibration-free estimation of blinking dynamics and model-based clustering to produce a corrected set of localization coordinates representing the true underlying fluorophore locations with enhanced localization precision, outperforming the state of the art. The corrected data can be reliably tested for spatial randomness or analyzed by other clustering approaches, and descriptors such as the absolute number of fluorophores per cluster are now quantifiable, which we validate with simulated data and experimental data with known ground truth. Using MBC, we confirm that the adapter protein, the linker for activation of T cells, is clustered at the T cell immunological synapse.

Single-molecule localization microscopy (SMLM) methods, such as photoactivated localization microscopy (PALM), circumvent the diffraction limit of light by separating fluorophore detections in time through stochastic activation and photobleaching, and then localizing the resulting sparse distribution of point spread functions (PSFs)[1]. The resulting point pattern is a purported realization of the underlying ground-truth positions of the fluorophores, but is corrupted by a number of artifacts resulting from the photophysical behavior of the probes as well as the imaging and localization steps. Most problematic is the multiple appearance (multiple-blinking) problem whereby fluorophores undergo multiple on-off cycles before permanently bleaching, making a single fluorophore appear as several localizations. This issue is exacerbated by the discretization effects that result from observing fluorescent signals on discrete camera frames[2]. The multiple-blinking problem produces datasets that are artificially clustered and overly populated (Fig. 1a). As such, quantitative cluster analysis of SMLM data, including testing for spatial randomness of the underlying fluorophores, remains a challenge.

The most commonly used method for correction of the multiple-blinking problem is to merge events that appear close in space and time[3–6], which we refer to as dark time thresholding (DTT). Such methods require a means of determining the best spatial and temporal thresholds for merging. This determination typically relies on heuristic methods, since the blinking behavior of the fluorescent probes is often unknown. Apart from the challenges involved in determining optimal thresholds, these methods have variable performance, depending on the underlying protein organization and fluorophore blinking characteristics. Instead of attempting to produce a corrected version of the data that can then be used for any subsequent analysis, other approaches have looked to correct specific spatial statistics to account for multiple blinking. For example, it is possible to use calibration data to estimate a multiple-blink corrected pair-correlation curve[7,8]. However, this cannot then be used to find a cluster map.

Here, we present a new method for correction of multiple-blinking artifacts in PALM data, which estimates, directly from the sample dataset, the parameters of a realistic model of fluorescent protein photophysics[9]. Cluster analysis of the spatial (x,y), temporal (t) and precision (σ) dataset then allows computation of the marginal likelihood of any given blink-merge proposal, under a full generative model for the data. We select the most likely of several proposals generated using a customized hierarchical clustering algorithm. Finally, each blink cluster is consolidated into a single position, now free from multiple blinking. As with other correction methodologies[5], an added benefit of this consolidation step is to improve localization precision. The overall effect is to convert the set of raw x,y,t,σ localization data into a new set, x,y,σ, with enhanced resolution.

We evaluate the method on simulated PALM data, varying both the ground-truth organization and photophysical properties of the fluorescent proteins. In each case, we compare to the state-of-the-art method of Bohrer et al.[10], and DTT, outperforming alternatives in speed and accuracy. Our method allows for testing the completely spatially random (CSR) hypothesis at the correct significance level, whereas DTT fails to do so. We also validate the method on experimental data in which a notion of ground truth is available, namely nuclear pore complex (NPC) data in which the number of proteins per complex is known to be 32.

PALM is increasingly used in the biological sciences and owing to the properties of commonly used total internal reflection fluorescence (TIRF) illumination, the distributions of membrane proteins have been especially well studied. Despite this, because of artificial clustering resulting from multiple blinking, the question of whether membrane proteins are randomly distributed or not has become

[1]Department of Mathematics, Aarhus University, Aarhus, Denmark. [2]Institute for Statistical Science, School of Mathematics, University of Bristol, Bristol, UK. [3]Randall Centre for Cell and Molecular Biophysics, King's College London, London, UK. [4]Laboratory of Experimental Biophysics, Institute of Physics, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [5]Biomedical Imaging Group, School of Engineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [6]Institute of Immunology and Immunotherapy, School of Mathematics and Centre of Membrane Proteins and Receptors, University of Birmingham, Birmingham, UK. [7]These authors contributed equally: Patrick Rubin-Delanchy, Dylan M. Owen. ✉e-mail: louis.gammelgaard@gmail.com; patrick.rubin-delanchy@bristol.ac.uk; d.owen@bham.ac.uk

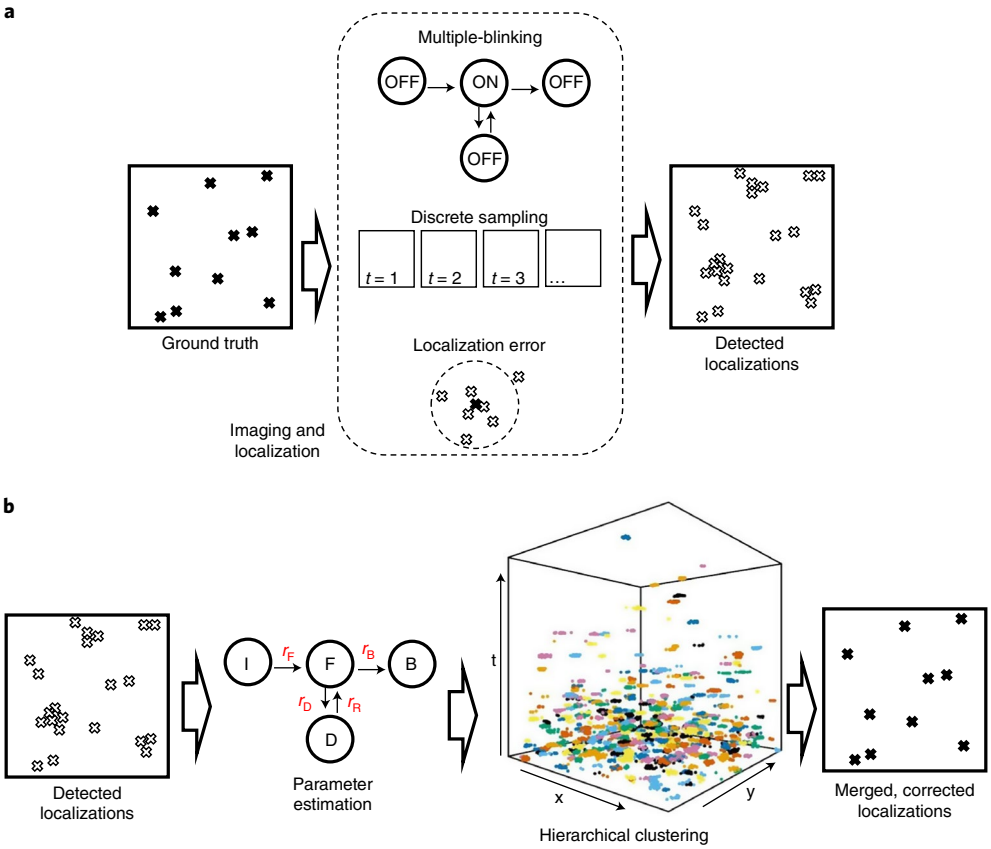


Fig. 1 | Illustration of the MBC workflow. a, During PALM image acquisition and subsequent localization steps, the ground-truth protein positions are corrupted by multiple blinking in combination with discretization by the camera frames and scrambling by the localization uncertainty, resulting in a dataset that is over-populated and over-clustered. **b**, Our algorithm (MBC) takes as input x,y,t,σ data and estimates the rate parameters of a four-state photophysical model, from which it derives the total number of molecules in the ROI. This is then used as input to a hierarchical clustering step (experimental data shown with colors representing the clusters found), after which clusters are merged to their centers, creating a new dataset free from multiple-blinking and with enhanced localization precision.

increasingly contentious[11]. Using our validated method combined with subsequent testing of the corrected protein locations, we show that the adapter protein Linker for Activation of T cells is clustered in the plasma membrane of T cell immune synapses[12,13].

Results

Description of the algorithm. We work with the space–time localizations and uncertainties that result from localization software (here, ThunderSTORM[14]) that is run on the raw microscope data. We apply drift correction, but otherwise no preprocessing is used. The data points are then modeled as a collection of independent and identically distributed fluorophore blinking clusters, with times following a realistic four-state model[15,16], discretized by the camera frames. The spatial locations for each cluster are independently drawn from a spherical Gaussian distribution of fixed center (the true molecule position) and variable but known standard deviation (the localization uncertainty). The centers are given a uniform prior over the region of interest (ROI).

We refer to our algorithm as model-based correction (MBC), and a schematic of its workflow is shown in Fig. 1b. We first estimate the temporal rates governing the switching behavior of fluorescent proteins under the four-state model[9], and the fraction of background noise points. This is done directly using the experimental data, requiring no additional calibration experiments. A recently developed mathematical technique extracts a component from the empirical mark and pair-correlation functions that depends only on the spatio-temporal dynamics of the multiple-blinking process, and not the underlying protein distribution. The parameters of the four-state model drive the theoretical shape of this component, and so they can be optimized to best fit the empirical version[9]. The rate estimates allow computation of the marginal likelihood of a sequence of time-points posited to correspond to one multiple-blinking fluorescent protein, and further yields an estimate on the total number, N, of proteins and noise points in the ROI. Using a custom agglomerative hierarchical clustering (AHC) algorithm[17], we split the data in the ROI into partitions with N categories. AHC takes as input a dissimilarity matrix and a linkage criterion. The dissimilarity matrix determines the distances between pairs of points, and the linkage criterion determines the way to generalize this distance to pairs of clusters. To favor groups likely to correspond to multiple-blinking clusters, we first scale the temporal dimension by a time-dilation hyperparameter, S, and then compute the sum of Euclidean distances in space and in time. For linkage, we choose Ward's minimum variance method[18], which is well-suited for Gaussian clusters, and consistently resulted in the most likely partitions across all tested linkage criteria. By varying S, we obtain a large sequence of blinking cluster proposals, and evaluate the marginal likelihood of each. Finally, using the best partition and the localization uncertainties, we optimally merge the clusters down to their estimated centers, using inverse-variance weighted averages, and update the uncertainty associated with that center (example shown in Supplementary Fig. 1 and three-dimensional (3D) rotatable version in Supplementary Data).

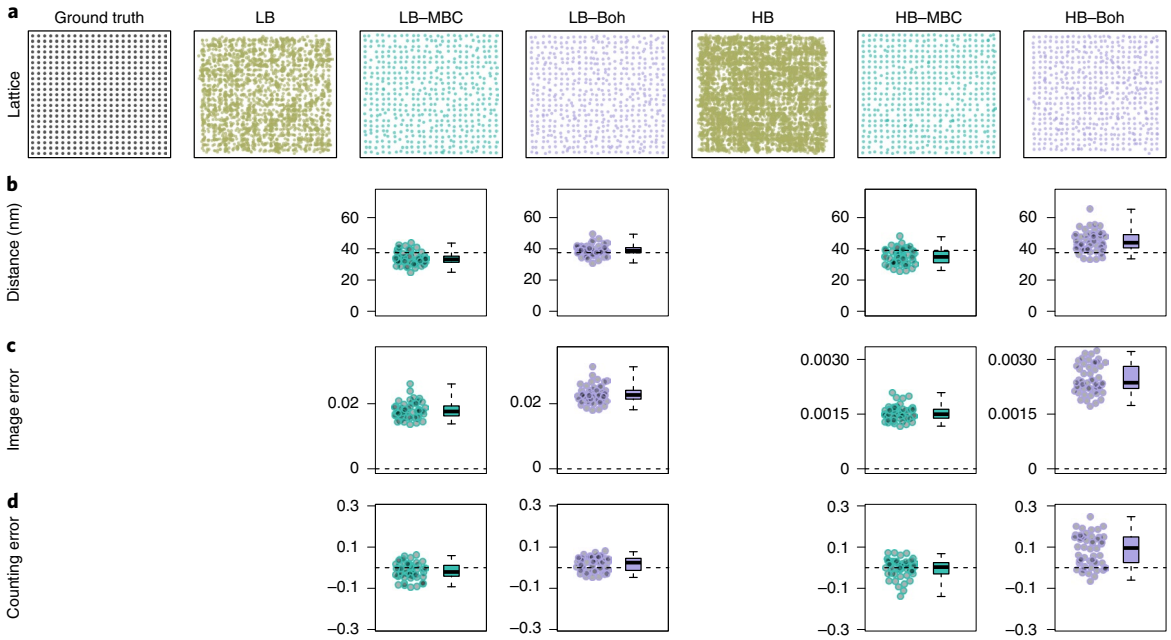


Fig. 2 | Comparison of MBC with the state of the art, with molecules on a fixed grid. a, Representative ground truth, blinking and corrected data (from one of $n = 50$ realizations). **b**–**d**, Wasserstein distances (**b**), image error (**c**), counting error (**d**), between corrected data and ground truth. **b**, The dashed horizontal line shows the 37 nm benchmark. **c**,**d**, The dashed horizontal line shows the optimal value 0. The columns show different blinking conditions (LB, light blinking and HB, heavy blinking) and correction methods (MBC, our method; Boh, Bohrer et al.'s method). Our method shows superior or comparable performance on all metrics. Box plots show median, 25th and 75th percentiles and minimum and maximum.

