nature methods

SUPPLEMENTARY INFORMATION

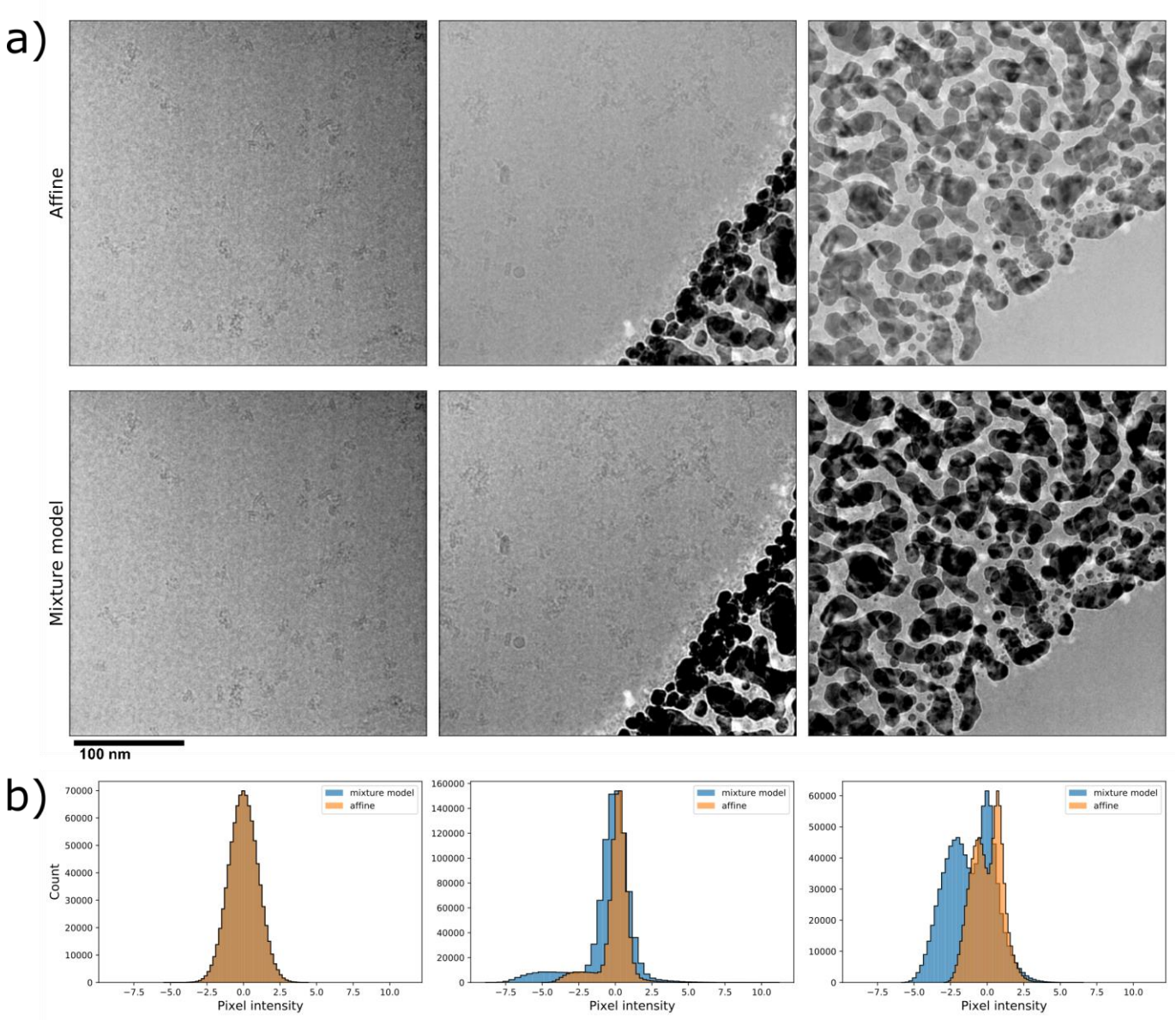
https://doi.org/10.1038/s41592-019-0575-8

In the format provided by the authors and unedited.

Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs

Tristan Bepler[1,2], Andrew Morin[2,3], Micah Rapp[4,5], Julia Brasch[4,5], Lawrence Shapiro[4], Alex J. Noble[5*] and Bonnie Berger[2,3*]

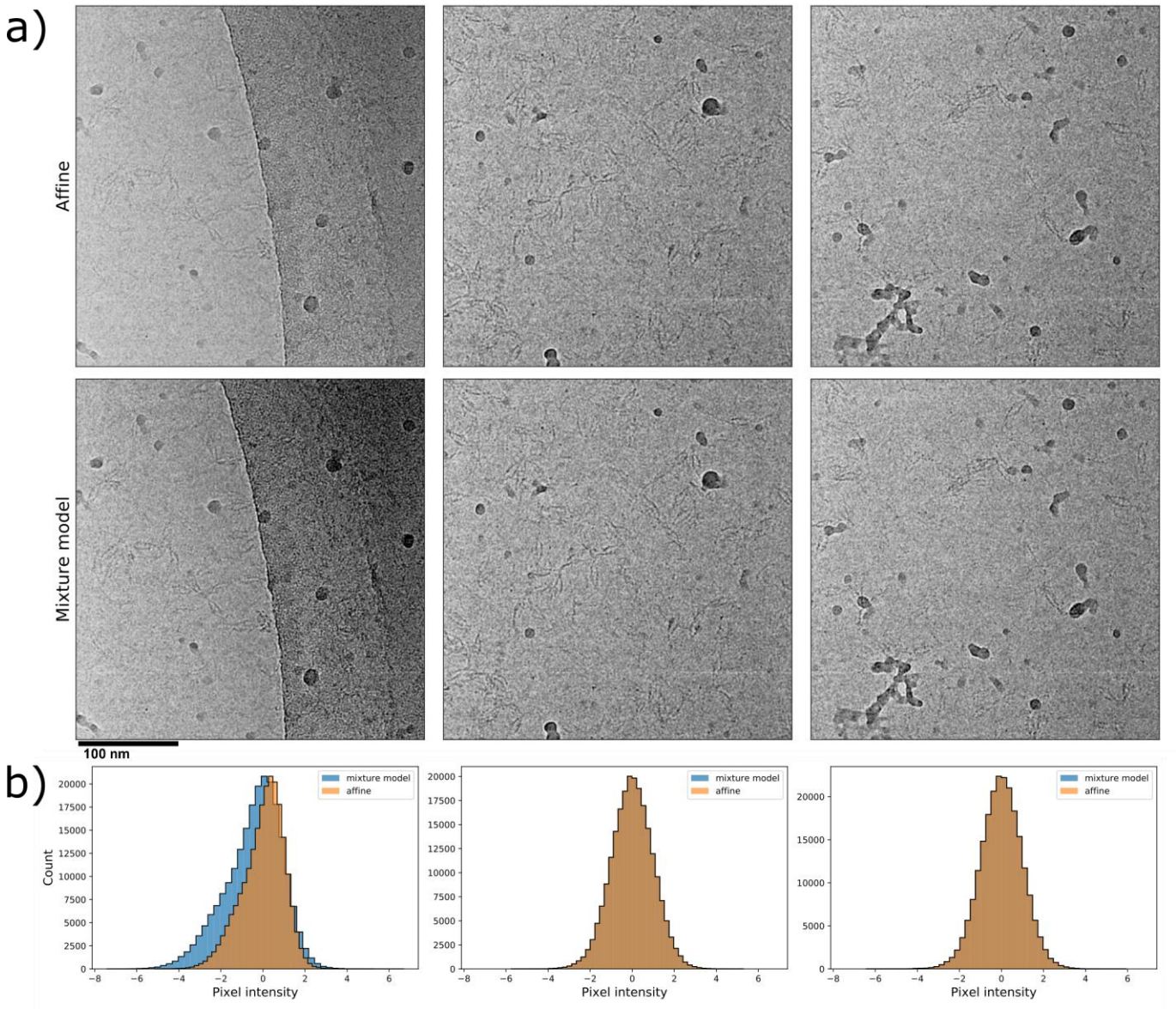
[1]Computational and Systems Biology, MIT, Cambridge, MA, USA. [2]Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA. [3]Department of Mathematics, MIT, Cambridge, MA, USA. [4]Department of Biochemistry and Molecular Biophysics, Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA. [5]National Resource for Automated Molecular Microscopy, Simons Electron Microscopy Center, New York Structural Biology Center, New York, NY, USA. *e-mail: anoble@nysbc.org; bab@mit.edu



Supplementary Figure 1

Normalization methods on EMPIAR-10261.

