



OPEN Cellpose as a reliable method for single-cell segmentation of autofluorescence microscopy images

Jeremiah M. Riendeau^{1,2}, Amani A. Gillette², Emmanuel Contreras Guzman², Mario Costa Cruz³, Aleksander Kralovec², Shirsa Udgata⁴, Alexa Schmitz⁴, Dustin A. Deming^{4,5,6}, Beth A. Cimini³ & Melissa C. Skala^{1,2}✉

Autofluorescence microscopy uses intrinsic sources of molecular contrast to provide cellular-level information without extrinsic labels. However, traditional cell segmentation tools are often optimized for high signal-to-noise ratio (SNR) images, such as fluorescently labeled cells, and unsurprisingly perform poorly on low SNR autofluorescence images. Therefore, new cell segmentation tools are needed for autofluorescence microscopy. Cellpose is a deep learning network that is generalizable across diverse cell microscopy images and automatically segments single cells to improve throughput and reduce inter-human biases. This study aims to validate Cellpose for autofluorescence imaging, specifically using multiphoton intensity images of NAD(P)H. Manually segmented nuclear masks of NAD(P)H images were used to train a new autofluorescence-trained model (ATM) in Cellpose for nuclear segmentation of NAD(P)H intensity images. These models were applied to PANC-1 cells treated with metabolic inhibitors and patient-derived cancer organoids (9 patients) treated with chemotherapies. These datasets include co-registered fluorescence lifetime imaging microscopy (FLIM) of NAD(P)H and FAD, so fluorescence decay parameters and the optical redox ratio (ORR) were compared between masks generated by the new ATM and manual segmentation. The Dice score between repeated manually segmented masks was significantly lower than that of repeated ATM masks ($p < 0.0001$) indicating greater reproducibility between ATM masks. There was also a high correlation ($R^2 > 0.9$) between ATM and manually segmented masks for the ORR, mean NAD(P)H lifetime, and mean FAD lifetime across 2D and 3D cell culture treatment conditions. Masks generated from ATM and manual segmentation also maintain similar means, variances, and effect sizes between treatments for the ORR and FLIM parameters. Overall, the Cellpose ATM provides a fast, reliable, reproducible, and accurate method to segment single cells in autofluorescence microscopy images such that functional changes in cells are accurately captured in both 2D and 3D culture.

Keywords Autofluorescence, Segmentation, Single-cell, Organoid, Deep-learning

Autofluorescence provides an innate source of molecular contrast that is attractive for imaging cells within intact, unaltered systems. Intrinsic sources of contrast such as autofluorescence also avoid the effects of fluorescent labels and genetic modifications, which may interfere with cell function^{1,2}. Therefore, autofluorescence microscopy continues to be used to study in vivo and in vitro models across numerous applications including cancer research, stem cell biology, and metabolism research³⁻⁶. Autofluorescence signals from reduced nicotinamide adenine dinucleotide or the phosphorylated counterpart, together referred to as NAD(P)H, are often imaged to track changes in cellular metabolism^{7,8}. While NAD(P)H exchange does occur between cell compartments, the nucleus is known to maintain a distinct pool of NAD(P)H that does not play a direct role in nutrient metabolism^{9,10}. The cytoplasm typically exhibits stronger NAD(P)H signal than the nucleus due to the higher

¹Department of Biomedical Engineering, University of Wisconsin, Madison, WI, USA. ²Morgridge Institute for Research, Madison, WI, USA. ³Broad Institute of Harvard and MIT, Imaging Platform, Cambridge, MA, USA. ⁴Division of Hematology, Medical Oncology and Palliative Care, Department of Medicine, School of Medicine and Public Health, University of Wisconsin, Madison, WI, USA. ⁵McArdle Laboratory for Cancer Research, Department of Oncology, University of Wisconsin, Madison, WI, USA. ⁶University of Wisconsin Carbone Cancer Center, Madison, WI, USA. ✉email: mcskala@wisc.edu

abundance of NAD(P)H involved in cellular metabolism. This greater NAD(P)H signal within the cytoplasm provides an opportunity to segment out the nucleus and investigate metabolically related NAD(P)H signals without the potential interference of nuclear NAD(P)H. However, autofluorescence microscopy suffers from relatively low signal-to-noise ratios (SNR)⁷ that makes nuclear segmentation challenging, especially within 3D samples.

Software such as CellProfiler, Fiji/ImageJ, and BioImageXD have been developed for cell segmentation in microscopy, though many such algorithms have been optimized for images that use extrinsic labels with high SNR^{11–14}. However, in cases where SNR is low such as in autofluorescence microscopy, segmentation methods including watershed and thresholding can be unreliable. Therefore, a new set of segmentation tools are needed for low SNR applications.

Recent advances in deep learning-based segmentation models have increased the speed, accuracy, and efficiency of segmentation with limited need for supervision. Cellpose is one such deep learning-based cell segmentation algorithm that uses training sets of manually segmented images to learn segmentation patterns for new, unseen images¹⁵. This automated method minimizes personnel time and human-to-human variances in cell segmentation, thereby increasing the throughput and reliability of image analysis. As autofluorescence microscopy becomes more widely adopted, it is important to develop and validate automated methods, such as Cellpose, to mitigate human biases, improve agreement and reproducibility, and increase the throughput of single cell analysis.

Given that convolutional neural networks are traditionally trained on images with clear staining, there is need to validate these approaches on more challenging data sets such as those acquired from autofluorescence microscopy. To achieve this, we trained a new Cellpose model on a selection of manually segmented nuclei from autofluorescence intensity images of the metabolic co-enzymes NAD(P)H. We then examined the reproducibility of the autofluorescence-trained model (ATM) compared to manual segmentation on new sets of images. We further compared changes in functional autofluorescence variables for manual and ATM segmented images, including fluorescence lifetime imaging microscopy (FLIM) of NAD(P)H and flavin adenine dinucleotide (FAD), another autofluorescent metabolic co-enzyme. Both NAD(P)H and FAD have known distinct lifetimes in the free and protein-bound states, therefore FLIM provides insight into protein-binding activities for these molecules, and as a result, the metabolic activity of the cells^{16,17}. In addition to quantifying the Dice score, which provides a measurement of mask similarity, direct comparisons of endpoint variables between segmentation methods ensures that the conclusions from each method are comparable. Here, we assessed the mean fluorescence lifetimes of NAD(P)H and FAD as well as the optical redox ratio (fluorescence intensities of NAD(P)H / [NAD(P)H + FAD]), which reflects changes in redox status within cells^{3,18–20}. FLIM is performed with two-photon (2P) microscopy to enable optical sectioning at deeper imaging depths in 3D samples compared to single-photon excitation²¹. We sought to investigate the accuracy of ATM segmentation performance on NAD(P)H intensity images in both two-dimensional (2D) cell lines and three-dimensional (3D) patient-derived cancer organoids (PDCOs) with varying morphologies. Through this study, we aim to provide a deeper understanding of the strengths and limitations of Cellpose for single-cell segmentation and how this ATM compares to manual segmentation of autofluorescence microscopy images.