PALM data simulation setup. For a given set of protein positions, corresponding PALM data were generated as follows. We simulated fluorescent protein time traces according to the four-state switching model (Fig. 1a), and the continuous signals were discretized to emulate a camera operating at 25 frames per second (40 ms integration time). This was done for two different sets of rates (given in Supplementary Table 1), with the light blinking resulting in 5.36 appearances per protein on average, and the heavy blinking resulting in 14.94 appearances. These were selected to approximate real blinking behavior of fluorescent proteins such as mEos2. For each of these appearances, the observed spatial coordinates were simulated by adding Gaussian localization noise to the ground-truth position of the associated fluorescent protein, with standard deviation following a Gamma distribution with mean 30 nm and standard deviation 13.4 nm, emulating the localization uncertainties that can be observed in real PALM data[12].

Recovery of the ground truth. We begin by comparing MBC to the method of Bohrer et al., which we take to be the state of the art at the task of recovering protein positions without calibration data. We will conduct this comparison with simulated data from regular (529 fluorescent proteins were regularly positioned on a roughly 3,000 × 3,000 nm grid), CSR (500 proteins were placed at random in a noiseless 3,000 × 3,000 nm ROI) and clustered distributions. For the latter, 500 ground-truth proteins were placed in a 3,000 × 3,000 nm ROI, with either ten clusters of ten molecules each, overlaid with 400 CSR molecules (light clustering) or ten clusters of 40 molecules each, overlaid with 100 CSR molecules (heavy clustering). The metrics used for the comparison are Wasserstein distance, image error and counting error. The Wasserstein distance can be thought of as the cost of transporting a standardized mass between two sets of points and is also known as the earth mover's distance. For a perfectly reconstructed dataset this distance is zero, whereas under a model in which each molecule was observed exactly once with localization error drawn from the Normal-Gamma distribution above, this distance would be roughly 37 nm. In this way, whenever we report a Wasserstein distance below 37 nm, we can conclude that we have exploited, rather than suffered from, the multiple-blinking artifact. Relevant plots show this 37 nm benchmark. The image error measures the protein count error in 100 × 100 nm pixels, whereas the counting error refers to the total number of proteins per ROI, calculated as (corrected-truth)/truth.

In the main text we show results at the two extremes, regular (Fig. 2) versus heavy clustering (Fig. 3), with the two other conditions shown in Extended Data Fig. 1. Our method is always superior in Wasserstein distance and image error, in counting error for heavy blinking and never substantially worse. Our gains are most substantial under heavy clustering with heavy blinking, where Bohrer et al.'s method overcounts proteins by a factor of at least two (Fig. 3). Our method is also much faster, taking 1 or 2 min per ROI on a standard desktop versus several hours.

As an alternative to Bohrer et al., we also consider an idealized version of DTT (iDTT), in which points are considered to have come from the same fluorophore if they were separated by at most r in space and T in time, chosen to minimize either the counting error (iDTT_N) or Wasserstein distance (iDTT_W). The reason we consider this choice, which is unavailable in practice, is to cover all possible methods for choosing r and T, including those involving calibration data. Results are shown for normally distributed data under light and heavy blinking in Extended Data Fig. 2, respectively. Our method is always superior on Wasserstein distance and image error (even when compared to iDTT_W), and superior in counting error when compared to iDTT_W.

To test MBC in dense scenarios, we repeat all experiments with five times the density, chosen to mimic the maximum density observed in any of the experimental data analyzed in this paper (2,500 proteins per 3,000 × 3,000 nm ROI). When it comes to denser scenarios, Bohrer et al.'s method can take 10 days per ROI, and optimizing over space and time thresholds to perform iDTT is also computationally demanding. However, MBC is still functional,

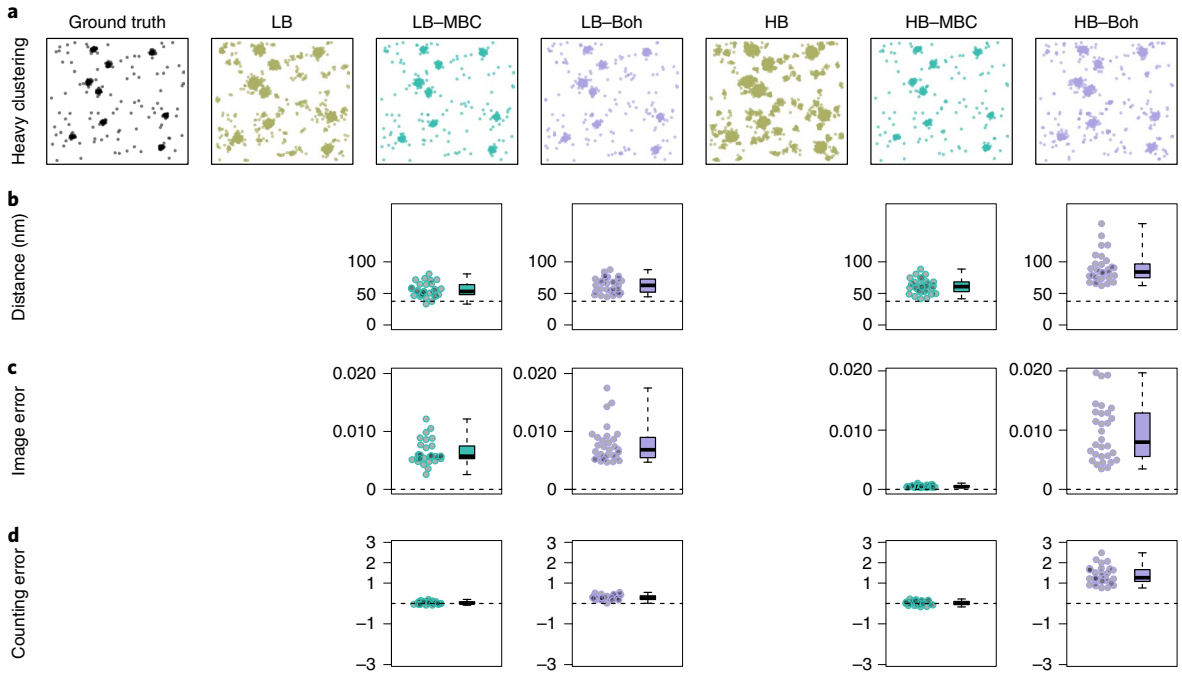


Fig. 3 | Comparison of MBC with the state of the art, with heavily clustered molecules. a, Representative ground truth, blinking and corrected data (from one of $n=30$ realizations). **b**–**d**, Wasserstein distances (**b**), image error (**c**), counting error (**d**), between corrected data and ground truth. **b**, The dashed horizontal line shows the 37 nm benchmark. **c**,**d**, The dashed horizontal line shows the optimal value 0. The columns show different blinking conditions and correction methods. Our method shows superior performance on all metrics. Box plots show median, 25th and 75th percentiles and minimum and maximum.

and its performance is compared to DTT across all conditions and metrics in Extended Data Figs. 3 and 4. For DTT, we use parameters r equal to four times the mean localization uncertainty in space, and T determined for each ROI using the method of Annibale et al.[4]. For each of four scenarios; regular and CSR (Extended Data Fig. 3) and light clustering and heavy clustering (Extended Data Fig. 4), MBC substantially outperforms DTT in all three metrics (Wasserstein distance, image error and counting error).

Having established that MBC outperforms Bohrer et al.'s method and iDTT for recovery of ground truth in all conditions, at an extremely reduced computational cost, we proceed with comparing our method only to DTT due to its simplicity, speed and wide uptake (including in commercial software).

Testing for complete spatial randomness. We next evaluate our algorithm for testing for complete spatial randomness of the underlying ground-truth proteins. In each run ($n=30$ per condition), 500 proteins were placed at random in a noiseless 3,000 × 3,000 nm ROI. For each ROI, we compute the function $L(r)-r$ (Fig. 4b), where L is Besag's L-function[19], testing its maximum (Fig. 4c) under a CSR null hypothesis. The standard DTT correction method was unable to recover the ground-truth functions and resulted in rejection of the CSR null hypothesis in 24 and 30 out of the 30 regions, for light and heavy blinking respectively. On the other hand, MBC resulted in the CSR null hypothesis being rejected for two and four of the regions for light and heavy blinking, respectively. These numbers are within the expected range at a 5% confidence level. Thus, we were able to reliably test the CSR hypothesis using MBC, but not using DTT. The estimated total number of fluorescent proteins in each ROI is shown in Fig. 4d. Under CSR, DTT tends to overestimate the number of proteins in the ROI whereas MBC closely recovers the ground truth.

Cluster analysis. In this experiment, we demonstrate that a clustering algorithm can extract correct cluster descriptions from underlying clustered ground-truth protein distributions when coupled with MBC, and we compare performance with DTT. We simulated data from protein distributions exhibiting light and heavy clustering ($n=30$ per condition), as described in the recovery of ground-truth section (Fig. 5a). We used Bayesian cluster analysis[20,21] for detection of clusters in MBC and DTT corrected datasets. Only under MBC could we consistently recover the ten clusters under varying degrees of blinking severity (Fig. 5b). The failure of DTT to recover the correct number of clusters is even more evident in the case of heavy clustering (Fig. 5c,d).

Determining optimal imaging conditions. As a final test using known ground-truth simulated data, we used Virtual-SMLM[22] to simulate raw camera frames. This allowed us to test the effect of varying both the camera frame rate and the intensity of the 405 nm activation laser on the performance of MBC. A ground truth of CSR fluorescent proteins was simulated (Extended Data Fig. 5a), imaged using the virtual microscope and output analysis with ThunderSTORM. The camera integration time was set to either 10 or 40 ms and the 405-nm laser intensity either kept constant, or ramped up to maintain a constant density of PSFs per frame over the course of the acquisition. Raw localizations (Extended Data Fig. 5a) were then corrected using MBC (Extended Data Fig. 5b). The Wasserstein distance shows marginally superior performance of the reconstruction when using a constant 405-nm laser power and when using longer 40 ms frames. We attribute this to the lower density of PSFs per frame in the constant 405 case leading to fewer overlapped PSFs during localization and to the increased localization precision offered by the longer frames (Extended Data Fig. 5c,d). The performance of MBC itself is only weakly dependent on the imaging conditions and, in each condition, we were able to recover the ground-truth number of molecules to within around 10% error (Extended Data Fig. 5e). We conclude therefore that when using MBC, PALM imaging conditions should be chosen to maximize

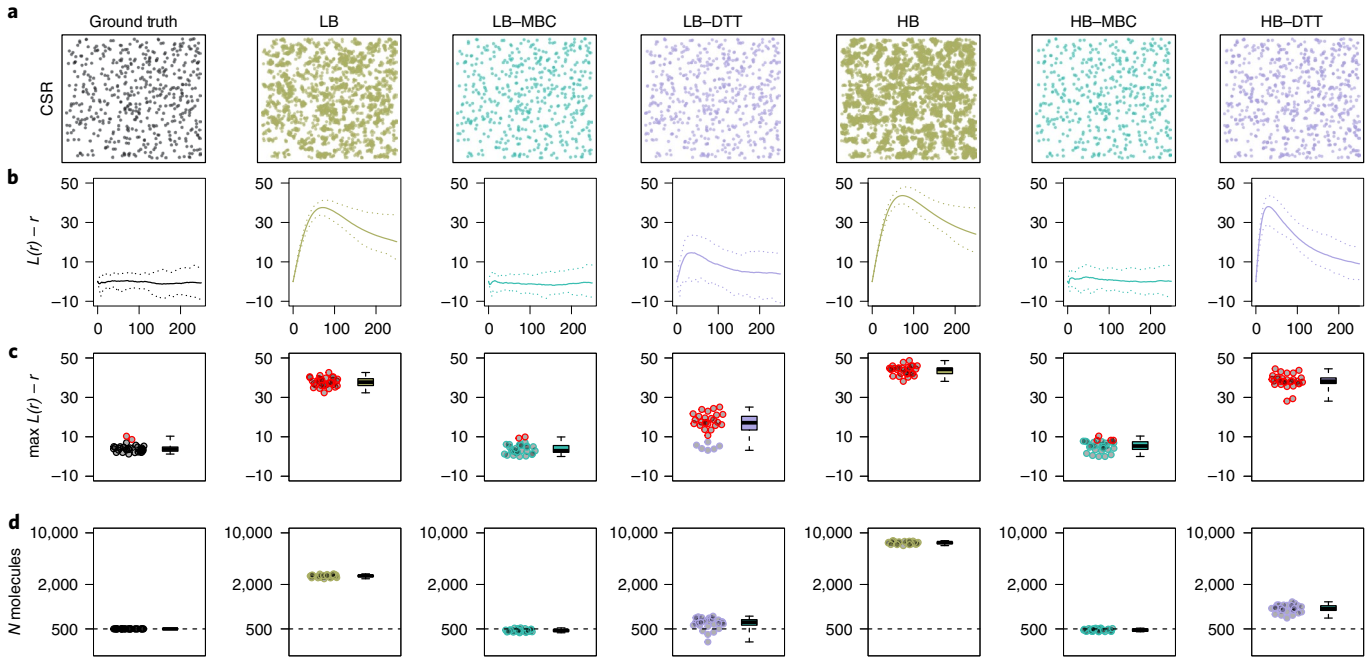


Fig. 4 | Testing for spatial randomness. a, Representative simulated data of ground-truth CSR points with light or heavy blinking either corrected by MBC or DTT as a comparison ($n = 30$ simulations). **b**, $L(r) - r$ (mean in solid line) with pointwise 95% quantile bands (dashed line). **c**, $\max(L(r) - r)$ derived from these functions. Points in red correspond to ROI that were rejected as CSR in a Monte Carlo test ($P < 0.05$). Note that DTT often (and sometimes always) incorrectly rejects the CSR null hypothesis, whereas MBC does not. **d**, Number of molecules per ROI (log-scaled) showing superior correction of MBC compared to DTT in light and heavy blinking cases. Box plots show median, 25th and 75th percentiles and minimum and maximum. *P* values calculated based on one-sided Monte Carlo test using 10,000 CSR simulations.

conventional notions of data quality—low density of PSFs and high signal-to-noise ratio. Because of this, we also conclude that MBC is also backward-compatible with all historically acquired PALM data.

