Multicolor single-particle reconstruction of protein complexes

Christian Sieben12,4*, Niccolò Banterle34, Kyle M. Douglass1, Pierre Gönczy23 and Suliana Manley12*

Single-particle reconstruction (SPR) from electron microscopy (EM) images is widely used in structural biology, but it lacks direct information on protein identity. To address this limitation, we developed a computational and analytical framework that reconstructs and coaligns multiple proteins from 2D super-resolution fluorescence images. To demonstrate our method, we generated multicolor 3D reconstructions of several proteins within the human centriole, which revealed their relative locations, dimensions and orientations.

Macromolecular complexes within cells usually contain multiple protein species, whose precise arrangement is required for proper functioning of molecular machines. Single-particle analysis of EM images can lead to 3D reconstructions of such complexes, recently with near-atomic resolution13. To deduce the organization of specific proteins, scientists have used computational methods to dock structures from X-ray crystallography or NMR spectroscopy in 3D reconstructions14. Alternatively, immunogold or nanobody labeling can be used to determine the location of target proteins15, and differences between electron density maps can reveal the position of mutant or missing proteins16. Nevertheless, it remains challenging to locate native proteins in 3D reconstructions, and the assembly mechanisms and functional modules of macromolecular complexes cannot be deciphered accurately without this information.

Fluorescence-based single-molecule localization microscopy (SMLM) can help address this challenge, as demonstrated by 2D averaging of nuclear pore complexes3. Extending into 3D, SMLM demonstrated isotropic reconstruction from 2D SMLM images of DNA origami and simulated data4. However, multicolor particle reconstruction of actual macromolecular complexes requires the generation of large image libraries of multiple proteins and a solution to the problem of 3D multichannel alignment. Here we present a systematic framework that addresses both of these challenges.

We used a dedicated high-throughput SMLM setup10 to collect large multicolor particle datasets, which we processed by using a semi-automated computational workflow to reconstruct and align multiple proteins onto a single 3D particle. Our reconstruction workflow comprises three steps (Supplementary Fig. 1): (i) SMLM imaging and particle extraction, (ii) reconstruction of separate 3D protein volumes (i.e., the reference and the protein(s) of interest), and (iii) mapping of one or more proteins of interest onto the reference. To facilitate the analysis, we developed SPARTAN, an SMLM single-particle analysis software that allows a user to carry out the major processing steps via a convenient graphical user interface (Supplementary Software, Supplementary Notes 1 and 2).

We applied our method to the human centriole, reconstructing and aligning both on- and off-axis structures. Centrioles are evolutionarily conserved diffraction-limited organelles that seed the formation of cilia, flagella and centrosomes17. The mature human centriole comprises ninefold-symmetrically arranged microtubule triplets and contains >100 different proteins organized into distinct substructures18. For instance, distal appendages harbor the protein Cep164 and are key for the formation of cilia and flagella15. A torus encircling the proximal part of the mature centriole and comprising the proteins Cep57, Cep63 and Cep152 acts as a nucleation platform for the new procentriole, whose assembly relies on the self-organization of the HsSAS-6 protein into a cartwheel structure11. The details of component dimensions within the Cep57–Cep63–Cep152 torus and of the emerging procentriole with respect to this torus remain unclear.

To demonstrate our multicolor 3D SMLM reconstruction workflow, we imaged proteins within centrioles and procentrioles. We isolated centrosomes from human KE37 cells arrested in S phase, concentrated them on coverslips by centrifugation, and then immunolabeled and stained them (Supplementary Note 3). Next, we used dual-color high-throughput SMLM to image ~100–300 centrioles per field of view (Supplementary Fig. 2). We segmented localizations belonging to centrioles and procentrioles (referred to here as particles) by using a mask generated through automated Otsu thresholding of the wide-field images. We then applied a density-based filter (DBSCAN16) to separate adjacent particles (Supplementary Figs. 3 and 4). Only densely labeled particles (typically 10–20% of the initial dataset) were rendered and used to populate the particle library.

