https://doi.org/10.1038/s42003-024-06772-8

MultiMatch: geometry-informed colocalization in multi-color super-resolution microscopy

Check for updates

Julia Naas[1,2], Giacomo Nies[3,4], Housen Li[3,4], Stefan Stoldt[4,5,6], Bernhard Schmitzer[7], Stefan Jakobs[4,5,6,8] & Axel Munk[3,4] ✉

With recent advances in multi-color super-resolution light microscopy, it is possible to simultaneously visualize multiple subunits within biological structures at nanometer resolution. To optimally evaluate and interpret spatial proximity of stainings on such an image, colocalization analysis tools have to be able to integrate prior knowledge on the local geometry of the recorded biological complex. We present *MultiMatch* to analyze the abundance and location of chain-like particle arrangements in multi-color microscopy based on multi-marginal optimal unbalanced transport methodology. Our object-based colocalization model statistically addresses the effect of incomplete labeling efficiencies enabling inference on existent, but not fully observable particle chains. We showcase that MultiMatch is able to consistently recover existing chain structures in three-color STED images of DNA origami nanorulers and outperforms geometry-uninformed triplet colocalization methods in this task. MultiMatch generalizes to an arbitrary number of color channels and is provided as a user-friendly Python package comprising colocalization visualizations.

Colocalization analysis aims to unravel the interconnection and interaction network between two or more groups of particles based on their spatial proximity in a microscopy image. By visualizing biological structures, like DNA, RNA and proteins, that are only a few nanometers in size, colocalization analysis makes it possible to study a wide range of biological processes, such as DNA replication and the transcription of genes[1], nuclear import of splicing factors[2] or the dynamics of cargo sorting zones in the trans-Golgi networks of plants[3], to name only a few.

In the following, we will denote any objects of interest that are depicted within a microscopy image, e.g., proteins as well as loci on DNA or RNA strands, as *particles*. In fluorescence light microscopy, such particles are stained, i.e., in case they do not already intrinsically fluoresce, they are labeled with fluorophores, which in turn are excited by an external light source. The emitted fluorescence radiation then can be imaged via several microscopy technologies.

Diffraction unlimited super-resolution fluorescence microscopy technologies, also called nanoscopy, are classified into two broad concepts[4]:

In coordinate-stochastic microscopy, fluorophores within the sample are stochastically excited resulting in a temporally resolved blinking dynamic[5–7], which allows to spatially separate fluorophores. Their coordinates are estimated by means of the detected radiation peak, yielding a list of coordinates of detected fluorophores as output data. If only one fluorophore is detected for one particle, the output translates into a list of particle coordinates. Else, fluorophore coordinates can be aggregated in order to localize the particle of interest in the imaged biological sample.

In scanning-based microscopy methods such as Stimulated Emission Depletion (STED)[8–10], the fluorescence distribution is stored as an intensity matrix, in which every entry encodes the detected radiation within a respective pixel of the microscopy image. To obtain coordinate estimates of particle positions, object detection algorithms have to be applied to the intensity matrix.

In order to study possible particle interactions or connections, stainings with different fluorescent markers are recorded in different color channels. Particles colocalize, if they are spatially closer than or equal to a

[1]Center for Integrative Bioinformatics Vienna (CIBIV), Max Perutz Labs, University of Vienna and Medical University of Vienna, Vienna, Austria. [2]Vienna Biocenter PhD Program, a Doctoral School of the University of Vienna and Medical University of Vienna, Vienna, Austria. [3]Institute for Mathematical Stochastics, University of Göttingen, Göttingen, Germany. [4]Cluster of Excellence 'Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells' (MBExC), University of Göttingen, Göttingen, Germany. [5]Department of NanoBiophotonics, Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany. [6]Clinic of Neurology, University Medical Center Göttingen, Göttingen, Germany. [7]Institute for Computer Science, University of Göttingen, Göttingen, Germany. [8]Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Translational Neuroinflammation and Automated Microscopy TNM, Göttingen, Germany.
✉e-mail: munk@math.uni-goettingen.de

colocalization distance, which heavily depends on the underlying biological setting and might be unknown prior to colocalization analysis[11].

Colocalization methods are divided in two categories based on the input data format they require:

Pixel-based colocalization methods take an intensity matrix as input and compare the pixel intensities across color channels, e.g., by utilizing overlap, correlation or intensity transport analysis. Such approaches are thus only applicable for scanning-based images and examples for well-established methods are Mander's Colocalization Coefficient[12,13], Pearson's Correlation Coefficient[14], BlobProb[15], SACA[16], and OTC curves[17].

Object-based colocalization methods, which our method MultiMatch classifies as, require the coordinates of particles and evaluate their distances, where pairwise particle distances can be defined in several ways[18]. Examples for other object-based tools are ConditionalColoc[18] and Ripley's K based methods[19,20] as SODA[21].

While nanoscopy for dual-color stainings is well studied for a long time, multi-color imaging including three or more stainings has received increased attention more recently since it allows simultaneous measurements of multiple particle types. There is a steadily increasing number of published multi-color STED microscopy datasets[22–28], of other super-resolution microscopy methods[29,30] and the development of appropriate labeling methods allowing for an ever-increasing number of channels is ongoing[24,30–33].

However, most pixel- and object-based colocalization tools are designed for and therefore limited to the analysis of two-color stainings. Applying them to multi-color images is not an obvious task: Particle arrangements with more than two different particle types can occur in different configurations and depending on the biological context, some may be of interest and others may simply not exist in the imaged sample. A geometry-uninformed, pairwise analysis of all possible channel combinations[34], as well as the few established methods that are explicitly presented as multi-color pixel-based[15,35–37] and object-based[18,21,38] colocalization tools are prone to overestimate colocalization, as soon as the biological complex of interest has a fixed geometry and stoichiometry, as we can show in a simulation study. To exploit the full potential of multi-color microscopy imaging in such a situation, it is therefore beneficial to actively incorporate prior knowledge of the local geometry into the colocalization analysis.

To this end, we introduce MultiMatch, a widely applicable colocalization methodology based on optimal transport theory, which is especially tailored to detect chain-like, one-to-one particle arrangements. Integrating this type of colocalization geometry optimizes the multi-color colocalization analysis of quadruples, triplets, pairs, and singlets, as they appear when marking different loci of a chain-like molecule with multi-color stainings. MultiMatch is able to statistically address the effect of incomplete labeling efficiencies on the detection results and includes statistical guarantees on the estimated number of structures of interest. It is provided as computationally efficient Python package allowing for a user-friendly visualization of colocalization results via colocalization curves and the exploratory napari viewer[39].

Results

Chain-like particle assembly detection with MultiMatch

One exemplary biological framework, in which the localization of chain-like particle arrangements is especially insightful, is the highly condensed mammalian mitochondrial genome: It is transcribed from both strands of the mitochondrial DNA as long polycistronic transcripts that have to undergo multiple processing steps, including endonucleolytic cleavage, in order to get to the different functional RNA species. Transcription of the heavy strand leads to polycistronic primary transcripts containing the premature mRNAs of 12 of the 13 oxidative phosphorylation (OXPHOS) subunits encoded in the mitochondrial genome. Labeling more than two of the mRNAs within such a primary construct, in combination with our colocalization approach, can significantly contribute to our understanding of the post-transcriptional processing steps and their dynamics, that lead to the generation of matured mRNA molecules[40,41].

We consider a particle arrangement as *chain-like* when exactly one particle of each type is stringed together in an ordered fashion and pairwise distances of chain-neighbors are smaller than or equal to a maximal colocalization threshold t. In MultiMatch we implemented the distance between reference points, i.e., the center of detected particles, as t by default (Fig. 1A). This approach is especially suited for particles of small size or in case the center of the particle is a suitable representation for its location on the microscopy image. However, we allow the user to also input arbitrary particle-to-particle distance matrices[18] as they are output by alternative particle detection and segmentation algorithms.

Even if the biological complex of interest itself is not chain-like, chain detection still can give substantial insights on the abundance and location of colocalization events inside a microscopy image as soon as the chain is a substructure of the colocalization geometry (Fig. 1B). The converse, on the other hand, does not hold true in general.

To fix the chain order of particles, we will refer to color channels, in which the respective particle type was imaged, as channel A, B, C, D etc. For simplicity, we will explain the main methodology for a three-color setting in what follows, but MultiMatch is applicable to an arbitrary number of color channels, which we showcase in the evaluation of simulated four-color STED images. We stress, that our software is already designed to process any number of channels ("Methods" section).

All configurations resulting from a three-color staining of an chain-like molecule are sketched in Fig. 1C, where we assume the following unknown abundances $\boldsymbol{n} = (n_{ABC}, n_{AB}, n_{BC}, n_A, n_B, n_C)$ of chain-like assemblies, where

n_{ABC} is the number of true ABC triplets,
n_{AB}, n_{BC} is the numbers of true AB and BC pairs,
n_A, n_B, n_C is the numbers of true A, B, and C singlets.

Optimal transport (OT) theory[42] has a wide range of applications throughout statistics[43], data science, and machine learning[44]. Generally, OT aims to allocate (transport plan) one mass distribution into another by minimizing the transportation cost arising from moving one mass unit from one location to another. Applied to fluorescence intensity distributions on a pixel grid and using the euclidean distances between pixels as transportation cost, OT introduces an intuitive distance between two microscopy images and could already successfully be utilized in the context of pixel-based, dual-color colocalization methods[17,45].

For object-based analysis, reference points of detected particles can also be interpreted as support points of mass one of a (discrete) two-dimensional distribution. For only two color channels with the same number of particles the standard OT problem simply assigns each particle from the first channel to one particle from the second channel while minimizing the total sum of Euclidean matching costs. This can be also generalized for other particle-to-particle distance matrices, in case the Euclidean distance between particle reference points is not suitable to represent particle proximity. We can obtain an optimal matching between more than two particle types by multi-marginal OT[46,47] and at the same time account for the not necessarily equal numbers of support points per channel by utilizing an unbalanced OT formulation[48] ("Methods" section). A combination of both OT generalizations, i.e., multi-marginal optimal unbalanced transport problems, have been recently discussed in the literature[49–52].

In this manner, the basic concept of MultiMatch can be interpreted as a linear assignment problem as described, e.g., in the field of object tracking[53–56]. In contrast to methods of this research field, we explicitly formulate the matching problem as a function of the colocalization threshold, allowing to plot the chain abundances dependent on a range of t ("Methods" section). We utilize the equivalence of the optimal transport methodology to a network flow problem to overcome the otherwise prohibitively high computational complexity of its corresponding linear program formulation[57] ("Methods" section, Supplementary Note 1, and Supplementary Fig. 1).

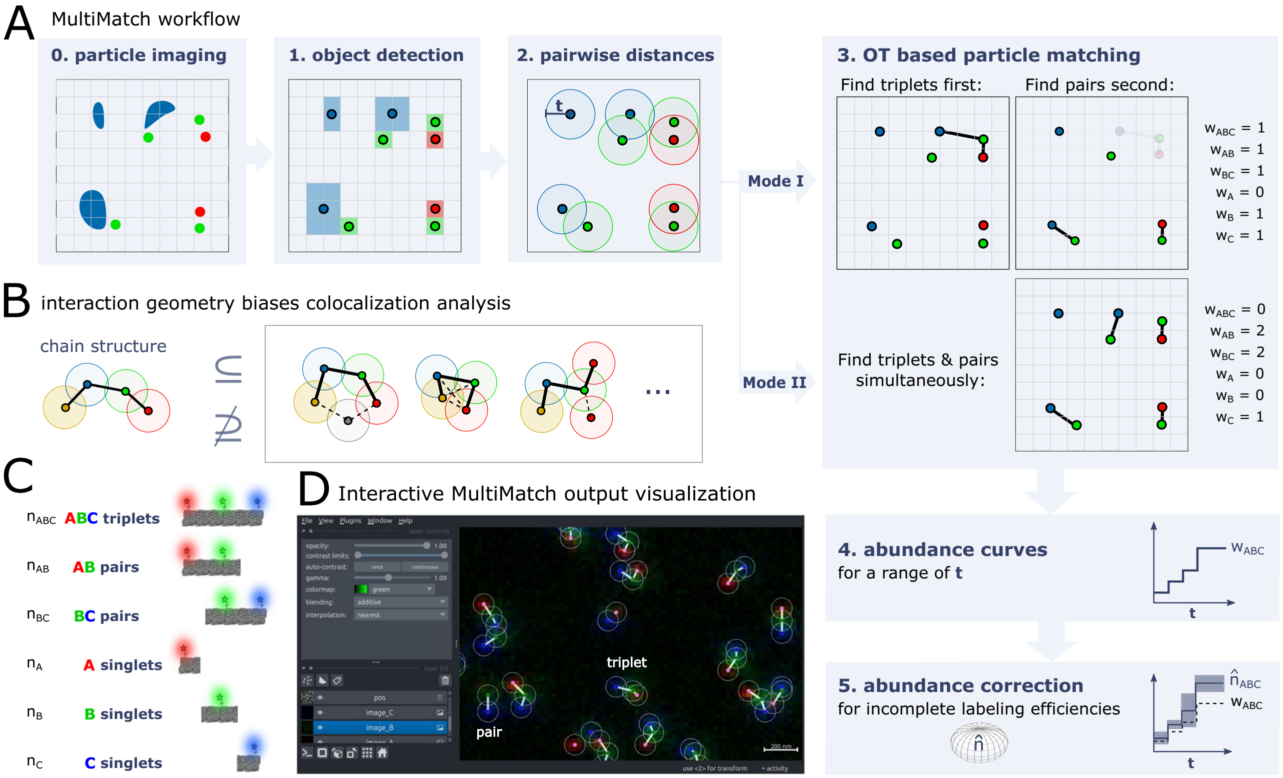

Fig. 1 | MultiMatch workflow to detect chain-like particle arrangements in multi-color microscopy images. A After microscopy imaging (0) and object detection (1), the distances between channel-specific lists of reference points or a user-defined distance matrix are input to the optimal matching procedure. Restricted on particle pairs with distance smaller or equal than colocalization distance t (2), MultiMatch either outputs the maximal number of triplets and subsequently pairs (Mode I) or simultaneously searches for triplets and pairs (Mode II) (3). MultiMatch provides the localization and number of detected chains for a known or abundance curves for a range of colocalization distances t (4). For known incomplete labeling efficiencies true abundances can be estimated with confidence statements (5) ("Methods" section). **B** If more than two different particle types are involved, multiple geometric colocalization patterns can emerge. In case the chain is a substructure of the colocalization geometry of interest, its detection will help to localize and quantify colocalization events. **C** Structures of interest in three-color colocalization analysis for chain-like, one-to-one particle interactions and fixed particle type order. All pairwise distances between neighboring particles in a chain are smaller or equal than colocalization distance t. **D** Exemplary MultiMatch output for an experimental STED image of DNA origami nanoruler structures (as sketched in **C**) in the interactive napari viewer[39].