Results

Similarity of ATM and manual masks

Representative NAD(P)H intensity images of 2D PANC-1 cells overlain with their respective masks indicate strong consensus between manual and ATM masks (Fig. 1a). To measure differences between manual and ATM masks, the Dice similarity coefficient, or Dice score was used. This metric measures the amount of overlap between two sets of masks by multiplying the area of overlap of the two masks by two, then dividing by the total masked area of both images, thus penalizing false positives (Fig. 1b). When the manually produced masks (M1) are compared with the ATM segmented masks (ATM1), an average Dice score of 0.74 is found for the M1-ATM1 comparison (Fig. 1c). To further categorize the reproducibility of masks, images were rotated by 90 degrees and segmented again (M2 and ATM2) resulting in an average Dice score of 0.78 for the manually segmented initial orientation and rotated images, M1-M2 (Fig. 1c). This is in agreement with prior work, where masks produced by different humans achieved a two-observer Dice score of 0.76 on a 0–1 scale²². Importantly, the ATM achieved a higher average Dice score of 0.87 for the automatic segmentation of initial orientation and rotated images, ATM1-ATM2 (Fig. 1c), indicating greater reproducibility between ATM segmentations compared to manual segmentations twice by the same observer.

While 2D cells are relatively homogenous in shape and size across treatment conditions, with easily identifiable cell boundaries, PDCOs lack clear cell boundaries which makes single cell identification difficult. Manually segmented nuclear masks in PDCO images show greater agreement with the ATM masks compared to the generalist cyto3 masks (Supp. Figure 1a-b), likely due to the relatively low SNR between the cytoplasm and nucleus (Supp. Figure 1c). Four distinct morphologies were found in the 3D PDCOs, which were characterized by their shape as either circular or irregular, and their contents as either densely filled with cells or void of cells. Variability in PDCO morphologies are well documented and can arise through a variety of mechanisms ranging from the genetic diversity in the sample to the randomness of the self-organization process as well as their developmental stage^{23–25}. These morphological variances can be clearly seen in 2P NAD(P)H intensity images (Fig. 2a). Combining all PDCO morphologies demonstrates similar reproducibility for manual vs. ATM segmentation (M1-ATM1 Dice score = 0.68) and manual vs. manual segmentation (M1-M2 Dice score = 0.66), with greater reproducibility between ATM masks (ATM1-ATM2 Dice score = 0.92) (Fig. 2b). Therefore, in these 3D PDCOs, the ATM masks are more reproducible than manual masks, which is consistent with the findings in 2D cells and with prior studies^{26,27} (Fig. 1).