Model validation. MBC is based on a statistical model for multiple blinking, camera discretization, localization error and a Bayesian prior distribution on the protein positions. In this section, we address concerns about model misspecification. First, we test our method on simulated data in which the multiple-blinking process follows a different model. Second, we test our method on NPC data in which there is a notion of ground truth, showing accurate recovery in a situation where the blinking dynamics have all the complexities of real fluorophores.

For the first experiment, the multiple blinking follows the three-dark-state photokinetic model shown in Jensen et al.[9] with parameters $r_F = 0.005$, $r_B = 2.5$, $r_{D1} = 4$, $r_{R1} = 0.25$, $r_{D2} = 4$, $r_{R2} = 1$, $r_{D3} = 4$ and $r_{R3} = 10$ (the numeric part of the subscript indicates the dark-state number). Overall, this produces slightly more total blinks than our previous heavy blinking case. Extended Data Fig. 6 shows the results of MBC on normally distributed data perturbed by this multiple-blinking process. MBC still performs well when the blinking shows the extra complexity of multiple dark states, still clearly outperforming DTT in this scenario. We believe the method should be robust, more generally, to misspecification of the blinking model as long as two assumptions can be made: first, the blinking behavior of different fluorophores is independent, and second, the process governing the blinking state switching is time-homogeneous.

Despite the performance on simulated data showing complex photoblinking behavior, real fluorophores in cells and experimental procedures such as imaging and localization have the potential to introduce many more complexities than we could simulate. We therefore investigate the performance of MBC on experimental reference data. We consider real biological cells from a recently developed cell line[23] expressing the nucleoporin Nup96, which forms NPCs. These complexes have a well-characterized ring shape, composed of 32 Nup96 positioned into eight identical corners. Knowing that each ring holds 32 proteins allows us, in principle, to compare corrected counts to a ground-truth value.

We consider 16 independent recordings of cells expressing the fluorophore Nup96-mMaple, as available on the BioImage Archive. Following Thevathasan et al.[23], we first filter out emitters of poor quality using the SMAP software[24]. In addition to filtering, SMAP also detects and segments out the NPCs present in each cell, and estimates the effective labeling efficacy (p_{ele}), the probability that a Nup96-mMaple is detected in the recording. For a given p_{ele}, which will vary for each recording, we can thus expect the number of detectable Nup96-mMaple in an NPC (N_{NPC}) to have a binomial distribution $N_{NPC} \cong \mathrm{Bin}(32, p_{ele})$.

After running MBC on each dataset, we cluster the corrected localizations into NPCs according to the closest NPC center. For each corrected NPC, we compute the number of localizations and divide by the p_{ele} relevant to that cell, to obtain a standardized count. Under an unbiased correction method this standardized count will have a mean value of 32 for all cells, allowing us to analyze and visualize all 16 cells simultaneously (Extended Data Fig. 7). By comparison to the theoretical distribution of ground-truth standardized counts (corresponding to perfect recovery) we see that MBC is closely emulating this optimal distribution. The percentage error on the mean number of proteins per NPC between MBC and the idealized case is only 1.1%, showing that MBC exhibits minimal bias. Our standard deviation is 6.16 and the ideal case is 4.6 (Bohrer et al. reported 11). Overall, experimental data acquired with a known ground truth show that MBC is accurate and unbiased in correcting multiple blinking, even through the full complexity of real fluorophore photophysics confounded by potential imaging and localization artifacts.

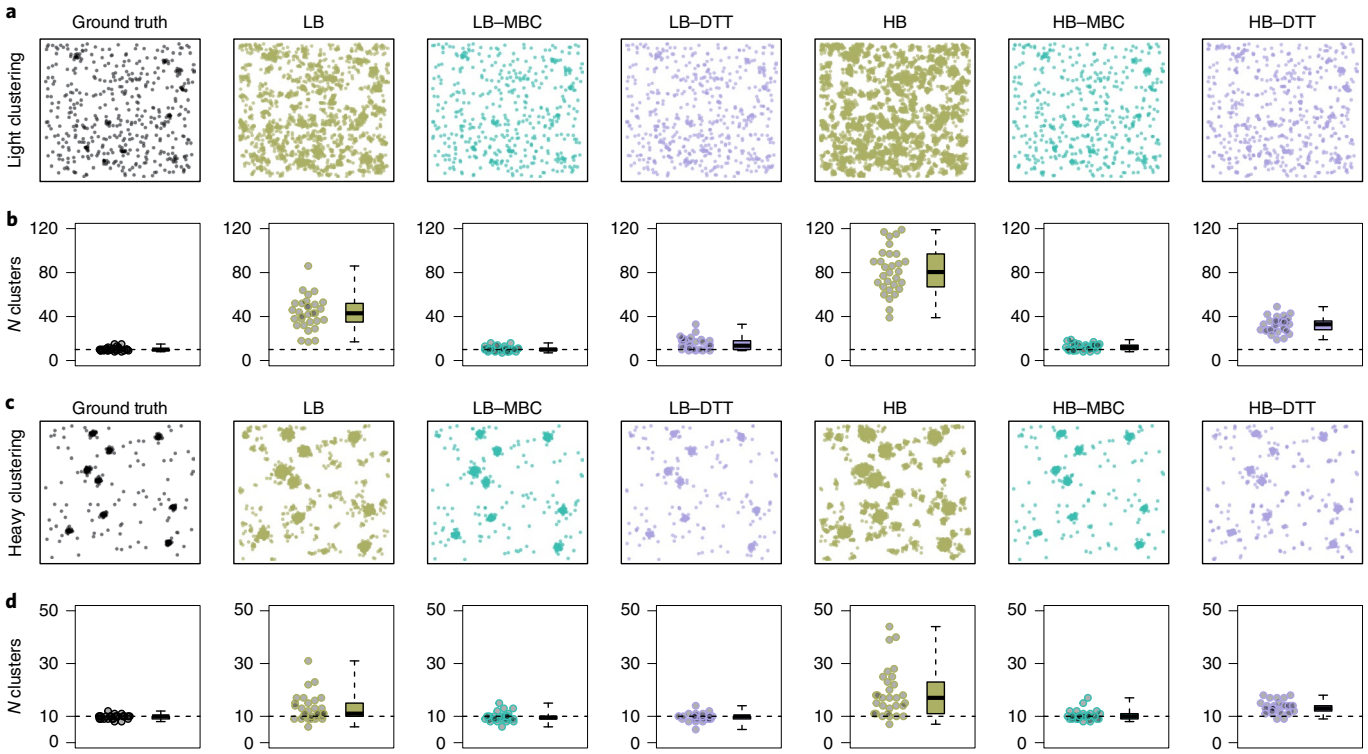


Fig. 5 | Testing on clustered ground-truth datasets. a, Low levels of clustering with either light or heavy blinking, corrected by MBC or DTT ($n = 30$ simulations). **b**, Number of detected clusters (true number of clusters in dashed line) by Bayesian analysis. **c**, High levels of clustering with either light or heavy blinking, corrected by MBC or DTT. **d**, Number of detected clusters (true number of clusters in dashed line) by Bayesian analysis. MBC has superior performance in all cases except heavy clustering/light blinking, where results are comparable. Box plots show median, 25th and 75th percentiles and minimum and maximum.

Analysis of experimental data. Nanoscale clustering is posited to play a role in regulating protein–protein interactions and therefore the efficiency of signal propagation along pathways[25]. T cell microclusters of proximal signaling molecules have been widely documented by conventional microscopy[26,27]. Many have recently been studied by SMLM and shown to also cluster on the nanoscale[12,13,28,29]. This has proved controversial, however, with counter-proposals that, in some circumstances, proteins may in fact be randomly distributed on the cell surface, with observed clustering attributed to multiple-blinking artifacts[11]. For PALM data, MBC should enable researchers to navigate this controversy.

We analyzed the distribution of an adapter protein, the linker for activation of T cells (LAT)[30], in the plasma membrane of the Jurkat CD4$^+$ Helper T cell line at an artificial immune synapse formed against an activating, antibody coated coverslip (Methods). To assess the role of intracellular phosphorylation in maintaining this distribution, we also mutated intracellular tyrosine residues to phenylalanine (YF LAT). Both wild-type (WT) LAT and YF LAT were fused to the photoconvertible fluorescent protein mEos3.2 with cells imaged under TIRF illumination. Raw localizations were obtained using ThunderSTORM and then corrected using MBC, with the average localization uncertainty across experiments reduced by 40.52%. The resulting corrected localizations were then tested for spatial randomness using the L-function, and any regions found to be clustered subjected to Bayesian cluster analysis[20] (using default parameters).

Figure 6 shows WT and YF LAT-mEos3.2 from representative regions (from $n = 12$–25 ROI from three to six cells) acquired from the central regions of the cell synapse and from the synapse periphery, both before (Fig. 6a) and after (Fig. 6b) correction using MBC. Clearly, the large, dense clusters evident in the uncorrected data in all conditions are reduced in the corrected regions. However, by analyzing the L-function curves from the ROI (Fig. 6c) and extracting the maximum value of those curves (Fig. 6d), we were able to perform significance testing on whether the LAT distributions were truly CSR. For the two WT LAT conditions, the null hypothesis that LAT is randomly distributed was rejected in most regions. Therefore, it is likely that WT LAT was clustered in most analyzed WT ROI. This was not true for the YF mutant, however, with the null hypothesis of randomly distributed LAT not rejected in most peripheral regions (Fig. 6d). This therefore may point to a role of intracellular tyrosine phosphorylation in maintaining LAT clustering.

For all regions where the CSR null hypothesis was rejected, we then further interrogated the data using Bayesian cluster analysis. For WT LAT, the data showed no statistically significant difference in cluster membership between central and peripheral regions. However, the YF mutant showed a significant decrease in the number of molecules per cluster in peripheral regions, both when compared to YF central regions ($P = 0.026$) and WT peripheral regions ($P = 0.001$) (Fig. 6e). Other outputs from the cluster analysis are shown in Extended Data Fig. 8, with P values summarized in Supplementary Table 2. The decrease in cluster membership and, in some ROI, the loss of clustering altogether, in peripheral regions of the T cell synapse resulting from the YF mutation, is a strong indication that intracellular tyrosine phosphorylation is involved in maintaining LAT signaling clusters. Signaling phosphorylation events are known to originate in the synapse periphery and it is therefore consistent that the effect of the mutation is most pronounced there, compared to the central region where signaling is terminated[31].