MultiMatch provides two different modes to solve the particle matching problem (Fig. 1A(3)):

Mode I: By restricting a k-marginal optimal unbalanced transport problem to particle pairs with a distance smaller than t and introducing a chain-cost that only considers distances between neighboring particle types ("Methods" section), the resulting OT plan encodes the *maximal* number of, for $k = 3$, triplets within the nanoscopy image. If requested, the matching process is subsequently repeated on the remaining particles to detect yet unresolved AB and BC pairs, respectively.

Mode II: This mode only detects AB, BC, etc. pairs by solving respective *two*-marginal unbalanced OT problems. Subsequently, the two-marginal OT matchings are coupled to chain structures: For $k = 3$, all pairs occupying the same intermediate particle are redefined as respective ABC triplet.

Depending on the underlying biological experiment, the user can select the appropriate mode for colocalization analysis: Mode I prioritizes the detection of a predefined chain structure of choice. For example, if a user aims to analyze triplets, Mode I will detect a triplet as soon as three particles A, B, and C are close enough to each other – even if another particle A or C is nearby that would allow to match two pairs instead of one triplet (as depicted in Fig. 1A(3)). If $k > 3$ and the user wants to detect multiple chain structures, one needs to set a prioritization order for Mode I. For example, for $k = 4$ and after ABCD quadruplet detection, one can search either for ABC or BCD triplets next. Depending on the order, the final matching results may change as soon as some particles cannot be uniquely assigned to one particle arrangement.

Mode II, on the other hand, does not need a predefined prioritization order of structures for subsequent matching steps, hence it does not overemphasize structures that are matched in the earlier steps. It is useful in case we do not have any prior knowledge on which structures might appear in the microscopy image and we do not want to prioritize any chain structures.

In the evaluation of experimental and simulated three-color STED microscopy images we show that for sparse particle distributions and mixed singlets, pairs, and triplet ratios the differences in detected abundances between the two modes is neglectable (Supplementary Note 2, Supplementary Fig. 2). However, in case of dense particle distributions (Supplementary Note 3 and Supplementary Figs. 3–5), or in case we know in advance that only one chain structure exists in the biological context, the multi-marginal approach of Mode I, which is also the default setting in the MultiMatch tool, outperforms the pairwise matching approach of Mode II.

MultiMatch outputs detected abundances $\boldsymbol{w} = (w_{ABC}, w_{AB}, w_{BC}, w_A, w_B, w_C)$ for a known colocalization distance t and depicts configuration positions on the respective microscopy image allowing further investigation on the spatial distribution of recorded biological complexes. If t is unknown (optionally channel-wise scaled) abundance curves $\boldsymbol{w}(t)$ are output for a user-defined range of t values. MultiMatch is compatible with the interactive Graphical User Interface of napari (Fig. 1D) enabling the visual evaluation of structure locations for different t values in form of a colocalization threshold slider.

The differentiation between triplets, pairs, and singlets within a microscopy image is additionally hindered by incomplete labeling efficiencies and point detection artifacts. This is a notorious problem in

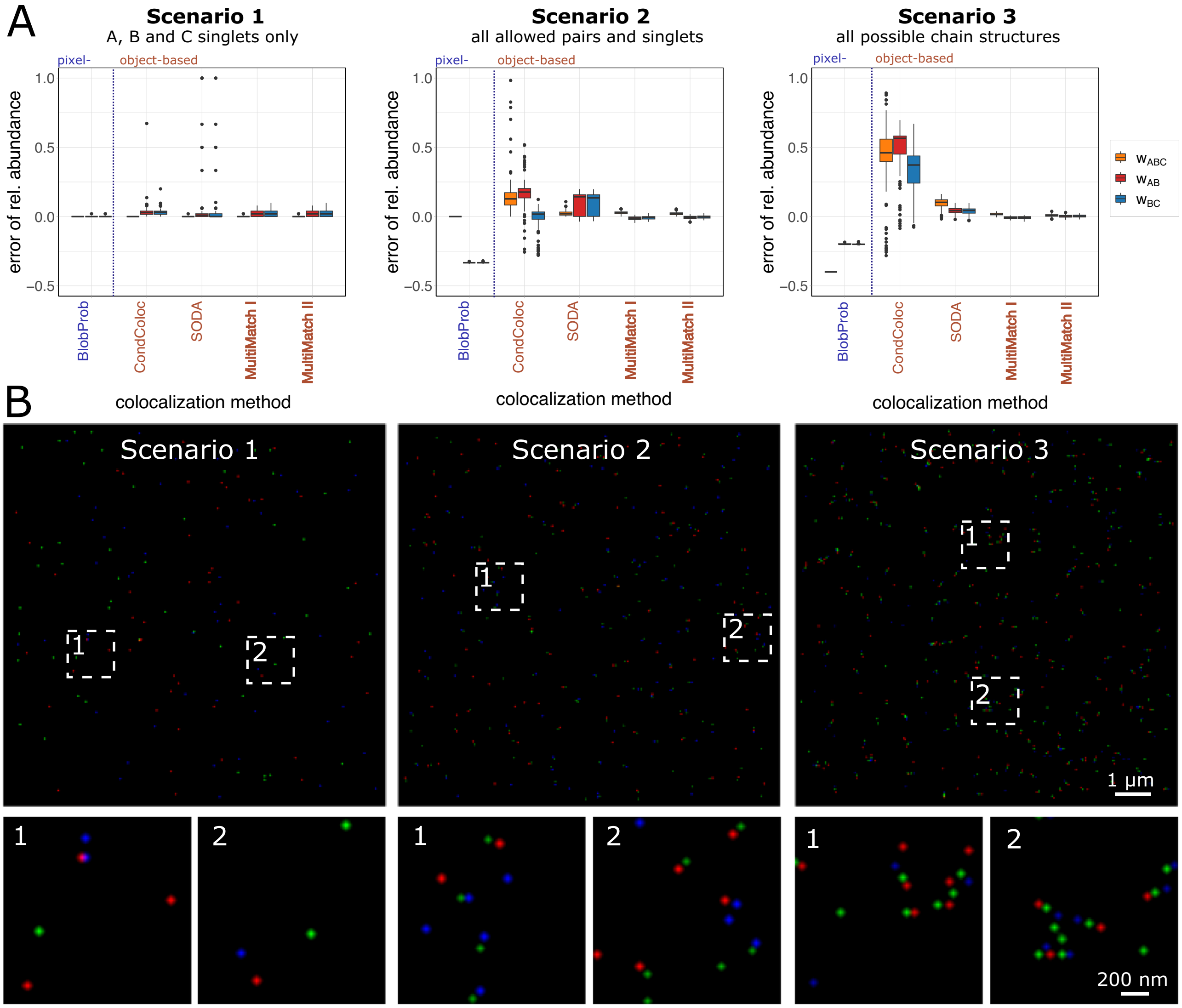

Fig. 2 | Simulation study for three-color microscopy images with three combinations of chain structures. In each scenario 100 independent STED images and different abundances of triplets, pairs, and singlets were simulated with 100% labeling efficiency. **A** Method specific boxplots of the errors in detected relative (scaled by the total number of points in channel B) structure abundances are displayed. The error is computed by subtracting true relative abundance from detected relative abundances. In *Scenario 1* only A, B, and C singlets, in *Scenario 2* all possible singlets as well as AB and BC pairs and in *Scenario 3* ABC triplets, AB, BC pairs and A, B, and C singlets were simulated. **B** Simulated STED images for Scenarios 1, 2, and 3 with respective image details. For visualization purposes, contrast stretching and increasement of image brightness was applied.

fluorescence microscopy, e.g., described in Hummert et al.[58], and missing detections can add an unpredictable bias toward systematic underestimation of triplet numbers and overestimation of singlet abundances, if not corrected. Currently, the problem of incomplete labeling efficiency is barely addressed in the field of colocalization analysis. Therefore, we propose a statistical framework to correct for incomplete labeling efficiencies and introduce an unbiased estimator $\hat{\boldsymbol{n}}(t)$ of true chain-structure abundances and confidence statements on the estimated quantities ("Methods" section, Supplementary Note 4, Supplementary Fig. 6, Supplementary Table 1).

An overview on the full workflow of MultiMatch from microscopy image to abundance curves is depicted in Fig. 1A.

Simulation study

To systematically evaluate the performance of MultiMatch against compatible colocalization methods, we simulated 100 microscopy images for each of three scenarios with different combinations of singlets, pairs, and triplet abundances. For this simulation study, we decreased the noise level to a minimum to allow a fair comparison despite different point detection tools implemented in the respective colocalization tools. Also, we amplified simulating linear triplet structures over randomly folded triplets ("Methods" section). For every simulated image,

Scenario 1: 50 singlets of each type A, B, and C were simulated.
Scenario 2: 50 A, B, and C singlets and 50 AB and BC pairs were simulated, respectively.
Scenario 3**:** 100 triplets and 50 AB and BC pairs and 50 A, B, and C singlets were simulated, respectively.

Exemplary, simulated images and the results of the simulation study for a fixed colocalization threshold of $t = 5$ pixels are shown in Fig. 2A, B. Analysis results for all considered methods across a range of colocalization thresholds are presented in Supplementary Note 3 and Supplementary Fig. 3.

As a representative of pixel-based methods, we include BlobProb[15], which counts the number of colocalized intensity blobs, i.e., groups of neighboring pixels with high intensity. In each channel, blobs are detected via image segmentation and for each blob the local intensity maximum is defined as reference particle coordinate. A blob pair colocalizes if the first reference point lies within the second blob and vice versa. Triplet colocalization is detected if all involved reference points are included in all three blobs. SODA[21] is an object-based method, which uses the Ripley's K function[19] and computes the coupling probability of point pairs based on marked-point process theory. In the most recently published method ConditionalColoc[18] particles are defined as colocalized as soon as their distance is below a maximal colocalization radius. Then, utilizing Bayes' Theorem, (conditional) probabilities are computed and assigned for triplet and pair colocalization. We experienced that ConditionalColoc, although aiming to output probabilities, in some cases yields values greater than one and hence the errors in relative abundance detection are not bounded by one as well. For a better comparison, we restricted the respective results to values between -0.5 and 1 in Fig. 2A and show ConditionalColoc outliers in Supplementary Note 5, Supplementary Fig. 7.

In none of the above methods triplet colocalization is restricted to one-to-one interactions. This has barely any negative effect on the detection of singlets in Scenario 1, where no additional pairs and triplets occur. Apart from few outliers of overestimation in pairs and triplet abundances in ConditionalColoc and SODA, all considered colocalization measures show consistently low errors with small variability. The maximal median error in relative abundances of 0.03 in Scenario 1 is obtained by ConditionalColoc in the detection of AB as well as BC pairs.

In Scenarios 2 and 3 on the other hand, we observe a consistent overestimation of relative pairs and triplet abundances in object-based methods SODA and ConditionalColoc, since one particle can be included in several structures at the same time. Additionally, in Scenario 2 SODA exhibits a larger variation in pairs abundances, resulting in median errors 0.14 in both AB and BC pairs with interquartile ranges of 0.16, respectively. In Scenario 3 the variation in abundance detection decreased and median errors are 0.1 for ABC triplets and 0.04 for AB as well as BC pairs. ConditionalColoc performances worst in Scenario 3 yielding a median error of 0.48 for ABC triplets.

The pixel-based method BlobProb mostly obtains zero relative abundances of triplets and pairs across all three scenarios and hence severely underestimates the triplet and pair configurations within the simulated images. This is due to the high resolution in the simulation setup, which was chosen to mimic conventional STED imaging. If particles are small and their respective intensity blobs do not overlap, BlobProb does not detect any colocalization.

MultiMatch on the other hand searches for optimal matches on a global scale while considering the local geometry of chain-like particle assemblies. It consistently recovers the ground truth abundances for each simulation scenario. The maximal median error across all scenarios and chain structures for both Modes of MultiMatch is 0.03 with a maximal interquartile range in errors of 0.04.

Apart from above considered, already established colocalization methods, we also implemented a Nearest Neighbor Matching as comparable object-based method. We can show that greedily matching particle pairs based on local optima leads to underestimation of ABC triplets in dense particle distributions (Supplementary Note 3 and Supplementary Fig. 4).

Incomplete labeling efficiencies and point detection errors

In experimental STED microscopy, typically it is impossible to record all existing particles of interest. This can, for example, be due to the fluorescent marker not being successfully attached to the probe or a flawed point detection. All such scenarios resulting in a failure of particle detection for simplicity will be summarized under incomplete labeling efficiency hereafter.

If only singlets were to be counted in multi-color images with the same labeling efficiency across channels, the relative abundance could still be estimated consistently. However, as soon as configurations of two or more particle types are to be recovered, incomplete labeling efficiencies can lead to under- and overestimation of structures. Figure 3A shows that a triplet can be erroneously detected as pair or singlet or not at all, which can introduce a severe bias. However, if the labeling efficiencies are known, the detection success of a particle can be modeled with a Bernoulli distribution, which allows the definition of an unbiased estimator $\hat{\boldsymbol{n}}$ for the vector of true chain structures abundances $\boldsymbol{n}$. This approach allows for constructing multi-dimensional joint confidence ellipsoids covering $\boldsymbol{n}$ with a given significance level, e.g., $\alpha = 0.1$ (Fig. 3B, C). The multi-dimensional confidence ellipsoid then can be respectively projected onto one dimension to obtain structure-specific confidence intervals or bands for a range of t values, while fixing the estimated abundances of all other considered structures ("Methods" section).