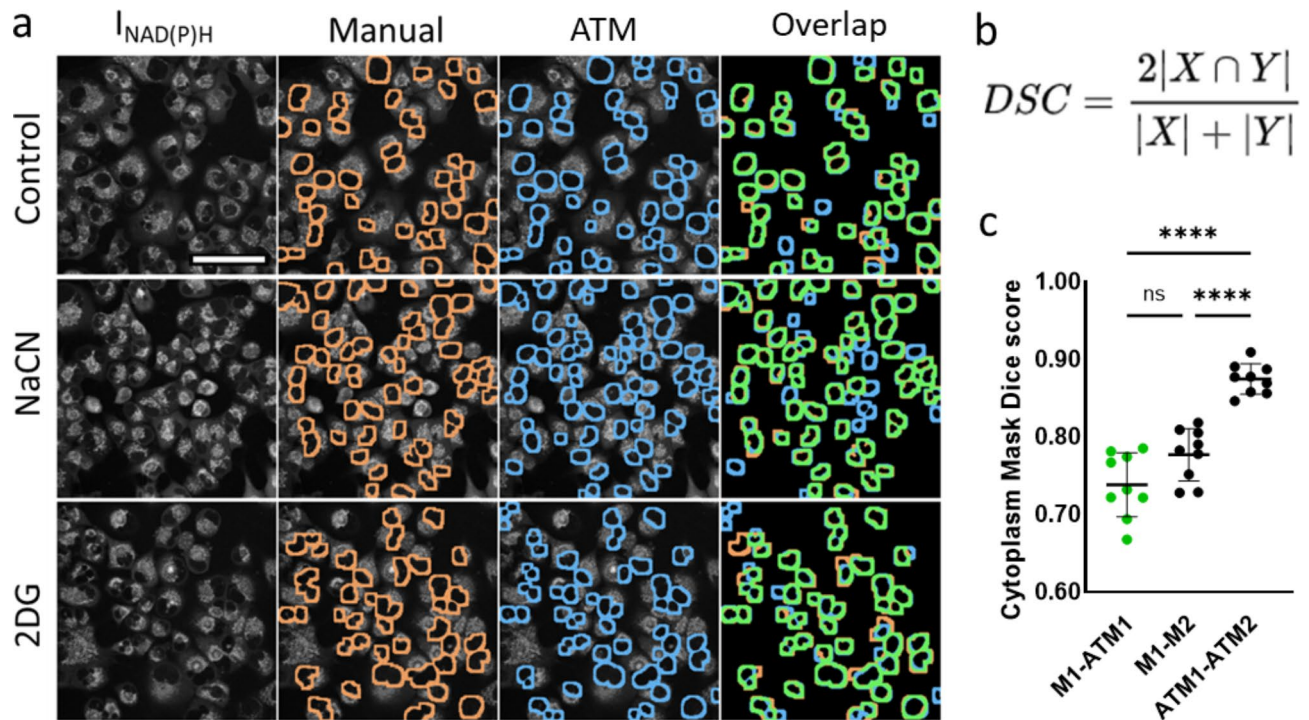


Fig. 1. ATM masks agree with manual masks while achieving greater reproducibility in 2D culture. **(a)** Representative two-photon microscopy images of NAD(P)H intensity (left) of PANC-1 cells with or without treatment with inhibitors of oxidative phosphorylation (NaCN 4mM; 30 min) or glycolysis (2DG 10mM; 120 min). Masks are overlain for manual segmentation (orange) and ATM segmentation (blue) where more overlap (green) results in a higher dice score. Cytoplasm area is defined as 8 μm propagated from the nuclear mask. Scale bar = 100 μm . **(b)** Dice Similarity Coefficient (DSC), or Dice Score, is calculated by multiplying the green pixel area by two, then dividing that value by the sum of the pixel areas of the blue and orange regions. **(c)** Dice scores compare Manual [M] and ATM generated masks where M2 and ATM2 images have been rotated 90° with respect to M1 and ATM1 images such that M1-M2 compares manual reproducibility and ATM1-ATM2 compares ATM reproducibility. Each dot represents one image. Center line is mean; error bars are standard deviation. Control $n = 3$ images, 146 nuclei (manual), 160 nuclei (ATM); NaCN $n = 3$ images, 143 nuclei (manual), 217 nuclei (ATM); 2DG $n = 3$ images, 143 nuclei (manual), 173 nuclei (ATM). **** $P < 0.0001$, ns = $P > 0.05$, Ordinary one-way ANOVA.

All PDCO images were acquired from a previously published dataset in Kratz et al.²⁸ Due to the varying morphologies and spatial arrangements of the PDCOs, the performance of the ATM between different morphologies was explored (Fig. 2c). Dice scores comparing the manual vs. ATM masks (M1-ATM1) found that three of the morphologies show similar agreement (Dice score = 0.54–0.58) while the circular void morphology, which contains fewer cells on average, has a slightly higher agreement (Dice score = 0.62). These varying PDCO morphologies can be found in many cancer types including breast and pancreatic cancers, providing a robust and varied testing set for the generalizability of this ATM across different cell shapes, sizes, and environments.^{29,30}

In addition to providing similar segmentation to manual segmentation, the ATM segments images several orders of magnitude faster (Supp. Figure 2). While our expert will spend roughly 11 min to segment 100 nuclei, the ATM in Cellpose can perform the same number of nuclei segmentations in less than a second (Supp. Figure 2). It is worth noting that these results were only compiled for one human and segmentation time will vary from person to person, while automation time depends on hardware and software systems. However, these results clearly indicate that the speed and reproducibility of automated deep-learning segmentation is unachievable by humans as demonstrated in numerous prior studies.^{15,31,32}

Metabolic parameters are preserved using ATM segmentation

While masks may be similar for manual and ATM segmentation, it is unclear whether changes in endpoint variables are similar for manual and ATM segmentation methods. Differences between the ATM and manual segmentation were investigated with treatment in 2D and 3D cultures using autofluorescence FLIM parameters as treatment response measurements, where a decrease in ORR and NAD(P)H mean lifetime (τ_m) indicates treatment response³³. Autofluorescence FLIM parameters include ORR, NAD(P)H τ_m , and FAD τ_m . PANC-1 cells were treated with either 4 mM sodium cyanide (NaCN) for 30 min or 10 mM 2-deoxy-glucose (2DG) for 2 h. These two metabolic inhibitors have known effects on autofluorescence FLIM parameters which we replicate in our 2D data (Fig. 3a–g)^{34–36}. Comparison between the values produced from the masks created by the ATM and those created manually indicates strong agreement in autofluorescence FLIM parameter changes