Next, we reconstructed a 3D volume from single particles of unknown orientation by using well-established EM routines (Supplementary Note 4). We classified particles from both fluorescence channels (Supplementary Fig. 5) by applying Clustering 2D (CL2D)19 or template-free maximum-likelihood multi-reference refinement (ML2D)14. Because of the high degree of radial symmetry within centrioles11, we chose a low number of classes (typically 8–15) to reduce the computational complexity (Fig. 1a, Supplementary Fig. 6). Although we applied symmetry information as a last step to overcome nonuniform angular coverage (Supplementary Fig. 7), our workflow is capable of reconstructing and recovering the symmetry of unknown objects given a number of particles sufficient to fully sample the orientational space, as verified in silico (Supplementary Fig. 8). Information loss resulting from under-represented orientations produces missing wedge artifacts, similarly to electron tomography. We then used the class averages most closely resembling the input particles (Supplementary Note 4) to compute an initial 3D model, and carried out structural refinement based on matching of the 2D projections to the input particles. In this manner, we reconstructed the torus protein Cep152, and measured its diameter.
as ~270 nm (Fig. 1a), in agreement with the ~242 nm measured for SNAP–Cep152 by stimulated emission depletion microscopy when accounting for antibody size. In addition, our 3D reconstruction showed that the height of the torus was ~160 nm (Supplementary Fig. 9). Using the same procedure, we reconstructed the well-known bacteriophage T4 (Supplementary Fig. 10), thus demonstrating the generality of this 3D SMLM reconstruction workflow. In this case, particles aligned preferentially parallel to the coverslip, which resulted in uneven angular sampling (Supplementary Fig. 7) that we compensated for with the known phage symmetry.

The general alignment of two volumes requires both translation and rotation of one of the volumes in three dimensions. The problem is less complex if the two proteins are symmetrically arranged and is further simplified if they share a symmetry axis. To achieve multicolor reconstruction, we first considered the latter case of proteins sharing a principal symmetry axis, where the only alignment parameter is the displacement along the symmetry axis, \( \Delta z \). We collected dual-color images of Cep152–Cep164, Cep152–Cep57 and Cep63–Cep57, using Cep152 and Cep57 as reference proteins. This reduced the problem of alignment to only two 3D volumes at a time. We assumed a consistent direction of the displacement, but a third marker could be used to determine particle orientation. We divided the alignment process into two steps: (i) co-orient both particles and reconstruct their volumes, and (ii) translate one volume by the correct distance \( \Delta z \) along the symmetry axis (Fig. 1b). Because both proteins are integrated into the same structure, the corresponding particles share the same relative orientation. Therefore, it suffices to find the orientation of particles in one channel (i.e., the reference), then preserve and assign the orientation to the second channel. Given that it can be challenging to image two or more proteins with super-resolution owing to low protein abundance and/or labeling efficiency, this procedure offers a great advantage in that only the reference protein images must contain enough information to be oriented. Application of this procedure results in two co-oriented volumes.

The last step for alignment of the two protein volumes is identification of the translational shift, \( \Delta z \), between their side-view (xz) projections (Fig. 1c,d, Supplementary Fig. 6). Thus, we carried out orientational filtering by using supervised machine learning to identify top-view and side-view projections of the reference protein from a combination of 12 calculated shape descriptors (Supplementary Fig. 11a, Supplementary Note 5). After training on ~10% of particles, the models successfully identified ~97% (true positive rate) of side-view projections (Supplementary Fig. 11b), typically yielding 50–100 side-view particles per imaged protein pair. This method offers the advantage that after the model is trained for a reference protein, it can be directly applied to other datasets using the same reference. Because individual two-color particles are subject to
heterogeneous labeling (Supplementary Fig. 12), we generated averaged side-view projections via a 2D alignment consisting of particle rotation in 3° increments followed by translational alignment and cross-correlation (Supplementary Fig. 13). These aligned averages allowed a more precise estimate of Δθ, and revealed the average particle dimensions and symmetry (Supplementary Figs. 9 and 14–16).