Note that microscopy images are also influenced by other sources of noise that complicates the detection of chain-like particles as we show in Fig. 3D: In this small study we simulated 100 STED images containing only one triplet ($n_{ABC} = 1$) and observe that the discrete nature of the pixel grid can effect on the accuracy of the measured distance between particles and hence the stabilization behavior of colocalization curves. For square pixels with side length l the worst case for pairwise particle comparisons is $\sqrt{2}l$.

Evaluation of experimental STED images

Chain-like particle structures occur within several biological complexes. To showcase the performance of our method on experimentally retrieved data we used one-, two-, and three-color nanorulers. Nanorulers are DNA-origamis with a predefined distance between spots at which 20 fluorophores are attached and hence, as their name suggests, can be used as rulers inside a microscopy image[1,59–61]. For this experimental setup, we chose nanorulers with pairwise distances between neighboring spots of 70 nanometers (nm). For each chain structure (as depicted in Fig. 1C), respective nanoruler origamis are available in separate solutions, which allows us to control whether in an experiment we record singlets, pairs or triplets only or a combination of those structures. We performed three experiments:

Setting 1: The experiment consists of all three single marker nanorulers (22 images in total). We expect to detect no pairs or triplets, i.e., $w_{ABC} = w_{AB} = w_{BC} = 0$.

Setting 2: The experiment consists of all three singlets, two pairs and triplet marker nanoruler solutions (22 images in total). We expect to detect all possible configurations, i.e., A, B, and C singlets, AB and BC pairs as well as ABC triplets.

Setting 3: The experiment consists of only triplet marker nanorulers (12 images in total). We expect to detect ABC triplets only, i.e., $w_{AB} = w_{AB} = w_A = w_B = w_C = 0$.

For each experimental setting we recorded STED images of size 400 × 400 pixels with a pixel size of 25 × 25 nm. In channel A, stainings with Star Red 640 nm are recorded, in channel B, stainings with Alexa 488 and in channel C, stainings with Alexa 594. Note, however, that the exact numbers of nanorulers within a recorded STED image is unknown. Due to misfolding and clumping of nanorulers and different nanoruler immobilization rates for each STED image one cannot compute a fixed unit of nanorulers per microscopy image and experiment.

The results of the colocalization analysis for all three settings (with default MultiMatch Mode I) are shown in Fig. 4A via relative abundance curves with standard deviation bands quantifying variation across images within the same setting. Here, we used MultiMatch Mode I and included the analysis with Mode II showing comparable results, but slightly underestimating the number of triplets in Setting 3, in Supplementary Note 2, Supplementary Fig. 2. Exemplary images for each setting are shown in Fig. 4B.

For Setting 1 we can appreciate that, as expected, across a range of t values only a few pairs and triplets are detected (Fig. 4A). The rise of relative abundance curves is unavoidable for large t, since the probability increases that randomly scattered particles are matched. In Setting 2, despite experimental variation, we clearly recover all supplied nanoruler structures. Even more, colocalization curves are still stabilizing for a colocalization threshold t greater than approximately 4 pixels (=100 nm): For $t > 100$ nm

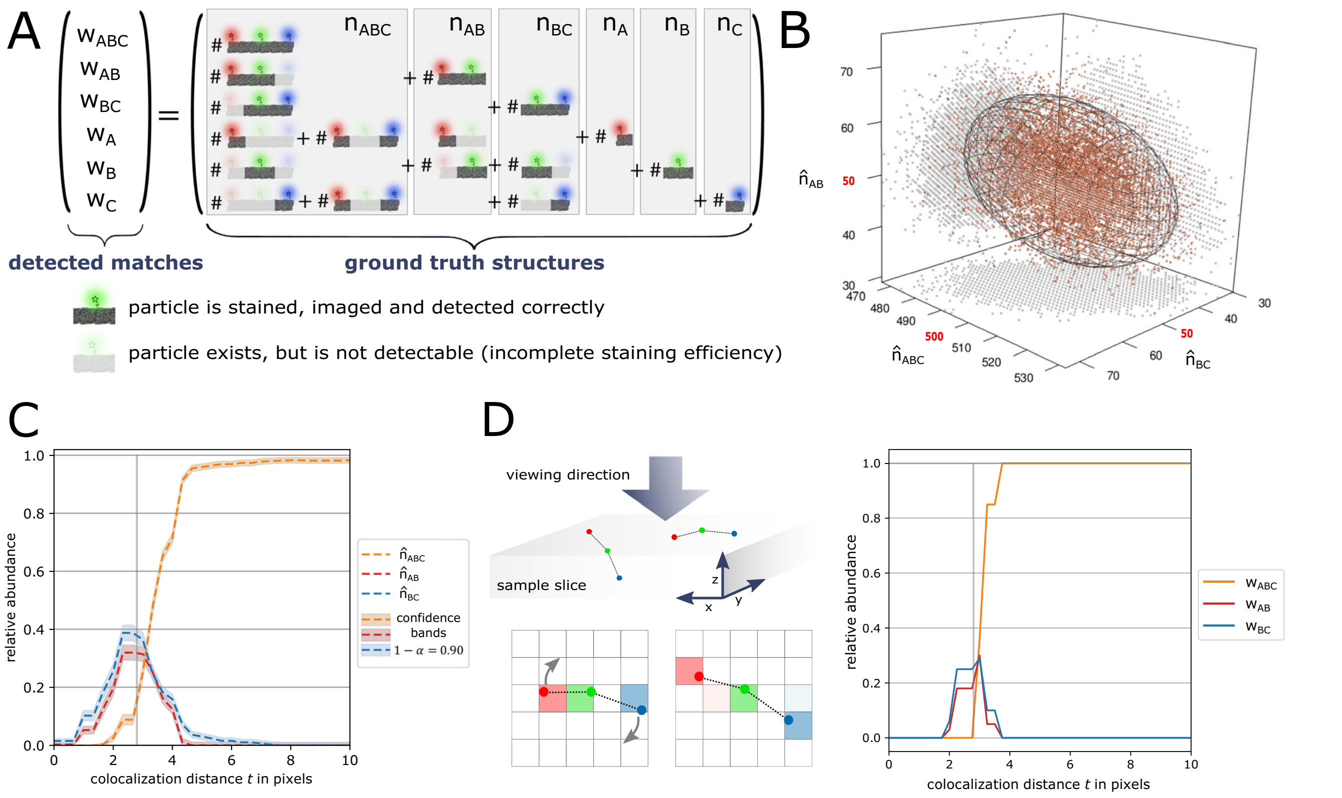

Fig. 3 | Chain-like particle structure detection is influenced by incomplete labeling efficiencies and structure rotation. **A** Because of channel-specific incomplete labeling efficiencies, triplets and pairs can erroneously counted to other structure abundances. **B** For entrywise large enough $\boldsymbol{n}$, estimator $\hat{\boldsymbol{n}}$ is approximately multi-dimensional normally distributed: Estimated abundances of 10,000 independent simulations with labeling efficiencies $s_A = s_B = s_C = 0.95$ and true abundances $n_{ABC} = 500$, $n_{AB} = n_{BC} = n_A = n_B = n_C = 50$ ("Methods" section). The respective 3-dimensional, normal 90% quantile ellipsoid is plotted. **C** Estimated abundance curves for one of the experimental multi-color STED images in Setting 3 with additional confidence bands for significance level $\alpha = 0.1$. **D** Restricted image resolution and 3-dimensional rotation of particle arrangements lead to variability in the observed colocalization thresholds: simulation study of 100 independent images only containing one triplet with pairwise distances set to 70 nm = 2.8 pixels per image (100% complete labeling efficiency, "Methods" section).

ABC triplets are approximately detected with relative abundance of 0.32, AB pairs with 0.16 and BC pairs with 0.42 relative abundance, yielding a relative amount of 0.1 unmatched B singlets. The relative abundance curves of all structures reach a plateau at approximately $t \geq 4$ pixels (= 100 nm), i.e., the slope of all curves within the same setting decreases rapidly. In Setting 3, as expected, the relative abundances of AB and BC pairs converge to zero while triplets are the dominantly detected structure for $t \geq 4$ pixels.

Notably, in Settings 2 and 3 stable abundance curves are reached at around 100 nm, which is 30 nm more than the experimentally fixed, maximal distance between neighboring fluorophore spots in the nanoruler structures. This effect can be explained by the still limited resolution in the microscopy image and can be reproduced via simulation (Fig. 3D).

Limited resolution alone does not explain why 20%–30% of detected B particles (for $t \geq 5$ pixels) are not matched to a triplet in Setting 3: The attachment of a single fluorophore to a nanoruler spot is expected to have a success probability of 85% to 90% and hence at least one fluorophore should be attached to each spot in almost 100% of all cases. Still, due to the above-described experimental variation in nanoruler imaging and additional errors in point detection, especially due to nanoruler clumping, the overall success rate of fluorophore spot detection is incomplete. Hence, we erroneously detect pairs instead of triplets or singlets due to noise. As in Setting 1 those artifacts will be matched into triplets for large enough t.

For simplicity, we model a 90% labeling efficiency across all three-color channels in the experimental STED setup. The estimated abundance curves $\hat{\boldsymbol{n}}(t)$ (dotted lines in Fig. 4A), in Setting 3 visibly correct the measurements towards the expected relative abundances. Additional confidence bands around $\hat{\boldsymbol{n}}$ allow to infer on the robustness of the abundance estimation as presented in (Fig. 3C) for one of the experimental STED images of Setting 3.

Evaluation of simulated four-color STED images

MultiMatch is applicable to an arbitrary number of color channels, which we showcase in a second simulation study with an adapted simulation setup for quadruples, triplets, pairs, and singlets in simulated four-color STED microscopy images. In contrast to the first simulation study simulating triplets, tuples, and singlets, we additionally challenged our MultiMatch tool with an increased noise level and by allowing arbitrarily curled chain structures ("Methods" section). In Fig. 5A–D we show the colocalization analysis results of two simulation scenarios:

Scenario I: We simulated 50 ABCD quadruples, 30 ABC triplets, 20 AB pairs and 30 C and D singlets, respectively, to mimic a chain-like molecule being split at loci C and D.

Scenario II: We simulated 100 ABCD quadruples and no triplets, pairs nor singlets

Exemplary images of both simulation scenarios are shown in Fig. 5E and three additional simulations setups are shown in Supplementary Note 3, Supplementary Fig. 5. For each scenario, we simulated 100 images with full labeling efficiencies ($s_A = s_B = s_C = s_D = 1$) and 100 images with incomplete labeling efficiencies ($s_A = s_B = s_C = s_D = 0.95$) by randomly deleting 5% of points simulated in the prior, full labeling efficiency simulation in each channel.

For this analysis we applied MultiMatch Mode II, i.e., allowing the detection of both ABC as well as BCD triplets and AB, BC and CD pairs without any prioritization order of chain structures. Again, also in the case of four-color images, we can appreciate that MultiMatch consistently recovers true abundances of quadruplets in case of full labeling efficiencies. Absolute abundance curves, as also described in the analysis of our experimental dataset in Fig. 4, stabilize for approximately $t = 4$

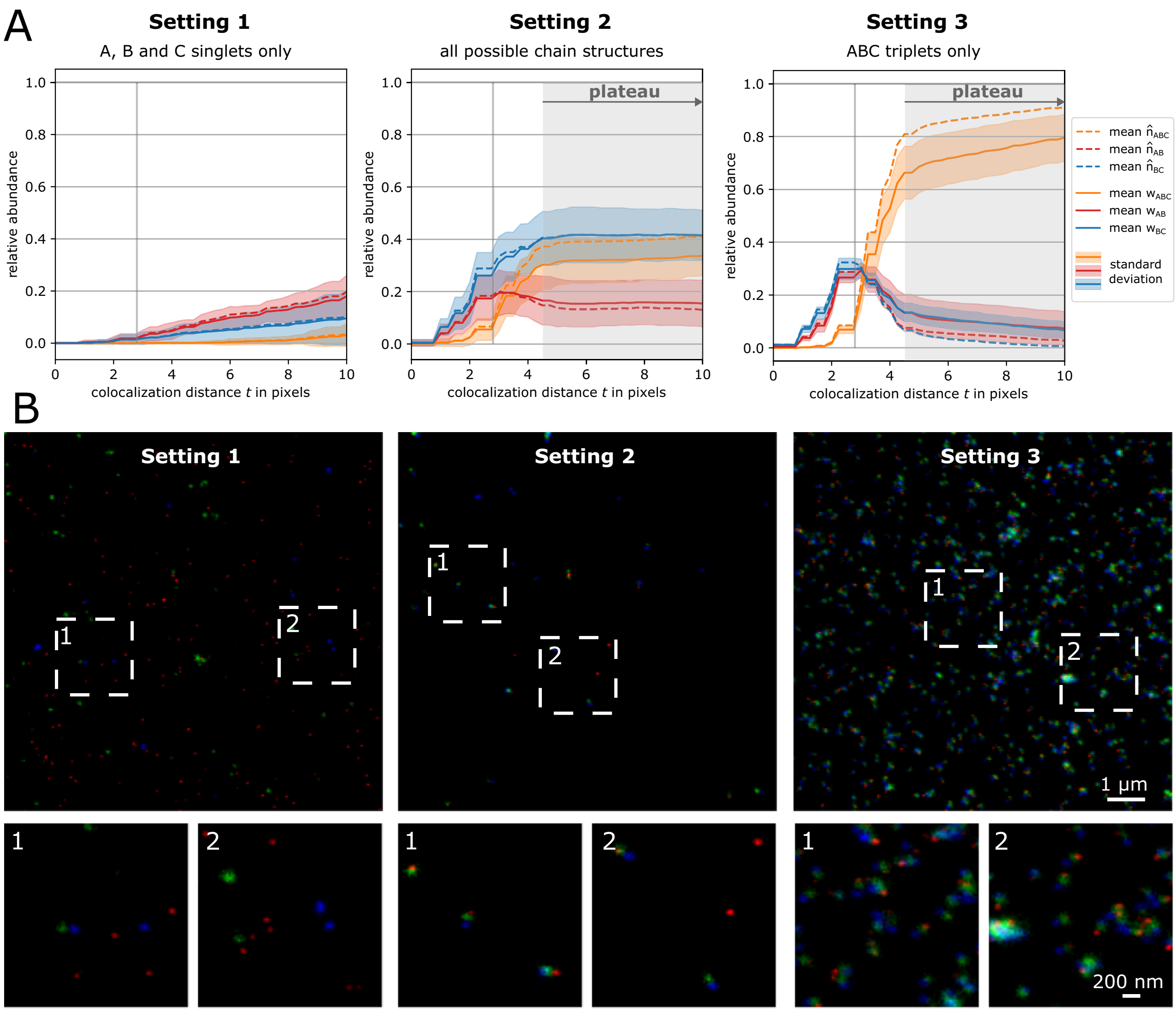

Fig. 4 | MultiMatch Mode I relative abundance curves *w*(*t*) for experimental STED images. For each setting the solid curves are mean relative abundances with standard deviation bands across a range of colocalization threshold *t* from 0 to 10 pixels (25 nm = 1pixel). The abundances are scaled by the total number of points detected in channel B. Additionally, incomplete labeling efficiency (90% in each channel) corrected abundances are plotted as dotted curves. The true colocalization distance of 70 nm within nanoruler structures is depicted as vertical line. **A** Setting 1: mean abundance curves for only singlets consistently show the expected 0% relative triplet and pair abundances (22 independent experimental STED images). Setting 2: triplets, pairs, and singlet nanoruler are detected with stable abundances for ~*t* ≥ 4 pixels (22 independent experimental STED images). Setting 3: mean abundance curves for analyzing the triplet nanoruler solution only. The incorporation of incomplete labeling efficiency clearly corrects the relative triplet abundance towards the in this setup expected 100% (12 independent experimental STED images). **B** Representative STED images for Settings 1, 2, and 3 with image details. For visualization purposes, contrast stretching and increasement of image brightness was applied.

pixels. For images simulated with incomplete labeling efficiencies, the colocalization curves show underestimation of quadruplets as expected. With our statistical framework we again can visibly correct the colocalization curves towards the true, simulated structures abundances and additionally gain confidence bands confirming the stability of our estimator.