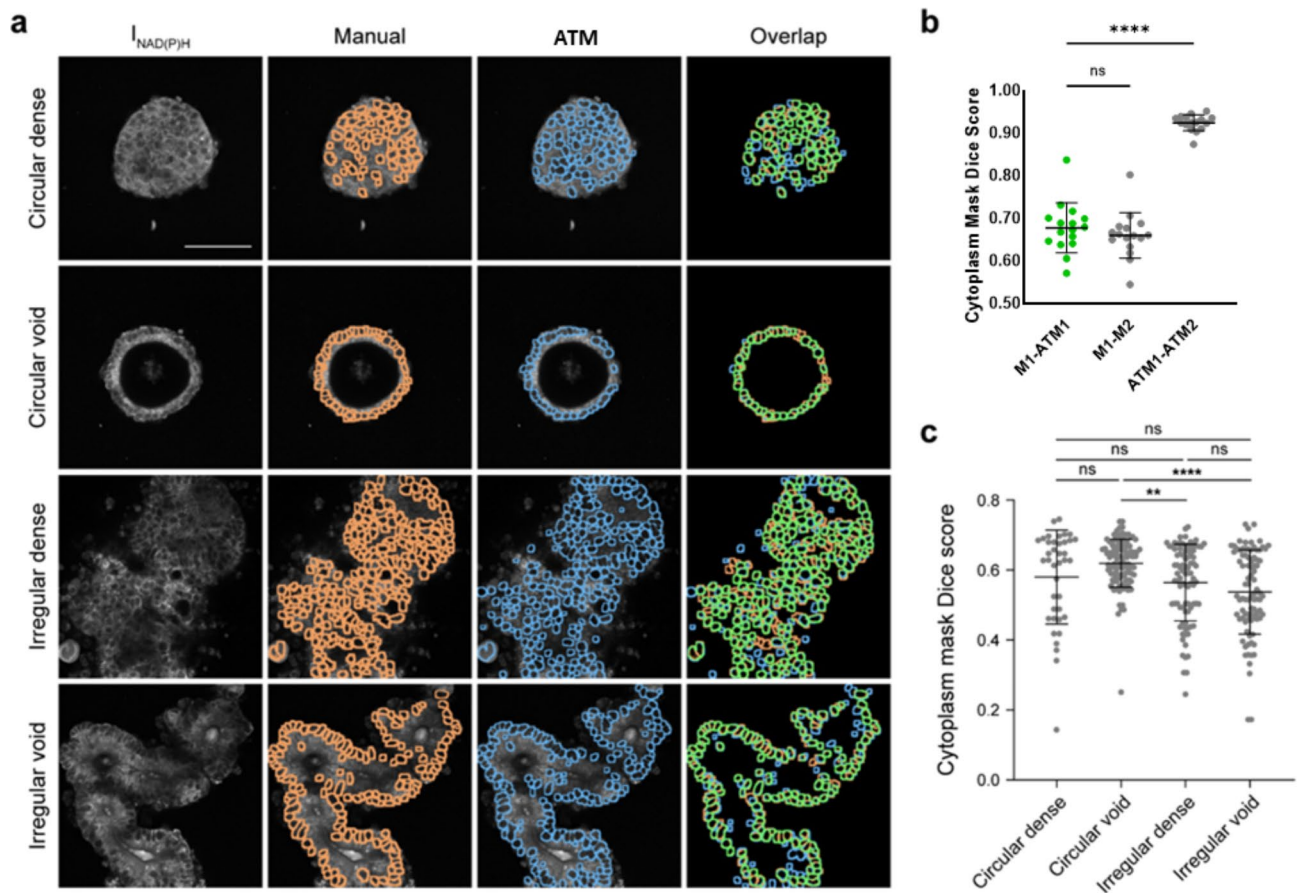


Fig. 2. ATM masks agree with manual masks while achieving greater reproducibility in optical cross sections of 3D PDCOs. **(a)** Representative NAD(P)H intensity images of rectal PDCOs (images acquired from Kratz et al.²⁸) with morphologies defined as circular or irregular and dense or void. Masks are overlain for manual segmentation (orange) and ATM segmentation (blue) where more overlap (green) results in a higher dice score. Cytoplasm area is defined as 8 μm propagated from the nuclear mask. Scale bar = 100 μm . **(b)** $n = 16$ organoids representative of the irregular dense morphology, chosen for their complexity. Dice scores compare manual [M] and ATM segmented masks where M2 and ATM2 images have been rotated 90°. ATM segmented masks display more similarity when the image is rotated and segmented again [ATM1-ATM2] than when a human performs the same task [M1-M2]. **(c)** Comparison of ATM segmentation and manual segmentation of PDCOs across PDCO morphologies. The two-observer Dice score measurement is 0.76 (see M1-ATM1 and M1-M2), similar to prior inter-observer Dice scores²². $n = 28$ circular dense images, 2973 nuclei (manual), 2544 nuclei (ATM); 81 circular void images, 7868 nuclei (manual), 7294 nuclei (ATM); 70 irregular dense images, 9318 nuclei (manual), 9048 nuclei (ATM); and 65 irregular void images, 7277 nuclei (manual), 8334 nuclei (ATM). Center line is mean and error bars are standard deviation. **** $P < 0.0001$, ns = $P > 0.05$, Ordinary one-way ANOVA.

(Fig. 3b-d). Outputs from ATM and manual segmentation are similar with an R^2 value of 0.984 for NAD(P)H τ_m , 0.989 for FAD τ_m , and 0.964 for ORR. Aside from maintaining near-identical values, the ATM also retains similar single-cell distributions (Fig. 3e-g) while maintaining the same effect sizes between treatment groups. This suggests that in 2D cultures, the ATM performs exceptionally well.

Autofluorescence FLIM is also a useful tool for measuring treatment response in 3D PDCO cultures, where cell borders become blurred, and cells become more irregular as they compact into dense PDCO structures (Fig. 4a)^{28,34}. Autofluorescence FLIM parameters across multiple treatments remain similar for ATM and manual segmentation of PDCOs (Fig. 4b-d) with an R^2 value of 0.983 for NAD(P)H τ_m , 0.981 for FAD τ_m , and 0.902 for ORR and appear to maintain similar agreement across PDCO morphological categories (Supp. Figure 3). Autofluorescence FLIM parameters are also plotted by treatment condition in PDCOs (Fig. 4e-j), which indicates that the ATM retains similar single-cell distributions to manually segmented images. Responses for individual patients with respect to treatment condition can be found in supplemental materials (Supp. Figure 4). Autofluorescence FLIM parameters for PDCOs segmented with the ATM and manual segmentation also maintain similar levels of effect size between treatment groups despite differences between the number of nuclei segmented manually or by the ATM, demonstrating a high level of performance even in 3D PDCOs (Fig. 4e-j, Supp. Table 1).