This workflow allowed us to reconstruct and coalign the toroidal complex Cep57–Cep152–Cep63, with the distal appendage protein Cep164 in a four-color volumetric reconstruction of the mature human centriole (Fig. 1e, Supplementary Video 1). This showed that although the Cep57 torus is aligned axially with the Cep152 and Cep63 volumes, as expected from their known association in cells20, it has the smallest dimensions of the three (~230 nm in diameter and ~130 nm in height), placing it near the outer microtubule wall. We also discovered a distribution of Cep57 and Cep152 with ninefold radial symmetry (Supplementary Fig. 15), further suggesting association with the ninefold-symmetric outer microtubule wall of the centriole, perhaps via the microtubule binding domain of Cep5720. We confirmed Cep164’s previously observed ninefold symmetry and located its N terminus more proximal and closer to the centriolar wall than previously reported21 (Supplementary Fig. 15).

The above approach works well for proteins that share a principal symmetry axis, but there are important exceptions. We therefore extended our method to the procentriole, marked by the protein HsSAS-6, which assembles from a single focus on the torus containing Cep57–Cep63–Cep15222 and was suggested by EM to initially adopt nonorthogonal orientations21. We collected dual-color images of Cep152–HsSAS-6 and generated average top and side views by following the workflow described above (Fig. 2a). The orientation of Cep152 was insufficient to define that of HsSAS-6, because the two proteins do not share a symmetry axis (Fig. 2a). Therefore, we combined the images from both proteins into a single image and carried out class averaging and alignment on the resulting dataset. However, we found that when we used a simple sum of the two channels, the signal from Cep152 dominated and prevented alignment of the smaller HsSAS-6 volume (Supplementary Fig. 17). To overcome this, we combined the two channels in a weighted sum (Supplementary Figs. 17 and 18) and used the combined particles for structural refinement of the initial Cep152 volume with no symmetry constraint (Fig. 2b). Finally, we fit the individually reconstructed protein volumes into the asymmetric global structure to achieve a two-color volumetric reconstruction of the nascent procentriole in the context of the centriolar torus (Fig. 2c, Supplementary Video 2). The combined reconstruction had a lower resolution than the individual structures (Fig. 2c), probably reflecting flexible relative positioning of the two entities. Indeed, we found that the angle θ between the two proteins measured from individual side-view particles (Fig. 2a) was variable, with an average value (±s.d.) of 15.4° ± 4.5° (n = 75), in agreement with the angle obtained from our 3D reconstruction (θ = 13°). Finally, in a three-color experiment, we used Centrin to mark the distal end of the centriole24 (Supplementary Fig. 19), which revealed a preferential orientation of the procentriole toward the distal end. Together, these findings support a loosely defined orientation between the torus and the emerging procentriole, with a broken distal–proximal symmetry.

In conclusion, we developed a framework that generates multicolor 3D volumes from dual-color 2D SMLM datasets, and used it to reveal unknown features of human centriole and procentriole architecture. Our approach is directly applicable to any single-particle dataset with sufficient angular coverage, although its extension to repetitive structures such as helices would require adaptation. Our flexible workflow is implemented in a software package that is suitable for other multiprotein complexes and imaging modalities. The combination of information from 3D SMLM reconstructions with EM particle reconstructions is likely to prove invaluable in the future, as are improvements in labeling, to permit higher fidelity of multicolor images to the underlying structure.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41592-018-0140-x.
Received: 22 December 2017; Accepted: 1 August 2018; Published online: 1 October 2018