For denser distributions, as shown in Supplementary Note 3, Supplementary Fig. 5C-F, we can observe that 1. MultiMatch II misses quadruples for the sake of closer particle pairs, and 2. similar to the experimental nanoruler analysis depends on the performance of the point detection and hence the noise level in the microscopy image. If consistent noise challenges the point detection, abundance curves still stabilize, but the plateau shows a smaller number of matched quadruples than simulated in absolute numbers. Hence, we advise user of MultiMatch to check the noise level of the microscopy image and the point detection result with the interactive napari viewer (Fig. 1D and Supplementary Fig 5G) and if necessary evaluate channel-wise scaled, relative instead of absolute abundances.

Discussion

In this article we introduce multi-marginal optimal unbalanced transport methodology for geometry-informed, multi-color colocalization analysis. We are able to show, that for the analysis of more than two color channels, it is crucial to take into account the colocalization geometry of the biological complex.

By either choosing chain costs in a multi-marginal OT problem (Mode I) or coupling consecutive two-marginal OT matchings (Mode II), MultiMatch successfully detects *k*-chain particle assemblies such as quadruples, triplets, pairs, and singlets, as they appear when staining multiple loci on chain-like molecules like DNA or RNA strands. Both modes have their advantages, which depend on the number of particles imaged and prior

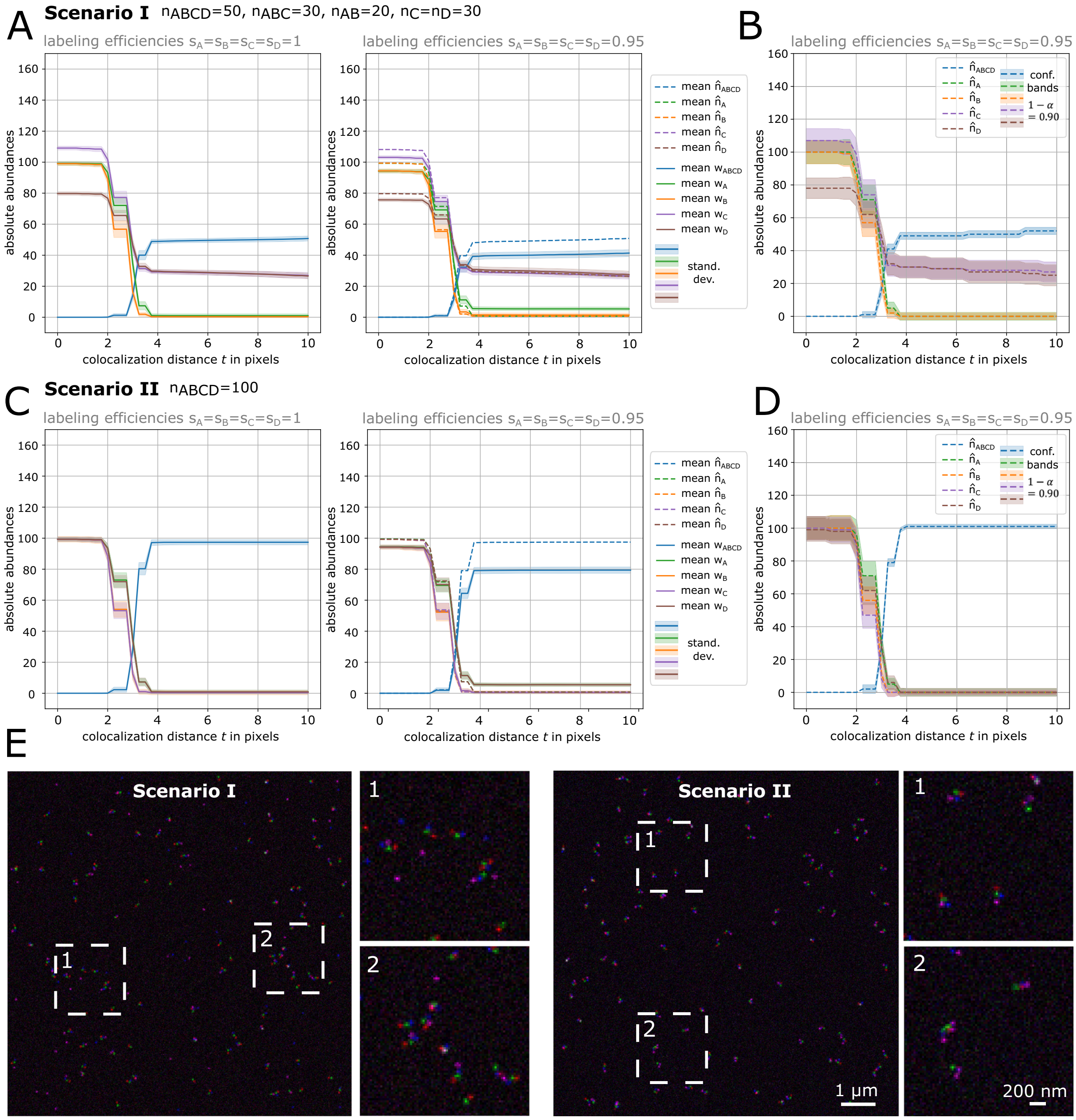

Fig. 5 | MultiMatch Mode II absolute abundance curves $w(t)$ and estimation results $\hat{n}(t)$ for simulated four-color STED images. For each simulated scenario 100 independent images were simulated with complete labeling efficiency ($s_A = s_B = s_C = s_D = 1$) and with incomplete labeling efficiency ($s_A = s_B = s_C = s_D = 0.95$), respectively. Solid curves are mean absolute detected abundances with standard deviation bands across a range of colocalization thresholds t from 0 to 10 pixels (25 nm = 1pixel). Corrected abundances are plotted as dotted curves. **A** Scenario I: a mixture of ABCD quadruplets, ABC triplets, AB pairs and C, D singlets were simulated. All curves stabilize at approximately $t = 4$ pixel close to the true simulated number of structures. For images with incomplete labeling efficiency uncorrected detected abundances plus standard deviation bands are plotted as solid curves showing consistent underestimation of quadruples. Corrected abundances recover the true number of simulated structures. **B** For one exemplary STED image of Scenario I simulated with incomplete labeling efficiency, corrected abundance curves and corresponding confidence bands are shown. **C**, **D** show the same analysis as shown in A and B but for Scenario II: only ABCD quadruplets were simulated. **E** Representative STED images for Scenarios I and Scenario II with image details. For visualization purposes, contrast stretching and increasement of image brightness was applied.

knowledge on the biological context: Mode I is best for detecting one chain structure of choice and is more robust in dense particle distributions. When the particle distribution is sparser and multiple chain structures in the imaged biological setting are of interest, Mode II is suited to detect them without any predefined prioritization order.

Since often the true colocalization distance is unknown, MultiMatch results can be output as structure-wise relative or absolute abundance curves across a range of colocalization thresholds t. In our simulation studies as well as our experimental settings we could show, that output curves stabilize close to ground truth abundances.

However, as for all object-based colocalization methods, the performance MultiMatch scales with the noise level of the microscopy image, the performance of the object detection and the resolution of the microscopy. Abundance curve plateaus can be less clear in case the microscopy image contains detected singlets of different particles types. In this case, the larger t, the more far away singlets are matched. In such cases it might be unclear, whether singlets truly exist in the biological sample or whether they are an artifact of the experiment and image processing. For such cases, we advise to observe the quality of the microscopy image with the MultiMatch compatible, interactive napari viewer.

Our network flow implementation significantly decreases computational costs compared to standard approaches solving comparable OT problems and comparable colocalization tools ("Methods" section). The simulation studies show that as soon as we have prior knowledge on the chain colocalization geometry, MultiMatch, in contrast to other triplet colocalization methods, is robust against overestimation of triplets with chain geometry since it only considers one-to-one interactions. MultiMatch is also tested on experimental STED images of different nanoruler combinations and can correct structure abundances for predefined incomplete labeling efficiencies and point detection errors, where confidence bands allow further inference on the estimated abundances.

All experimental studies have been performed for $k = 3$ color channels. However, in many scientific fields the detection of k-chains for larger k is of interest. The mathematical and statistical frameworks allow straightforward generalization ("Methods" section) and we exemplarily show successful detection results for simulated four-color STED images. With current technical standards, the experimental setup of multi-color nanoscopy imaging is still challenging, costly and time consuming, but in view of further technological improvements our algorithm is already applicable for the evaluation of this type of experimental setups, and especially promising in view of recent developments in super-resolution microscopy with a resolution of a few nanometers and below[62,63].

In the same way channel specific colocalization thresholds as t_{AB}, t_{BC} and t_{CD} can be considered within the OT problem. Although we only present the evaluation of 2D STED images with constant labeling efficiencies across channels, our software package can directly be applied to multi-color 3D microscopy images with channel-specific labeling efficiencies.

Limitations: If the microscopy image shows especially dense point clouds, MultiMatch necessarily will have difficulties in differentiating between random and biological reasonable proximity. Note, however, that this is not a specific weakness of MultiMatch, but any other method will face this identifiability problem, which is caused by missing linkage information. It can only be overcome with additional prior information of the underlying biological sample. However, MultiMatch Mode I is especially robust against dense particle distribution in comparison to pairwise matching approaches as implemented in MultiMatch Mode II or greedy Nearest Neighbor Matchings. An adaption to tree like particle arrangements and the inclusion of additional constraints, e.g., incorporating regions of interest are future research objectives.

Methods

Optimal chain-matching

In the following we will denote the sets of two-dimensional particle coordinates in the image domain for each of the k color channels as

$$X^{(1)} := \left\{\boldsymbol{x}_l^{(1)}\right\}_{l=1}^{n_1}, \ldots, X^{(k)} := \left\{\boldsymbol{x}_l^{(k)}\right\}_{l=1}^{n_k} \subseteq \mathbb{R}^2, \tag{1}$$

where number of particles $n_j \in \mathbb{N}_{\geq 0}$ for $j \in \{1, \ldots, k\}$. For simplicity and related to the considered data in this article, we will only consider the cases $k = 2, 3$ in the following. Generalization to larger k is straight-forward. In a chain-like particle arrangement of the form $\left(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(k)}\right)$ with $\boldsymbol{x}^{(j)} \in X^{(j)}$, all neighbors $\boldsymbol{x}^{(j)}, \boldsymbol{x}^{(j+1)}$ have to be closer than the colocalization threshold t and we will denote according tuples as $\mathbf{d}_t^k$-chains:

Definition 1. ($\mathbf{d}_t^k$-chain). Fix $k \geq 2$. For sets $X^{(1)}, \ldots, X^{(k)}$, a distance $\mathbf{d} : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ and a predefined maximal threshold $t \geq 0$ a tuple of k points

$$\left(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(k)}\right) \in \mathbb{R}^{2\times k} \quad \text{with} \quad \boldsymbol{x}^{(j)} \in X^{(j)} \text{ for } j \in \{1, \ldots, k\} \tag{2}$$

is a $\mathbf{d}_t^k$-chain, if pairwise point distances along the fixed tuple point order are smaller or equal than t, i.e.,

$$\mathbf{d}\left(\boldsymbol{x}^{(j)}, \boldsymbol{x}^{(j+1)}\right) \leq t \quad \text{for} \quad j \in \{1, \ldots, k-1\}. \tag{3}$$

In the context of our colocalization problem, $\mathbf{d}$ is the Euclidean distance (this can easily be generalized), a $\mathbf{d}_t^3$-chain is a *triplet* and a $\mathbf{d}_t^2$-chain a *pair*. For given t, we now aim to detect as many $\mathbf{d}_t^k$-chains as possible:

Definition 2. (Optimal $\mathbf{d}_t^k$-matching). A collection of pairwise disjoint $\mathbf{d}_t^k$-chains is called $\mathbf{d}_t^k$-matching. It is called *optimal* if its number of chains is maximal among all matchings.