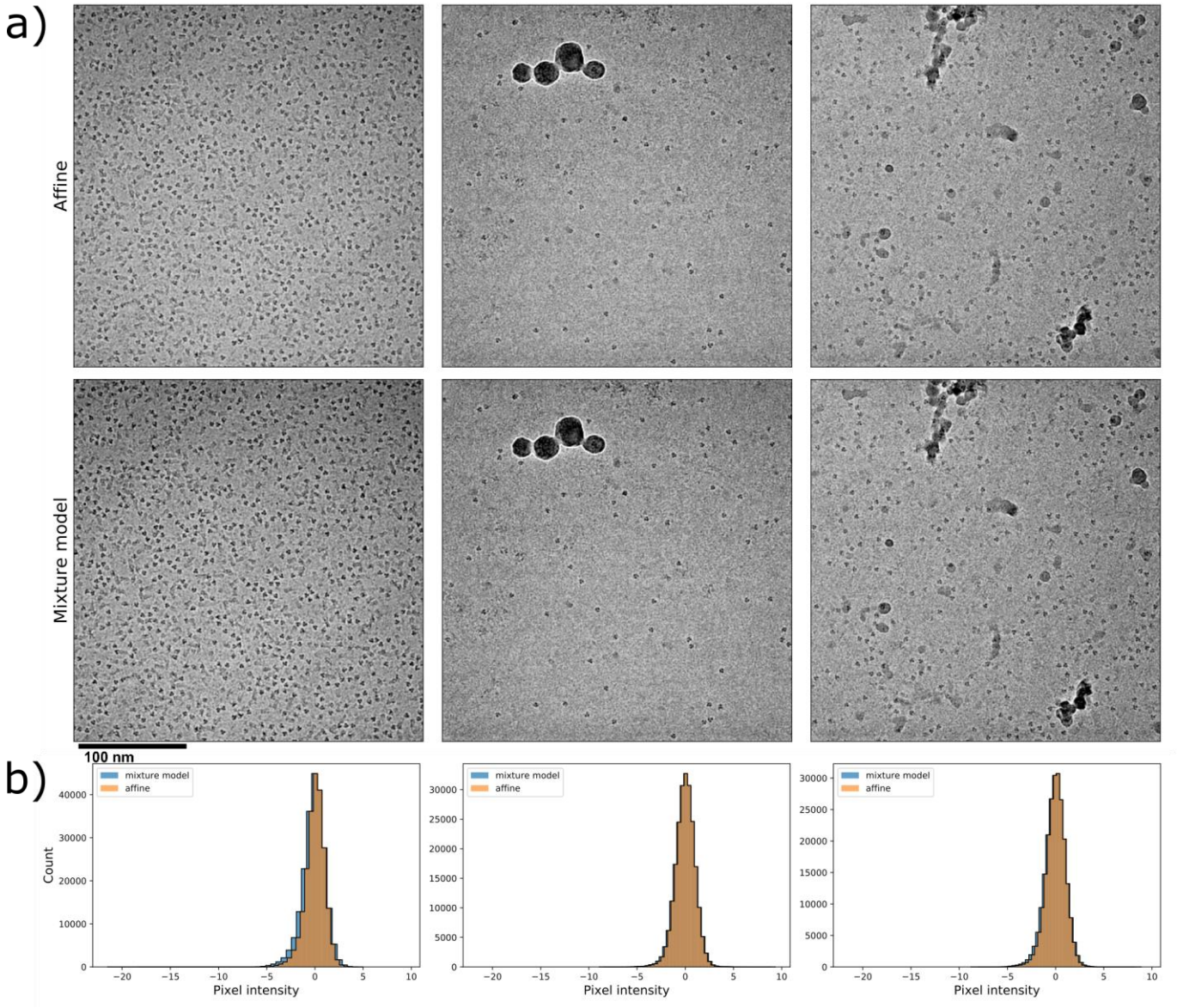
Comparison of standard affine normalization and our proposed mixture model normalization on EMPIAR-10261 micrographs downsampled 4x. For affine normalization, micrographs are transformed by subtracting the mean and dividing by the standard deviation of the pixel values. **(a)** Visualization of three example micrographs with either affine (top) or GMM (bottom) normalization. Affine normalized micrographs are washed out when there are dark grid regions present in the micrographs. **(b)** Histograms of the pixel intensities of the same three micrographs after affine or GMM normalization. GMM normalization correctly centers the pixel intensities around the high intensity peak. Results are consistent across >20 micrographs examined.



Supplementary Figure 2

Normalization methods on EMPIAR-10234.

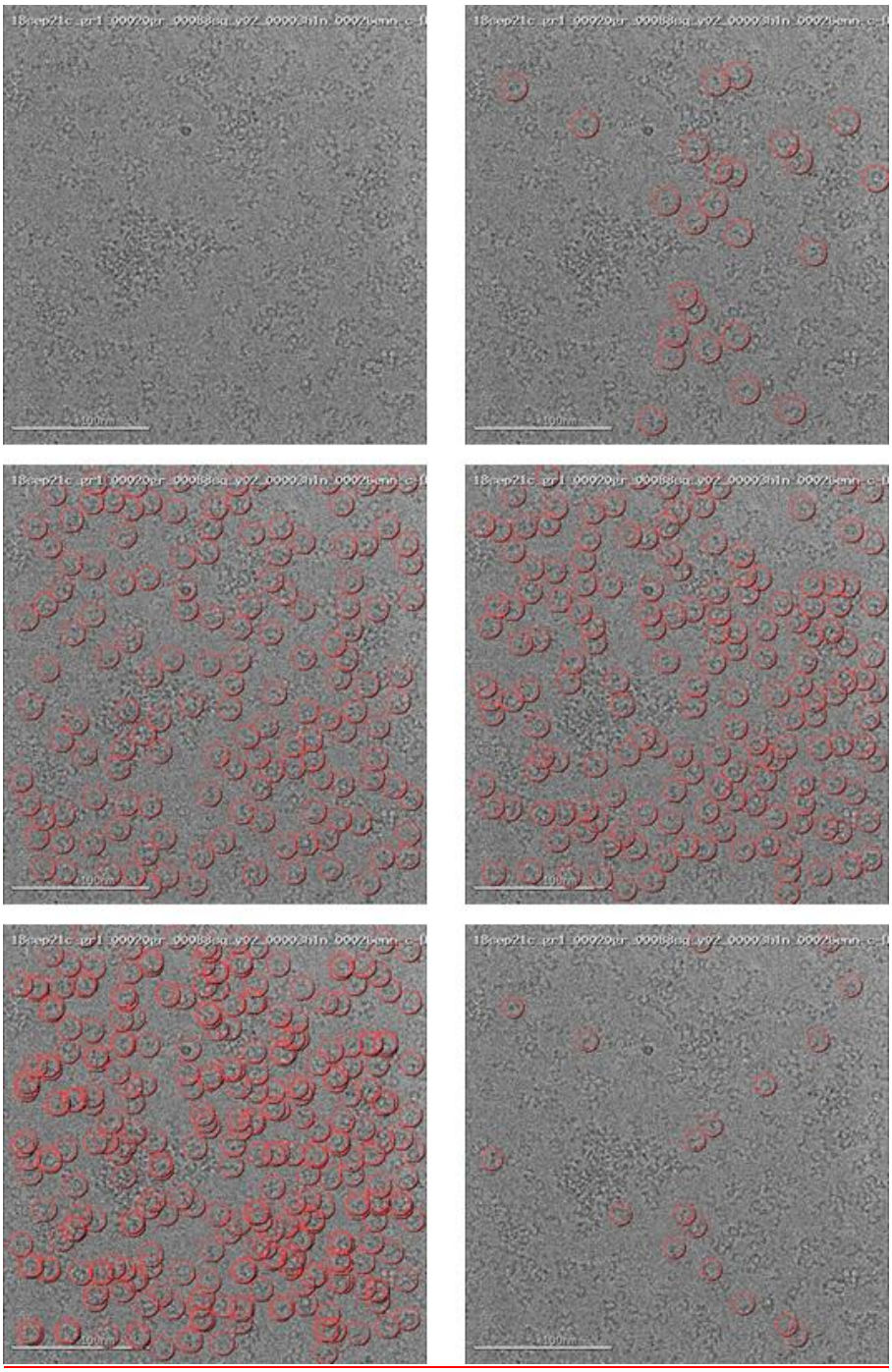
Comparison of standard affine normalization and our proposed mixture model normalization on EMPIAR-10234 micrographs downsampled 8x. For affine normalization, micrographs are transformed by subtracting the mean and dividing by the standard deviation of the pixel values. **(a)** Visualization of three example micrographs with either affine (top) or GMM (bottom) normalization. The left-most micrograph contains light carbon grid that is correctly removed by GMM normalization. **(b)** Histograms of the pixel intensities of the same three micrographs after affine or GMM normalization. Results are consistent across >20 micrographs examined.