References
1. Campbell, M. G., Veesler, D., Cheng, A., Potter, C. S. & Carragher, B. eLife 4, e06380 (2015).
2. Jiang, J., Pentelute, B. L., Collier, R. J. & Zhou, Z. H. Nature 521, 545–549 (2015).
3. Byeon, I.-J. L. et al. Cell 139, 780–790 (2009).
4. Beck, M., Lucić, V., Förster, F., Baumeister, W. & Medalia, O. Nature 449, 611–615 (2007).
5. Strauss, M., Schotte, L., Karunatilaka, K. S., Filman, D. J. & Hogle, J. M. J. Virol. 91, e01443-16 (2017).
6. Chang, Y.-W. et al. Science 351, aad2001 (2016).
7. Chang, Y.-W. et al. Nat. Microbiol. 2, 16269 (2017).
8. Szymborska, A. et al. Science 341, 655–658 (2013).
9. Salas, D. et al. Proc. Natl. Acad. Sci. USA 114, 9273–9278 (2017).
10. Douglass, K. M., Sieben, C., Archetti, A., Lambert, A. & Manley, S. Nat. Photonics 10, 705–708 (2016).
11. Bornens, M. Science 335, 422–426 (2012).
12. Bauer, M., Cubizolles, F., Schmidt, A. & Nigg, E. A. EMBO J. 35, 2152–2166 (2016).
13. Graser, S. et al. J. Cell. Biol. 179, 321–330 (2007).
14. Kitagawa, D. et al. Cell 144, 364–375 (2011).
15. Görnitz, P. Nat. Rev. Mol. Cell Biol. 13, 425–435 (2012).
16. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. in Proc. 2nd International Conference on Knowledge Discovery and Data Mining (eds Simoudis, E., Han, J. & Fayyad, U.) 226–231 (AAAI Press, Palo Alto, CA, 1996).
17. Sorzano, C. O. S. et al. J. Struct. Biol. 171, 197–206 (2010).
18. Scheres, S. H. W. et al. J. Mol. Biol. 348, 139–149 (2005).
19. Lukinačičus, G. et al. Curr. Biol. 23, 265–270 (2013).
20. Momotani, K., Khromov, A. S., Miyake, T., Stukenberg, P. T. & Somlyo, A. V. Biochem. J. 412, 265–273 (2008).
21. Sonnen, K. F., Schermelleh, L., Leonhardt, H. & Nigg, E. A. Biol. Open 1, 965–976 (2012).
22. Banterle, N. & Görnitz, P. Annu. Rev. Cell. Dev. Biol. 33, 23–49 (2017).
23. Loncarek, J., Hergert, P., Magidson, V. & Khodjakov, A. Nat. Cell Biol. 10, 322–328 (2008).
24. Paoletti, A., Moujou, M., Paintrand, M., Salisbury, J. L. & Bornens, M. J. Cell Sci. 109, 3089–3102 (1996).

Acknowledgements
We thank C. Lehmann for growth, purification and AFM characterization of T4 bacteriophages; J.M. de la Rosa-Trevín for support regarding the optimal use of Scipion; and C.G. Morrison (Centre for Chromosome Biology, School of Natural Sciences and National University of Ireland Galway, Galway, Ireland) for providing the Cep164 antibody (1F3G10). We thank B. Huang and X. Shi for providing their 2D particle-alignment code. We also thank R. Laine and C. Kaminski for providing an HSV SMLM dataset. This work was supported by the European Research Council (ERC; grant AdG 340227 to P.G.), the EPFL Fellows postdoctoral fellowship program funded by the European Union’s Horizon 2020 Framework Programme for Research and Innovation (grant agreement 665667 to N.B., MSCA-COFUND), and the National Centre for Competence in Research (NCCR) Chemical Biology (S.M.). We thank M. Dąszykowski for providing a public DBSCAN implementation.

Author contributions
C.S., N.B., P.G. and S.M. conceived and designed the project. C.S., P.G. and S.M. supervised the project. C.S. and N.B. performed all experiments and data analysis. C.S. and K.M.D. wrote analysis code. All authors wrote, revised and contributed to the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41592-018-0140-x.
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to C.S. or S.M.
Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Consequently, we estimated the total uncertainty $\sigma_{\text{total}}$ as the sum of the individual error components $\sigma_{\text{local}} = (\sigma_{\text{loc}} + \sigma_{\text{Reg}} + \sigma_{\text{Ab}})$, where $\sigma_{\text{loc}}$ is the localization precision (Supplementary Fig. 20), $\sigma_{\text{Reg}}$ is the final TRE after both registration steps, and $\sigma_{\text{Ab}}$ is the uncertainty added by a primary–secondary antibody conjugate: $\sigma_{\text{total}} = \sqrt{\sigma_{\text{loc}}^2 + \sigma_{\text{Reg}}^2 + \sigma_{\text{Ab}}^2} \approx 24$. This total uncertainty sets a lower boundary for the resolution of the obtained structure. The final resolution of the 3D model is further affected by the error introduced though nonhomogeneous and/or sparse labeling. Considering this, the obtained value is in agreement with the obtained 3D resolution as measured by Fourier shell correlation (Supplementary Table 2).