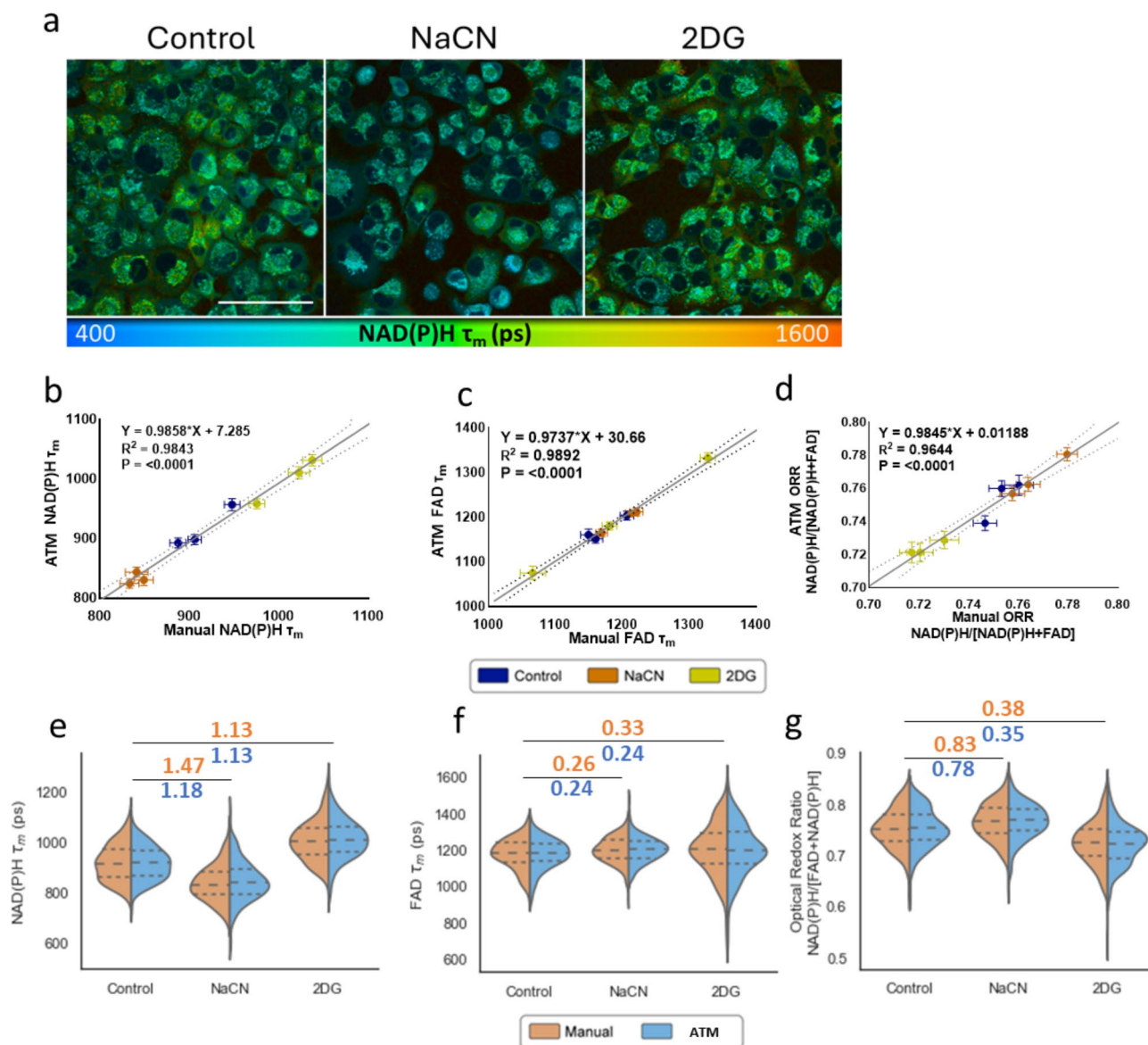
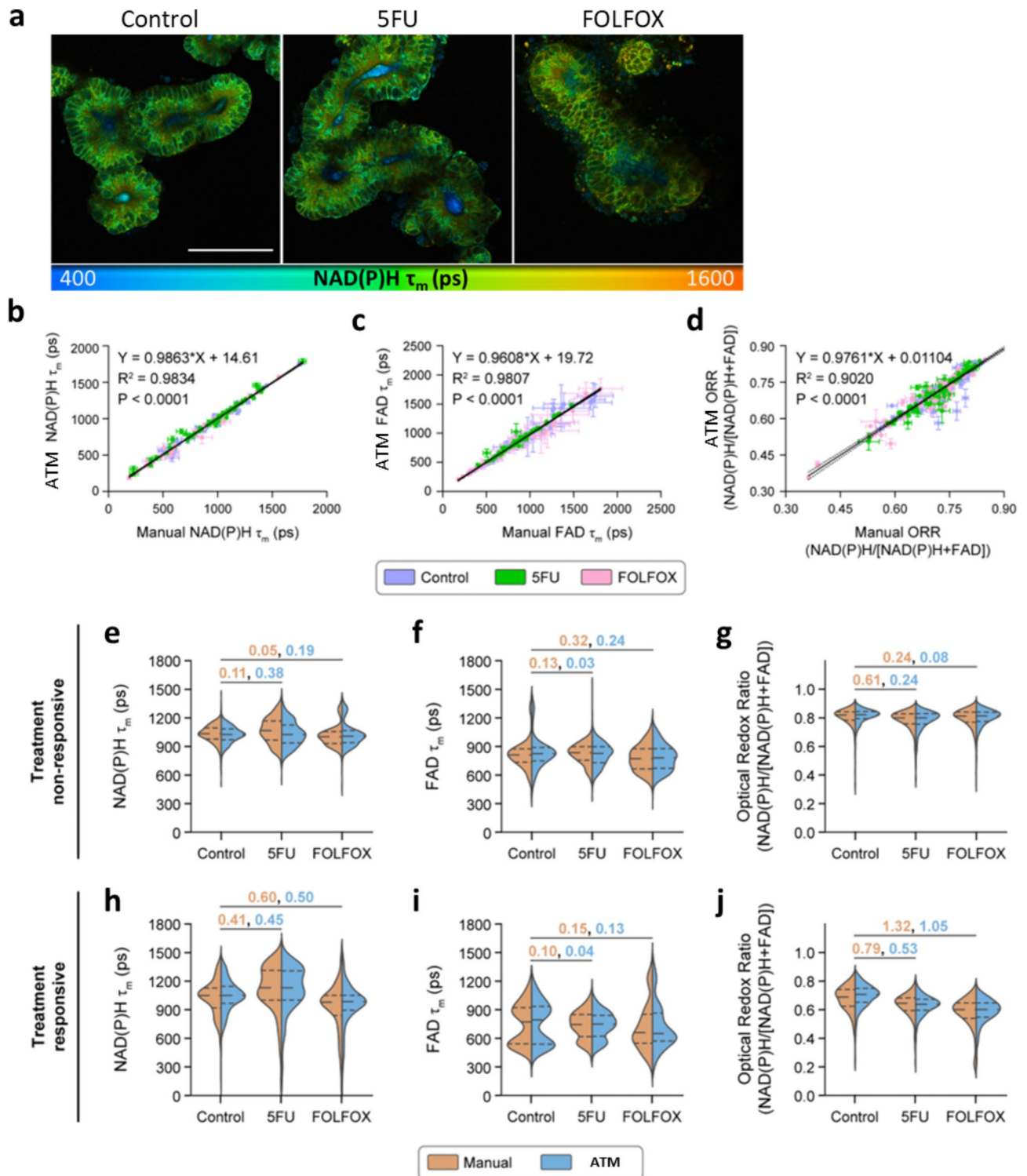


Fig. 3. Autofluorescence FLIM parameters from ATM and manual generated cytoplasm masks are highly correlated for 2D cultured PANC-1 cells under 3 treatment conditions. PANC-1 cells were left untreated or treated with an inhibitor of oxidative phosphorylation (NaCN 4 mM; 30 min) or glycolysis (2DG 10 mM; 120 min). **(a)** Representative NAD(P)H mean lifetime images. Scale bar = 100 μ m. **(b–d)** Autofluorescence FLIM parameters generated from manually segmented images and ATM segmented images are highly correlated ($R^2 > 0.964$). Each dot represents one image, colored by treatment condition, where error bars represent standard error of the mean. Solid black diagonal line is linear regression and dashed lines represent 95% confidence bands. P value determined by ratio T test of Spearman correlation values. **(e–g)** Glass's Delta effect sizes between control and treatment groups are maintained with both ATM segmentation and manual segmentation. ATM segmentation retains cell distributions observed with manual segmentation. Dashed lines separate quartiles. Control $n = 3$ images, 146 nuclei (manual), 160 nuclei (ATM); NaCN $n = 3$ images, 143 nuclei (manual), 217 nuclei (ATM); 2DG $n = 3$ images, 143 nuclei (manual), 173 nuclei (ATM). Orange numbers are effect size between manually segmented cells and blue numbers are effect size between ATM segmented cells. No statistical difference is measured between the manual and ATM segmented groups for the same condition (side-by-side orange and blue violin plots).

Discussion

Autofluorescence microscopy is an attractive tool due to its single cell resolution, sensitivity to cellular function, and nondestructive nature. Due to these advantages, autofluorescence microscopy is a popular method to characterize cell heterogeneity, so it is important to standardize single cell segmentation methods to improve reproducibility and reduce human variability and error. However, cells can be difficult to segment in autofluorescence images where SNR can be low. In such cases, cell borders become less clear, resulting in



increased processing times. Fortunately, recent advancements in deep learning have created a host of new tools for single-cell segmentation including Ilastik, Stardist, and Cellpose^{15,37–39}. These technologies are easy to train, require minimal adjustments, and promise to automate the task of single-cell segmentation to minimize human error and improve the speed of autofluorescence image analysis.