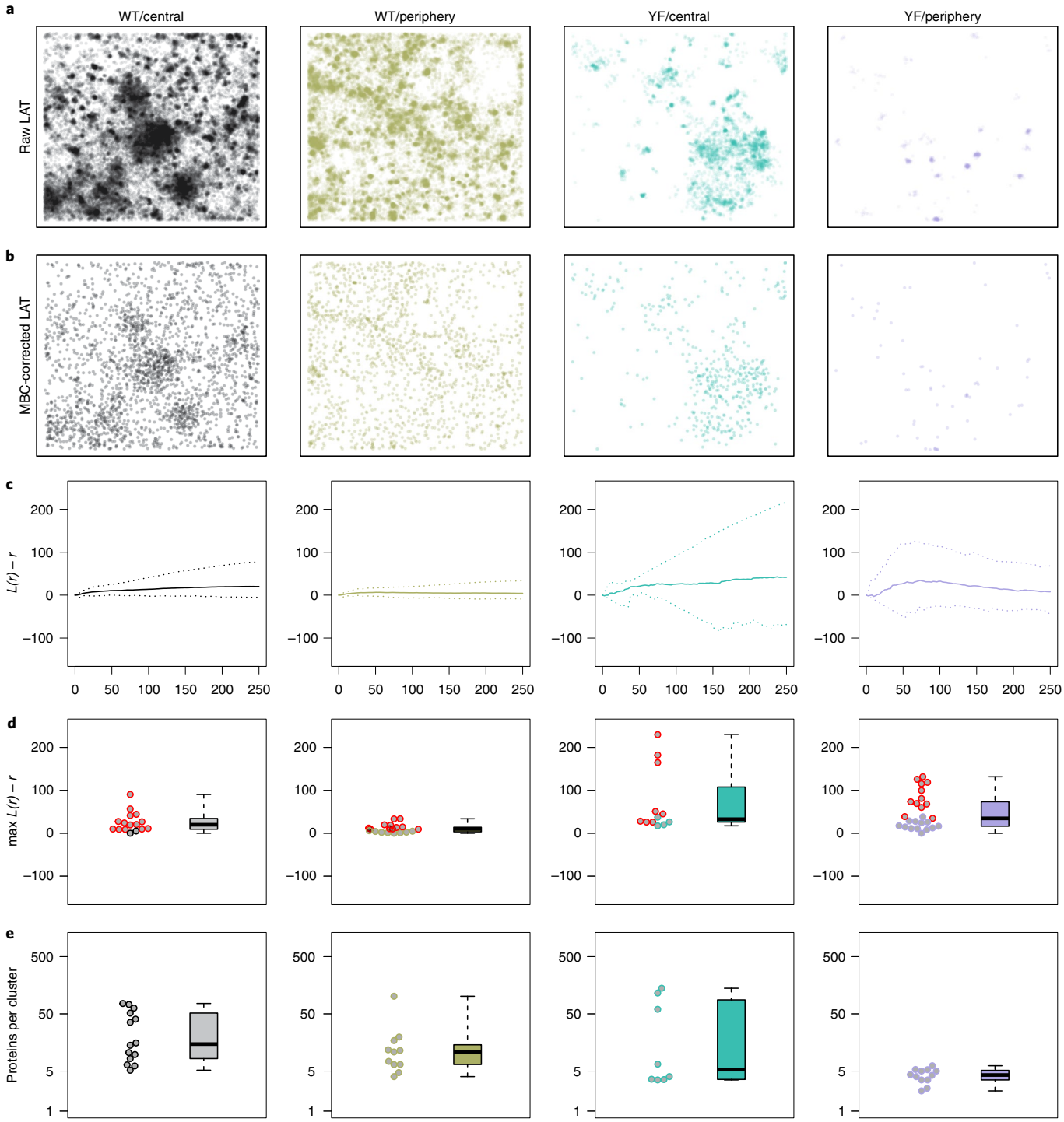


Fig. 6 | Cluster analysis of LAT-mEos3.2 at the T cell immunological synapse. a, Representative, raw 3,000 × 3,000 nm ROI from each of the four conditions (WT center ($n = 16$ regions), WT periphery ($n = 19$ regions), YF center ($n = 12$ regions) and YF periphery ($n = 25$ regions)). **b**, Representative MBC-corrected ROI, on which analysis was conducted. **c**, $L(r) - r$ (mean in solid line) with pointwise 95% quantile bands (dashed line). **d**, $\max(L(r) - r)$ derived from these functions. Points colored in red correspond to ROI where the CSR null hypothesis was rejected in a Monte Carlo test ($P < 0.05$). These ROI were then retained for subsequent Bayesian cluster analysis. **e**, Number of proteins per cluster detected by Bayesian analysis. Box plots show median, 25th and 75th percentiles and minimum and maximum. *P* values calculated based on one-sided Monte Carlo test using 10,000 CSR simulations.

Discussion

Super-resolution fluorescence microscopy by SMLM, such as PALM, results in a pointillist dataset representing an attempted realization of the underlying ground-truth fluorophore locations[1]. A common goal in the biological sciences is to test whether such underlying distributions are clustered or randomly distributed and, if clustered, to determine their clustering properties. Achieving this has proved difficult, however, because the generated localizations are corrupted by artifacts, principally the repeated localization of the same fluorophore due to multiple-blinking[2]. This has led to controversy about

whether proteins are truly clustered in cells, hindering our understanding of the causes and function of nanoscale protein clustering.

Here, we develop an algorithm, MBC, for correcting multiple blinking that requires no user input, no additional calibration data, and is not limited to a specific analysis goal. We show that it can be used to reliably test for spatial randomness or recover other clustering properties from the ground truth.

A number of methods have been put forward to test for spatial randomness in SMLM data. These include methods based on varying the labeling density and observing the effects on specific cluster analysis outputs[32] or by labeling the same species with two different fluorophores allowing a cross-comparison to be made[33]. These, however, require multiple sample preparation rounds and are therefore more complex and time-consuming. Correction can also be made by measuring blinking behavior in a separate sample of well-isolated fluorophores[8], but this again adds complexity and experimental effort and requires the assumption that probe photophysics are maintained between the sample and the calibration. It is also possible to measure or simulate multiple blinking using realistic photophysical models and use these to derive new CSR confidence intervals for the L-function curves[34]. It should be noted, however, that none of these methods produces a new set of corrected positions.

An approach with the same goal as ours is the method of Bohrer et al.[10]. We perform a comprehensive comparison to this approach, showing MBC achieves superior accuracy at a fraction of the computational cost. Performance gains are often substantial in both measures, for example, the method of Bohrer et al. can overcount the number of proteins by up to a factor of two, and can take more than a week per ROI (where our method takes 2 minutes). MBC demonstrates a sufficiently accurate recovery of ground truth to reliably test the CSR hypothesis. MBC therefore represents a new capability: that of obtaining a set of corrected ground-truth locations of sufficient quality so that any subsequent statistical analysis can be conducted with assurance.

The limitations of MBC are as follows. The method is only applicable to photophysical models typical of PALM acquisitions, and therefore cannot be used to correct dSTORM or other SMLM modalities. Performance of the correction will decrease as the clustering of the ground-truth increases, however, it tested favorably with realistic and heavily clustered scenarios. The method also adds computational time to any analysis pipeline. For a $3{,}000 \times 3{,}000$ nm ROI containing 500 ground-truth proteins, we estimate the MBC step to take 2–4 minutes per ROI on a standard desktop computer. Of course, as it results in fewer points per ROI, subsequent analysis will typically be accelerated. While here the correction is limited to two-dimensional (2D) data, it can in principle be adapted to 3D (x,y,z) coordinates. In this case, the algorithm would need to take account of the differing localization precision in z and potentially the lower detection efficiency of proteins deeper into the sample. Of course, there are many other sources of potential error in SMLM other than multiple blinking. These include low detection efficiency[35,36], endogenous expression, overlapped PSFs, drift and so on, which are not addressed by MBC. The resulting point patterns generated by MBC should therefore still be interpreted with these in mind.

In conclusion, MBC allows for accurate recovery of ground-truth fluorophore positions, with enhanced precision, from PALM datasets subjected to multiple-blinking artifacts. Corrected sets are of sufficient quality to allow accurate cluster analysis and the statistical testing for complete spatial randomness. We therefore believe that PALM combined with MBC will be an invaluable tool for addressing questions on the existence, determinants and functions of protein nanoscale clustering.

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/s41592-022-01463-w.

Received: 24 March 2021; Accepted: 18 March 2022;
Published online: 11 May 2022

References

1. Betzig, E. et al. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).
2. 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).
3. Annibale, P., Vanni, S., Scarselli, M., Rothlisberger, U. & Radenovic, A. Identification of clustering artefacts in photoactivated localisation microscopy. *Nat. Methods* **8**, 527–528 (2011).
4. Annibale, P., Vanni, S., Scarselli, M., Rothlisberger, U. & Radenovic, A. Quantitative photo activated localisation microscopy: unraveling the effects of photoblinking. *PLoS ONE* **6**, e22678 (2011).
5. Lee, S.-H., Shin, J. Y., Lee, A. & Bustamante, C. Counting single photoactivatable fluorescent molecules by photoactivated localisation microscopy (PALM). *Proc. Natl Acad. Sci. USA* **109**, 17436–17441 (2012).
6. Levet, F. et al. SR-Tesseler: a method to segment and quantify localisation-based super-resolution microscopy data. *Nat. Methods* **12**, 1065–1071 (2015).
7. Veatch, S. L. et al. Correlation functions quantify super-resolution images and estimate apparent clustering due to over-counting. *PLoS ONE* **7**, e31457 (2012).
8. Sengupta, P. et al. Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis. *Nat. Methods* **8**, 969–975 (2011).
9. Jensen, L. G., Williamson, D. J. & Hahn, U. Semiparametric point process modelling of blinking artefacts in PALM. Preprint at *bioRxiv* https://arxiv.org/abs/2101.12285 (2021).
10. Bohrer, C. H. et al. A pairwise distance distribution correction (DDC) algorithm to eliminate blinking-caused artefacts in SMLM. *Nat. Methods* **18**, 669–677 (2021).
11. Rossboth, B. et al. TCRs are randomly distributed on the plasma membrane of resting antigen-experienced T cells. *Nat. Immunol.* **19**, 821–827 (2018).
12. Williamson, D. J. et al. Pre-existing clusters of the adaptor Lat do not participate in early T cell signaling events. *Nat. Immunol.* **12**, 655–662 (2011).
13. 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).
14. Ovesný, M., Křížek, P., Borkovec, J., Švindrych, 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).
15. Fricke, F., Beaudouin, J., Eils, R. & Heilemann, M. One, two or three? Probing the stoichiometry of membrane proteins by single-molecule localisation microscopy. *Sci. Rep.* **5**, 14072 (2015).
16. Coltharp, C., Kessler, R. P. & Xiao, J. Accurate construction of photoactivated localisation microscopy (PALM) images for quantitative measurements. *PLoS ONE* **7**, e51725 (2012).
17. Cormack, R. M. A review on classification. *J. R. Stat. Soc.* **134**, 321–367 (1971).
18. Ward, J. H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**, 236–244 (1963).
19. Diggle, P. J. On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrics* **35**, 87–101 (1979).
20. Rubin-Delanchy, P. et al. Bayesian cluster identification in single-molecule localisation microscopy data. *Nat. Methods* **12**, 1072–1076 (2015).
21. Griffié, J. et al. A Bayesian cluster analysis method for single-molecule localisation microscopy data. *Nat. Protoc.* **11**, 2499–2514 (2016).
22. Griffié, J. et al. Virtual-SMLM, a virtual environment for real-time interactive SMLM acquisition. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/2020.03.05.967893v1 (2020).
23. Thevathasan, J. V. Nuclear pores as versatile reference standards for quantitative super-resolution microscopy. *Nat. Methods* **16**, 1045–1053 (2019).
24. Ries, J. SMAP: a modular super-resolution analysis platform for SMLM data. *Nat. Methods* **17**, 870–872 (2020).
25. Roob, E. 3rd, Trendel, N., Rein Ten Wolde, P. & Mugler, A. Cooperative clustering digitizes biochemical signaling and enhances its fidelity. *Biophys. J.* **110**, 1661–1669 (2016).
26. Grakoui, A. et al. The immunological synapse: a molecular machine controlling T cell activation. *Science* **285**, 221–227 (1999).

27. Purbhoo, M. A. et al. Dynamics of subsynaptic vesicles and surface microclusters at the immunological synapse. *Sci. Signal.* **3**, ra36 (2010).
28. 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).
29. Razvag, Y., Neve-Oz, Y., Sajman, J., Reches, M. & Sherman, E. Nanoscale kinetic segregation of TCR and CD45 in engaged microvilli facilitates early T cell activation. *Nat. Commun.* **9**, 732 (2018).
30. Balagopalan, L., Kortum, R. L., Coussens, N. P., Barr, V. A. & Samelson, L. E. The linker for activation of T cells (LAT) signaling hub: from signaling complexes to microclusters. *J. Biol. Chem.* **290**, 26422–26429 (2015).
31. Varma, R., Campi, G., Yokosuka, T., Saito, T. & Dustin, M. L. T cell receptor-proximal signals are sustained in peripheral microclusters and terminated in the central supramolecular activation cluster. *Immunity* **25**, 117–127 (2006).
32. Baumgart, F. et al. Varying label density allows artefact-free analysis of membrane-protein nanoclusters. *Nat. Methods* **13**, 661–664 (2016).
33. Arnold, A. M. et al. Verifying molecular clusters by 2-color localisation microscopy and significance testing. *Sci. Rep.* **10**, 4230 (2020).
34. Platzer, R. et al. Unscrambling fluorophore blinking for comprehensive cluster detection via photoactivated localisation microscopy. *Nat. Commun.* **11**, 4993 (2020).
35. Dursic, N. et al. Single-molecule evaluation of fluorescent protein photoactivation efficiency using an in vivo nanotemplate. *Nat. Methods* **11**, 156–162 (2014).
36. 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

Sample preparation. For LAT images, Jurkat E6.1 cells (ECACC 88042803) expressing LAT-mEos3.2 (WT LAT, or signaling deficient mutant, YF LAT) were introduced to anti-CD3 (at 2 μg ml^{-1}; eBioscience clone OKT3, 16-0037-81) and anti-CD28 (at 5 μg ml^{-1}; RnD Systems, clone CD28.2, 16-0289-85) coated glass-bottomed chamber slides (no. 1.5 glass, ibidi μSlides) at 50×10^3 cells per cm^2 in warm Hanks' balanced salt solution and incubated at 37 °C for 5 min to allow for synapse formation. The chamber wells were gently washed with warm Hanks' balanced salt solution and then fixed in 3% paraformaldehyde in phosphate-buffered saline (PBS) for 20 min at 37 °C. Fixed cells were washed five times in PBS and used immediately for PALM imaging.

Imaging. PALM image sequences were acquired on a Nikon N-STORM system in a TIRF configuration using a ×100 1.49 numerical aperture (NA) CFI Apochromat TIRF objective for a pixel size of 160 nm and running NIS Elements software v.4.6. Samples were continuously illuminated with 561 nm laser light at approximately 2 kW cm^{-2} and 405 nm laser light (to induce photo-conversion) at approximately 2 W cm^{-2}. Images were recorded on an Andor IXON Ultra 897 EMCCD with an electron multiplier gain of 200 and pre-amplifier gain profile 3 to a centered 256×256 pixel region at 40 ms per frame for 5,000 to 15,000 frames.

Virtual microscope simulations. Raw camera frames were generated using Virtual-SMLM[22] operating in PALM mode (that is, using a four-state photophysical model). The frame rate was set to 25 or 100 frames per second. The activation laser (that is, initial state transition) was either fixed or ramped up over the acquisition. In the first case, the number of fluorophores emitting per frame decreases over time. In the second case, it remains constant over the acquisition. Emission traces were generated independently for each fluorophore and imaging continued until all fluorophores had been imaged and bleached. All other state transition probabilities and photophysics properties were fixed to mimic mEos blinking characteristics. The PSFs were recorded on a virtual EMCCD camera, with an electron microscopy gain fixed at 300. Virtual-SMLM took as input ground-truth maps of mEos2 positions. Then 5,556 mEos proteins were placed randomly over a $10{,}000\times10{,}000$ nm 2D area. Generated camera frames were then analyzed using ThunderSTORM and the data cropped into nonoverlapping $3{,}000\times3{,}000$ nm regions.