Particle segmentation and 3D reconstruction. After channel registration, the two localization datasets were ready for particle segmentation (SPARtan>Particles>Particle Segmentation). The localization maps for each field of view were loaded into MATLAB together with the corresponding wide-field (WF) images taken before the SMLM stack acquisition. A WF image was used for automatic Otu segmentation to identify the approximate location of individual particles within each field of view. Here we used the higher-contrast WF image (typically from the reference protein). To accommodate small shifts between WF image and localization data, we expanded each identified region by up to five pixels on all sides.

Overlapping regions were removed and the localizations from both channels were extracted for each segmented particle. Particles were filtered for a minimum number of localizations (typically $>100$) to ensure good particle labeling. We also applied an upper cut-off to reject clusters of particles and misidentified gold fiducials. During the next step, labeling noise was removed by applying a Gaussian mask to the WF images in a same region were separated via density-based clustering (DBSCAN$^-$). An example is shown in Supplementary Fig. 3. A low-density protein of interest was imaged (e.g., HsSAS-6), we used an additional filter selecting only the largest cluster (Supplementary Fig. 4). We then calculated a number of particle quality and shape descriptors (Supplementary Note 5), as well as the resolution (using Fourier ring correlation$^-$) for each particle, which allowed for efficient particle filtering. Finally, particles from both channels were rendered into a pixel image via a 2D histogram function with a bin size of 10 nm and blurred with a Gaussian filter with $\sigma$ corresponding to the measured localization precision. The final image approximated the probability density distribution of the fluorescent labels on the underlying structure and represented a widely used approach for visualization of SMLM data. We stitched the particle images together using the Montage function in ImageJ (Miji for MATLAB), which yielded the final input image for the 3D reconstruction (example shown in Supplementary Fig. 5).

SPR was done with Scipion, a freely available software package that integrates several widely distributed and well-developed 3D EM particle-reconstruction routines$^\dagger$. A brief tutorial of the required steps is provided in Supplementary Note 4. The particle montage images were imported into Scipion. Depending on the size of the dataset ($400 \times 6000$, Supplementary Table 2), each montage contained $\sim 500$ particles; thus, each reconstruction required the generation and import of multiple montages. During particle extraction, we removed labeling noise around the detected label particles (Supplementary Fig. 3), which resulted in a high-contrast particle montage that facilitated automatic particle picking (Xmipp$^3$). The particles were then aligned by 2D clustering (CL2D, Xmipp$^3$) and classified on the basis of template-free multi-reference maximum likelihood (ML2D, Xmipp$^3$) or 2D clustering (CL2D, Xmipp$^3$). Class averages were visually inspected. Some classes (typically particle two) accumulated particle fragments, which were removed in this stage. The remaining classes were used (Supplementary Note 4) to generate the initial model. For symmetric centriolar reconstructions (Cep164, Cep57, Cep152, Cep63), we used between 8 and 15 classes. To allow reconstruction with a limited number of particles, we applied rotational or ninefold symmetry at the final stage to fill in missing angular information (Supplementary Note 4). For bacteriophage T4, we used six class averages and used the weighted sum of the intended sixfold symmetry ($c6$). Initial models were calculated with either Xmipp$^2$ or Eman$^2$, and the two classifications provided similar results. Finally, the initial model was refined by particle back projection (Xmipp$^3$). Fourier shell correlation was calculated during particle refinement (particle back projection, Xmipp$^3$).