Commonly, tools for biological image analysis are trained on high contrast images of stained cells, which work well with typical methods of thresholding and edge-detection. In cases of low contrast and low SNR like autofluorescence imaging of molecules with low quantum yields, these traditional methods are often lacking, require significant amounts of fine tuning, and cannot be used broadly across similar types of images. Metabolic studies based on NAD(P)H autofluorescence can also be confounded by signal in the nucleus that is related to transcriptional activity rather than nutrient metabolism, so nuclear segmentation is preferred (Supp. Figure 5)^{9,40}. This presents additional challenges due the relatively small intensity differences between the cytoplasm and the nucleus, especially in PDCOs (Supp. Figure 1c). Though our study chose to segment solely on the intensity

◀ **Fig. 4.** Autofluorescence FLIM parameters from ATM and manual segmentation are highly correlated for all 3D colorectal PDCOs across 4 treatment conditions. PDCOs were left untreated or treated with 10 $\mu\text{mol/L}$ of 5FU, or FOLFOX (10 $\mu\text{mol/L}$ of 5FU and 40 mmol/L of oxaliplatin) for 4 days (images acquired from Kratz et al.²⁸). **(a)** Representative NAD(P)H mean lifetime images. Scale bar = 100 μm . **(b–d)** Autofluorescence FLIM parameters generated from manually segmented images and ATM segmented images are highly correlated ($R > 0.90$). Each dot represents one organoid image, where error bars represent standard error of the mean. Black diagonal line is linear regression and dashed lines represent 95% confidence bands. P value determined by ratio T test of Spearman correlation values. **(e–j)** Significant differences between control and treatment groups are maintained with both ATM segmentation and manual segmentation of **(e–g)** PDCOs that do not respond to treatment or **(h–j)** PDCOs that are treatment responsive. ATM segmentation retains cell distributions observed within manual segmentation, where dashed lines separate quartiles. Nonresponsive PDCOs (patient 6) Control $n = 8$ images, 797 nuclei (manual), 979 nuclei (ATM); 5FU $n = 8$ images, 983 nuclei (manual), 1009 nuclei (ATM); FOLFOX $n = 7$ images, 730 nuclei (manual), 816 nuclei (ATM). Responsive PDCOs (patient 3) Control $n = 9$ images, 672 nuclei (manual), 1161 nuclei (ATM); 5FU $n = 9$ images, 635 nuclei (manual), 1042 nuclei (ATM); FOLFOX $n = 9$ images, 611 nuclei (manual), 808 nuclei (ATM). Orange numbers are effect size between manually segmented cells and blue numbers are effect size between ATM segmented cells. No statistical difference is measured between the manual and ATM segmented groups for the same condition (side-by-side orange and blue violin plots).

of these images due to the decreased NAD(P)H intensity of the nucleus compared to the cytoplasm, future developments in autofluorescence image segmentation may take advantage of the NAD(P)H lifetime differences between the nucleus and cytoplasm (Supp. Figure 5) to further improve segmentation accuracy^{7,9}.

Cellpose, through its generalist approach, aims to make these models more broadly applicable to a vast array of diverse data sets¹⁵. While Cellpose is just one algorithm in a crowded landscape of new deep learning tools for image analysis, we can presume that other methods would perform similarly, though Cellpose has made significant effort to ensure that training models is simple and accessible^{41,42}. Recently, Cellpose has demonstrated segmentation of three-dimensional objects, demonstrating continued effort to broaden its application in biomedical research^{43,44}. In its latest iteration, Cellpose 3 has been refined to segment noisy, blurred, and under sampled images, all of which are common in autofluorescence images⁴⁵. However, the base model, cyto3, did not perform well on NAD(P)H intensity images of PDCOs (Supp Fig. 1). Therefore, we trained a new Cellpose model, referred to as ATM, on a small set of manually segmented NAD(P)H autofluorescence images of PDCOs.

The capacity of the ATM to produce comparable masks to human segmentation while significantly reducing the segmentation time demonstrates its ability to automate single-cell segmentation in autofluorescence microscopy. While the Dice Scores may appear low where one would expect a value of 1.0 for a perfect match, a value of 0.76 was previously shown to represent a high level of agreement between two independent human observers²². Therefore, in terms of the Dice score, the ATM appears to perform as well as another human observer (Figs. 1 and 2). In terms of speed, the ATM also greatly outperforms manual segmentation as expected. While manual observers can easily become fatigued segmenting cells and require roughly 11 min to segment only 100 regions, Cellpose can segment cells substantially faster (Supp. Figure 2). Altogether, the speed and reproducibility of ATM cell masks greatly reduce the time needed to segment autofluorescence microscopy images.

Importantly, R^2 values for the autofluorescence FLIM parameters of both 2D and 3D cultures are high, indicating good agreement between manual and ATM segmentation for quantitative analysis of autofluorescence FLIM images (Figs. 3 and 4). While R^2 values for the ATM vs. manual segmentation were higher for 2D culture compared to 3D PDCOs, this is likely due to the increased presence of irregular cell shapes and blurred boundaries between cells in 3D culture. However, despite the more challenging morphologies associated with 3D organoid culture, the ATM still retains a similar data distribution as manual masks for each of the three autofluorescence FLIM parameters. While masking of 2D cells was performed across only 9 images, a large selection of images was tested for PDCOs which provided a large sample size for statistically robust results.

While the ATM provides a high throughput method for single-cell segmentation, it is still important to verify mask accuracy as there is potential for patient-specific differences that may reduce the accuracy of the single-cell masks. As no human-in-the-loop step was incorporated in the ATM, some inconsistencies were found in images such as occasional segmentation of the necrotic core and over-segmentation leading to large multi-cell “cells”. Additionally, the ATM commonly identifies more nuclei than manual segmentation, which could be attributed to human fatigue. Overall, it is likely the relatively small number of errors are overcome by the large volume of accurate masks. In addition to incorporating a manual check, further refining the ATM may also increase confidence in single-cell segmentation⁴⁶.