Supplementary Figure 3

Normalization methods on EMPIAR-10096.

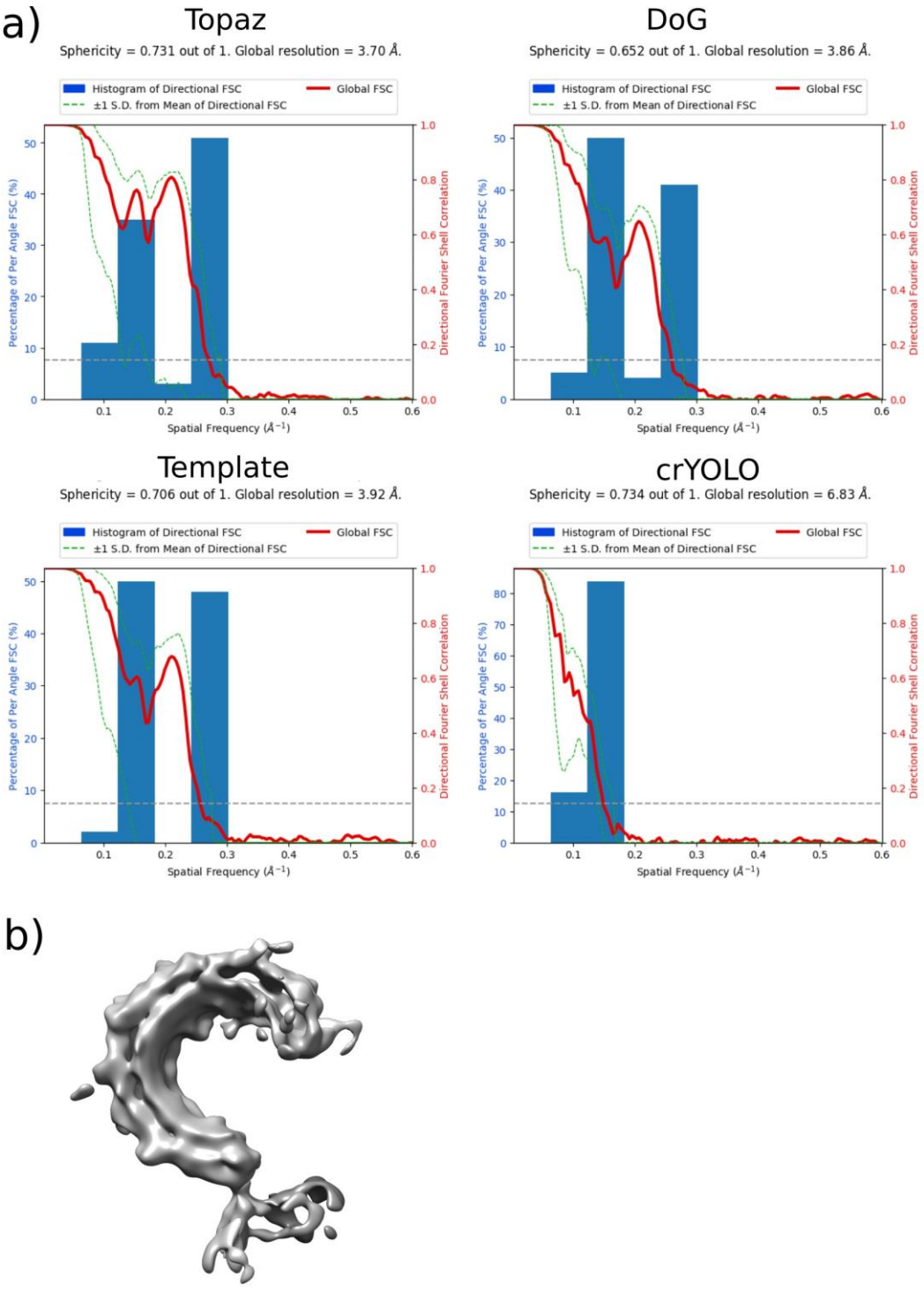
Comparison of standard affine normalization and our proposed mixture model normalization on EMPIAR-10096 micrographs downsampled 8x. For affine normalization, micrographs are transformed by subtracting the mean and dividing by the standard deviation of the pixel values. **(a)** Visualization of three example micrographs with either affine (top) or GMM (bottom) normalization. No grid is present in these micrographs, so affine and GMM normalization given nearly identical results. **(b)** Histograms of the pixel intensities of the same three micrographs after affine or GMM normalization. Results are consistent across >20 micrographs examined.



Supplementary Figure 4

Toll receptor example micrograph with picks.

