

SUPPLEMENTARY METHODS

Preparation of segmentation input channels

Nuclear and cell membrane-associated channels were prepared as inputs for the segmentation methods that use membrane-associated channels for whole-cell segmentation (including Cellpose3, Cellpose4-SAM, Mesmer, and napari-CellSAM), as summarized in **Supplementary Table S10**. For the MIBI datasets, dsDNA was used as the nuclear input, while Na-K-ATPase $\alpha 1$ and HLA class I signals were acquired in a single channel and used as the membrane-associated input. For the Lunaphore COMET dataset, DAPI was used as the nuclear input, while cytokeratin, CD45, CD68, CD3, CD20, and HLA-DR channels were summed into a single membrane-associated channel using the Image Calculator function in FIJI (version 2.14). For methods undergoing dataset-specific fine-tuning, identical channel preparation was applied to the training and testing images. Other phenotypic channels were not used as inputs for model training or segmentation.

Open-source Software Tested

Cellprofiler, version 5b1: Both the native CellProfiler segmentation routine and the StarDist segmentation plugin were executed on all nine images. The settings for primary and secondary object identification are summarized in the table below (**Supplementary Table S11**).

When the StarDist¹⁻³ plugin was applied, it replaced the IdentifyPrimaryObjects module. Within the StarDist model, a probability threshold of 0.5 and an overlap threshold of 0.4 were run on all three datasets. Cell boundaries were detected with the IdentifySecondaryObjects module with the same settings as the CellProfiler native approach and label images were generated with the ConvertObjectsToImage function and saved.

Supplementary Table S11. Settings of cell segmentation via CellProfiler.

Module	Settings	MIBI-Melanoma	MIBI-Breast cancer	Lunaphore-Breast cancer
Primary object (nucleus) identification	Object diameters	17-400 pixels	11-400 pixels	13-400 pixels
	Thresholding method	Adaptive, minimum cross-entropy	Adaptive, minimum cross-entropy	Adaptive, minimum cross-entropy
	Adaptive window size	10 pixels	20 pixels	20 pixels
	Threshold smoothing scale	1.3488 with a correction factor of 1.0	1.3488 with a correction factor of 1.0	1.3488 with a correction factor of 1.0
	Threshold bounds	0.0- 1.0	0.0-1.0	0.0-1.0
	Declumping	Based on intensity and local maxima	Based on intensity and local maxima	Based on intensity and local maxima

		fluorescence intensity with an allowed distance was 8 pixels	smaller than the minimum allowed distance of 9 pixels were suppressed	with distances smaller than 11 pixels were suppressed
Secondary object (the whole cell) identification	Method	Propagation	Propagation	Propagation
	Thresholding method	Adaptive thresholding, minimum cross-entropy	Adaptive thresholding, minimum cross-entropy	adaptive two-class Otsu
	Adaptive window	50 pixels, with a threshold correction factor of 1.0 (0.3-1.0)	100 pixels, correction factor equaled 1.0, threshold range of 0.0-1.0	100 pixels, threshold correction factor was set to 1.0, and had bounds set to 0.0-10
	Regularization factor	0.05	0.05	0.05

Cellpose3 Segmentation and Model Training: Cell segmentation using Cellpose version 3.1.1.2 was performed using its underlying U-net architecture, which predicts spatial gradients derived from simulated diffusion beginning from the center of a ground-truth (GT) cell mask⁴. A topological map of the spatial gradients is formed and vector flows are tracked to the center of the GT cell mask in the X and Y directions. During inference, the U-net model predicts gradient vector flows for each pixel and pixels whose trajectories converge to the same central point are grouped to form individual cell masks.

Cellpose3 featured a new generalist model, Cyto3, trained from nine different image sets, which amounted to ~8,000 images⁵. Training images consisted of brightfield and phase contrast yeast images, phase contrast and fluorescent bacteria images, images from the LiveCell, TissueNet, and cyto2 datasets, plus fluorescent images of nuclei. In addition, Cellpose3 includes a TissueNet segmentation model, trained from a broad dataset consisting of over 1 million paired whole-cell and nuclear annotations in tissue images⁶. Given the similarity between our datasets and the TissueNet training data, both models were included in our evaluation.

To comprehensively assess Cellpose3 performance, three segmentation strategies were executed with Cellpose3: 1) inference using the pretrained Cyto3 model with default parameters; 2) inference using the pretrained TissueNet model with default parameters; 3) dataset-specific models trained from scratch without pretrained weights.