In the final stage, we used the reference protein as a template and reconstructed the reference protein using the steps described above, then applied the final alignment from each reference particle to the corresponding particle of the protein of interest (function alignment assign). The co-oriented particles of the protein of interest could then be reconstructed (function reconstruct in Relion$^3$ or Eman$^3$). Supplementary Note 4 includes a more detailed description. For the generation of a two-color volume of proteins sharing the same principal symmetry axis, it is also possible to use two independently reconstructed volumes. Here only the volume alignment (i.e., determination of $\Delta z$) requires a two-color SMLM dataset. For asymmetric reconstruction we used an adapted workflow in Scipion (Supplementary Fig. 17). First, we reconstructed each protein separately. Then we refined the symmetric volume of the reference (Cep164, Cep57, Cep152, Cep63) images (Cep152+2 HsSAS-6) without a symmetry constraint. Into the resulting asymmetric joint volume, we fit both individual protein volumes to obtain a high-resolution dual-color model. The volume fitting was done with Chimera$^7$, using the ‘Fit in Map’ tool (Tools>Volume Data>Fit in Map).
2D particle averaging and volume alignment. In general, the relative offset between the centers of mass of two distinct 3D particles can be determined by triangulation between any two projected views. In the case of the centriole, which has a principal rotational symmetry axis, only one projected orientation is needed to align two volumes. Any orientation that is not orthogonal to the symmetry axis (top view, \(xy\)), together with knowledge of the angle between the projection and the symmetry axis, is sufficient to determine the axial distance. In the simplest case, we can directly determine the axial distance between two volumes (\(\Delta z\)) from the projection into a plane parallel to the z axis (\(xz\)). We used particle projections of centriolar side \((xz)\) views for this purpose (Supplementary Fig. 14). Below, we describe the orientational filtering of top \((xy)\) and side \((xz)\) views. The volume alignment required only the side-view orientation, but we used the top views to characterize the protein’s symmetry properties (Supplementary Fig. 15). To efficiently identify particles with these orientations among a large number of individual particles with different orientations, we calculated 12 shape parameters whose values could be used as a characteristic signature for top-view \((xy)\) and side-view \((xz)\) projections (Supplementary Note 5). A similar approach was used recently to filter out centriole top-view particles \(^3\). Next, a subset of 200 particles was selected and manually filtered into top, side or intermediate views \((i.e., the \text{response})\). The shape descriptors and the results of the manual sorting were copied into a data table that was used as a training dataset to generate models via supervised machine learning (Supplementary Fig. 11). We used MATLAB’s Classification Learner to identify the best model for predicting the classified outcome \((\text{response})\) on the basis of the shape parameters. The best model was subsequently saved and could later be applied to other datasets. We found consistently good performance with support vector machine models, which are now also implemented into SPARtan \((\text{Particles} > \text{Manual Classifier} \text{ and } \text{Train SVM Classifier})\). The model generally identified certain shapes more accurately for top \((xy)\) orientations, for which it required little manual selection/filtering. Importantly, only one of the two imaged centriolar proteins \((\text{i.e., the reference})\) needed to be classified into top/side view.

All of the following operations were then performed on both channel datasets. Notably, whereas we used all particles for the determination of \(\Delta z\), we used only a visually filtered subset for the investigation of the ninefold symmetry. Many particles were over- or underlabeled and were thus not considered. The identified side-view particles were registered to the center of mass of the reference protein and aligned via an extended version of efficient subpixel registration by cross-correlation \(^4\) as described previously \(^5\) \((\text{see also Supplementary Note 2})\). Specifically, during the first iteration, we rotated each image from 1° to 359° in 3° steps, which resulted in 120 cross-correlations, from which we picked the orientation with the maximum root-mean-square error, thereby obtaining the optimal angle of rotation. The sum of all images was used as a reference for the first iteration. We refined the alignment over three to ten iterations, using the sum of all aligned particles from the previous iteration as the reference (Supplementary Fig. 13).