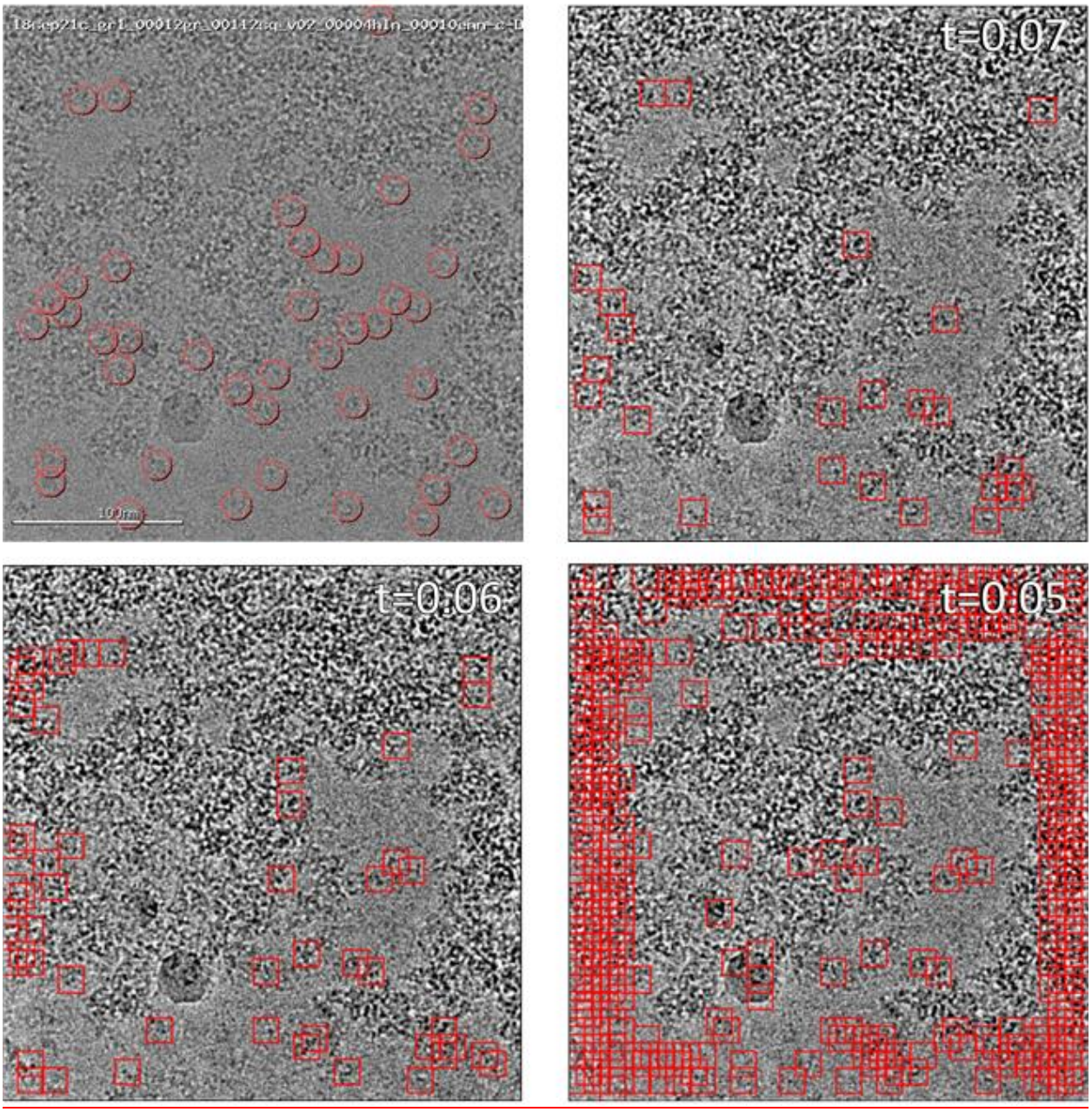
Example micrograph for the Toll receptor dataset with the unlabeled micrograph **(top left)**, training picks **(top right)**, DoG Picker picks **(middle left)**, Topaz picks **(middle right)**, FindEM picks **(bottom left)**, and crYOLO picks **(bottom right)**. This result is consistent over >100 micrographs examined.



Supplementary Figure 5

Toll receptor 3DFSC curves and crYOLO reconstruction.

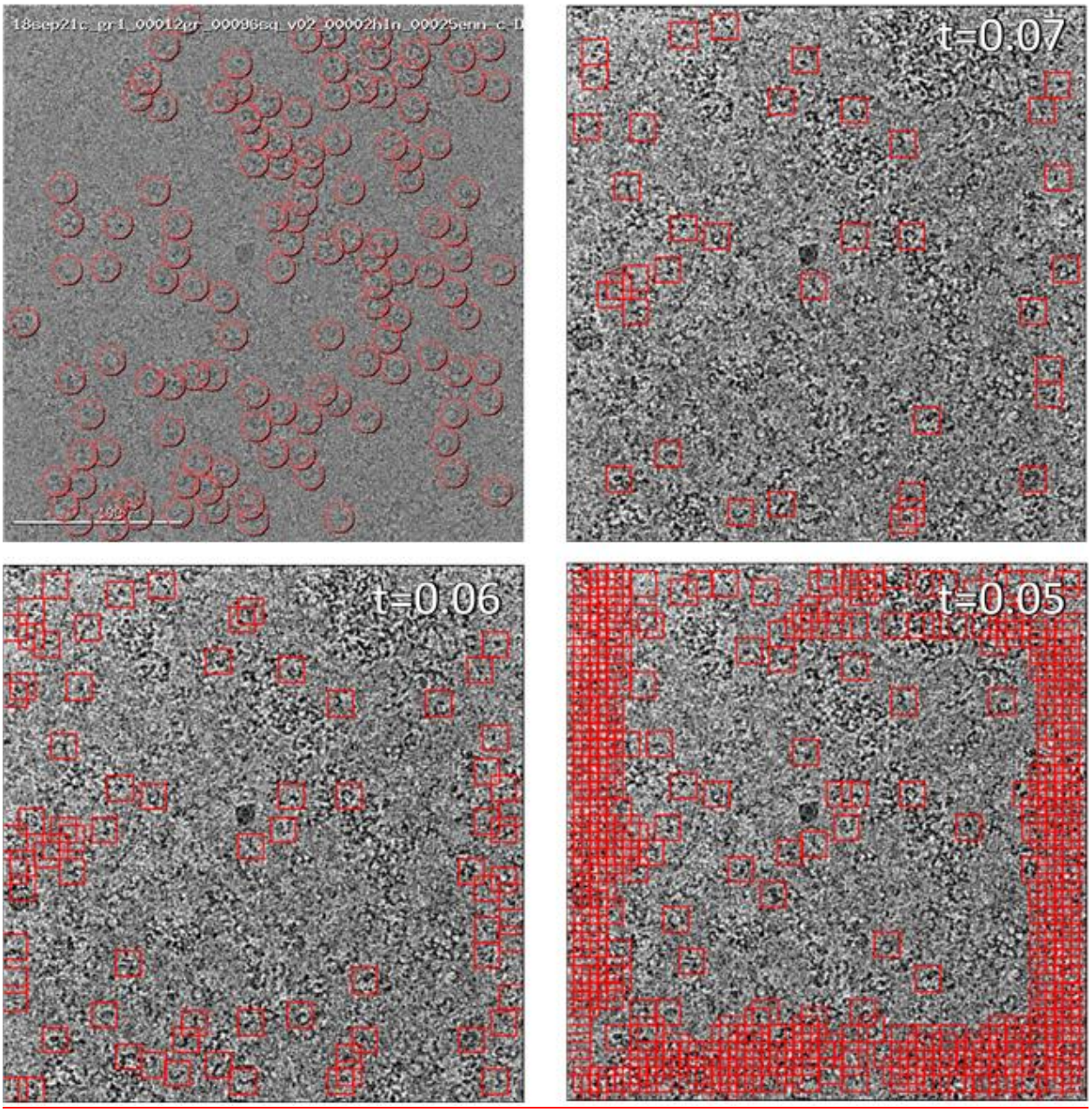
Toll receptor 3DFSC plots and 3D reconstruction of the Toll receptor using 131,300 particles picked using crYOLO. **(a)** 3DFSC plots for Toll receptor structures solved using particles from Topaz, DoG, Template, and crYOLO. **(b)** Density map of the crYOLO structure. The crYOLO structure reaches a resolution of 6.83 Å at FSC0.143 with a sphericity of 0.734.



Supplementary Figure 6

Sparse picking in aggregates: Topaz vs crYOLO.