Localization. Localizations of fluorophore coordinates were reconstructed using ThunderSTORM[14] and corrected for sample drift using cross-correlation of images from five bins at a magnification of five. No further postprocessing was performed.

Mathematical details. *Marginal likelihood of clusters.* We represent the observed process by a series of localizations $(X_i, Y_i)_{i=1}^n \in R$ with associated 'blink' times $T_1, \ldots, T_n$, and localization uncertainties $\sigma_1^2, \ldots, \sigma_n^2$, where $R = [x_0, x_1] \times [y_0, y_1]$ is the ROI. For a given partition of the localizations into groups, we compute the marginal likelihood of the data as follows. Consider a group comprising the observations $1, \ldots, m$, with $1 \le m \le n$, posited to correspond to one, distinct molecule. In particular, we defer until later the treatment of background noise. The independence assumptions set out in the main article result in the following marginal likelihood factorization:

$$p\left\{(X_i)_{i=1}^m, (Y_i)_{i=1}^m, (T_i)_{i=1}^m\right\} = p\left\{(X_i)_{i=1}^m\right\} p\left\{(Y_i)_{i=1}^m\right\} p\left\{(T_i)_{i=1}^m\right\}.$$

Denoting by $\mu = (\mu_X, \mu_Y)$ the true position of the molecule, the spatial components above have a likelihood (given only for $(X_i)_{i=1}^m$)

$$\begin{aligned} p\left\{(X_i)_{i=1}^m \mid \mu_X\right\} &= \prod_{i=1}^m \sqrt{(2\pi\sigma_i^2)} \exp\left\{-\tfrac{1}{2}\left(\tfrac{X_i-\mu_X}{\sigma_i}\right)^2\right\} \\ &= (2\pi)^{-m/2}\left(\textstyle\prod \sigma_i\right)^{-1} \exp\left\{-\tfrac{1}{2}\textstyle\sum \eta_i (X_i-\mu_X)^2\right\}, \end{aligned}$$

where $\eta_i = 1/\sigma_i^2$. Defining the weighted mean

$$\tilde{X} = \frac{\sum \eta_i X_i}{\sum \eta_i}$$

We find

$$\begin{aligned} &p\left\{(X_i)_{i=1}^m \mid \mu_X\right\} \\ &= (2\pi)^{-m/2}\left(\textstyle\prod \sigma_i\right)^{-1} \exp\left\{-\tfrac{1}{2}\textstyle\sum \eta_i \left(X_i-\tilde{X}\right)^2\right\} \times \exp\left\{-\tfrac{\sum \eta_i}{2}\left(\mu_X-\tilde{X}\right)^2\right\}. \end{aligned}$$

Placing a uniform prior on μ_X, we find

$$\begin{aligned} &p\left\{(X_i)_{i=1}^m \mid \mu_X\right\} = (2\pi)^{-m/2}\left(\textstyle\prod \sigma_i\right)^{-1} \exp\left\{-\tfrac{1}{2}\textstyle\sum \eta_i \left(X_i-\tilde{X}\right)^2\right\} \times (x_1-x_0)^{-1} \\ &\int_{x_0}^{x_1} \exp\left\{-\tfrac{\sum \eta_i}{2}\left(\mu_X-\tilde{X}\right)^2\right\} d\mu_X = (2\pi)^{-m/2}\left(\textstyle\prod \sigma_i\right)^{-1} \exp\left\{-\tfrac{1}{2}\textstyle\sum \eta_i \left(X_i-\tilde{X}\right)^2\right\} \\ &\times (x_1-x_0)^{-1}\left(2\pi\textstyle\sum \eta_i\right)^{1/2}\left\{\Phi\left(\tfrac{x_1-\tilde{X}}{\sqrt{\sum \eta_i}}\right) - \Phi\left(\tfrac{x_0-\tilde{X}}{\sqrt{\sum \eta_i}}\right)\right\}. \end{aligned}$$

The temporal component has likelihood

$$p\left((T_i)_{i=1}^m\right) = p(u)\, p(T_{\min}) \prod_k p(f_k) \prod_k p(d_k),$$

where each term is computed as follows. The blink times $T_1, \ldots, T_m$ are not typically observed exactly, and instead one has access only to associated frame numbers $F_1, \ldots, F_m$, taken to represent (small) windows of time containing them. We therefore consider a visit to the fluorescent state to be a block of $L \ge 1$ contiguous fluorescent frames (or consecutive frame numbers), and impute the length of this visit to be the time elapsed over $L-1$ frames, to obtain auxiliary quantities

$$f_k : \text{ time spent in fluorescent state } (k\text{th visit}).$$

Up to discrete approximation error, each f_k represents the minimum of two exponential random variables with respective rates r_D and r_B, with likelihood contribution

$$p(f_k) = (r_D + r_B)\, e^{-(r_D+r_B)f_k}.$$

Similarly, let d_k denote the time elapsed over the kth interval between noncontiguous frames, taken to represent d_k: time spent in dark state (kth visit). The likelihood contribution is

$$p(d_k) = r_R e^{-r_R d_k} \frac{r_D}{r_D + r_B}.$$

The initial switch of the fluorophore to the activated state happens at time $T_{\min}$, computed simply as the minimum T_i value, and this contributes

$$p(T_{\min}) = r_F e^{-r_F T_{\min}},$$

to the likelihood.Finally, let u denote the time since the last blink (a period during which it is unknown whether the process has entered a dark or bleached state). The final contribution is

$$p(u) = \frac{r_D}{r_B + r_D} e^{-r_R u} + \frac{r_B}{r_B + r_D}$$

To finalize calculations, one must account for background noise (in the case of $m=1$). Such points are assumed to be uniform in space–time. The complete marginal likelihood is

$$\begin{aligned} &\pi\left((X_i)_{i=1}^m, (Y_i)_{i=1}^m, (T_i)_{i=1}^m\right) \\ &= \begin{cases} (1-\alpha)\, p\left((X_i)_{i=1}^m, (Y_i)_{i=1}^m, (T_i)_{i=1}^m\right) & m > 1 \\ \alpha V^{-1} + (1-\alpha)\, p\left((X_i)_{i=1}^m, (Y_i)_{i=1}^m, (T_i)_{i=1}^m\right) & m = 1 \end{cases} \end{aligned}$$

where $V = T(x_1 - x_0)(y_1 - y_0)$, T is the length of the period of observation and α is the background probability.

Identifying and summarizing clusters. For both MBC and DTT clustering, an expected number of clusters, N, is first estimated, and a version of AHC is then used to partition the dataset into N clusters. In AHC, each point is initially considered to be a distinct cluster. Using a user-specified metric and a linkage criterion, a stepwise greedy merging of the closest clusters is repeated until a partition with a predetermined number of clusters is obtained, or until no more clusters can be merged with a distance less than some specified number. The metric determines the distances between pairs of points, and the linkage criterion generalizes these to a distance between clusters. Once the final partition has been identified, we merge each cluster down to its estimated center, and the uncertainty of the center is computed. In the following, we use the notation

$$L = (X, Y)$$

$$d_l(L_1, L_2) = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$$

$$d_T(T_1, T_2) = |T_1 - T_2|$$

MBC clustering. For MBC, the number of desired clusters, N, is an output of the rate-estimation step, and is thus decoupled from the clustering problem. For the AHC step, we use the family of metrics

$$d_s((L_1, T_1), (L_2, T_2)) = \frac{d_l(L_1, L_2)}{(\sigma_1 + \sigma_2)} + S d_t(T_1, T_2)$$

For $S \ge 0$, which is simply the sum of the Euclidean distance between the locations and (scaled) times. For the linkage criterion we chose Ward's Minimum Variance Method, as implemented via the Lance–Williams formula[18], as it tends to find

homogeneous clusters of spherical shape.By varying S along a grid, we obtain a sequence of partitions, each slightly different but all chosen to have N clusters. By evaluating the marginal likelihood for the partitions obtained at each value of S, we then pick the most likely partition among those considered.

DTT clustering. DTT is a general idea in SMLM blinking correction literature, but implementation details are rarely discussed. The general principle is to merge locations that are close in space and time, with hard thresholds on the maximally allowed bridging distances in space and time. As a way to implement this idea in the AHC framework, we define the distance between two observations as

$$d_{\tau_s,\tau_t}\left((L_1, T_1), (L_2, T_2)\right) = d_l(L_1, L_2) + d_t(T_1, T_2) + \infty \cdot 1\left(d_l(L_1, L_2) > \tau_s \text{ or } d_t(T_1, T_2) > \tau_t\right)$$

where $\infty \cdot 0 = 0$. Although not strictly a metric, this distance measure allows us to implement the DTT idea. We use the single-linkage criterion for cluster merging, which considers the distance between two clusters to be the smallest pairwise distance between them. Combined with our metric, this means that the clustering algorithm is allowed to merge points and clusters, so long as they can be combined via paths that do not violate the hard thresholds. Finally, a clustering is achieved by continuing to merge clusters until only infinite distances between clusters remain (no more legal merges can be made). For τ_s, we used four times the mean localization uncertainty. The temporal threshold, τ_t, was determined as follows. First, the method of Annibale et al. was used to determine N[4]. Next, τ_t was increased incrementally until the AHC algorithm produced a partition with N clusters, or as close to N as possible.

Cluster centers and uncertainty

Let $(X_i, Y_i)_{i=1}^m$ be the coordinates of an arbitrary cluster with center μ. Once a particular clustering is given, it makes sense to treat the cluster centers as fixed parameters to be estimated. Thinking therefore of μ as fixed, the maximum likelihood estimator, $\hat{\mu}$, maximizes the likelihood of the cluster coordinates

$$\log(p)\left\{(X_i, Y_i)_{i=1}^m\right\} = -\frac{\sum \eta_i}{2}\left(\left(\mu_X - \tilde{X}\right)^2 + \left(\mu_Y - \tilde{Y}\right)^2\right) + C$$

where C does not depend on μ and it follows immediately that

$$\hat{\mu} = \left(\tilde{X}, \tilde{Y}\right) = \left(\frac{\sum \eta_i X_i}{\sum \eta_i}, \frac{\sum \eta_i Y_i}{\sum \eta_i}\right).$$

Using $\hat{\mu}$, we can estimate the position of the molecule associated with a given cluster. As the coordinates of $\hat{\mu}$ are i.i.d., the covariance matrix of $\hat{\mu}$ is given as $\mathbb{V}[\hat{\mu}] = \tilde{\sigma}^2 I_2$ where I_2 is the 2×2 identity matrix, and

$$\tilde{\sigma}^2 = \mathbb{V}\left[\tilde{X}\right] = \frac{\sum \eta_i^2 \sigma_i^2}{\left(\sum \eta_i\right)^2} = \frac{1}{\sum \frac{1}{\sigma_i^2}}$$

And the updates localization uncertainty is then simply the associated standard deviation

$$\tilde{\sigma} = \frac{1}{\sqrt{\sum \frac{1}{\sigma_i^2}}}$$

Justification of optimization strategy. It may seem wasteful to correct the localizations in two steps rather than a more integrated approach. However, if, in a single step, we optimize the model likelihood for blinking parameters and data partitions simultaneously, the maximum likelihood clustering overfits in a catastrophic way, giving each position its own cluster and fitting a degenerate blinking model (infinite bleach rate). A more sophisticated attempt would be to take a fully Bayesian approach, with priors on the blinking dynamics, but this has two disadvantages: the first is added and potentially catastrophic sensitivity to model misspecification, for example, multiple dark states; second, this approach introduces substantial computational issues, which can only be addressed by further layers of approximation (for example, Monte Carlo or variational inference) and analysis parameters (burnin, runtime, step-size and so on). The convergence of Markov Chain Monte Carlo is highly uncertain in any Bayesian clustering problem: to quote an influential source, 'Although we may be presumptuous, we consider that almost the entirety of Markov chain Monte Carlo samplers implemented for mixture models has failed to converge!'[37].

We now demonstrate that, once the blinking parameters are estimated and held fixed, (1) the likelihood does not overfit the data partition, and (2) our simple search strategy gets close to the optimum. To do this, we implement a more flexible clustering approach, Bayesian hierarchical clustering (BHC), in which clusters are greedily merged according to the largest increase in likelihood. Despite an efficient implementation, for example exploiting the Lance–Williams update, this algorithm is too time costly to be suggested for general use, but it provides a benchmark for the performance that could be obtained with a more involved approach and more computational power. In Supplementary Fig. 2, we show the likelihood obtained by BHC in a range of different conditions, noting that the algorithm starts with a partition separating all points and ends with all points in one partition (so obtains a likelihood for every possible number of clusters). We observe that (1) the highest likelihood partition found by BHC does not routinely overfit (overestimate the number of clusters) and (2) the likelihood of the MBC and BHC solutions are close. MBC and BHC are also comparable in terms of other metrics (for example, each wins roughly half the time in Wasserstein distance).

Significance testing. The P values reported in Supplementary Table 2 are based on a two-sided permutation test of the absolute difference of means, using 10,000 simulations.

