Published in final edited form as:

Nat Struct Mol Biol. 2018 December ; 25(12): 1062–1064. doi:10.1038/s41594-018-0162-1.

Reply to ‘Deconstructing transport-distribution reconstruction in the nuclear-pore complex’

Ruba Andrew¹, Kelich Joseph¹, Ma Jiong², Yang Weidong^{1,*}

¹Department of Biology, Temple University, Philadelphia, PA, USA

²Department of Optical Science and Engineering, Fudan University, Shanghai, China

Ruba et al. reply — In their Correspondence, Tu et al.¹ explore the mathematical principles and the reliability of the 2D-to-3D transformation in SPEED microscopy that we have developed^{2–8}. The authors have validated the mathematical principle, but they argue that the SPEED technique is unable to reliably distinguish among peripheral, central, bimodal, and uniform distributions of molecules within the nuclear-pore complex (NPC) unless thousands of high-precision single-molecule locations are collected. We disagree with this conclusion and present a categorization routine and accompanying bin-size optimization that readily distinguish those distributions with high reliability (>90%) after collection of only a few hundred single-molecule locations, with localization precisions comparable to those in our previous publications^{2–8}.

To provide a more direct comparison with the analysis by Tu et al.¹, we also simulated datasets from the four distinct distributions mentioned above, then compared the reconstructed distributions to the ground-truth input distributions via our categorization routine (Fig. 1). There are several differences in our approaches. First, instead of a regime of Gaussian fitting and statistical analysis, our routine calculates the sum of the absolute-valued residuals (SAR) between each trial and the ground-truth input distributions (Fig. 1e). Second, before each simulation, the routine dynamically calculates an optimal categorization bin size for each set of parameters via a minimized chi-square error algorithm (Fig. 1b and Supplementary Fig. 1). The source code for our categorization routine and accompanying simulations is provided in Supplementary Note and at <https://github.com/andrewruba/YangLab/>. Importantly, this optimized bin size (10 nm under the simulation conditions in Fig. 1) was typically much larger than the bin size of 1.5–5 nm used in the simulations performed by Tu et al. (Supplementary Fig. 1h,i). Bin-size optimization contributed to the robustness of our routine against statistical variation and is important in experimental settings in which hundreds of single-molecule locations are typically collected. Moreover, our optimization routine can be performed without any prior knowledge of the single-molecule distribution. With these measures, ~100 locations allowed us to reach >90%

* weidong.yang@temple.edu.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41594-018-0162-1>.

successful categorization for peripheral, central, and uniform distributions within the NPC, when the single-molecule localization precision was set to 5 nm. A bimodal distribution requires more localizations for accurate categorization, but nonetheless only ~200 locations were sufficient to achieve >90% successful categorization (Fig. 1g–s). When the simulation was applied to the case of 10-nm localization precision, ~100–300 points were sufficient to distinguish the four distributions with >90% success in the NPC (Fig. 1t). In contrast, the best-case 5-nm categorization bin size used by Tu et al. substantially lowered the categorization success (Fig. 1u), and smaller bin sizes further decreased the success rate. Overall, we believe that Tu et al. were unable to reliably distinguish these distributions with a few hundred single-molecule locations because of the lack of an accurate categorization routine as well as not using optimal categorization bin sizes in their simulations.

In general, a bin size that is too small would yield substantial noise, whereas a bin size that is too large would mask underlying trends. In response to our reply outlining the use of an optimized bin size for successful categorization, Tu et al. varied the bin size in their simulations but found that it played only a minor role in increasing the successful categorization rate. To examine the underlying cause of their results, we replicated their simulations by selecting a fixed bin size, ranging from 4 to 20 nm. As shown in Supplementary Fig. 2, we obtained significantly higher rates of categorization success than that of the simulation performed by Tu et al. After examining their categorization algorithm more closely, we found that Gaussian fitting, a critical component of their categorization algorithm, is probably the cause of the lower successful categorization rates, because the reliability of a Gaussian fit strongly depends on the number of localizations and the localization precision used in their simulations. In contrast, our categorization algorithm is more robust to these effects; however, Tu et al. chose not to use our algorithm, claiming that it makes use of a priori knowledge that is not available in an experimental setting.

To respond to this claim and further highlight the importance and feasibility of using our categorization algorithm in experimental settings, we expanded our simulations by following the process illustrated in Fig. 1t, but changing the ground-truth distributions across all solution space (Supplementary Fig. 3a). As shown in Supplementary Fig. 3b, the s.d. of both peak means and peak widths for all simulated dataset trials converged on the simulation parameters within a few nanometers, with only a few hundred points, thereby suggesting the high robustness of our categorization approach. Therefore, we maintain that a proper bin size and a robust, well-implemented categorization routine are essential to obtain highly successful categorization rates. Moreover, after verification of the accuracy of an experimental dataset through the above method, other issues mentioned by Tu et al. regarding the registration precision between individual datasets, precision in determining NPC rotation, and the degree of symmetry in an experimental dataset can be assessed.