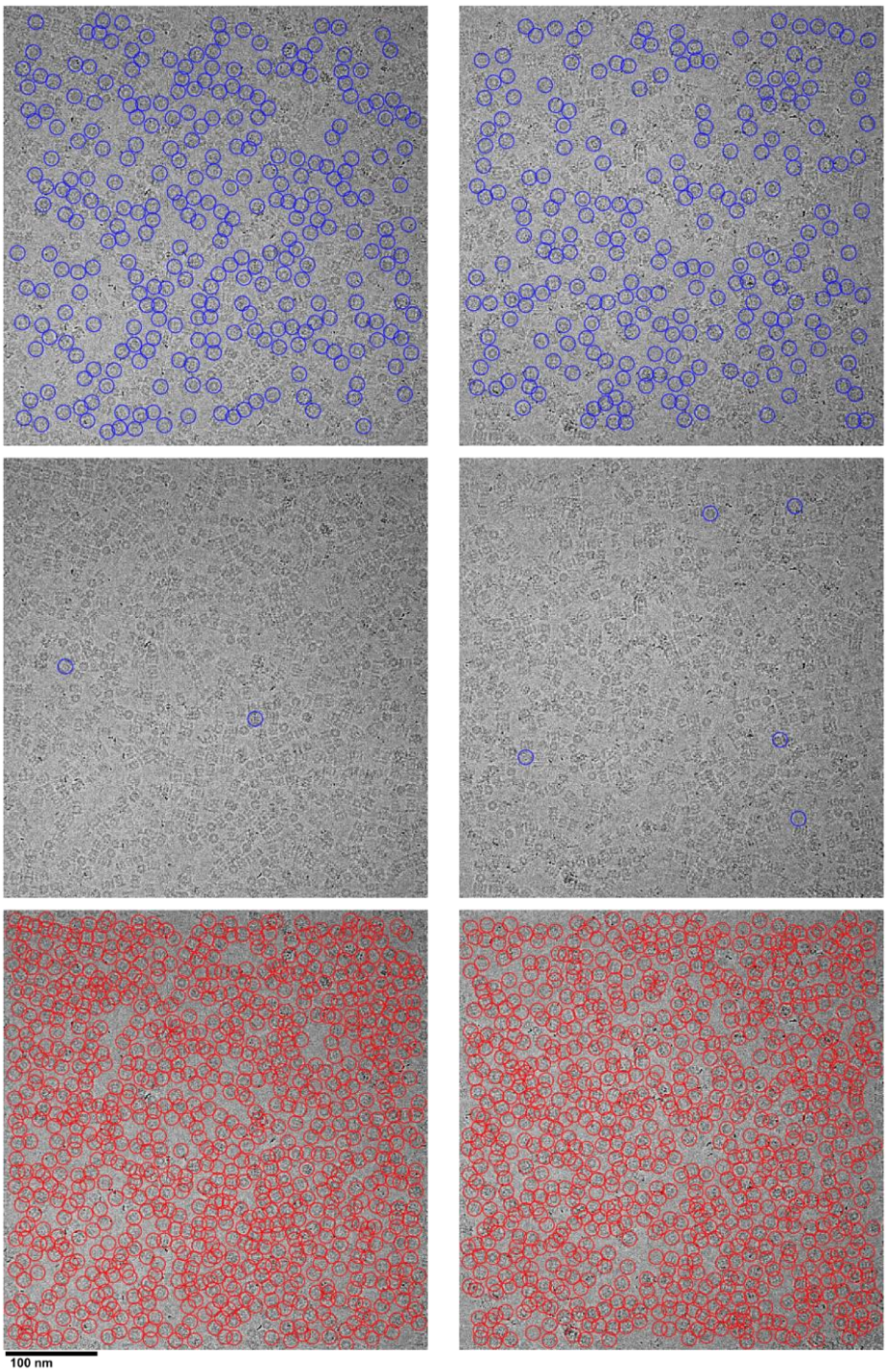
Example micrograph with a large amount of aggregate but still a number of real particles, comparing Topaz with crYOLO after following crYOLO's sparse picking procedure. Topaz **(top left)** picks 40 particles while avoiding aggregation and ice contamination at the default threshold of 0.0. crYOLO picks 1 particle at its default threshold of 0.3. Lowering the threshold to 0.07 **(top right)** yields 26 particles with a small cluster of picks in the aggregation near the bottom right hand corner of the image. Further decreasing the threshold to 0.06 **(bottom left)** yields even more particles, but now it is clear the network is picking a large number of pixels within the aggregation. At a threshold of 0.05 **(bottom right)**, crYOLO is no longer able to avoid aggregation or ice contamination. Note that the particle picking threshold increment in crYOLO is 0.01. We found these results to be consistent across >100 micrographs examined.



Supplementary Figure 7

Sparse picking with low SNR: Topaz vs. crYOLO.

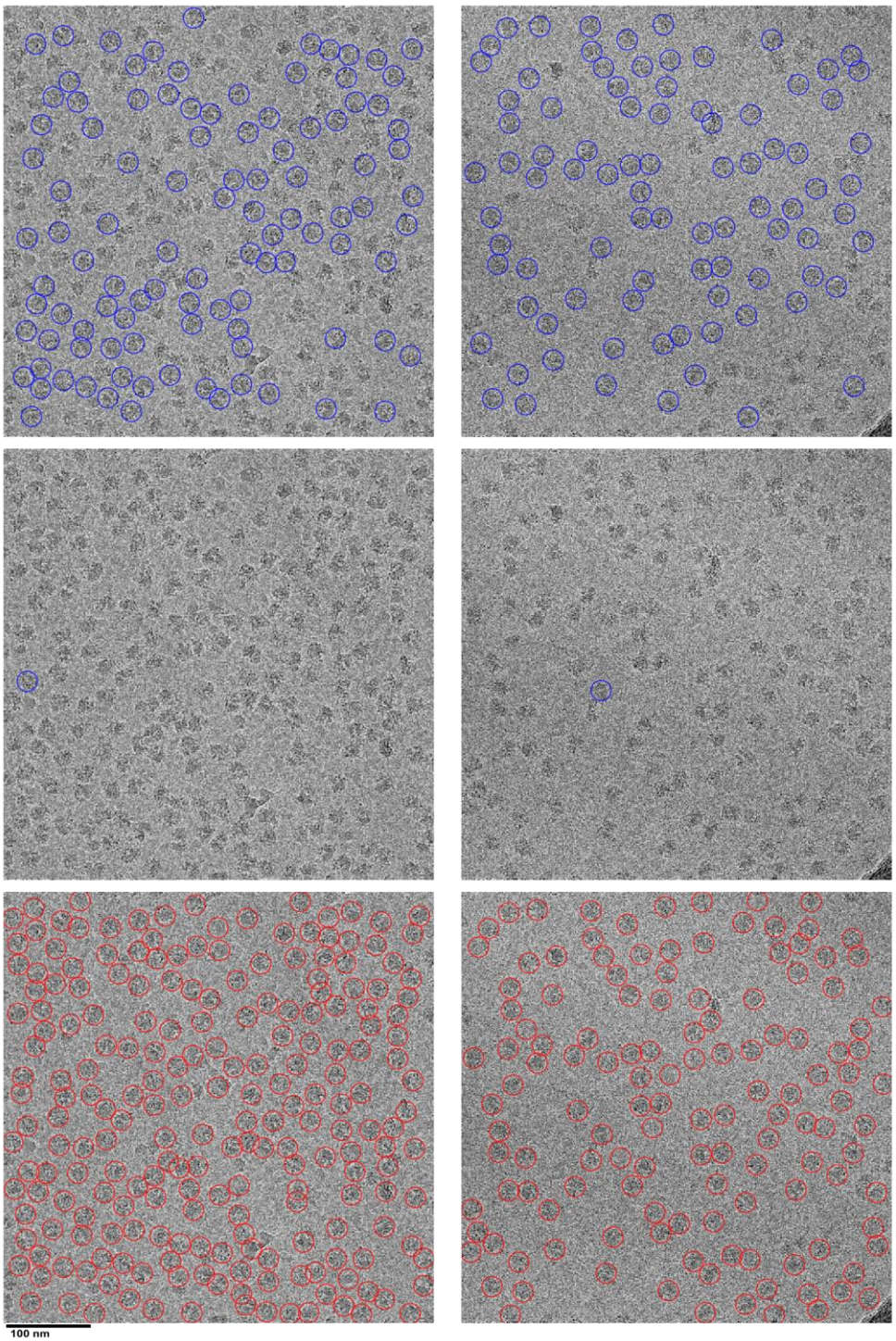
Example micrograph with real particles but thicker ice and thus lower contrast, comparing Topaz with crYOLO after following crYOLO's sparse picking procedure. Topaz (**top left**) picks 127 particles at the default threshold of 0.0. crYOLO picks 0 particle at its default threshold of 0.3. Lowering the threshold to 0.07 (**top right**) yields 37 particles. Further decreasing the threshold to 0.06 (**bottom left**) yields more particles, but the network is still missing real particles while starting to select some background pixels. At a threshold of 0.05 (**bottom right**), there are clear artifacts at the edges of the micrograph. We found these results to be consistent across >100 micrographs examined.



Supplementary Figure 8

T20S proteasome example micrographs.