Such an optimal $\mathbf{d}_t^k$-matching can be found by utilizing a multi-marginal and unbalanced formulation of OT. For example, if $k = 3$, for each channel $i = 1, 2, 3$, we interpret coordinates of detected particles as support points with mass 1 of a respective discrete distribution. Due to this discrete structure, the resulting optimization problem will be finite-dimensional. Since in our measurements the number of detected particles per channel might differ, we require an unbalanced formulation to compare distributions with different total masses. A wide variety of penalty terms for mass discrepancies has been studied in the literature, see for instance[64]. Our problem formulation is closely related to an ℓ^1-penalty for unmatched particles, see also[52]. We first consider the problem of finding optimal d_t^2-matchings between two point clouds, i.e. $k = 2$. This can be solved via the following optimization problem:

Definition 3. (Optimal $\mathbf{d}_t^2$-matchings via unbalanced optimal transport). Let $\lambda \in \mathbb{R}_{\geq 0}$, set the cost function

$$\begin{aligned} c : \mathbb{R}^2 \times \mathbb{R}^2 &\to \mathbb{R}_{\geq 0} \cup \{\infty\}, \\ (x_1, x_2) &\mapsto \begin{cases} \mathbf{d}(x_1, x_2) - \lambda & \text{if } \mathbf{d}(x_1, x_2) \leq t, \\ +\infty & \text{otherwise}, \end{cases} \end{aligned} \tag{4}$$

and $\boldsymbol{c} \in \mathbb{R}^{n_1 \times n_2}$ the pairwise cost between all points in $X^{(1)}$ and $X^{(2)}$, defined by $c_{i_1,i_2} = c(\boldsymbol{x}_{i_1}^{(1)}, \boldsymbol{x}_{i_2}^{(2)})$. The optimal unbalanced transport problem of interest can now be stated as the following linear program

$$\begin{aligned} &\underset{\pi \in \mathbb{R}^{n_1 \times n_2 \times n_3}}{\arg\min} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} c_{i_1 i_2} \pi_{i_1 i_2} \\ &\quad s.t. \sum_{i_2=1}^{n_2} \pi_{i_1 i_2} \leq 1 \textit{ for all } i_1 = 1, \ldots, n_1 \\ &\qquad \sum_{i_1=1}^{n_1} \pi_{i_1 i_2} \leq 1 \textit{ for all } i_2 = 1, \ldots, n_2 \\ &\qquad \pi_{i_1 i_2} \geq 0 \textit{ for all } (i_1, i_2) \in \{1, \ldots, n_1\} \times \{1, \ldots, n_2\}. \end{aligned} \tag{5}$$

Entries of an optimal $\boldsymbol{\pi}$ indicate which particles have been matched. The constraints enforces that each particle can at most be part of one matching, but it may also be discarded. By the definition of the cost vector $\boldsymbol{c}$, the solution of Equation (5) does not match points $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ as soon as they are farther apart than t, but for each matching below distance t there is an incentive by the parameter λ. For λ sufficiently large in comparison to t one can show that the solution yields an optimal $\mathbf{d}_t^2$-matching. Among all optimal matchings the above problem prefers one with the lowest sum of pairwise particle distances among matched particles.

We now generalize this to $k = 3$ via a multi-marginal transport problem.

Definition 4. (Optimal $\mathbf{d}_t^3$-matchings via unbalanced multi-marginal optimal transport). Let $\lambda \in \mathbb{R}_{\geq 0}$, set the cost function

$$\begin{aligned} c : \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 &\to \mathbb{R}_{\geq 0} \cup \{\infty\}, \\ (x_1, x_2, x_3) &\mapsto \begin{cases} \mathbf{d}(x_1, x_2) + \mathbf{d}(x_2, x_3) - \lambda & \textit{if } \mathbf{d}(x_1, x_2) \leq t \wedge \mathbf{d}(x_2, x_3) \leq t, \\ +\infty & \textit{otherwise}, \end{cases} \end{aligned} \tag{6}$$

and let $\boldsymbol{c} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, be the cost tensor between all triplets in $(X^{(1)}, X^{(2)}, X^{(3)})$, defined by $c_{i_1 i_2 i_3} = c(\boldsymbol{x}_{i_1}^{(1)}, \boldsymbol{x}_{i_2}^{(2)}, \boldsymbol{x}_{i_3}^{(3)})$. Then the unbalanced multi-marginal OT problem can be stated as the following linear program:

$$\begin{aligned} &\underset{\pi \in \mathbb{R}^{n_1 \times n_2 \times n_3}}{\arg\min} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \sum_{i_3=1}^{n_3} c_{i_1 i_2 i_3} \pi_{i_1 i_2 i_3} \\ &s.t. \sum_{i_2=1}^{n_2} \sum_{i_3=1}^{n_3} \pi_{i_1 i_2 i_3} \leq 1 \quad \textit{for all } i_1 \in [n_1] \\ &\sum_{i_1=1}^{n_1} \sum_{i_3=1}^{n_3} \pi_{i_1 i_2 i_3} \leq 1 \quad \textit{for all } i_2 \in [n_2] \\ &\sum_{i_1=1}^{n_1} \sum_{i_3=1}^{n_3} \pi_{i_1 i_2 i_3} \leq 1 \quad \textit{for all } i_3 \in [n_1] \\ &\pi_{i_1 i_2 i_3} \geq 0 \quad \textit{for all } (i_1, i_2, i_3) \in [n_1] \times [n_2] \times [n_3], \end{aligned} \tag{7}$$

where we used the notation $[n] = \{1, \ldots, n\}$. As mentioned above, note that per the marginal constraints, particles may be matched at most once and can also be discarded. Likewise, by definition of the cost vector $\boldsymbol{c}$ only allows matchings between points that are valid $\mathbf{d}_t^3$-chains. Analogously there is a matching incentive via the parameter λ and for sufficiently high values (relative to t) one can show that the above problem provides an optimal $\mathbf{d}_t^3$-matching. Among all these matchings, one with minimal sum of pairwise distances is selected by the problem.

Generalization of Definition 4 to arbitrary k is now obvious, leading to a multi-marginal problem with k marginals. In general, multi-marginal problems quickly become numerically impractical due to the large number of variables. The cost function c in (6) has a chain structure, i.e. it can be written as a sum of functions only depending on (x_1, x_2) and (x_2, x_3). This chain structure allows the reformulation of the problem as a much more compact network flow problem (see Section below), and it implies the existence of optimal binary matchings. Problems where the cost exhibits a tree-structure can still be solved efficiently, see ref. 51 and references therein, but they cannot be formulated as network flow problems and do not exhibit binary minimizers in general.

Network flow formulation

In this section, we show that the multi-marginal optimal unbalanced transport problem corresponds to a min cost flow problem if the cost function has a chain structure as in (6). This has two relevant consequences:

1. It guarantees that (7) has integer solutions and thus indeed corresponds to a matching problem, which in general does not hold true for discrete OT problems;
2. It allows us to solve the multi-marginal optimal unbalanced transport problem efficiently.

Definition 5. Let (V, E) be a directed graph with a source node $S \in V$, a target node $T \in V$, an edge capacity function $l_E : E \to \mathbb{R} \cup \infty$ and an edge cost function $c_E : E \to \mathbb{R} \cup \infty$. Then we call (V, E, c_E, l_E) a flow network. Given an amount of flow, $m \in \mathbb{R}_+$ the min cost flow problem consists in finding a function $f : E \to \mathbb{R}$ that solves the following optimization problem:

$$\begin{aligned} &\min_f \sum_{(u,v) \in E} f(u, v) c_E(u, v) \\ &s.t.\ 0 \leq f(u, v) \leq l_E(u, v) \textit{ for all } (u, v) \in E \quad (\textit{capacity constraints}) \\ &\sum_{\{u:(u,v) \in E\}} f(u, v) - \sum_{\{w:(v,w) \in E\}} f(v, w) = 0 \quad \textit{for all } v \neq S, T \\ &\sum_{\{u:(S,u) \in E\}} f(S, u) - \sum_{\{v:(v,S) \in E\}} f(v, S) = m \quad (\textit{flow source}) \\ &\sum_{\{u:(u,T) \in E\}} f(u, T) - \sum_{\{v:(T,v) \in E\}} f(T, v) = m \quad (\textit{flow sink}). \end{aligned}$$

Notably, due to the total unimodularity of the constraint matrix, the min cost flow problem with integer total flow m and integer capacity function l_E has an integer solution (Theorem 13.11 in[65]). In the following, we recast (7) to a min cost flow problem (see sketch in Supplementary Fig. 1):

- Node set V**:** Define source node $S \in V$ and target node $T \in V$ and add two nodes $v_l^{(j)}$ and $\hat{v}_l^{(j)}$ for each detected particle position $\mathbf{x}_l^{(j)}$ in Equation (1).
- Edge set E**:**
  Connect nodes referring to the same detected point and set edge costs $c_E(v_l^{(j)}, \hat{v}_l^{(j)}) = -\frac{\lambda}{k}$ where k is the number of point clouds as in (1).
  Add all possible edges of form $(\hat{v}^{(j)}, v^{(j+1)}) \in E$ for $j = 1, \ldots, k - 1$. Set edge costs

  $$c_E(\hat{v}^{(j)}, v^{(j+1)}) = \begin{cases} \infty, & \text{if } \boldsymbol{d}(\boldsymbol{x}^{(j)}, \boldsymbol{x}^{(j+1)}) > t \\ \boldsymbol{d}(\boldsymbol{x}^{(j)}, \boldsymbol{x}^{(j+1)}), & \text{otherwise.} \end{cases}$$

  Include source and target nodes via edges of form $(S, v^{(1)}), (\hat{v}^{(k)}, T), (S, T) \in E$, and set its costs to 0.
  Define edge capacities

  $$l_E(v_i, v_j) = \begin{cases} \infty, & \text{if } v_i = S \text{ and } v_j = T \\ 1, & \text{otherwise.} \end{cases}$$

Proposition 1. Let $f : E \to \mathbb{R}$ be an integer solution of the min cost flow problem for the flow network (V, E, c_E, l_E) defined above with transported mass $m = \min(n_1, n_2, n_3)$. Then one of the optimal solutions $\boldsymbol{\pi}^*$ of the multi-marginal optimal unbalanced transport problem (7) is given by,

$$\pi^*_{i_1 i_2 i_3} = f(\hat{v}_{i_1}^{(1)}, v_{i_2}^{(2)}) f(\hat{v}_{i_2}^{(2)}, v_{i_3}^{(3)}), \tag{8}$$

for $i_1 \in [n_1]$, $i_2 \in [n_2]$ and $i_3 \in [n_3]$ with notation $[n] = \{1, \ldots, n\}$.

Proof. First we show that $\boldsymbol{\pi}^*$ as defined in Eq. (8) is in fact a valid transport plan for Eq. (7). For any $i_3 \in [n_3]$ we have that, using the conservation constraint,

$$\begin{aligned} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \pi^*_{i_1 i_2 i_3} &= \sum_{i_2=1}^{n_2} f(\hat{v}_{i_2}^{(2)}, v_{i_3}^{(3)}) \left(\sum_{i_1=1}^{n_2} f(\hat{v}_{i_1}^{(1)}, v_{i_2}^{(2)}) \right) \\ &= \sum_{i_2=1}^{n_2} f(\hat{v}_{i_2}^{(2)}, v_{i_3}^{(3)}) f(\hat{v}_{i_2}^{(2)}, v_{i_2}^{(2)}) \\ &\leq \sum_{i_2=1}^{n_2} f(\hat{v}_{i_2}^{(2)}, v_{i_3}^{(3)}) = f(\hat{v}_{i_3}^{(3)}, v_{i_3}^{(3)}) \leq 1 \end{aligned}$$

Analogously it is easy to verify that $\boldsymbol{\pi}^*$ satisfies

$$\sum_{i_1=1}^{n_1}\sum_{i_3=1}^{n_3}\pi^*_{i_1 i_2 i_3} \leq 1 \qquad \text{for all } i_2 \in [n_2]$$
$$\sum_{i_2=1}^{n_2}\sum_{i_3=1}^{n_3}\pi^*_{i_1 i_2 i_3} \leq 1 \qquad \text{for all } i_1 \in [n_1].$$

Hence, $\boldsymbol{\pi}^*$ is a feasible solution of (7). Further, since the source node S is directly connected to the target node T with an edge of infinite capacity and finite cost, the total flow cost must be finite. This implies that for any $i_1 \in [n_1]$, $i_2 \in [n_2]$ and $i_3 \in [n_3]$, we have that $f(\hat{v}^{(1)}_{i_1}, v^{(2)}_{i_2}) = 0$ if $\boldsymbol{d}(\boldsymbol{x}^{(1)}_{i_1}, \boldsymbol{x}^{(2)}_{i_2}) > t$ and $f(\hat{v}^{(2)}_{i_2}, v^{(3)}_{i_3}) = 0$ if $\boldsymbol{d}(\boldsymbol{x}^{(2)}_{i_2}, \boldsymbol{x}^{(3)}_{i_3}) > t$. Hence, using the shorthand notation

$$\langle \boldsymbol{c}, \pi \rangle = \sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\sum_{i_3=1}^{n_3} c_{i_1 i_2 i_3}\pi_{i_1 i_2 i_3},$$

we can rewrite the total cost of the transport problem as

$$\langle \boldsymbol{c}, \boldsymbol{\pi}^* \rangle = \sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\sum_{i_3=1}^{n_3}\left(\boldsymbol{d}(\boldsymbol{x}^{(1)}_{i_1}, \boldsymbol{x}^{(2)}_{i_2}) + \boldsymbol{d}(\boldsymbol{x}^{(2)}_{i_2}, \boldsymbol{x}^{(3)}_{i_3}) - \lambda\right)$$
$$\cdot f(\hat{v}^{(1)}_{i_1}, v^{(2)}_{i_2}) f(\hat{v}^{(2)}_{i_2}, v^{(3)}_{i_3}).$$

By the flow conservation constraints and the fact that f is an integer solution, we can simply reformulate the sum above in terms of the network flow cost function to obtain

$$\langle \boldsymbol{c}, \pi^* \rangle = \sum_{(u,v)\in E} c_E(u,v) f(u,v).$$

Let us now assume that there exists a feasible solution of (7), $\tilde{\boldsymbol{\pi}}$, such that

$$\langle \boldsymbol{c}, \tilde{\boldsymbol{\pi}} \rangle < \langle \boldsymbol{c}, \boldsymbol{\pi}^* \rangle.$$

Then we can define the flow $\tilde{f} : E \to \mathbb{R}$ by setting:

$$\tilde{f}(\hat{v}^{(1)}_{i_1}, v^{(2)}_{i_2}) = \sum_{i_3=1}^{n_3}\tilde{\pi}_{i_1 i_2 i_3} \text{ and} \tilde{f}(\hat{v}^{(2)}_{i_2}, v^{(3)}_{i_3}) = \sum_{i_1=1}^{n_1}\tilde{\pi}_{i_1 i_2 i_3},$$

for $i_1 \in [n_1]$, $i_2 \in [n_2]$ and $i_3 \in [n_3]$. The value of the flow on the remaining nodes of E can then be determined by the conservation constraints. In particular, we have $\tilde{f}(S, T) = \min\{n_1, n_2, n_3\} - \sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\sum_{i_3=1}^{n_3}\pi_{i_1 i_2 i_3}$. This flow is a feasible solution of the given min cost flow problem and hence, by the definition of the cost function for the edges we can derive a contradiction:

$$\sum_{(u,v)\in E} c_E(u,v)\tilde{f}(u,v) = \langle \boldsymbol{c}, \tilde{\pi} \rangle < \sum_{(u,v)\in E} c_E(u,v) f(u,v). \qquad \square$$

As a result of Proposition 1, we immediately obtain that the multi-marginal optimal unbalanced transport problem Eq. (7) has an integer solution and hence provides one-to-one point matchings.