The translation between both channels along the z axis was determined from a line profile measurement of the two-color reconstruction \((\text{Fig. 1d, Supplementary Fig. 14})\). To generate a final multicolor volumetric representation, we loaded the co-oriented volumes into Chimera \(^6\) and centered them on top of each other. The \(\Delta z\) axial transformation was applied with the ‘transform coordinates’ tool \((\text{Tools} > \text{Movement} > \text{Transform coordinates})\). The volume obtained from the lower-resolution SMLM channel \((\text{i.e., DyLight 755 channel; Supplementary Fig. 5})\) was then replaced by a higher-resolution volume of the same structure \((\text{taken in the Alexa Fluor 647 channel})\). To this end, the high-resolution volume was loaded into Chimera and aligned to the low-resolution volume using the ‘Fit in Map’ tool \((\text{Tools} > \text{Volume Data} > \text{Fit in Map})\).

SMLM simulations. To evaluate the contribution of labeling noise and efficiency, as well as to test the particle-processing workflow, we developed a particle simulator that generates localization maps from ground truth models. To define a ground truth model, we used the geometric dimensions of the complex as obtained from SMLM. The ground truth model was then randomly rotated and projected onto the \(xy\) plane. A random number of molecules were selected according to the labeling efficiency and a defined number of noise molecules placed at random positions around each particle. Localizations \((\text{single frame appearances})\) originating from each fluorophore were assigned parameters drawn from measured distributions for photon count, localization precision, and on and off time. The distributions were obtained from single-molecule measurements of Alexa Fluor 647 obtained under experimental conditions. The resulting simulated particles were analyzed as described for experimental SMLM datasets. All simulations were carried out with custom-written MATLAB code supplied as part of the Supplementary Software package.

Statistics and reproducibility. Figures show representative data from \(\geq 3\) \((\text{Fig. 2, Supplementary Figs. 2, 10, 13, 14 and 16})\) or 2 \((\text{Fig. 1, Supplementary Figs. 8, 9, 15, 19 and 20})\) representative experiments, or from single high-throughput experiments \((\text{Supplementary Figs. 9, 12, 14 and 15})\). Supplementary Figs. 8, 13, 15 and 16 show representative data from two similar independent simulations.

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

Code availability. The version of SPARtan used in this paper is available as Supplementary Software. Supplementary Table 3 presents a comparison of the current SPARtan-based study with other studies demonstrating single-particle processing. The most recent version of SPARtan, including the source code, can be downloaded from GitHub \(\text{https://github.com/christian-7/MultiColorSPR}\) \((\text{download instructions can be found in the wiki})\).

Data availability

The datasets generated and analyzed in this study are available from the corresponding authors upon reasonable request. Sample datasets are available on Zenodo \(\text{https://doi.org/10.5281/zenodo.1288783}\).

References

25. Gogendeau, D., Guichard, P. & Tassin, A.-M. Methods Cell Biol. 129, 171–189 (2015).
26. Bornens, M., Paintrand, M., Berges, J., Marty, M.-C. & Karsenti, E. Cell Motil. Cytoskeleton 8, 238–249 (1987).
27. Bourdin, G. et al. Appl. Environ. Microbiol. 80, 1469–1476 (2014).
28. Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Curr. Protoc. Mol. Biol. 92, 14.20.1–14.20.17 (2010).
29. Olivier, N., Keller, D., Gönçzy, P. & Manley, S. PLoS One 8, e69004 (2013).
30. Huang, F. et al. Nat. Methods 10, 653–658 (2013).
31. Churchman, L. S. & Spudich, J. A. Cold Spring Harb. Protoc. 2012, 141–149 (2012).
32. Niewenhuizen, R. P. J. et al. Nat. Methods 10, 557–562 (2013).
33. Banterle, N., Bui, K. H., Lemke, A. E. & Beck, M. J. Struct. Biol. 183, 363–367 (2013).
34. de la Rosa-Trevín, J. M. et al. J. Struct. Biol. 195, 93–99 (2016).
35. Sorzano, C. O. S. et al. J. Struct. Biol. 148, 194–204 (2004).
36. Ludtke, S. J., Baldwin, P. R. & Chiu, W. J. Struct. Biol. 128, 82–97 (1999).
37. Scheres, S. H. W. J. Struct. Biol. 180, 519–530 (2012).
38. Pettersen, E. F. et al. J. Comput. Chem. 25, 1605–1612 (2004).
39. Gartenmann, L. et al. Curr. Biol. 27, R1054–R1055 (2017).
40. Guizar-Sicairos, M., Thurman, S. T. & Fienup, J. R. Opt. Lett. 33, 156–158 (2008).
41. Shi, X. et al. Nat. Cell Biol. 19, 1178–1188 (2017).
Reporting Summary