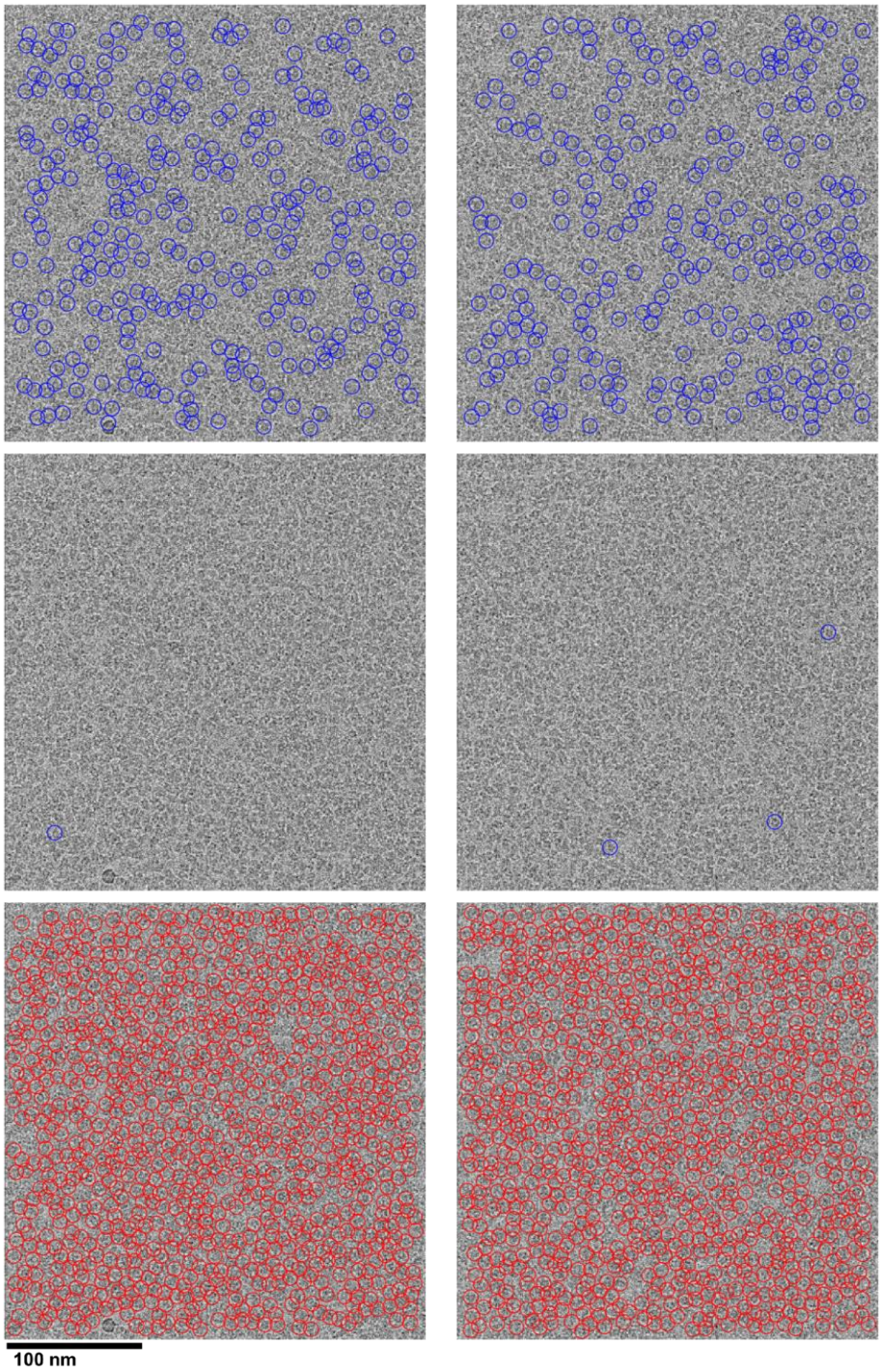
Two example micrographs for the T20S proteasome dataset (EMPIAR-10025) with **(top)** published particles circled in blue, **(middle)** training particles sampled from the published particles circled in blue, and **(bottom)** Topaz particles circled in red. Each column is a different micrograph. The circled training particles illustrates how sparse the Topaz picks are for this dataset. The PU learning framework allows picking to be performed with high accuracy despite the sparsity of examples, as seen by the Topaz picks in red. Furthermore, Topaz recovers many more real particles than are present in the published set. We found this result to be consistent across >100 micrographs examined.



Supplementary Figure 9

80S ribosome example micrographs.

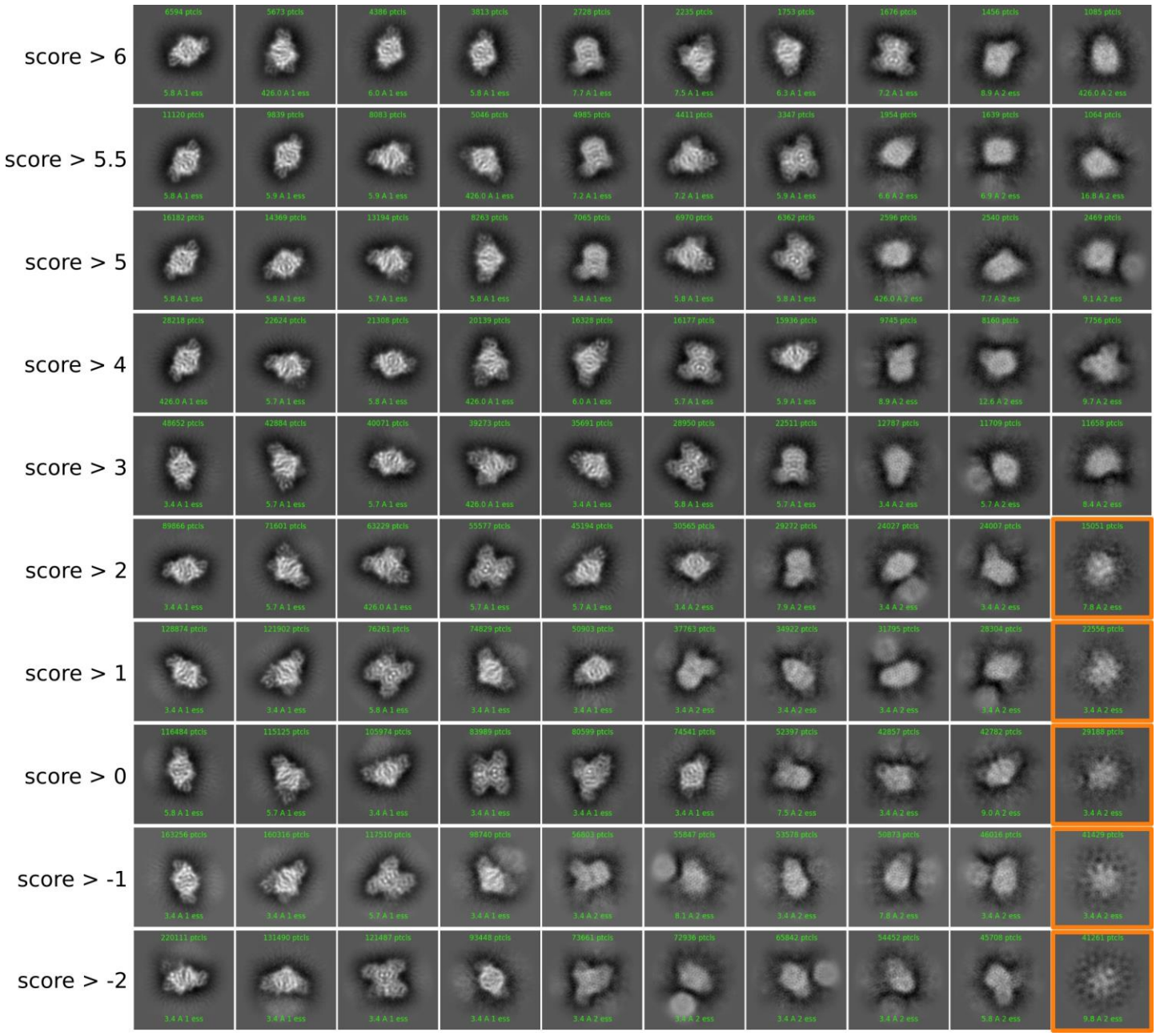
Two example micrographs for the 80S ribosome dataset (EMPIAR-10028) with **(top)** published particles circled in blue, **(middle)** training particles sampled from the published particles circled in blue, and **(bottom)** Topaz particles circled in red. Each column is a different micrograph. The circled training particles illustrates how sparse the Topaz picks are for this dataset. The PU learning framework allows picking to be performed with high accuracy despite the sparsity of examples, as seen by the Topaz picks in red. Furthermore, Topaz recovers many more real particles than are present in the published set. We found this result to be consistent across >100 micrographs examined.