For dataset-specific training from scratch, we applied a leave-one-image-out cross-validation strategy: for each dataset, models were trained with two out of three images, and then the models were tested on the held-out images. Thus, in total, three independent models were trained for each of the three datasets (3 independent models/dataset). Training was initialized from scratch, with the membrane-associated marker input (e.g., Na⁺/K⁺ ATPase or combined CK, CD45, CD68, CD3, CD20, HLA-DR channels, depending on the dataset as indicated in **Supplementary Table S11**) specified as the primary segmentation channel and the nuclear channel as the secondary input. The segmentation channel preparation is the same for training and testing data (**Supplementary Table S10**). The model training was performed for 1,000 epochs with a learning rate of 0.1 and a weight decay of 0.0001. Segmentation performance was evaluated independently for each cross-validation fold, and

the resulting metrics from the three models were averaged to obtain the final performance estimates for each dataset.

To assess the impact of data augmentation, we also trained Cellpose 3 models from scratch using augmented data. These models were trained and tested following the same leave-one-image-out cross-validation strategy. These new models were trained in 1,000 epochs at a learning rate of 0.1 and a weight decay of 0.0001.

Cellpose4-SAM Segmentation Model: In 2025, Cellpose 4.0.4 was released and it introduced a new unified segmentation model, Cellpose4-SAM, which replaced the pretrained models available in earlier versions (e.g., Cyto3 and TissueNet in version 3.1.1.2)⁷ with a modified Segment Anything Model (SAM)⁸. This model is derived from the original Segment Anything Model (SAM), which employs a vision transformer (ViT) to extract information-rich features from input images. Using these features, SAM can generate instance masks based on user-provided prompts, which can either be spatial, such as an object centroid or bounding box, or semantic, such as an object's visual description. User-provided prompts enabled efficient dataset annotation by allowing rapid generation of accurate instance masks with minimal user input.

To adapt SAM for automated cell segmentation, the Cellpose framework integrates the SAM image encoder while replacing the original prompt-based decoder with the Cellpose vector flow-based segmentation head⁷. Specifically, the SAM encoder is used to generate feature embeddings, from which Cellpose predicts spatial gradient (flow) fields that define pixel-wise trajectories toward object centers. These flows are then used to reconstruct individual cell masks, consistent with the Cellpose segmentation paradigm.

For computational efficiency, Cellpose4-SAM modified the SAM encoder by reducing the size of the default image input from 1024×1024 to 256×256 pixels and patch sizes from 16×16 to 8×8 pixels, which improved run-time performance. The combined architecture was trained on 22,826 images with over 3 million training annotations, enabling robust generalization across diverse imaging modalities.

We evaluated the Cellpose4-SAM model on all three datasets. In addition, dataset-specific models were trained using a leave-one-image-out cross-validation strategy, consistent with the approach described for Cellpose3 (using the same segmentation channel preparation strategy as indicated in **Supplementary Table S10**). Essentially, new models were trained on images and tested on the held-out image, resulting in three independently trained models per dataset. A second set of models trained with augmented image data was made and tested as well. Data augmentation was performed via rotations of 90° , 180° , and 270° in both clockwise and counterclockwise directions using FIJI software. Models were trained in 1,000 epochs at a learning rate of 0.1 and a weight decay of 0.0001. Segmentation outputs were generated and compared with the corresponding ground-truth annotations for performance evaluation.

Mesmer: The Mesmer segmentation algorithm was based on a ResNet50 backbone with a feature pyramid network⁶. The model employs two semantic segmentation heads: one predicts pixel-level class probabilities corresponding to cell interior, cell boundary, and background, while the other head predicts cell center probability maps. These outputs are subsequently integrated using a watershed-based post-processing step to generate instance-level cell segmentation masks. The Mesmer algorithm was trained on the TissueNet dataset, which comprises large-scale, annotated tissue imaging data⁶.

For segmentation analysis, 32-bit composite MIBI images consisting of the nucleus (dsDNA antibody) and membrane [Na⁺/K⁺-ATPase/HLA class 1 (HLA-A, HLA-B, and HLA-C) antibody, as indicated in **Supplementary Table S10**

] channels were converted to 16-bit images before processing. Segmentations were performed using the Mesmer deep learning model via the DeepCell Kiosk FIJI plugin (version 0.3.2; <https://github.com/vanvalenlab/kiosk-imageJ-plugin>) using default settings. Similarly, 16-bit Lunaphore COMET images containing nucleus (DAPI) and membrane channels (combination of CK, CD45, CD68, CD3, CD20, HLA-DR channels, as indicated in **Supplementary Table S10**) were processed with the same plugin and settings. Output segmentation label images were acquired and saved for comparison with the matching GT label images.