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

Data availability

A data simulator to recapitulate simulated data conditions and raw experimental data (point clouds) are available at https://github.com/Louis-Jensen/MBC-for-PALM. Raw experimental data (camera frames) available upon request. Source data are provided with this paper.

Code availability

MBC code is available as Supplementary Material together with installation instructions and example simulated datasets. MBC code is also available at https://github.com/Louis-Jensen/MBC-for-PALM.

References

37. Celeux, G., Hurn, M. & Robert, C. P. Computational and inferential difficulties with mixture posterior distributions. *J. Am. Stat. Assoc.* **95**, 957–970 (2000).

Acknowledgements

L.G.J. was supported by the Center for Stochastic Geometry and Advanced Bioimaging, funded by grant no. 8721 from the Villum Foundation. D.O. acknowledges funding from the Biotechnology and Biological Sciences Research Council (BBSRC) grant no. BB/R007365/1. P.R.D. acknowledges funding from the BBSRC grant no. BB/R007837/1. We also acknowledge the use of the King's College London Nikon Imaging Center. We thank U. Hahn (Aarhus University) for motivating this project in initial discussions. We acknowledge J. Ries and the European Molecular Biology Laboratory for supplying NPC data.

Author contributions

L.G.J. developed the software. L.G.J., T.Y.H., D.M.O. and P.R.-D. conceived the experiments. L.G.J. and T.Y.H. ran the analysis. L.G.J. and T.Y.H. provided simulated data. D.J.W., J.G. and D.S. provided additional simulated and experimental data. L.G.J., T.Y.H., D.M.O. and P.R-D. wrote the manuscript. L.G.J. and P.R.-D. conceived the method.

Competing interests

The authors declare no competing interests.

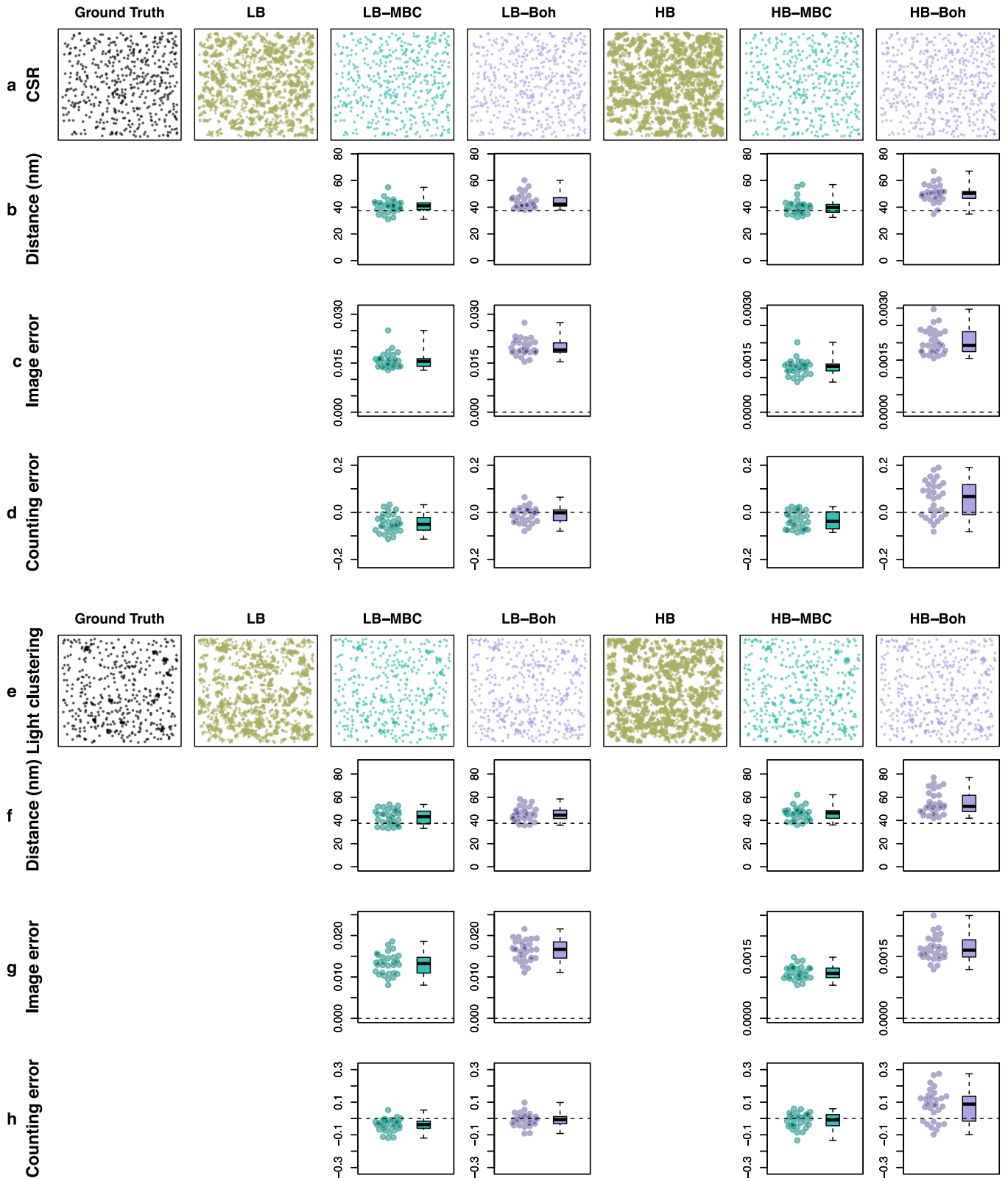
Additional information

Extended data are available for this paper at https://doi.org/10.1038/s41592-022-01463-w.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41592-022-01463-w.

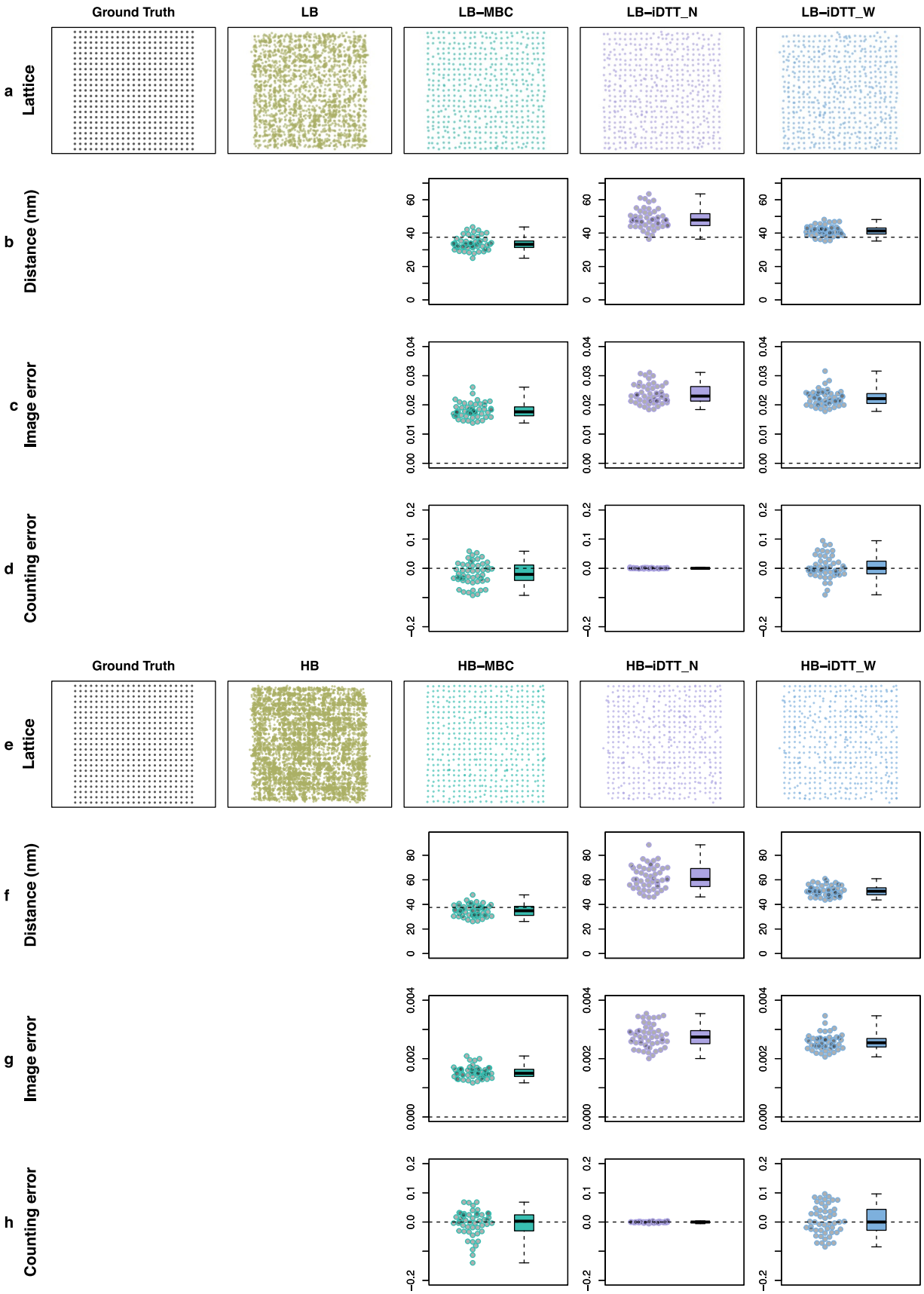
Correspondence and requests for materials should be addressed to Louis G. Jensen, Patrick Rubin-Delanchy or Dylan M. Owen.

Peer review information *Nature Methods* thanks Antony Lee, Aleksandra Radenovic and Jie Xiao for their contribution to the peer review of this work. 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. Peer reviewer reports are available.



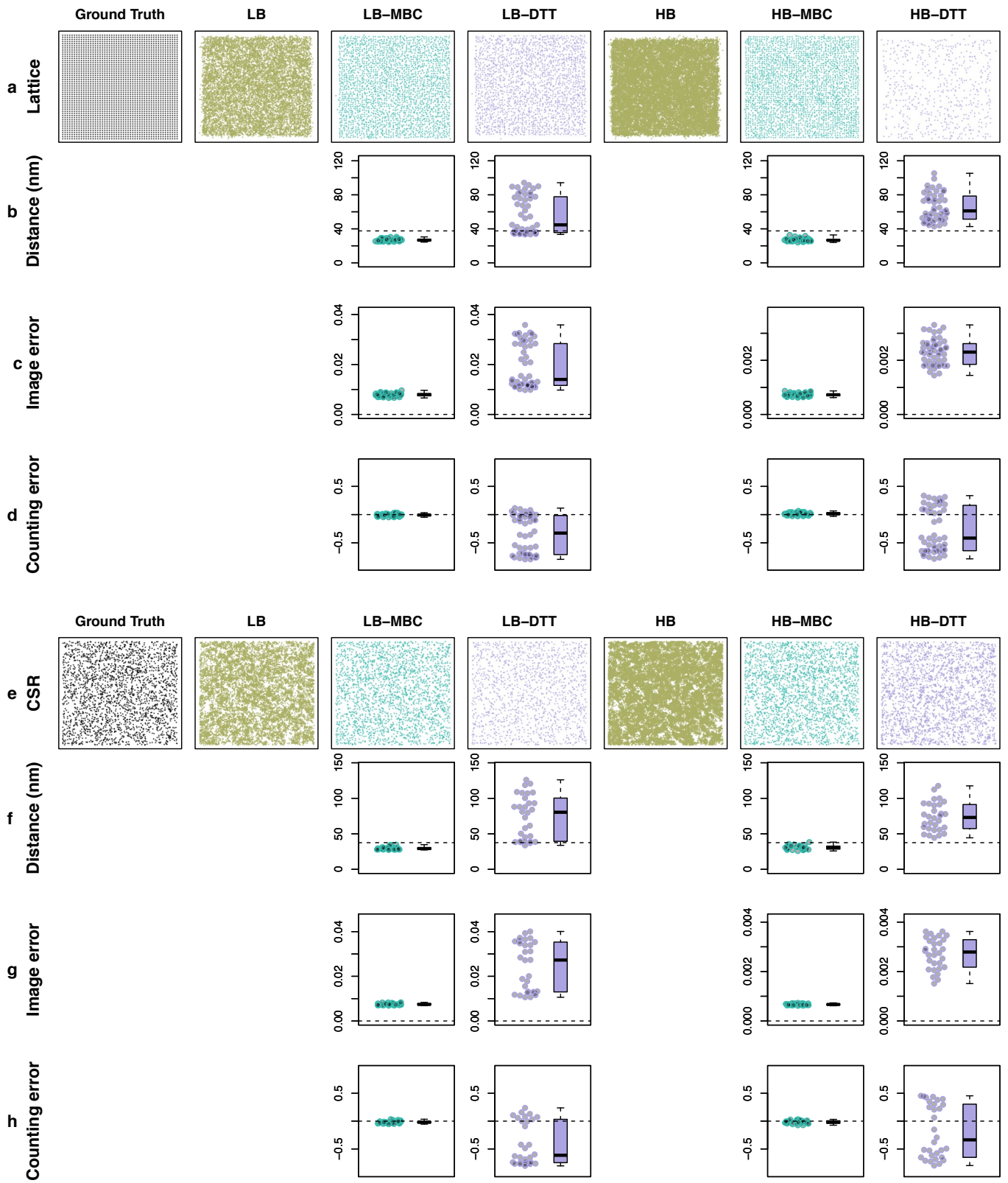
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Comparison of MBC with the state-of-the-art for the case of CSR and clustered molecules. a) Representative ground truth, blinking and corrected data (from one of $n = 30$ realisations) for CSR molecules. b,c,d) Wasserstein distances (**b**), image error (**c**), counting error (**d**), between corrected data and ground truth. In b), the dashed horizontal line shows the 37 nm benchmark and c,d) the dashed horizontal line shows the optimal value 0. The columns show different blinking conditions (LB: light blinking, HB: heavy blinking) and correction methods (MBC: our method, Boh: Bohrer et al.'s method). Our method has superior performance on all metrics except counting error for the light blinking case. **e**) Representative ground truth, blinking and corrected data (from one of $n = 30$ realisations) for clustered molecules. f,g,h) Wasserstein distances (**f**), image error (**g**), counting error (**h**), between corrected data and ground truth. In f), the dashed horizontal line shows the 37 nm benchmark and c,d) the dashed horizontal line shows the optimal value 0. The columns show different blinking conditions (LB: light blinking, HB: heavy blinking) and correction methods (MBC: our method, Boh: Bohrer et al.'s method). Our method has superior performance on all metrics except counting error for the light blinking case. Box plots show median, 25^{th} and 75^{th} percentiles and min, max.