Supplementary Figure 10

Rabbit muscle aldolase example micrographs.

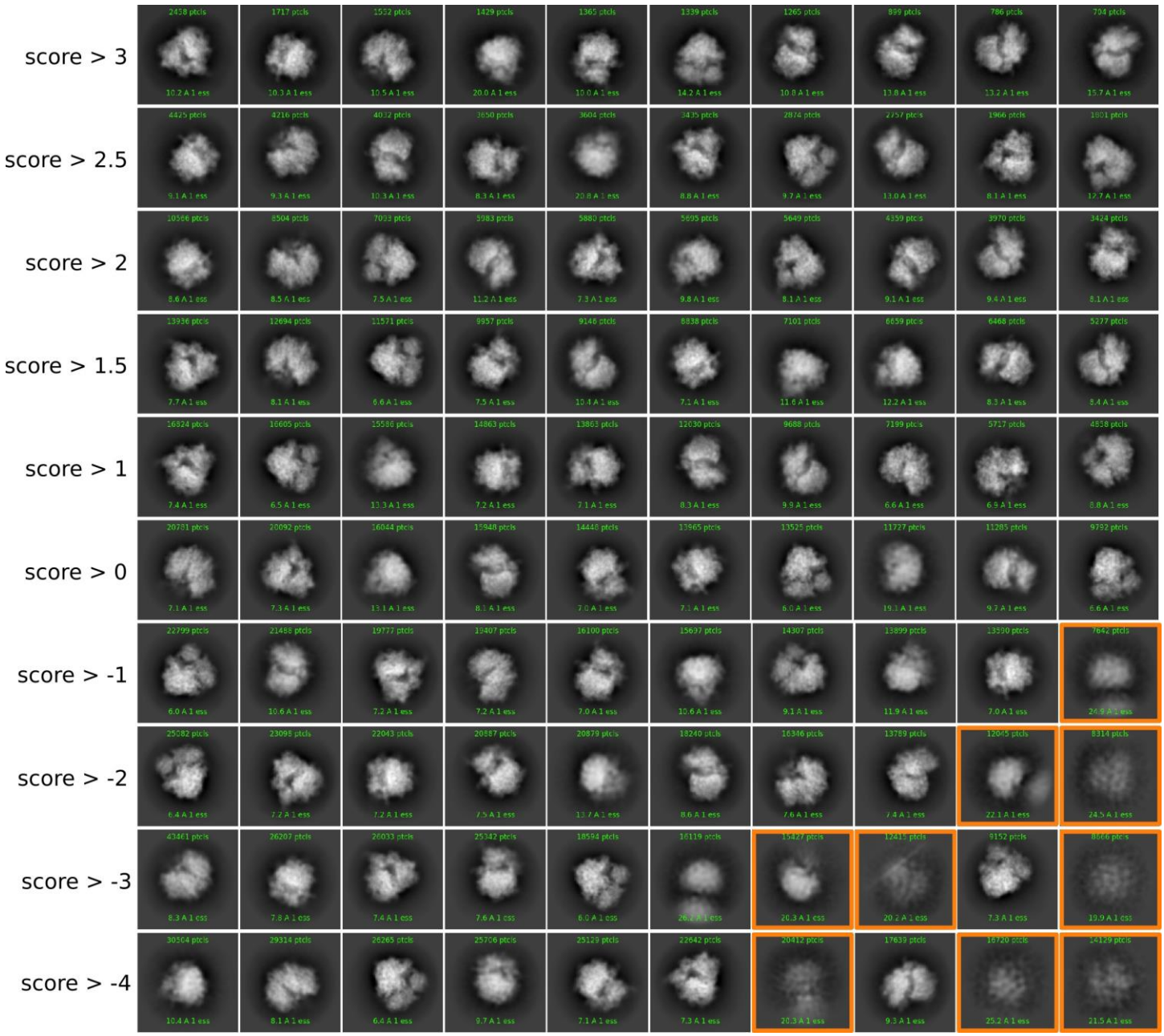
Two example micrographs for the rabbit muscle aldolase dataset (EMPIAR-10215) with **(top)** published particles circled in blue, **(middle)** training particles sampled from the published particles circled in blue, and **(bottom)** Topaz particles circled in red. Each column is a different micrograph. The circled training particles illustrates how sparse the Topaz picks are for this dataset. The PU learning framework allows picking to be performed with high accuracy despite the sparsity of examples, as seen by the Topaz picks in red. Furthermore, Topaz recovers many more real particles than are present in the published set. In this aldolase dataset, the particles are tightly packed, but Topaz correctly identifies and centers the particles. We found this result to be consistent across >100 micrographs examined.



Supplementary Figure 11

EMPIAR-10215 2D class averages.

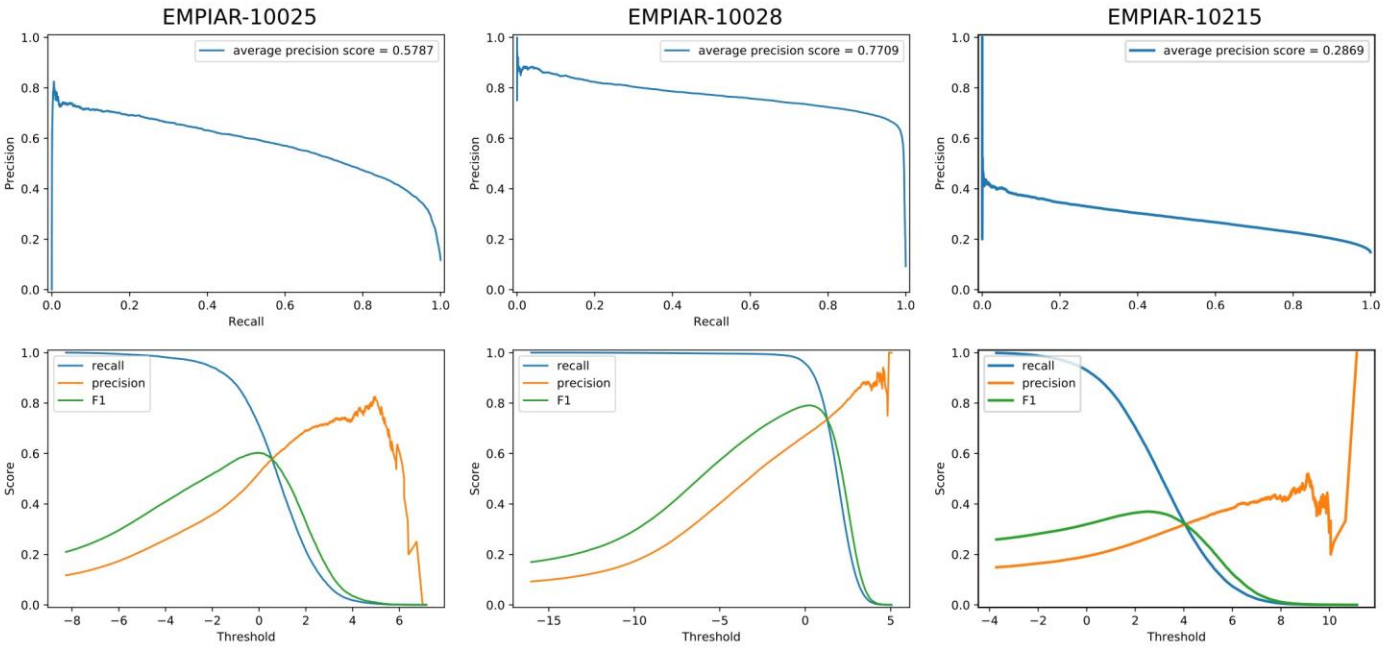
2D class averages of Topaz particles with decreasing score threshold for the aldolase dataset. Classes identified as false positives for quantification in Figure 5 are indicated by orange boxes.