As a practical example, we found that ~100 points with an effective localization precision of 8.6 nm determined experimentally in the scaffold region of the NPC were sufficient to generate a consistent and reliable 3D probability-density map for the nuclear-transport route of importin β 1 (Supplementary Fig. 4). Moreover, the 100-point-based 3D map was further confirmed with a reproducibility rate of 100% and a mean peak localization error of ~1 nm after the number of data points was increased from 100 to 450 in our experimental

measurements⁹ (Supplementary Fig. 4e). At present, the acquisition of 3D super-resolution information with high temporal resolution in both live and fixed samples is still challenging. SPEED microscopy and its 2D-to-3D-transformation algorithm provide an alternative approach to obtain 3D sub-diffraction-limited information for fast macromolecular trafficking occurring in rotationally symmetric, subcellular organelles, such as the NPC^{2–8} and primary cilia¹⁰.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The project was supported by grants from the National Institutes of Health (NIH GM116204 and GM122552 to W.Y.).

References

1. Tu L-C, Huisman M, Chung Y-C, Smith CS & Grunwald D Nat. Struct. Mol. Biol 10.1038/s41594-018-0161-2 (2018).
2. Ma J & Yang W Proc. Natl Acad. Sci. USA 107, 7305–7310 (2010). [PubMed: 20368455]
3. Ma J, Goryaynov A, Sarma A & Yang W Proc. Natl Acad. Sci. USA 109, 7326–7331 (2012). [PubMed: 22529346]
4. Ma J et al. Nat. Commun 4, 2414 (2013). [PubMed: 24008311]
5. Goryaynov A & Yang W PLoS One 9, e88792 (2014). [PubMed: 24558427]
6. Ma J, Goryaynov A & Yang W Nat. Struct. Mol. Biol 23, 239–247 (2016). [PubMed: 26878241]
7. Ma J, Kelich JM & Yang W Methods Mol. Biol 1411, 503–518 (2016). [PubMed: 27147062]
8. Goryaynov A, Ma J & Yang W Integr. Biol. (Camb.) 4, 10–21 (2012). [PubMed: 22020388]
9. Ruba A, Kelich J, Luo W & Yang W Preprint at <https://www.biorxiv.org/content/early/2017/10/11/188060> (2017).
10. Luo W et al. Sci. Rep 7, 15793 (2017). [PubMed: 29150645]

Fig. 1 |. Route-categorization simulation.

a, For each set of simulation parameters (distribution type, localization precision, peripheral radius, central radius, and number of single-molecule locations), 1,000 datasets were generated in the y and z dimensions. To clearly visualize the distribution from the y, z scatter data, displayed here is a representative dataset from a peripheral distribution with a localization precision of 5 nm, peripheral radius of 23 nm, central radius of 0 nm, and 500 single-molecule locations. y, z coordinates can also be transformed to analogous r, θ coordinates. **b**, The bin-size optimization algorithm was performed according to the

protocol in Supplementary Fig. 1 by using the given simulation parameters. **c**, The y -dimensional histogram was obtained with the optimized bin size for every dataset. **d**, The 2D-to-3D-transformation algorithm was performed on every dataset. **e**, The resultant 3D density histograms were compared via SAR to the central, peripheral, bimodal, and uniform ground-truth distributions, which were obtained by simulating a dataset with all the same parameters except for 1,000,000 single-molecule locations. The lowest SAR indicates which distribution a given dataset is most similar to and thus which ground-truth distribution it is classified as. **f**, The count of all classifications for the ground-truth peripheral parameter set of 5-nm localization precision, 23-nm peripheral radius, 0-nm central radius, and 100 single-molecule locations. **g–j**, Sample datasets from simulation runs for the peripheral, central, bimodal, and uniform distributions respectively, with the same parameters as in **a**. **k–n**, y -dimensional histograms from **g–j**. **o–r**, 3D density histograms calculated from **k–n**. **s**, Resultant success rates for the different distributions. **t**, Simulation results for the peripheral, central, bimodal, and uniform distributions with parameters of 10-nm localization precision, 23-nm peripheral radius, 0-nm central radius, and bin sizes of 10–20 nm, obtained from the optimization algorithm. **u**, With the same parameters as in **t**, with the exception of a best-case 5-nm bin size used in simulations by Tu et al., the performance of the optimization algorithm was significantly lower.