Another significant consequence of Proposition 1 is that we can solve the unbalanced optimal transport problem given in Eq. (7) efficiently. While it is often unfeasible to compute directly the solution of the $n_1 \cdot n_2 \cdots \cdot n_k$-dimensional linear programming problem in Eq. (7), the min cost flow problem can be solved by the Scaling Minimum-Cost Flow Algorithm in ref. 66 in $O(|V|^2|E|\log(|V|))$ elementary operations, where $|V|$ is the number of nodes, $|E|$ is the number of edges. In our case the number of nodes is of the order $O(n_1 \cdot \dots \cdot n_k)$ and the number of edges can be upper bounded by an expression of the order $O(n_1 \cdot n_2^2 \cdot \dots \cdot n_{k-1}^2 \cdot n_k)$. In practice, it is further possible to omit all edges with infinite cost, since the source S and the sink T are connected through an edge of cost 0 and with infinite capacity. This implies that for small t much fewer edges to the network are added which results in better computational performance.

For an image containing around 1,000 points in each color channel, a solution of the min cost flow problem can be computed for about 10 different values of t in ~1 s on a standard laptop.

Estimating the true chain-like particle abundances

The quality of fluorescence microscopy suffers from non-optimal labeling efficiencies and point detection errors. This will be addressed by a statistical framework to infer on how many of the detected structures in the image actually concur with the ground truth biological structure and how many detections represent only incomplete parts of the underlying particle assembly. For color channels $i \in \{1, \dots, k\}$ let

$$\left\{\xi^{(i)}_j\right\}_{j=1}^{n_i} \subset \mathbb{R}^2 \tag{9}$$

be the pairs of coordinates of all particles that lie within the scope of the microscope. Note that these point clouds do not necessarily equal those defined in Eq. (1) describing the coordinates of detected particles, since we might not be able to measure all of the existing particles to do unsuccessful labeling or point detection errors.

Definition 6. (Labeling Efficiency). For each color channel $i \in \{1, \dots, k\}$ we assume that there is a specific probability $s_i \in (0, 1]$ quantifying whether a particle of this channel is successfully imaged and detected. For simplicity in the following we will always call probabilities s_i*labeling efficiencies*.

We further assume that the random event of successful detection is statistically independent for each point. Accordingly, the detection success can be described by independent Bernoulli variables

$$\left\{Z^{(i)}_j\right\}_{j=1}^{n_i} \sim \text{ Ber}\,(s_i), \tag{10}$$

where $s_i \in (0, 1]$ and $\xi^{(i)}_j$ is detectable, if and only if $Z^{(i)}_j = 1$.

If there exists a true $\mathbf{d}^k_t$-chain of form $(\xi^{(1)}, \dots, \xi^{(k)})$, then this can only be correctly identified as such, if each of the included particles was detected, i.e., if and only if $\prod_{i=1}^k Z^{(i)} = 1$. From independence it follows that

$$\prod_{i=1}^{k} Z^{(i)} \sim \text{ Ber}\left(\prod_{i=1}^k s_i\right). \tag{11}$$

Detecting an ABC triplet correctly is $\text{Ber}(s_A s_B s_C)$ distributed. Therefore, all possible substructures that can be detected conditioned on the true underlying ABC triplet, i.e.,

1. ABC triplet, if we see all particles
2. AB pair, if we do not see C
3. BC pair, if we do not see A
4. *AC substructure, if we do not see B – which is detected as A and C singlets*
5. A singlet, if we do not see B and C
6. B singlet, if we do not see A and C
7. C singlet, if we do not see A and B
8. ∅, *if we do not see A,B and C which can not be detected at all,*

can accordingly be modeled as Multinomial random variable

$$\boldsymbol{W}_{\cdot|ABC} = \begin{bmatrix} W_{ABC|ABC} \\ W_{AB|ABC} \\ W_{BC|ABC} \\ W_{AC|ABC} \\ W_{A|ABC} \\ W_{B|ABC} \\ W_{C|ABC} \\ W_{\emptyset|ABC} \end{bmatrix}. \tag{12}$$

This can be done in the same manner for all other structures of interest, i.e., true AB and BC pairs and A, B, and C singlets (and their respective substructures) yielding random variables $W_{\cdot|AB}$, $W_{\cdot|BC}$, $W_{\cdot|A}$, $W_{\cdot|B}$, $W_{\cdot|C}$. The actual detectable numbers of those structures are

$$\begin{aligned} W_{ABC} &= \sum W_{ABC|\cdot}, \\ W_{AB} &= \sum W_{AB|\cdot}, \\ W_{BC} &= \sum W_{BC|\cdot}, \\ W_{A} &= \sum W_{A|\cdot} + \sum W_{AC|\cdot}, \\ W_{B} &= \sum W_{B|\cdot}, \\ W_{C} &= \sum W_{C|\cdot} + \sum W_{AC|\cdot}, \end{aligned} \tag{13}$$

which define a random variable $\boldsymbol{W} = (W_{ABC}, W_{AB}, W_{BC}, W_A, W_B, W_C)^T$. This leads to a statistical framework, that allows us to estimate the true underlying structures abundances from the detected number of structures.

Theorem 2. Let known, positive labeling efficiencies $s_A > 0$, $s_B > 0$ and $s_C > 0$ and unknown structure abundances $\boldsymbol{n} = (n_{ABC}, n_{AB}, n_{BC}, n_A, n_B, n_C)^T$ and define $N = \sum_{i \in \{ABC, \ldots, C\}} n_i$. Assume the multinomial model as described in Equation (12) and Equation (13).

Part 1: An unbiased estimator $\hat{\boldsymbol{n}}$ of true abundances $\boldsymbol{n}$ is given as

$$\hat{\boldsymbol{n}} = \begin{bmatrix} \frac{1}{s_A s_B s_C} & 0 & 0 & 0 & 0 & 0 \\ \frac{s_C - 1}{s_A s_B s_C} & \frac{1}{s_A s_B} & 0 & 0 & 0 & 0 \\ \frac{s_A - 1}{s_A s_B s_C} & 0 & \frac{1}{s_B s_C} & 0 & 0 & 0 \\ \frac{s_B - 1}{s_A s_B} & \frac{s_B - 1}{s_A s_B} & 0 & \frac{1}{s_A} & 0 & 0 \\ \frac{(1-s_A)(1-s_C)}{s_A s_B s_C} & \frac{s_A - 1}{s_A s_B} & \frac{s_C - 1}{s_B s_C} & 0 & \frac{1}{s_B} & 0 \\ \frac{s_B - 1}{s_B s_C} & 0 & \frac{s_B - 1}{s_B s_C} & 0 & 0 & \frac{1}{s_C} \end{bmatrix} \boldsymbol{W}. \tag{14}$$

Part 2: For $\boldsymbol{n} \longrightarrow \infty$ entrywise, $n_j/N \longrightarrow f_j$ with $\infty > f_j > 0$ constant for each $j \in \{ABC, \ldots, C\}$, and $\Theta\Sigma(\hat{\boldsymbol{n}})\Theta^T$ invertible,

$$P\left(\Xi \le \chi^2_{6,\alpha}\right) \le 1 - \alpha, \tag{15}$$

where

$$\Xi = (\hat{\boldsymbol{n}} - \boldsymbol{n})^T (\Theta\mu)^T \left(\Theta\Sigma(\hat{\boldsymbol{n}})\Theta^T\right)^{-1} (\Theta\mu)(\hat{\boldsymbol{n}} - \boldsymbol{n}) \tag{16}$$

and $\chi^2_{6,\alpha}$ is the α-quantile of a chi-squared distribution with 6 degrees of freedom and with Θ, μ and $\Sigma(\hat{\boldsymbol{n}})$ defined as in the following proof. If $\left(\Theta\Sigma(\hat{\boldsymbol{n}})\Theta^T\right)^{-1}$ does not exist, we get Equation (15) with $\chi^2_{r,\alpha}$ plugging its pseudoinverse $\left(\Theta\Sigma(\hat{\boldsymbol{n}})\Theta^T\right)^{+}$ in Equation (16), where $r = \operatorname{rank}\left(\Theta\Sigma(\hat{\boldsymbol{n}})\Theta^T\right)$.

Proof. *Part 1*: conditioned on a true ABC triplet, the number of (mis) specifications resulting from incomplete labeling efficiencies is multinomially distributed:

$$\boldsymbol{W}_{\cdot|ABC} = \begin{bmatrix} W_{ABC|ABC} \\ W_{AB|ABC} \\ W_{BC|ABC} \\ W_{AC|ABC} \\ W_{A|ABC} \\ W_{B|ABC} \\ W_{C|ABC} \\ W_{\emptyset|ABC} \end{bmatrix} \sim \text{Mnom}\left(n_{ABC}, \boldsymbol{p}_{ABC}\right) \tag{17}$$

with probability vector

$$\boldsymbol{p}_{ABC} = \begin{bmatrix} s_A s_B s_C \\ s_A s_B (1 - s_C) \\ (1 - s_A) s_B s_C \\ s_A (1 - s_B) s_C \\ s_A (1 - s_B)(1 - s_C) \\ (1 - s_A) s_B (1 - s_C) \\ (1 - s_A)(1 - s_B) s_C \\ (1 - s_A)(1 - s_B)(1 - s_C) \end{bmatrix}, \tag{18}$$

where $\sum_{j=1}^{8} \boldsymbol{p}_{ABC}[j] = 1$. Accordingly, the abundances of (mis)detections of a true AB pair are

$$\begin{bmatrix} W_{ABC|AB} \\ W_{AB|AB} \\ W_{BC|AB} \\ W_{AC|AB} \\ W_{A|AB} \\ W_{B|AB} \\ W_{C|AB} \\ W_{\emptyset|AB} \end{bmatrix} \sim \text{Mnom}\left(n_{AB}, \boldsymbol{p}_{AB}\right) \tag{19}$$

with

$$\boldsymbol{p}_{AB} = \begin{bmatrix} 0 \\ s_A s_B \\ 0 \\ 0 \\ s_A (1 - s_B) \\ (1 - s_A) s_B \\ 0 \\ (1 - s_A)(1 - s_B) \end{bmatrix}. \tag{20}$$

This can be done accordingly for all other structures of interest, i.e. BC pairs and A, B, and C singlets yielding

$$\begin{aligned} \boldsymbol{W}_{\cdot|ABC} &\sim \text{Mnom}\left(n_{ABC}, \boldsymbol{p}_{ABC}\right) \\ \boldsymbol{W}_{\cdot|AB} &\sim \text{Mnom}\left(n_{AB}, \boldsymbol{p}_{AB}\right) \\ \boldsymbol{W}_{\cdot|BC} &\sim \text{Mnom}\left(n_{BC}, \boldsymbol{p}_{BC}\right) \\ \boldsymbol{W}_{\cdot|A} &\sim \text{Mnom}\left(n_{A}, \boldsymbol{p}_{A}\right) \\ \boldsymbol{W}_{\cdot|B} &\sim \text{Mnom}\left(n_{B}, \boldsymbol{p}_{B}\right) \\ \boldsymbol{W}_{\cdot|C} &\sim \text{Mnom}\left(n_{C}, \boldsymbol{p}_{C}\right) \end{aligned} \tag{21}$$

with

$$\boldsymbol{p}_{BC} = \begin{bmatrix} 0 \\ 0 \\ s_B s_C \\ 0 \\ 0 \\ s_B (1 - s_C) \\ (1 - s_B) s_C \\ (1 - s_B)(1 - s_C) \end{bmatrix}, \boldsymbol{p}_A = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ s_A \\ 0 \\ 0 \\ (1 - s_A) \end{bmatrix}, \boldsymbol{p}_B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ s_B \\ 0 \\ (1 - s_B) \end{bmatrix}, \boldsymbol{p}_C = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ s_C \\ (1 - s_C) \end{bmatrix}. \tag{22}$$

Note, that Ø can not be detected at all and substructure AC is counted as a separate A and C singlet Hence, the total numbers of detected triplets, pairs and singlets are defined as the following sums

$$\begin{aligned} W_{ABC} &= W_{ABC|ABC} \\ W_{AB} &= W_{AB|ABC} + W_{AB|AB} \\ W_{BC} &= W_{BC|ABC} + W_{BC|BC} \\ W_A &= W_{A|ABC} + W_{AC|ABC} + W_{A|AB} + W_{A|A} \\ W_B &= W_{B|ABC} + W_{B|AB} + W_{B|BC} + W_{B|B} \\ W_C &= W_{C|ABC} + W_{AC|ABC} + W_{C|BC} + W_{C|C}. \end{aligned} \tag{23}$$

This can be rewritten as

$$\boldsymbol{W} = \begin{bmatrix} W_{ABC} \\ W_{AB} \\ W_{BC} \\ W_A \\ W_B \\ W_C \end{bmatrix} = \Theta\left(\boldsymbol{W}_{\cdot|ABC} + \boldsymbol{W}_{\cdot|AB} + \boldsymbol{W}_{\cdot|BC} + \boldsymbol{W}_{\cdot|A} + \boldsymbol{W}_{\cdot|B} + \boldsymbol{W}_{\cdot|C}\right), \tag{24}$$

using the transformation matrix

$$\Theta = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix} \in \mathbb{R}^{6\times 8}. \tag{25}$$

With this definition of Θ we delete the last entry in each binomial distributed vector and add an AC substructure appearance to singlet detections A and B. By Eq. (24) we get that

$$\mathbb{E}[\boldsymbol{W}] = \Theta\mu\boldsymbol{n} \tag{26}$$

with

$$\mu = \begin{bmatrix} \boldsymbol{p}_{ABC} & \boldsymbol{p}_{AB} & \boldsymbol{p}_{BC} & \boldsymbol{p}_A & \boldsymbol{p}_B & \boldsymbol{p}_C \end{bmatrix} \in \mathbb{R}^{8\times 6}. \tag{27}$$

Hence, with positive labeling efficiencies $s_A > 0, s_B > 0$ and $s_C > 0$, multiplying

$$(\Theta\mu)^{-1} = \begin{bmatrix} \frac{1}{s_A s_B s_C} & 0 & 0 & 0 & 0 & 0 \\ \frac{s_C - 1}{s_A s_B s_C} & \frac{1}{s_A s_B} & 0 & 0 & 0 & 0 \\ \frac{s_A - 1}{s_A s_B s_C} & 0 & \frac{1}{s_B s_C} & 0 & 0 & 0 \\ \frac{s_B - 1}{s_A s_B} & \frac{s_B - 1}{s_A s_B} & 0 & \frac{1}{s_A} & 0 & 0 \\ \frac{(1-s_A)(1-s_C)}{s_A s_B s_C} & \frac{s_A - 1}{s_A s_B} & \frac{s_C - 1}{s_B s_C} & 0 & \frac{1}{s_B} & 0 \\ \frac{s_B - 1}{s_B s_C} & 0 & \frac{s_B - 1}{s_B s_C} & 0 & 0 & \frac{1}{s_C} \end{bmatrix} \tag{28}$$

with $\boldsymbol{W}$ introduces an unbiased estimator $\hat{\boldsymbol{n}}$.