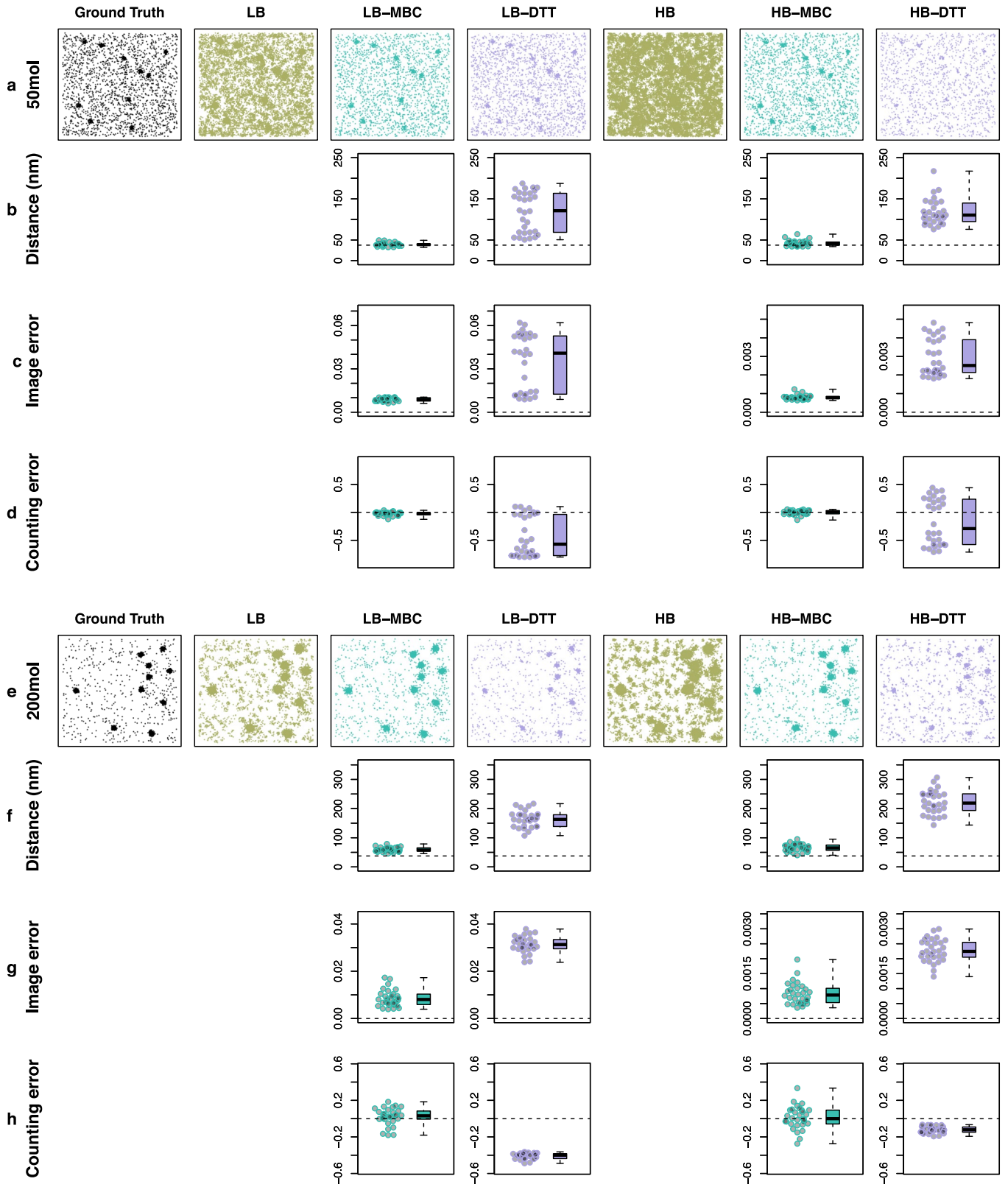
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Comparison of MBC with idealized DTT, with molecules on a fixed grid with light (LB) and heavy (HB) blinking. a) Representative ground truth, blinking and corrected data (from one of $n = 50$ realisations) with light blinking. b,c,d) Wasserstein distances (**b**), image error (**c**), counting error (**d**), between corrected data and ground truth. In b), the dashed horizontal line shows the 37 nm benchmark and c,d) the dashed horizontal line shows the optimal value 0. The columns show different correction methods (MBC: our method, iDTT_N: DTT minimising counting error; iDTT_W: DTT minimising Wasserstein distance). Our method is always superior on Wasserstein distance and image error, and comparable in counting error when compared to iDTT_W. **e**) Representative ground truth, blinking and corrected data (from one of $n = 50$ realisations) for heavy blinking. f,g,h) Wasserstein distances (**f**), image error (**g**), counting error (**h**), between corrected data and ground truth. In f), the dashed horizontal line shows the 37 nm benchmark and c,d) the dashed horizontal line shows the optimal value 0. The columns show different correction methods (MBC: our method, iDTT_N: DTT minimising counting error; iDTT_W: DTT minimising Wasserstein distance). Our method is always superior on Wasserstein distance and image error, and comparable in counting error when compared to iDTT_W. Box plots show median, 25^{th} and 75^{th} percentiles and min, max.



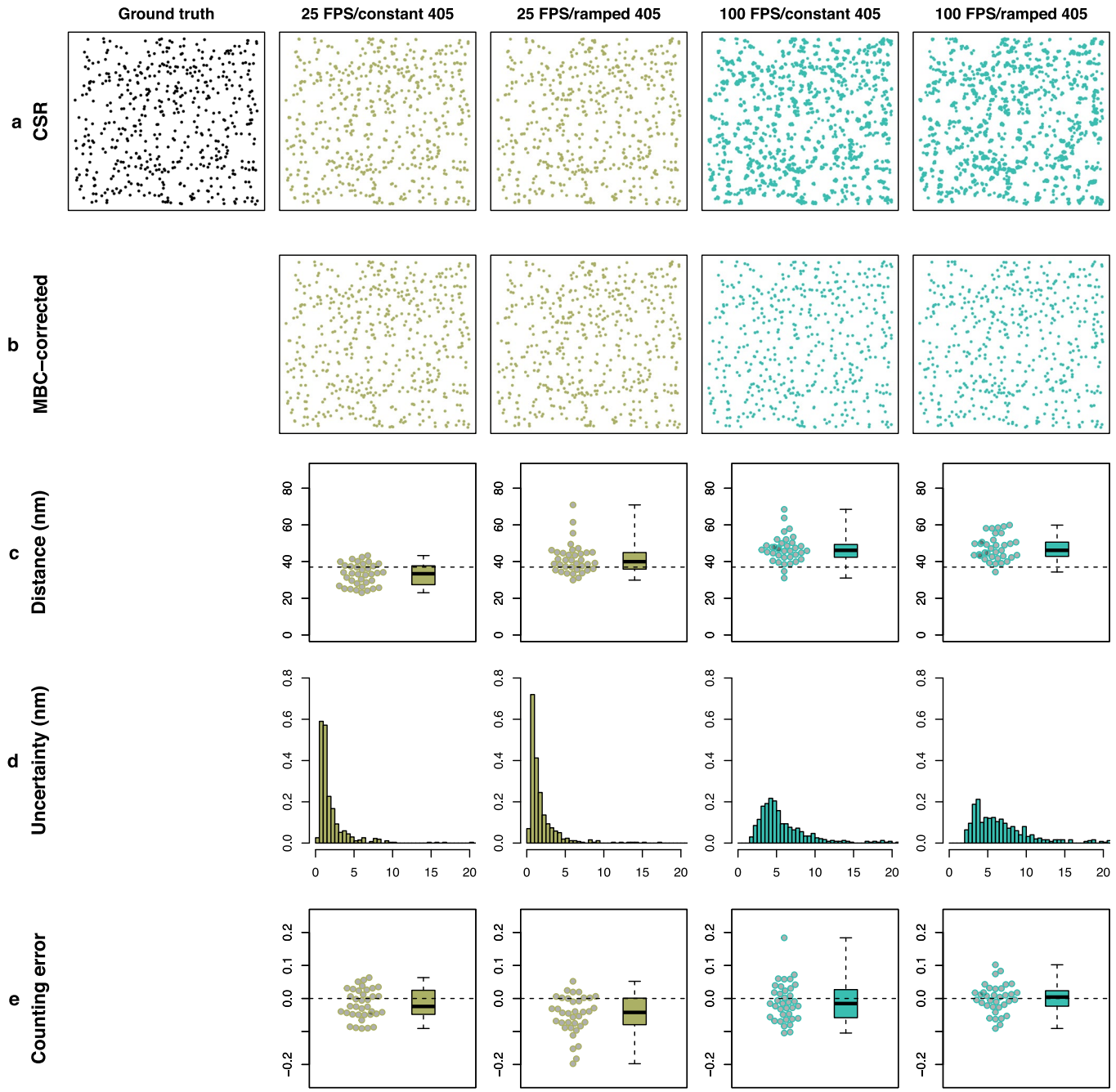
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Comparison of MBC with DTT for regular and CSR molecules. a) Representative ground truth for 2500 molecules on a fixed grid, blinking and corrected data (from one of $n = 50$ realisations). **b**–**d**) Wasserstein distances (b), image error (c), counting error (d), between corrected data and ground truth. In b), the dashed horizontal line shows the 37 nm benchmark and c,d) the dashed horizontal line shows the optimal value 0. The columns show different blinking conditions (LB: light blinking, HB: heavy blinking) and correction methods. Our method has superior performance on all metrics. **e**) Representative ground truth for 2500 CSR molecules, blinking and corrected data (from one of $n = 30$ realisations). **f**-**h**) Wasserstein distances (f), image error (g), counting error (h), between corrected data and ground truth. In b), the dashed horizontal line shows the 37 nm benchmark and g,h) the dashed horizontal line shows the optimal value 0. The columns show different blinking conditions (LB: light blinking, HB: heavy blinking) and correction methods. Our method has superior performance on all metrics. Box plots show median, 25^{th} and 75^{th} percentiles and min, max.

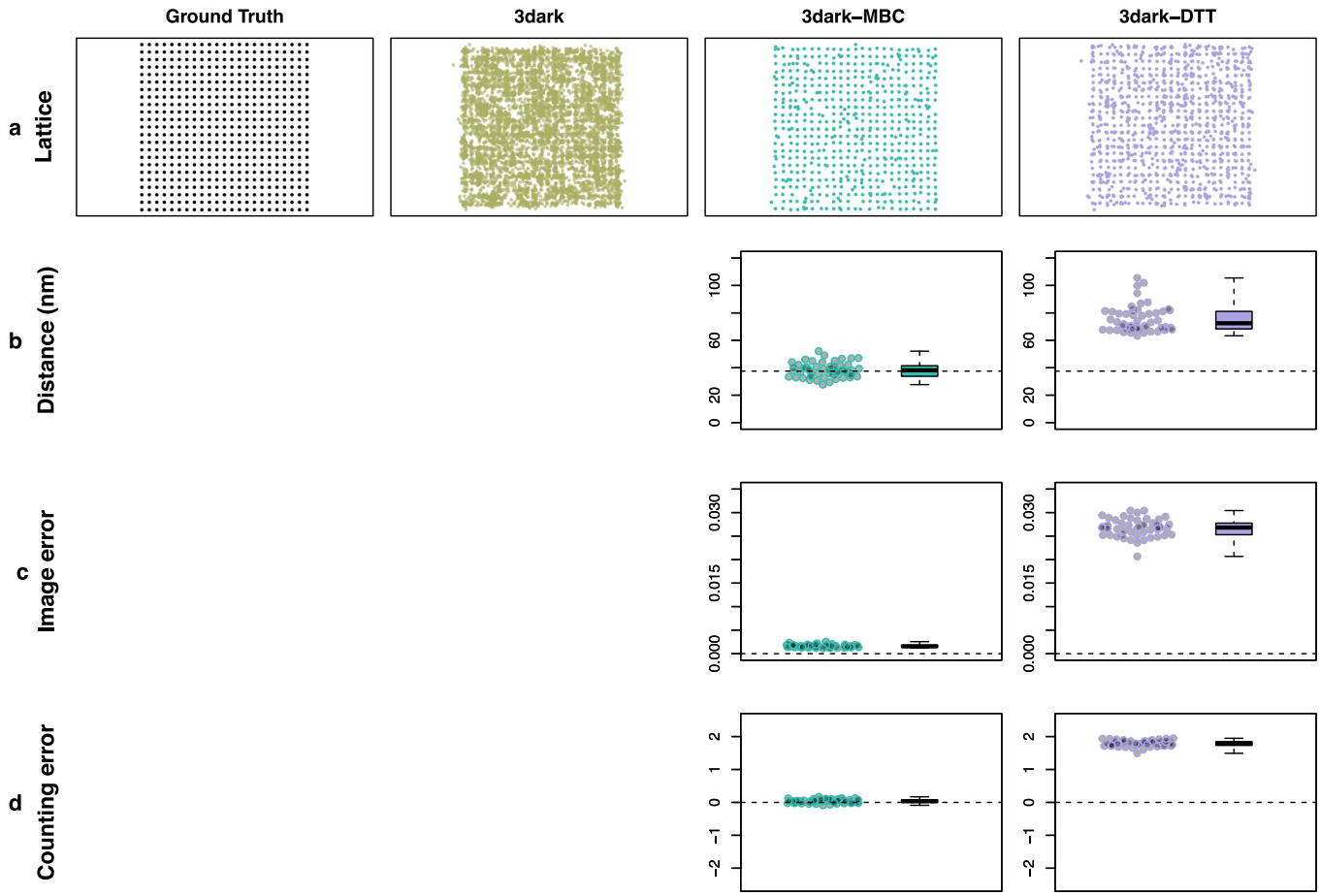


Extended Data Fig. 4 | See next page for caption.