Here, we find that decreasing cell segmentation time through use of a trained Cellpose model for autofluorescence microscopy can increase the throughput of image analysis with little impact on accuracy. This could enable greater sample sizes in autofluorescence microscopy studies beyond FLIM of NAD(P)H and FAD, including fats like lipofuscin, structural components such as collagen, and commonly occurring pigments like chlorophyll for plant studies. This study provides confidence that trained Cellpose models can reliably segment cell nuclei with high accuracy in autofluorescence images of both 2D and 3D cell systems. Overall, we demonstrated that single-cell distributions of functional variables are highly conserved between manual and ATM segmentation methods for autofluorescence microscopy, which provides high confidence in deep learning segmentation methods for low SNR applications such as autofluorescence microscopy.

Methods

Cell culture

2D cell line. PANC-1 cells were obtained from the American Type Culture Collection (ATCC) and cultured in Dulbecco's minimum essential medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS) and 1% penicillin-streptomycin (Fisher Scientific). For 2P FLIM, 1.8×10^5 cells were seeded 48 h prior to imaging in 35 mm glass-bottom dishes (MatTek). The cells were treated with solutions in Dulbecco's phosphate-buffered saline (DPBS) of either NaCN for 30 min or 2DG for 2 h prior to imaging, resulting in final concentrations of 4 mM NaCN or 10 mM 2DG. Cells were imaged at three different locations in each dish for an average of 150 cells per treatment group.

3D PDCOs. All PDCO images were acquired from a previously published dataset in Kratz et al.²⁸ All studies were completed following Institutional Review Board (IRB) approval with informed consent obtained from subjects through the University of Wisconsin (UW) Molecular Tumor Board Registry (UW IRB#2014–1370) or UW Translational Science BioCore (UW IRB#2016–0934). Cell isolation and organotypic culture techniques were performed as described previously by Kratz et al.^{28,47}. Briefly, tissue samples were obtained through endoscopic biopsy or primary surgical resection and treated with chelation buffer. Digestion was performed in DMEM stock with collagenase and dispase. The samples were resuspended in Advanced DMEM/F-12 stock after which organoid suspensions were mixed with Matrigel matrix and plated. Plated cultures on 35 mm glass-bottom dishes (MatTek) were overlaid with 450 μ L of feeding medium supplemented with tissue dependent components⁴⁷. PDCO cultures were incubated at 37 °C in 5% CO₂ with medium changes every 48–72 h and treatment with 10 μ mol/L of 5-fluorouracil (5FU), 10 μ mol/L 5FU and 40 mmol/L of oxaliplatin (FOLFOX), or control.

Two-photon FLIM

2D cell line. Two-photon imaging was performed on a custom-built Multiphoton Imaging System (Bruker) consisting of an inverted microscope (TI-E, Nikon) coupled to an Insight DeepSee + solid state laser (Spectra Physics). Images were acquired in photon-counting mode using time-correlated single-photon counting electronics (SPC-150, Becker & Hickl GmbH) and GaAsP photomultiplier tubes (H7422P-40, Hamamatsu). Imaging was performed using Prairie View Software (Bruker). The tunable multiphoton laser allowed sequential excitation of NAD(P)H and FAD at 750 and 890 nm, respectively. The emission filter used for NAD(P)H was a bandpass 460/80 nm, whereas FAD used a bandpass 550/100 nm. All samples were imaged using a 40x water immersion, 1.15 NA objective (Plan Apo, Nikon) with an image scan speed of 4.8 μ s per pixel, 60 s integration time, and image size of 256 \times 256 pixels (2D) or 512 \times 512 pixels (3D). The laser power at the sample was maintained at < 10 mW. The instrument response function (IRF) was collected by recording the second harmonic generation signal of urea crystals. A Fluoresbrite YG microsphere (Polysciences Inc.) was imaged as a daily standard for fluorescence lifetime. The lifetime decay curves for these beads were fit to a single exponential decay, and the fluorescence lifetime was measured to be 2.1 ns ($n=7$), which is consistent with published values^{4,6}.

3D PDCOs. All PDCO images were acquired from a previously published dataset in Kratz et al.²⁸ Briefly, imaging protocol was the same as for PANC-1, though differences existed for laser [Mai Tai DeepSee Ti: sapphire laser (Coherent, Inc)] and NAD(P)H bandpass filter (440/80 nm).

ATM development

Cellpose is a versatile algorithm for cell and nucleus segmentation, featuring pre-trained models developed on a wide range of microscopy images. It also supports a human-in-the-loop pipeline for the rapid prototyping of new custom models⁴¹. However, the pre-trained models were unable to accurately segment the nuclei and cells in our autofluorescence microscopy images.

To train a new deep learning model in Cellpose, annotated data is required. Labeled images where each object has a unique label (masks) paired with the corresponding autofluorescence image were used for this purpose. To streamline the annotation process, we cropped, without rescaling, segments from at least three different images from 3D cell cultures, each containing 5–10 cells. This approach was intended to save time during annotation. For each cell in these cropped images, we drew the nucleus masks using the graphical user interface (GUI) of Cellpose.

These pairs of images and masks (cropped image with nucleus mask) were used to train a new model for segmentation of nuclei from 3D cell culture images. Following the methodology described in the “Cellpose 2.0: how to train your own model” paper⁴¹. The training parameters were as follows: --pretrained_model cyto, --chan 2, --chan2 1, --learning_rate 0.01, --weight_decay 0.0001, and --n_epochs 500.

Image analysis

Fluorescence lifetimes were extracted using SPCImage software (SPCImage v8.1, Becker & Hickl). The fluorescence lifetime decay curve was convolved with the IRF, using an iterative parameter optimization to obtain the lowest sum of the squared differences between model and data (weighted least squares algorithm), the goodness of fit was monitored to ensure a chi-squared value < 1.3. The two-component exponential decay model is $I(t) = \alpha_1^{-t/\tau_1} + \alpha_2^{-t/\tau_2} + C$, where $I(t)$ is the fluorescence intensity at time t , α is the fractional contribution of each component, τ is the lifetime of each component, and C accounts for background signal. A bin of 3 \times 3 pixels was used to increase photon counts for decays. The two lifetime components are used to distinguish between the free and bound forms of NAD(P)H and FAD^{16,17}. The mean fluorescence lifetime was calculated using $\tau_m = \alpha_1\tau_1 + \alpha_2\tau_2$. Additionally, the decay curves for NAD(P)H and FAD were integrated for each pixel to obtain intensity values. These intensity values are used to calculate the ORR using the equation $ORR = \frac{I_{NAD(P)H}}{I_{NAD(P)H} + I_{FAD}}$, where the intensities of NAD(P)H and FAD are given by $I_{NAD(P)H}$ and I_{FAD} , respectively.