Part 2: we utilize that by the central limit theorem for a multinomially distributed random variable $\boldsymbol{M} \sim \mathrm{Mnom}(m, \boldsymbol{p})$ with probability vector $\boldsymbol{p} = (p_1, p_2, \cdots, p_k)^T$

$$\frac{1}{\sqrt{m}}(\boldsymbol{M} - m\boldsymbol{p}) \to^{\mathcal{D}} \mathcal{N}_k\left(\boldsymbol{0}_k, \operatorname{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^T\right) \quad \text{for} \quad m \to \infty, \tag{29}$$

where

$$\operatorname{diag}(\boldsymbol{p}) = \begin{bmatrix} p_1 & 0 & \cdots \\ 0 & p_2 & \cdots \\ \vdots & \vdots & p_k \end{bmatrix} \tag{30}$$

and $\boldsymbol{0}_k = (0, \ldots, 0)^T \in \mathbb{R}^k$ (see, e.g., ref. 67). Hence, for $\boldsymbol{n}$ entrywise large enough, we can approximate properly scaled independent, multinomial random vectors

$$\boldsymbol{W}_{\cdot|ABC}, \ \boldsymbol{W}_{\cdot|AB}, \ \boldsymbol{W}_{\cdot|BC}, \ \boldsymbol{W}_{\cdot|A}, \ \boldsymbol{W}_{\cdot|B}, \ \boldsymbol{W}_{\cdot|C} \tag{31}$$

with multi-dimensional normal distributions, respectively. In the following assume $\boldsymbol{n} \longrightarrow \infty$ entrywise and $n_j/N \longrightarrow f_j$ with $\infty > f_j > 0$ constant for each $j \in \{ABC, \ldots, C\}$, where $N = \sum_{i\in\{ABC,\ \ldots,\ C\}} n_i$. Then, it holds that

$$\begin{aligned} \sum_{i\in\{ABC,\ldots,C\}} \sqrt{\tfrac{n_i}{N}} \tfrac{1}{\sqrt{n_i}}\left(\boldsymbol{W}_{\cdot|i} - n_i\boldsymbol{p}_i\right) &= \sqrt{\tfrac{1}{N}} \sum_{i\in\{ABC,\ldots,C\}} \left(\boldsymbol{W}_{\cdot|i} - n_i\boldsymbol{p}_i\right) \\ &\to^{\mathcal{D}} \mathcal{N}_8\left(\boldsymbol{0}_8, \sum_{i\in\{ABC,\ldots,C\}} f_i\left(\operatorname{diag}(\boldsymbol{p}_i) - \boldsymbol{p}_i\boldsymbol{p}_i^T\right)\right). \end{aligned} \tag{32}$$

For now, suppose $\sum n_i\left(\operatorname{diag}(\boldsymbol{p}_i) - \boldsymbol{p}_i\boldsymbol{p}_i^T\right)$ is invertible. Then in the limit

$$\begin{aligned} &\left(\sum f_i\left(\operatorname{diag}(\boldsymbol{p}_i) - \boldsymbol{p}_i\boldsymbol{p}_i^T\right)\right)^{-1/2} \sqrt{\frac{1}{N}} \sum\left(\boldsymbol{W}_{\cdot|i} - n_i\boldsymbol{p}_i\right) \\ =&\left(\sum N f_i\left(\operatorname{diag}(\boldsymbol{p}_i) - \boldsymbol{p}_i\boldsymbol{p}_i^T\right)\right)^{-1/2} \sum\left(\boldsymbol{W}_{\cdot|i} - n_i\boldsymbol{p}_i\right) \\ =&\left(\sum n_i\left(\operatorname{diag}(\boldsymbol{p}_i) - \boldsymbol{p}_i\boldsymbol{p}_i^T\right)\right)^{-1/2} \sum\left(\boldsymbol{W}_{\cdot|i} - n_i\boldsymbol{p}_i\right) \end{aligned} \tag{33}$$

and hence

$$\left(\sum n_i\left(\operatorname{diag}(\boldsymbol{p}_i) - \boldsymbol{p}_i\boldsymbol{p}_i^T\right)\right)^{-1/2} \sum\left(\boldsymbol{W}_{\cdot|i} - n_i\boldsymbol{p}_i\right) \to^{\mathcal{D}} \mathcal{N}_8\left(\boldsymbol{0}_8, I_{8\times 8}\right), \tag{34}$$

where $I_{8\times8}$ is the 8-dimensional identity matrix. In the following we denote

$$\Sigma(\boldsymbol{n}) = \left(\sum n_i\left(\operatorname{diag}(\boldsymbol{p}_i) - \boldsymbol{p}_i\boldsymbol{p}_i^T\right)\right). \tag{35}$$

Multiplying $(\Theta\mu)^{-1}\Theta$ with Eq. (32) consequently yields

$$\begin{aligned} &\left((\Theta\mu)^{-1}\Theta\Sigma(\boldsymbol{n})\Theta^T\left((\Theta\mu)^{-1}\right)^T\right)^{-1/2}\left((\Theta\mu)^{-1}\Theta\sum \boldsymbol{W}_{\cdot|i} - (\Theta\mu)^{-1}\Theta\sum n_i\boldsymbol{p}_i\right) \\ &= \left((\Theta\mu)^{-1}\Theta\Sigma(\boldsymbol{n})\Theta^T\left((\Theta\mu)^{-1}\right)^T\right)^{-1/2}(\hat{\boldsymbol{n}} - \boldsymbol{n}) \to^{\mathcal{D}} \mathcal{N}_6\left(\boldsymbol{0}_6, I_{6\times 6}\right) \end{aligned} \tag{36}$$

with $\hat{\boldsymbol{n}} = (\Theta\mu)^{-1}\Theta\sum \boldsymbol{W}_{\cdot|i}$ and $\boldsymbol{n} = (\Theta\mu)^{-1}\Theta\mu\boldsymbol{n} = (\Theta\mu)^{-1}\Theta\sum n_i\boldsymbol{p}_i$. By law of large numbers, it holds that

$$\frac{1}{N}(\hat{\boldsymbol{n}} - \boldsymbol{n}) = \frac{\hat{\boldsymbol{n}}}{N} - \frac{\boldsymbol{n}}{N} \to^{\mathcal{P}} \boldsymbol{0}_6. \tag{37}$$

and hence for all $j \in \{ABC, \ldots, C\}$

$$\frac{\hat{n}_j}{N} \to^{\mathcal{P}} f_j. \tag{38}$$

By Slutsky's Lemma we can use Eq. (38) to replace $\boldsymbol{n}$ in $\Sigma(\boldsymbol{n})$ with $\hat{\boldsymbol{n}}$. For $\boldsymbol{n} \longrightarrow \infty$ entrywise this yields

$$\Xi = (\hat{\boldsymbol{n}} - \boldsymbol{n})^T(\Theta\mu)^T\left(\Theta\Sigma(\hat{\boldsymbol{n}})\Theta^T\right)^{-1}(\Theta\mu)(\hat{\boldsymbol{n}} - \boldsymbol{n}) \to^{\mathcal{D}} \chi^2_6. \tag{39}$$

In case $\Theta\Sigma(\hat{\boldsymbol{n}})\Theta^T$ is not invertible, one can use its pseudoinverse yielding convergence to a chi-square distribution with r degrees of freedom, i.e., χ_r^2 in Equation (39), where $r = \operatorname{rank}\left(\Theta\Sigma(\hat{\boldsymbol{n}})\Theta^T\right)$ □.

With Part 2 of Theorem 2 we can construct a confidence ellipsoid around $\hat{\boldsymbol{n}}$ in a straight-forward manner. To show that Ξ in our setting is approximately chi-square distributed for finite sample sizes and to compare simulated and theoretical coverages of $\hat{\boldsymbol{n}}$, we performed a simulation study as described in the following section.

Simulation study setup

In the first simulation study a predefined number of triplets, pairs, and singlets are generated as follows:

Step 1: Draw the coordinate for channel B as $b \sim \mathcal{U}([0, 400 \cdot r]^2)$, where $\mathcal{U}$ is the continuous uniform distribution.
Step 2a: Draw angle $\alpha \sim \mathcal{U}[0, 2\pi]$ and normally distributed distance $d_A \sim \mathcal{N}(t, 0.5)$. Set $a = b\left(\cos(\alpha)d_A + \sin(\alpha)d_A\right)$.
Step 2b: Draw $\epsilon \sim \mathcal{N}(0, 0.2)$ and set angle $\beta = \alpha + \pi + \epsilon$. Draw $d_C \sim \mathcal{N}(t, 0.5)$ and set $c = b\left(\cos(\beta)d_C + \sin(\beta)d_C\right)$.
Step 3: Round a, b and c to match the pixel grid $[0, 400]^2 \subseteq \mathbb{N}^2_{\geq 0}$.

This design favors to simulate triplets of an approximately linear structure. Pairs are simulated by skipping either Step 2a or 2b. Singlets are drawn as in Step 1.

In the second simulation study quadruples, triplets, pairs, and singlets n are generated similarly, but replacing and adding

Step 2b: Draw angle $\beta \sim \mathcal{U}[0, 2\pi]$ and $d_C \sim \mathcal{N}(t, 0.5)$ and set $d = b\left(\cos(\beta)d_C + \sin(\beta)d_C\right)$.
Step 2c: Draw angle $\gamma \sim \mathcal{U}[0, 2\pi]$ and $d_D \sim \mathcal{N}(t, 0.5)$ and set $d = c\left(\cos(\gamma)d_D + \sin(\gamma)d_D\right)$.

This simulation setup allows arbitrarily curved chain-structures. The distance threshold is always fixed to $t = 70$ nm.

To obtain intensity images close to an experimental STED setup from the simulated point sets we followed the simulation setup introduced in Tameling et al.[17], to mimic experimental STED images of 400 × 400 pixels with full-width at half-maximum (FWHM) value of 40 nm (approximately the resolution of the STED microscope) and pixel size 25 nm = 1 pixel). In the second simulation study (including quadruples) the Poisson noise level was on average increased by a factor of 10.

Methods included in the simulation study

For the Ripley's K based Statistical Object Distance Analysis (SODA)[21] we used the triplet colocalization protocol SODA 3 Colors in ICY[68] (version 2.4.0.0). For the analysis we used default input parameters and set scale threshold per channel to be 100. The plugin BlobProb[15] was called in ImageJ/Fiji[69] (version 2.3.0/1.53q) and the number of colocalized blobs were considered. We set voxel size to 25 nm in every dimension and the threshold per channel to 100. The ConditionalColoc[18] from GitHub (https://github.com/kjaqaman/ConditionalColoc) was executed on MATLAB (version R2023a). Particles were detected using the "point-source detection" algorithm provided via the integrated u-track package (https://github.com/DanuserLab/u-track).

For all implementations but ConditionalColoc the detected chain-structure abundances were output as integers. Therefore, we scaled abundances, i.e., divided them by the total number of particles detected in channel B. ConditionColoc already aims to output probabilities that are scaled by detected particles per channel, hence no further transformation of the output was performed by us. Since for all simulated Scenarios the same number of particles was generated in every channel, we ensured that both scaling procedures are comparable. The maximal colocalization threshold is set to $t = 5$ pixels = 125 nm throughout all considered methods.

Nanoruler samples

Custom-made DNA nanoruler samples featuring one, two, or three fluorophore spots, each consisting of 20 fluorophores (Alexa Fluor488, Alexa Fluor594, Star Red), with a distance between the spots of 70 nm, were purchased from Gattaquant - DNA Nanotechnologies (Gräfelfing, Germany). The biotinylated nanorulers were immobilized on a BSA-biotin-neutravidin surface according to the manufacturer's specifications.

Stimulated emission depletion super-resolution light microscopy

Image acquisition was done using a quad scanning STED microscope (Abberior Instruments, Göttingen, Germany) equipped with a UPlanSApo 100x/1,40 Oil objective (Olympus, Tokyo, Japan). Excitation of Alexa Fluor 488, Alexa Fluor 594 and Star Red was achieved by laser beams featuring wave lengths of 485 nm, 561 nm, and 640 nm, respectively. For STED imaging, a laser beam with an emission wavelength of 775 nm was applied. For all experimental STED images, a pixel size of 25 nm was utilized. For visualization purposes, contrast stretching and increasement of image brightness was applied to exemplary STED images within the figures of this manuscript. No image processing was applied prior to the application of the MultiMatch analysis workflow.