Napari-CellSAM (version 0.5.6): CellSAM is a generalist foundation model for cell instance segmentation that extends the Segment Anything Model architecture for microscopy applications⁹. The framework incorporated an object detector called CellFinder, based on the Anchor DETR framework, which shares the same vision transformer backbone as SAM. In this pipeline, image features are used to generate bounding boxes, which are then fed into SAM as prompts, enabling the model to generate instance-level segmentation masks via its decoder.

CellSAM was trained on a diverse collection of datasets: TissueNet, DeepBacs, BriFiSeg, Cellpose, Omnipose, YeastNet, YeaZ, the 2018 Kaggle Science Bowl dataset, a collection of hematoxylin-eosin-stained datasets, and Phase400⁹, supporting broad generalizability across imaging modalities.

Segmentation was implemented using a napari CellSAM plugin, in which users specify nucleus and cell membrane channels through the graphical interface. In our analysis, the Na⁺/K⁺-ATPase/HLA class 1 channel (MIBI datasets) or the combined channel of CD45, CD68, CD3, CD20, and HLA-DR staining (Lunaphore-Breast Cancer dataset) was used as the membrane-associated channel, and the nuclear channel was defined by the corresponding nucleus staining (as indicated in **Supplementary Table S10**). Segmentation was executed using the “Segment All” function with default settings, generating instance-level segmentation masks. The resulting label images were converted to region-of-interest (ROI) representations using the BIOP FIJI plugin (“Label Images to ROIs” function). These ROIs were overlaid onto the denoised and contrast-enhanced test images, and segmentation boundaries were compared with GT annotations using the Mask Instant Comparator plugin, as described in the main Methods.

QuPath Segmentation (version 0.6.0): Cell segmentation in QuPath was performed using three approaches: the native cell detection algorithm¹⁰ and two deep learning-based extensions, StarDist¹⁻³ and InstanSeg¹¹. These three approaches represented complementary segmentation paradigms, including rule-based nucleus expansion and deep learning-based instance segmentation. Channel usage was determined by the requirements of each method and reflects their intended design.

Native QuPath Cell Detection: The native QuPath cell detection pipeline is based on nucleus-first segmentation. Nuclei are identified by detecting local intensity maxima, followed by watershed-based segmentation to delineate nuclear boundaries¹⁰. Whole-cell boundaries were subsequently approximated by radially expanding from the nucleus to a user-defined distance (in our case, 5 μm). This boundary extension is independent of fluorescence signal and is constrained by neighboring cells through an additional watershed-based separation step.

For segmentation of the MIBI-Melanoma dataset, parameters included a pixel size of 0.3906 μm , background radius of 8 μm , background subtraction by reconstruction enabled, Gaussian smoothing with $\sigma = 1.8$, nucleus area range of 10–400 μm^2 , threshold of 2.0, and watershed post-processing enabled. For the MIBI-Breast Cancer dataset, parameters were similar, with $\sigma = 2.0$, threshold = 0.1, and all other parameters unchanged. For the Breast Cancer Lunaphore dataset, the same parameters were applied except for the threshold, which was set to 1250.0. Threshold values were adjusted empirically for each dataset to account for differences in signal intensity and imaging modality.

StarDist Segmentation: StarDist is a deep learning-based segmentation method that employs a U-Net-like architecture to represent objects as star-convex polygons¹². The model predicts distances from the object center to its boundary along a fixed set of radial directions (32 rays), along with object probability maps. These predictions are combined to reconstruct polygonal object masks, followed by non-maximum suppression to remove overlapping candidates. QuPath employs two pretrained models for StarDist, one for hematoxylin and eosin-stained images and another for fluorescently labeled nuclei, the latter of which was utilized in this study. Segmentation was performed using the QuPath StarDist extension with the pretrained `dsb2018_heavy_augment` model, designed for 2D fluorescent nucleus segmentation. Parameters included normalization percentiles of 1 and 99, a probability threshold of 0.5, and a pixel size of 0.3906 μm . Whole-cell boundaries were subsequently approximated by radial expansion of 5 μm from the detected nuclei.

InstanSeg Segmentation: InstanSeg is a deep learning-based instance segmentation model that utilizes a modified U-Net backbone (ChannelNet) to cluster pixel embeddings into individual objects^{11,13}. The model generates a seed map with positional and conditional embeddings, from which seed points are identified and used to assign pixels to individual instances. Final segmentation masks are produced by aggregating pixel-wise probability maps corresponding to each instance. The ChannelNet backbone was trained with the TissueNet and CPDMI¹⁴ datasets. Our segmentation was performed using the pretrained `fluorescent_nuclei_and_cells` model (version 0.1.0), with inference conducted on CPU (4

threads), tile size of 256, and padding of 16 pixels. All available image channels were used as input for segmentation.