Supplementary Figure 12

EMPIAR-10028 2D class averages.

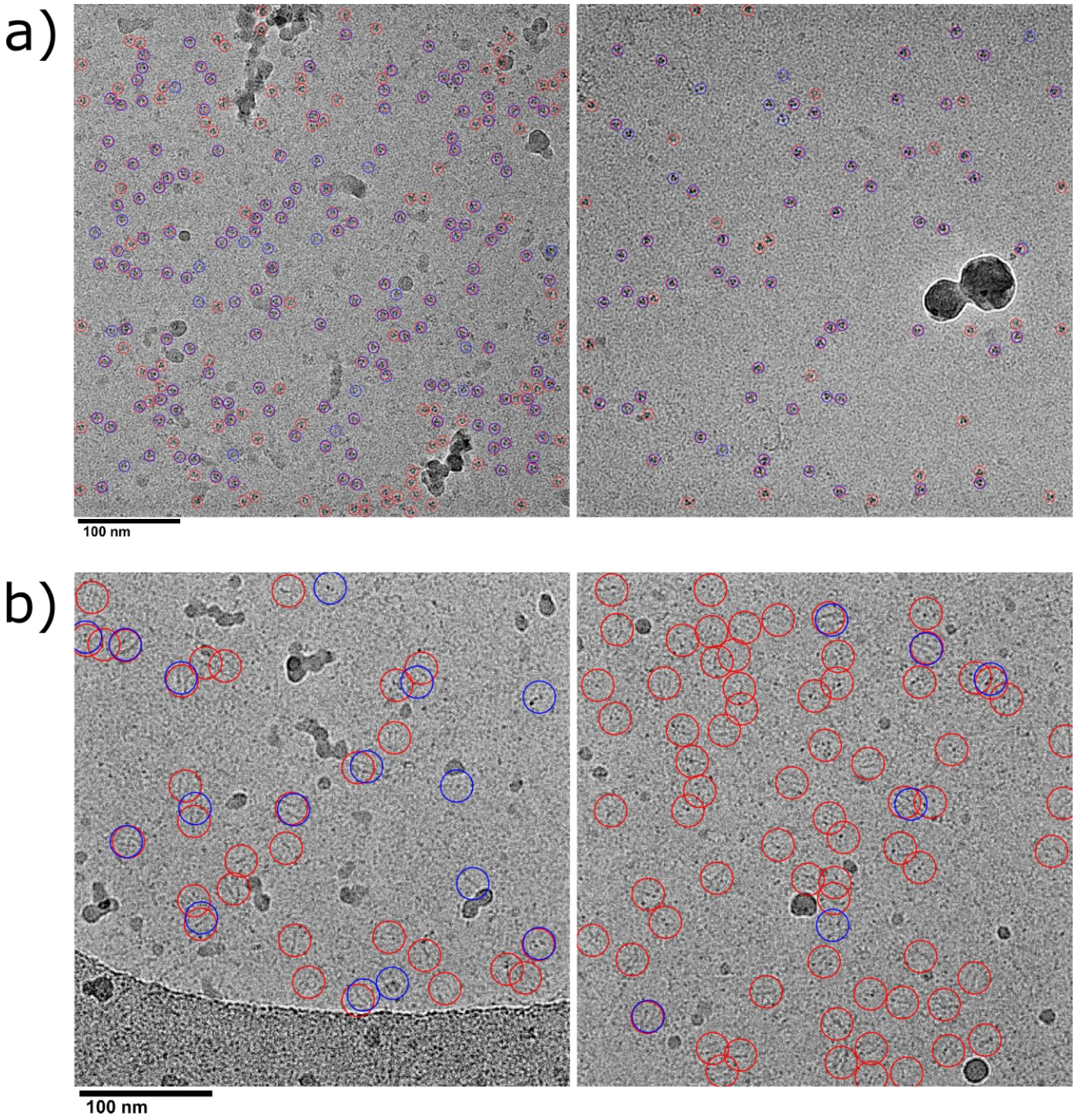
2D class averages of Topaz particles with decreasing score threshold for the 80S ribosome. Classes identified as false positives for quantification in Figure 5 are indicated by orange boxes.



Supplementary Figure 13

Precision-recall and F1 curves for EMPIAR-10025, EMPIAR-10028, and EMPIAR-10215.

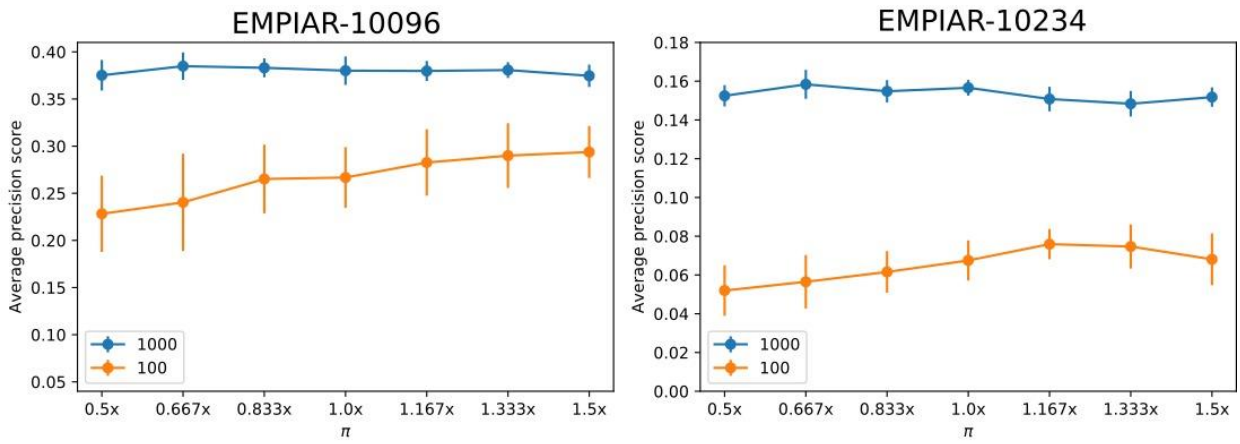
Precision-recall curves and threshold vs precision, recall, and F1 score curves for classifiers trained on the EMPIAR-10025, EMPIAR-10028, and EMPIAR-10215 datasets. Curves were calculated by matching the particles predicted by the Topaz models on the test set micrographs of each dataset to the published particle annotations on those micrographs. We note that the precision and average-precision scores are *underestimates* of the true precision and true average-precision of the Topaz models due to incompleteness of the published particle set.



Supplementary Figure 14

EMPIAR-10096 and EMPIAR-10234 example micrographs with Topaz picks.

Representative micrographs from **(a)** the EMPIAR-10096 and **(b)** the EMPIAR-10234 test sets. For EMPIAR-10096, curated particles from EMPIAR (blue) and particles predicted by Topaz (red) are circled. For EMPIAR-10234, manually selected (blue) and predicted (red) particles are circled. Topaz avoids ice chunks, particles in proximity to the edge of the hole, and particles on carbon and correctly identifies many particles missing from the manually labeled/curated particles sets. We found this result to be consistent over >20 micrographs examined.



Supplementary Figure 15

Sensitivity to π of GE-binomial.

Sensitivity of GE-binomial objective function to the setting of π. For EMPIAR-10096 and the EMPIAR-10234 datasets we report average-precision scores for classifiers trained with 100 and 1,000 labeled particles and values of π varying from 0.5x to 1.5x the values reported in table 1. We report the mean and standard deviation of 10 runs.