Statistics and reproducibility

The statistical framework developed and applied in this manuscript and the settings of simulation studies performed are presented in the Method sections. All sample sizes and significance levels of the confidence bands are listed in the respective figure legends. Experimental and simulated data and analysis scripts to reproduce results and figures are provided on Zenodo (https://doi.org/10.5281/zenodo.7221879)[70].

Data availability

Datasets generated and analyzed in this manuscript can be accessed via *Zenodo* (https://doi.org/10.5281/zenodo.7221879)[70].

Code availability

The Python package MultiMatch is available on GitHub repository https://github.com/gnies/multi_match. All scripts used to create the main and Supplementary Figs. are implemented in R (version 4.1.0) and Python (version 3.8.5) and are available via Zenodo (https://doi.org/10.5281/zenodo.7221879)[70]. In order to locate the positions of the particles in STED images, we perform point detection via the Python package scikit-image[71] (version 0.19.1). This is provided as an optional analysis step in our MultiMatch implementation for the evaluation of intensity matrices. Multicolor microscopy images, point detection results and MultiMatch output can be loaded into the interactive napari viewer. MultiMatch is compatible with Python package napari[39] (version 0.4.18) and an exemplary use-case is described on our repository https://github.com/gnies/multi_match. We utilize the minimum-cost flow solver provided in the package ortools[72] (version 9.4.1874).

Received: 28 February 2024; Accepted: 22 August 2024;
Published online: 13 September 2024

References

1. Cainero, I. et al. Measuring nanoscale distances by structured illumination microscopy and image cross-correlation spectroscopy (SIM-ICCS). *Sensors* **21**, 2010 (2021).
2. Costa, R. et al. Morphological study of TNPO3 and SRSF1 interaction during myogenesis by combining confocal, structured illumination and electron microscopy analysis. *Mol. Cell. Biochem.* **476**, 1797–1811 (2021).
3. Shimizu, Y. et al. Cargo sorting zones in the trans-Golgi network visualized by super-resolution confocal live imaging microscopy in plants. *Nat. Commun.* **12**, 1901 (2021).
4. Sahl, S. J., Hell, S. W. & Jakobs, S. Fluorescence nanoscopy in cell biology. *Nat. Rev. Mol. Cell Biol.* **18**, 685–701 (2017).
5. Betzig, E. et al. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).

6. 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).
7. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–796 (2006).
8. Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19**, 780–782 (1994).
9. Hell, S. W. Far-field optical nanoscopy. *Science* **316**, 1153–1158 (2007).
10. Klar, T. A., Jakobs, S., Dyba, M., Egner, A. & Hell, S. W. Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proc. Natl Acad. Sci. USA* **97**, 8206–8210 (2000).
11. Malkusch, S. et al. Coordinate-based colocalization analysis of single-molecule localization microscopy data. *Histochem. Cell Biol.* **137**, 1–10 (2012).
12. Manders, E. M. M., Verbeek, F. J. & Aten, J. A. Measurement of co-localization of objects in dual-colour confocal images. *J. Microsc.* **169**, 375–382 (1993).
13. Xu, L. et al. Resolution, target density and labeling effects in colocalization studies - suppression of false positives by nanoscopy and modified algorithms. *FEBS J.* **283**, 882–898 (2016).
14. Adler, J. & Parmryd, I. Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytom. Part A* **77A**, 733–742 (2010).
15. Fletcher, P. A., Scriven, D. R. L., Schulson, M. N. & Moore, E. D. W. Multi-image colocalization and its statistical significance. *Biophys. J.* **99**, 1996–2005 (2010).
16. Wang, S. et al. Spatially adaptive colocalization analysis in dual-color fluorescence microscopy. *IEEE Trans. Image Process.* **28**, 4471–4485 (2019).
17. Tameling, C. et al. Colocalization for super-resolution microscopy via optimal transport. *Nat. Comput. Sci.* **1**, 199–211 (2021).
18. Vega-Lugo, J., da Rocha-Azevedo, B., Dasgupta, A. & Jaqaman, K. Analysis of conditional colocalization relationships and hierarchies in three-color microscopy images. *J. Cell Biol.* **221**, e202106129 (2022).
19. Ripley, B. D. The second-order analysis of stationary point processes. *J. Appl. Probab.* **13**, 255–266 (1976).
20. Mukherjee, S., Gonzalez-Gomez, C., Danglot, L., Lagache, T. & Olivo-Marin, J.-C. Generalizing the statistical analysis of objects' spatial coupling in bioimaging. *IEEE Signal Process. Lett.* **27**, 1085–1089 (2020).
21. Lagache, T. et al. Mapping molecular assemblies with fluorescence microscopy and object-based spatial statistics. *Nat. Commun.* **9**, 698 (2018).
22. Winter, F. R. et al. Multicolour nanoscopy of fixed and living cells with a single STED beam and hyperspectral detection. *Sci. Rep.* **7**, 46492 (2017).
23. Spahn, C., Grimm, J. B., Lavis, L. D., Lampe, M. & Heilemann, M. Whole-cell, 3D, and multicolor STED imaging with exchangeable fluorophores. *Nano Lett.* **19**, 500–505 (2019).
24. Butkevich, A. N. et al. Photoactivatable fluorescent dyes with hydrophilic caging groups and their use in multicolor nanoscopy. *J. Am. Chem. Soc.* **143**, 18388–18393 (2021).
25. Glogger, M. et al. Synergizing exchangeable fluorophore labels for multitarget STED microscopy. *ACS Nano* **16**, 17991–17997 (2022).
26. Gonzalez Pisfil, M. et al. Stimulated emission depletion microscopy with a single depletion laser using five fluorochromes and fluorescence lifetime phasor separation. *Sci. Rep.* **12**, 14027 (2022).
27. Wang, J., Fan, Y., Sanger, J. M. & Sanger, J. W. STED analysis reveals the organization of nonmuscle muscle II, muscle myosin II, and F-actin in nascent myofibrils. *Cytoskeleton* **79**, 122–132 (2022).
28. Saal, K. A. et al. Heat denaturation enables multicolor X10-STED microscopy. *Sci. Rep.* **13**, 5366 (2023).
29. Andronov, L., Genthial, R., Hentsch, D. & Klaholz, B. P. SplitSMLM, a spectral demixing method for high-precision multi-color localization microscopy applied to nuclear pore complexes. *Commun. Biol.* **5**, 1–13 (2022).
30. Unterauer, E. M. et al. Spatial proteomics in neurons at single-protein resolution. *Cell* **187**, 1785–1800.e16 (2024).
31. Beater, S., Holzmeister, P., Lalkens, B. & Tinnefeld, P. Simple and aberration-free 4color-STED - multiplexing by transient binding. *Opt. Express* **23**, 8630–8638 (2015).
32. Willig, K. I., Rizzoli, S. O., Westphal, V., Jahn, R. & Hell, S. W. STED microscopy reveals that synaptotagmin remains clustered after synaptic vesicle exocytosis. *Nature* **440**, 935–939 (2006).
33. Reinhardt, S. C. M. et al. Ångström-resolution fluorescence microscopy. *Nature* **617**, 711–716 (2023).
34. Smallcombe, A. Multicolor imaging: The important question of co-localization. *BioTechniques* **30**, 1240–1246 (2001).
35. Sastre, D., Estadella, I., Bosch, M. & Felipe, A. *Methods in Molecular Biology* (Springer, 2019).
36. Goucher, D. R., Wincovitch, S. M., Garfield, S. H., Carbone, K. M. & Malik, T. H. A quantitative determination of multi-protein interactions by the analysis of confocal images using a pixel-by-pixel assessment algorithm. *Bioinformatics* **21**, 3248–3254 (2005).
37. Humpert, F., Yahiatène, I., Lummer, M., Sauer, M. & Huser, T. Quantifying molecular colocalization in live cell fluorescence microscopy. *J. Biophoton.* **8**, 124–132 (2015).
38. Haas, K. T. & Peaucelle, A. Protocol for multicolor three-dimensional dSTORM data analysis using MATLAB-based script package Grafeo. *STAR Protoc.* **2**, 100808 (2021).
39. napari contributers. *Napari: A Multi-dimensional Image Viewer For Python* https://doi.org/10.5281/zenodo.8115575 (2019).
40. Boettiger, A. N. et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).
41. Miron, E. et al. Chromatin arranges in chains of mesoscale domains with nanoscale functional topography independent of cohesin. *Sci. Adv.* **6**, eaba8811 (2020).
42. Villani, C. *Optimal Transport* (Springer Berlin, 2009).
43. Panaretos, V. M. & Zemel, Y. Statistical aspects of Wasserstein distances. *Annu. Rev. Stat. Appl.* **6**, 405–431 (2019).
44. Peyré, G. & Cuturi, M. Computational Optimal Transport: With Applications to Data Science. *Foundations and Trends® in Machine Learning* **11**, 355–607 (2019).
45. Zaritsky, A. et al. Decoupling global biases and local interactions between cell biological variables. *eLife* **6**, e22323 (2017).
46. Kim, Y.-H. & Pass, B. A general condition for monge solutions in the multi-marginal optimal transport problem. *SIAM J. Math. Anal.* **46**, 1538–1550 (2014).
47. Pass, B. Multi-marginal optimal transport: theory and applications. *ESAIM: Math. Model. Numer. Anal.* **49**, 1771–1790 (2015).
48. Chizat, L., Peyré, G., Schmitzer, B. & Vialard, F.-X. Unbalanced optimal transport: Dynamic and Kantorovich formulations. *J. Funct. Anal.* **274**, 3090–3123 (2018).
49. Friesecke, G., Matthes, D. & Schmitzer, B. Barycenters for the Hellinger–Kantorovich Distance Over $\mathbb{R}^d$. *SIAM Journal on Mathematical Analysis* **53**, 62–110 (2021).
50. Heinemann, F., Klatt, M. & Munk, A. Kantorovich-Rubinstein distance and barycenter for finitely supported measures: foundations and algorithms. *Appl. Math. Optim.* **87**, 4 (2022).
51. Beier, F., von Lindheim, J., Neumayer, S. & Steidl, G. Unbalanced multi-marginal optimal transport. *J. Math. Imag. Vis.* **65**, 394–413 (2023).
52. Le, K., Nguyen, H., Nguyen, K., Pham, T. & Ho, N. On multimarginal partial optimal transport: Equivalent forms and computational complexity. *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics* 4397–4413 (2022).

53. Schulter, S., Vernaza, P., Choi, W. & Chandraker, M. Deep network flow for multi-object tracking. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2730–2739 (2017).
54. Chari, V., Lacoste-Julien, S., Laptev, I. & Sivic, J. On pairwise costs for network flow multi-object tracking. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 5537–5545 (2015).
55. Jaqaman, K. et al. Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods* **5**, 695–702 (2008).
56. Zhang, L., Li, Y. & Nevatia, R. Global data association for multi-object tracking using network flows. *2008 IEEE Conference on Computer Vision and Pattern Recognition* 1–8 (2008).
57. Lin, T., Ho, N., Cuturi, M. & Jordan, M. I. On the complexity of approximating multimarginal optimal transport. *J. Mach. Learn. Res.* **23**, 1–43 (2022).
58. Hummert, J., Tashev, S. A. & Herten, D.-P. An update on molecular counting in fluorescence microscopy. *Int. J. Biochem. Cell Biol.* **135**, 105978 (2021).
59. Schmied, J. J. et al. DNA origami-based standards for quantitative fluorescence microscopy. *Nat. Protoc.* **9**, 1367–1391 (2014).
60. Schmied, J. J. et al. Fluorescence and super-resolution standards based on DNA origami. *Nat. Methods* **9**, 1133–1134 (2012).
61. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
62. Balzarotti, F. et al. Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. *Science* **355**, 606–612 (2017).
63. Gwosch, K. C. et al. MINFLUX nanoscopy delivers 3D multicolor nanometer resolution in cells. *Nat. Methods* **17**, 217–224 (2020).
64. Liero, M., Mielke, A. & Savaré, G. Optimal transport in competition with reaction: the Hellinger–Kantorovich distance and geodesic curves. *SIAM J. Math. Anal.* **48**, 2869–2911 (2016).
65. Alexander, S. *Combinatorial optimization: Polyhedra and efficiency*, 24 edn (Springer, 2003).
66. Goldberg, A. V. An efficient implementation of a scaling minimum-cost flow algorithm. *J. Algorithms* **22**, 1–29 (1997).
67. Morris, C. Central limit theorems for multinomial sums. *Ann. Stat.* **3**, 165–188 (1975).
68. de Chaumont, F. et al. Icy: an open bioimage informatics platform for extended reproducible research. *Nat. Methods* **9**, 690–696 (2012).
69. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
70. Naas, J. et al. Source Data and Scripts - MultiMatch: Geometry-Informed Colocalization in Multi-Color Super-Resolution Microscopy (v0.0.2). *Zenodo* https://doi.org/10.5281/zenodo.7221879 (2024).
71. Walt, Svd et al. Scikit-image: Image processing in Python. *PeerJ* **2**, e453 (2014).
72. Perron, L. & Furnon, V. OR-tools https://developers.google.com/optimization/ (2022).

Acknowledgements

We would like to thank Leo Lehmann for his help in software implementation, Jan-Niklas Dohrke for the graphic illustration of the nanorulers in Fig. 1C, and Christiane Elgert and Arndt von Haeseler for constructive criticism on the manuscript. J.N. is supported by the Austrian Science Fund (FWF) project number F78 to Arndt von Haeseler. This work was supported by the European Research Council (ERC AdG no. 835102) (to S.J.). G.N., H.L., S.J., and A.M. are supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC 2067/1-390729940 and H.L., B.S., S.J., and A.M. by the DFG CRC 1456 "Mathematics of Experiment" (Project Number B04, C06).

Author contributions

S.S. and S.J. designed the experimental work and S.S. acquired the experimental data. J.N. designed the MultiMatch algorithm with support from H.L., B.S., and A.M. G.N. implemented the MultiMatch Python package. J.N. and G.N. performed data analysis and method evaluation and developed the statistical methodology with input from H.L. and A.M.; A.M. initiated and S.J. and A.M. supervised the project. J.N. wrote the manuscript with contributions from all co-authors. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-024-06772-8.

Correspondence and requests for materials should be addressed to Axel Munk.

Peer review information *Communications Biology* thanks Thibault Lagache and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editor: Ophelia Bu.