Post-processing and Export: For all three segmentation methods, detected cell boundaries were exported as ROI files using custom QuPath scripts. ROIs were subsequently imported into FIJI, converted into label images using the BIOP plugin (“ROIs to Label Image” function), and used for downstream comparison with GT annotations.

Equations for Pixel- and Object-based Segmentation Analysis

Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{FP + FN}$
Jaccard Index	$\frac{TP}{TP + FP + FN}$
F1 Score	$\frac{2TP}{2TP + FP + FN}$
Accuracy (pixel only)	$\frac{TP + TN}{TP + TN + FP + FN}$

Mask Instant Comparator Macro

```
// Ask for ground truth image
gtPath = File.openDialog("Select Ground Truth Image");

myDir = File.getParent(gtPath) + File.separator; // store the
folder for later

open(gtPath);

name1 = File.nameWithoutExtension;

run("Duplicate...", "title=GT_Image");


// Ask for test image
testPath = File.openDialog("Select Test Image");

open(testPath);

name2 = File.nameWithoutExtension;

run("Duplicate...", "title=Test_Image");
```

```
// Run comparator
run("Mask instant Comparator",

    "truth_mask_image=GT_Image  test_mask_image=Test_Image  pixel
object_(iou=0.5)  object_(varying_iou)  minimum_iou_threshold=0
maximum_iou_threshold=1  increment_of_iou_threshold=0.01
show_composite_images  show_graphs  show_summary_graph
show_gt_objects_correspondence_table  minimum_size_for_objects=4
minimum_distance_to_border=23");

// Run CLIJ2

run("CLIJ2 Macro Extensions", "cl_device=[NVIDIA GeForce RTX 5090
Laptop GPU]");

image5 = "GT_Image";
Ext.CLIJ2_push(image5);
image6 = "Test_Image";
Ext.CLIJ2_push(image6);
Ext.CLIJ2_sorensenDiceCoefficient(image5, image6);

// Save results in same folder as GT image
saveAs("Tiff", myDir + name2 + "_Plot.tif");
selectImage("display of masks GT_Image__VS__Test_Image");
saveAs("Tiff", myDir + name2 + "_IoUStk.tif");
```

References

- 1 Weigert, M. & Schmidt, U. in *2022 IEEE International Symposium on Biomedical Imaging Challenges (ISBIC)*. 1-4 (IEEE).
- 2 Weigert, M., Schmidt, U., Haase, R., Sugawara, K. & Myers, G. in *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 3655-3662.
- 3 Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II* 265–273 (Springer-Verlag, Granada, Spain, 2018).
- 4 Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods* **18**, 100-106 (2021).
<https://doi.org/10.1038/s41592-020-01018-x>

- 5 Stringer, C. & Pachitariu, M. Cellpose3: one-click image restoration for improved cellular segmentation. *Nat Methods* **22**, 592-599 (2025). <https://doi.org/10.1038/s41592-025-02595-5>
- 6 Greenwald, N. F. *et al.* Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat Biotechnol* **40**, 555-565 (2022). <https://doi.org/10.1038/s41587-021-01094-0>
- 7 Pachitariu, M., Rariden, M. & Stringer, C. Cellpose-SAM: superhuman generalization for cellular segmentation. *BioRxiv*, 2025.2004.2028.651001 (2025).
- 8 Kirillov, A. *et al.* in *Proceedings of the IEEE/CVF international conference on computer vision*. 4015-4026.
- 9 Marks, M. *et al.* CellSAM: a foundation model for cell segmentation. *Nat Methods* **22**, 2585-2593 (2025). <https://doi.org/10.1038/s41592-025-02879-w>
- 10 Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878 (2017). <https://doi.org/10.1038/s41598-017-17204-5>
- 11 Goldsborough, T. *et al.* A novel channel invariant architecture for the segmentation of cells and nuclei in multiplexed images using InstanSeg. *bioRxiv*, 2024.2009.2004.611150 (2024). <https://doi.org/10.1101/2024.09.04.611150>
- 12 Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. in *International conference on medical image computing and computer-assisted intervention*. 265-273 (Springer).
- 13 Thibaut Goldsborough, B. P., Alan O'Callaghan, Fiona Inglis, Leo Leplat, Andrew Filby, Hakan Bilen, Peter Bankhead. InstanSeg: an embedding-based instance segmentation algorithm optimized for accurate, efficient and portable cell segmentation. *arXiv* **2408.15954** (2024). <https://doi.org/https://doi.org/10.48550/arXiv.2408.15954>
- 14 Aleynick, N. *et al.* Cross-platform dataset of multiplex fluorescent cellular object image annotations. *Sci Data* **10**, 193 (2023). <https://doi.org/10.1038/s41597-023-02108-z>