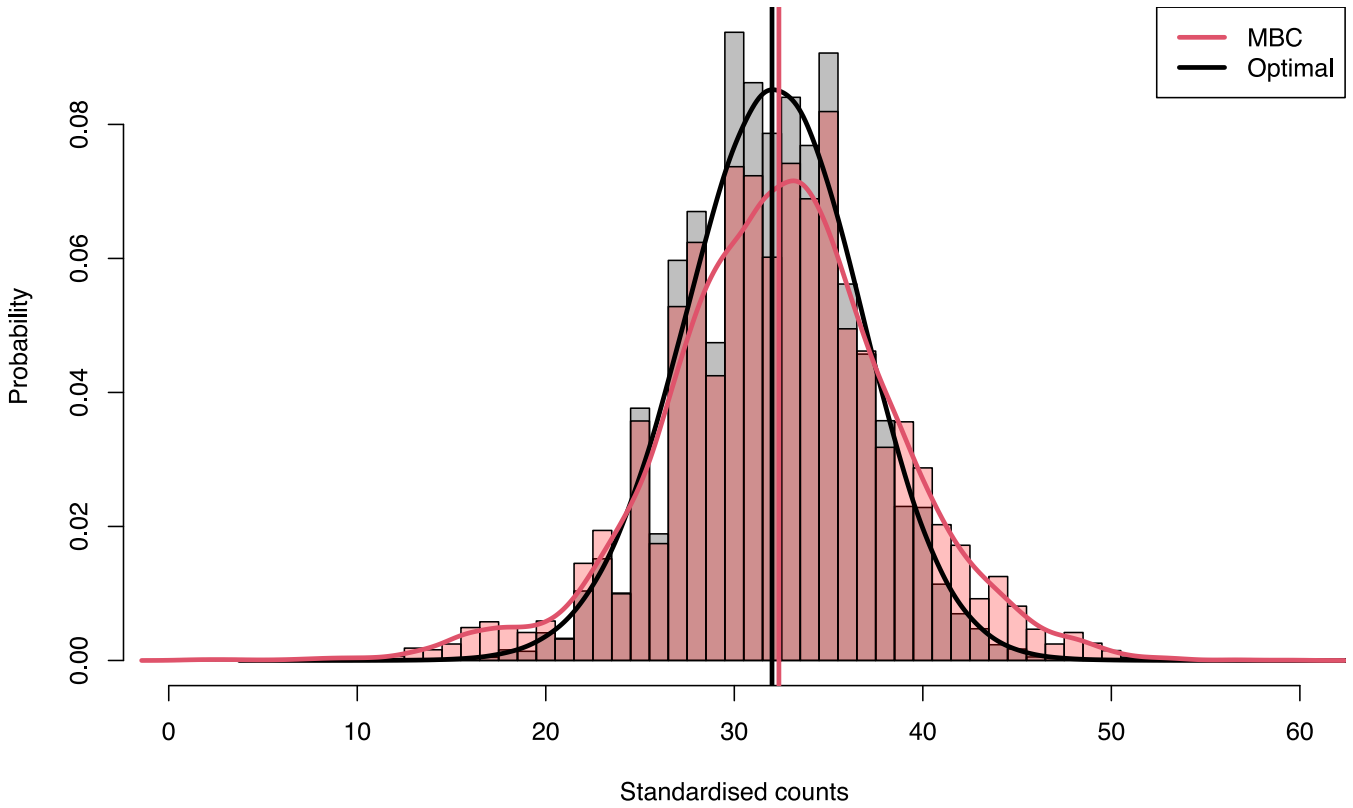
Extended Data Fig. 4 | Comparison of MBC with DTT for clustered molecules. a) Representative ground truth for 10 clusters of 50 molecules and 2000 CSR molecules, blinking and corrected data (from one of $n = 30$ realisations). **b**-**d**) Wasserstein distances (b), image error (c), counting error (d), between corrected data and ground truth. In b), the dashed horizontal line shows the 37 nm benchmark and c,d) the dashed horizontal line shows the optimal value 0. The columns show different blinking conditions (LB: light blinking, HB: heavy blinking) and correction methods. Our method has superior performance on all metrics. **e**) Representative ground truth for 10 clusters of 200 molecules and 500 CSR molecules, blinking and corrected data (from one of $n = 30$ realisations). **f**-**h**) Wasserstein distances (f), image error (g), counting error (h), between corrected data and ground truth. In b), the dashed horizontal line shows the 37 nm benchmark and g,h) the dashed horizontal line shows the optimal value 0. The columns show different blinking conditions (LB: light blinking, HB: heavy blinking) and correction methods. Our method has superior performance on all metrics. Box plots show median, 25^{th} and 75^{th} percentiles and min, max.



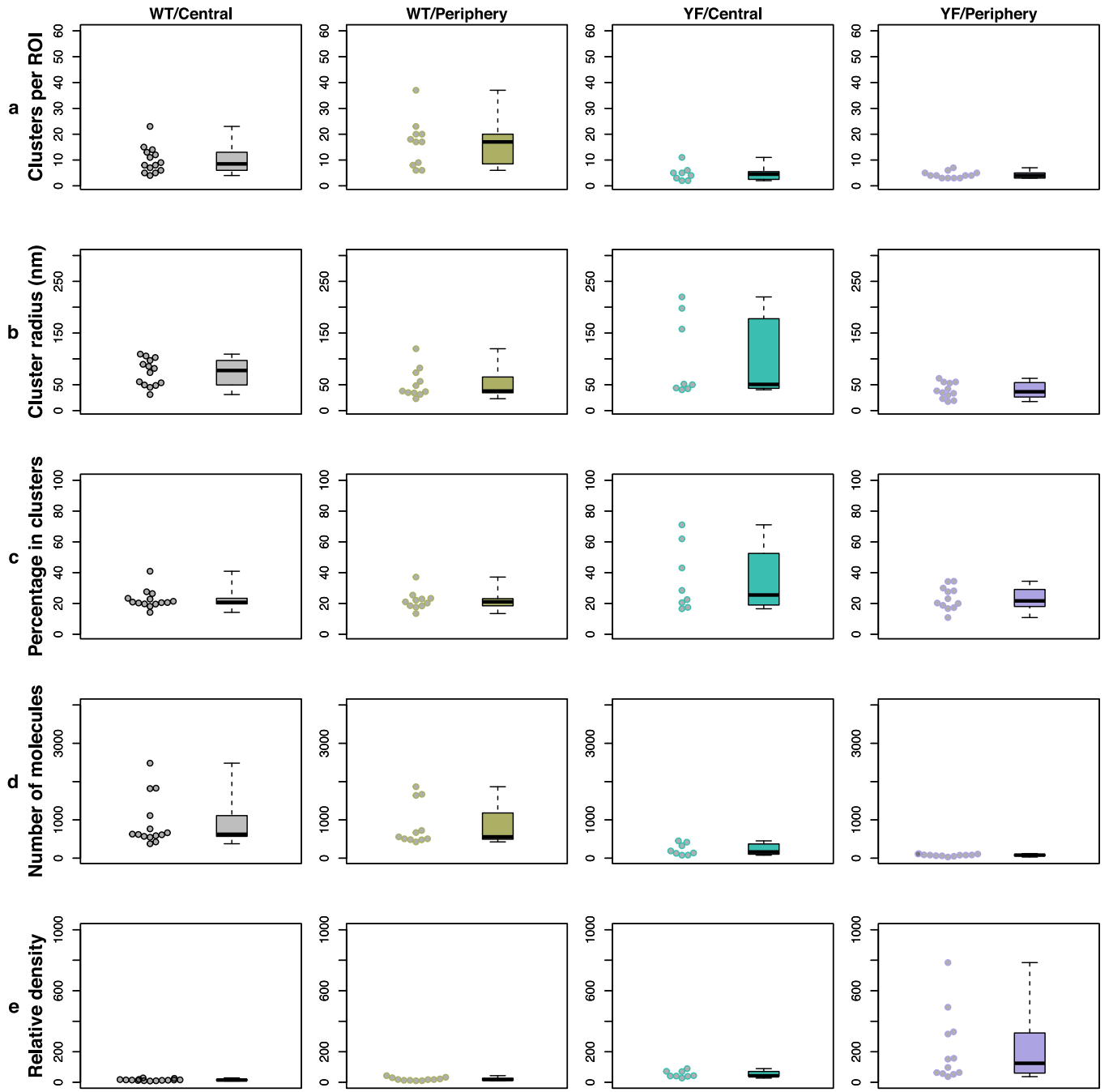
Extended Data Fig. 5 | MBC Performance as a function of camera frame rate and the activating, 405 nm laser power. a) Example ground-truth and raw localisation maps for the different conditions (n = 30 relisations). **b**) Example MBC-corrected maps. **c**) Wasserstein distances. **d**) Normalised histograms of localisation uncertainty. **e**) Counting error in estimated number of ground-truth molecules (mean in dashed line). Box plots show median, 25th and 75th percentiles and min, max.



Extended Data Fig. 6 | Comparison of MBC with DTT, with molecules on a fixed grid with blinking dynamics following a three-dark-state model. **a**) Representative ground truth, blinking and corrected data (from one of $n = 50$ realisations). **b**–**d**) Wasserstein distances (b), image error (c), counting error (d), between corrected data and ground truth. In b), the dashed horizontal line shows the 37 nm benchmark and c,d) the dashed horizontal line shows the optimal value 0. The columns show different correction methods. Our method shows superior performance on all metrics. Box plots show median, 25th and 75th percentiles and min, max.



Extended Data Fig. 7 | Correction of NPC counts, after accounting for effective labeling efficacy. Histogram of the number of MBC-recovered proteins per NPC ($n = 8146$), corrected for effective labeling efficacy (red), with mean indicated by the red vertical line, and a kernel density estimate shown as the red curve. For comparison, a histogram of an equally-sized sample of counts under perfect correction is shown in black, with mean indicated by the black vertical line, and kernel density estimate as the black curve.



Extended Data Fig. 8 | Additional statistics from the Bayesian cluster analysis of non-CSR LAT-mEos3.2 regions. a) Number of detected clusters, **b**) cluster radii (nm), **c**) percentage of molecules in clusters, **d**) number of molecules per ROI and **e**) relative density of molecules located in clusters as compared to the surrounding region. Box plots show median, 25th and 75th percentiles and min, max. n-numbers = 14 (WT centre), 11 (WT periphery), 8 (YF centre) and 12 (YF periphery).

nature research

Corresponding author(s): Dylan M Owen and Patrick Rubin-Delanchy

Last updated by author(s): Feb 16, 2022

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	The microscope hardware was operated by Nikon NIS Elements v4.6 software.
Data analysis	Custom code was used for analysis, written in R and available for testing in the supplementary material. The localisation step was performed in the freely available ThunderSTORM software. SMAP software used was v201217 from the Ries lab.

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

A data simulator to recapitulate simulated data conditions and raw experimental data (point clouds) are available at https://github.com/Louis-Jensen/MBC-for-PALM. Raw experimental data (camera frames) 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	The paper concerns the development of a new analysis methodology. The main basis of comparison is therefore against previously published methods. Sample size was therefore chosen to be comparable to previously published data sets e.g. Williamson, Nature Immunology 2011
Data exclusions	No data was excluded
Replication	Experimental regions of interest were acquired from 5 technical replicates. Data was consistent between replicates
Randomization	No randomization was carried out as this is not relevant to this study
Blinding	Investigators were not blinded as they have no input into the operation of the analysis algorithm, which is fully automated

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	anti-CD3 (eBioscience clone OKT3, 16-0037-81) anti-CD28 (RnD Systems, clone CD28.2, 16-0289-85)
Validation	Antibodies were only used for cell stimulation. Cell stimulation confirmed by morphology and signalling and antibodies. Antibodies were verified by the manufacture using known Relative Expression.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Jurkat E6.1 cells (ECACC 88042803)
Authentication	Not authenticated
Mycoplasma contamination	Not tested for mycoplasma
Commonly misidentified lines (See ICLAC register)	None