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

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

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

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

Microscopy data was collected using the freely available software Micromanager 2.0 beta. All localization data was analyzed using MATLAB 2016b for Windows 10 and macOS 10.12.6. 3D volumes where then reconstructed using the free software package Scipion 1.0 and visualized with UCSF Chimera 1.11.2.

Data analysis

All data analysis was developed, performed and tested in MATLAB 2016b for Windows 10 and macOS 10.12.6. All developed code is provided as a Supplementary Software package and via GitHub, where also updates will be made available. Link: https://github.com/christian-7/MultiColorSPR

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

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding authors upon request. Software sample data sets can be downloaded from Zenodo https://zenodo.org/record/1288783#.W1DbJ9gzaRs.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

| Sample size | No hypothesis-based experiment was performed. Therefore, the sample size was not predetermined. Sample sizes were chosen based on how many images were required for 3D reconstruction and could be collected on a given day. Each sample was analyzed separately. For all simulations, the data size was chosen large enough that statistical errors are minimal compared to the effect. |
| Data exclusions | Datasets with antibodies that stained in a manner inconsistent with the published literature was excluded from the analysis. Data were filtered for quality as described using common metrics in the field (number of localizations, size, eccentricity, FRC). Exclusion criteria were not pre-established. |
| Replication | Multiple replicates were performed with each antibody and the staining results were in general reproducible. If a dense labelling could not be achieved or was inconsistent with published literature, the dataset was excluded. If good and replicable, the staining was still variable, and sometimes did not yield data that could be used for further analysis and was excluded. |
| Randomization | This is not relevant to our study. There was no group allocation component to our study. |
| Blinding | There was not a group allocation component to the study. |

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a Involved in the study
☐ Unique biological materials
☐ Antibodies
☐ Eukaryotic cell lines
☐ Palaeontology
☐ Animals and other organisms
☐ Human research participants

Methods

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

Unique biological materials

Policy information about availability of materials

Obtaining unique materials

All materials are readily available from standard commercial sources as indicated in the manuscript and supplementary information. For any custom, unique or non-commercial material, we also provide the source for direct requests. Of course the authors would be happy to make contact or provide further assistance.
Antibodies used

All used antibodies are summarized in Supplementary Table 1. Primary antibodies against Cep152, Cep164 and Cep57 and HsSAS-6 are commercial and information regarding validation and application can be found in the manufacturers website. Cep152, Sigma, HPA039408; Cep57, Abcam, 169301, 1:500; HsSAS-6, Santa Cruz Bt, SC98506(H-300), 1:500; Cep63, Merck Millipore, 06-1292, 1:500; Centrin-1, Santa Cruz Bt, SC49622, 1:500.

Validation

Information regarding validation and application can be found in the manufacturers website. The antibody against the N terminus of Cep164 (named Cep164-2 in the manuscript) was obtained from Ciaran Morrison. The antibody was first described in Daly, O., Gaboriau, D., Karakaya, K., King, S., Dantas, T.J., Lalor, P., Dockery, P., Krämer, A. and Morrison, C.G. (2016) Genetargeted CEP164-deficient cells show a ciliation defect with intact DNA repair capacity. J. Cell Sci. 129: 1769-1774.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

KE37 cells were obtained from Bornens Laboratory, Institut Curie, France

Authentication

The cell line used in this study has not been authenticated by the authors.

Mycoplasma contamination

The used cell line is routinely tested for mycoplasma contamination and cells are typically discarded if positive.

Commonly misidentified lines

(See ICLAC register)

We did not use commonly misidentified cell lines.