Image segmentation

Manual segmentation was performed by human observers who identified and manually circled each nucleus within each image using CellProfiler software. The individuals tasked with performing image segmentation underwent a comprehensive training regimen designed to enhance accuracy and precision in segmenting relevant features within NAD(P)H intensity images. The training encompassed observing a series of annotated NAD(P)H intensity images spanning 2D and 3D cultures and then segmenting those images with guidance given through comparisons of the annotated images with their segmentation. After manual identification of nuclei, the cytoplasm region was isolated using a Python script to propagate out from the masked nuclei by 8 μm (10 pixels) to match the protocol of prior studies where we employed manual segmentation^{28,33}. The reported segmentation time was determined using a stopwatch from the opening of an image set to the closing of an image set and dividing by the number of images in the set.

ATM segmentation was performed by loading images into Cellpose where the ATM or cyto3 models identified nuclear regions. No further corrections were made to the masks. The cytoplasm region was then recognized using a Python script to propagate out from the nuclei by 8 μm (10 pixels) to match the procedure of the manual segmentation above. Segmentation time was determined using the Python DateTime module to measure the execution time of the segmentation script⁴⁸.

For additional validation, after initial segmentation, all PANC-1 images were rotated 90 degrees and segmented again by the same observer. These rotated images were presented as new images to the observer, who was unaware that they were re-segmenting the same images. Patient 6 PDCO images were also rotated and manually segmented again, though not by the same observer.

To obtain fluorescence lifetime features, Cell-analysis-tools 0.0.14 was used to overlay the generated masks on the NAD(P)H and FAD images (τ_1 , τ_2 , α_1 , α_2 , and I) produced from Kratz et al.²⁸. Values were averaged across all pixels within the cytoplasm mask of each segmented cell, for both manual and ATM generated masks, providing average values of each feature for each cell⁴⁹. This method was used for both the manual and ATM segmented images.

Statistics

Differences between Dice scores were tested using an Ordinary one-way ANOVA test. Effect sizes between groups in NAD(P)H τ_m , FAD τ_m , and ORR were calculated with Glass's Delta because comparisons of very large sample sizes of individual cells always pass traditional significance tests unless the population effect size is truly zero⁵⁰. Glass's Delta is defined as $\Delta = (\mu_{control} - \mu_{test}) / \sigma_{control}$ where values ranging from 0 to 0.2 represent no change, 0.2–0.5 small change, 0.5–0.8 represent moderate change, and anything higher than 0.8 signifies a large change^{51,52}. Simple linear regression modeling was performed using GraphPad Prism 9.4.1.

Data availability

The repository with the initial models and pipelines is located here: <https://github.com/COBA-NIH/SkalaLab>.

Received: 7 June 2024; Accepted: 6 December 2024

Published online: 14 February 2025

References

1. Cho, Y., An, H. J., Kim, T., Lee, C. & Lee, N. K. Mechanism of Cyanine5 to Cyanine3 photoconversion and its application for high-density single-particle Tracking in a living cell. *J. Am. Chem. Soc.* **143**, 14125–14135 (2021).
2. Zheng, Q., Jockusch, S., Zhou, Z. & Blanchard, S. C. The contribution of reactive oxygen species to the photobleaching of Organic fluorophores. *Photochem. Photobiol.* **90**, 448–454 (2014).
3. Chance, B., Schoener, B., Oshino, R., Itshak, F. & Nakase, Y. Oxidation-reduction ratio studies of mitochondria in freeze-trapped samples. NADH and flavoprotein fluorescence signals. *J. Biol. Chem.* **254**, 4764–4771 (1979).
4. C. Skala, M. et al. In vivo multiphoton microscopy of NADH and FAD redox states, fluorescence lifetimes, and cellular morphology in precancerous epithelia. *Proc. Natl. Acad. Sci. U S A* **104**, 19494–19499 (2007).
5. Qian, T. et al. Label-free imaging for quality control of cardiomyocyte differentiation. *Nat. Commun.* **12**, 4580 (2021).
6. Bird, D. K. et al. Metabolic mapping of MCF10A human breast cells via multiphoton fluorescence lifetime imaging of the coenzyme NADH. *Cancer Res.* **65**, 8766–8773 (2005).
7. Leben, R., Köhler, M., Radbruch, H., Hauser, A. E. & Niesner, R. A. Systematic enzyme Mapping of Cellular Metabolism by Phasor-analyzed label-free NAD(P)H fluorescence lifetime imaging. *Int. J. Mol. Sci.* **20**, 5565 (2019).
8. Schaefer, P. M., Kalinina, S., Rueck, A. & von Arnim, C. A. F. Einem, B. NADH Autofluorescence—A marker on its way to Boost Bioenergetic Research. *Cytometry A* **95**, 34–46 (2019). von.
9. Wright, B. K. et al. Phasor-FLIM analysis of NADH distribution and localization in the nucleus of live progenitor myoblast cells. *Microsc. Res. Tech.* **75**, 1717 (2012).
10. Rück, A., Breytmayer, J. & Kalinina, S. 8 TCSPC FLIM and PLIM for metabolic imaging and oxygen sensing. in *Multiphoton Microscopy and Fluorescence Lifetime Imaging*.
11. Schindelin, J. et al. Fiji - an Open source platform for biological image analysis. *Nat. Methods.* **9**. <https://doi.org/10.1038/nmeth.2019> (2012).
12. Kankaanpää, P. et al. BioImageXD: an open, general-purpose and high-throughput image-processing platform. *Nat. Methods* **9**, 683–689 (2012).
13. Dobson, E. T. A. et al. ImageJ and CellProfiler: complements in Open-Source Bioimage Analysis. *Curr. Protoc.* **1**, e89 (2021).
14. Carpenter, A. E. et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
15. Cellpose. a generalist algorithm for cellular segmentation | Nature Methods. <https://www.nature.com/articles/s41592-020-01018-x>
16. Lakowicz, J. R., Szmajdzinski, H., Nowaczyk, K. & Johnson, M. L. Fluorescence lifetime imaging of free and protein-bound NADH. *Proc. Natl. Acad. Sci. U S A* **89**, 1271–1275 (1992).
17. Nakashima, N., Yoshihara, K., Tanaka, F. & Yagi, K. Picosecond fluorescence lifetime of the coenzyme of D-amino acid oxidase. *J. Biol. Chem.* **255**, 5261–5263 (1980).

18. Varone, A. et al. Endogenous two-photon fluorescence imaging elucidates metabolic changes related to enhanced glycolysis and glutamine consumption in precancerous epithelial tissues. *Cancer Res.* **74**, 3067–3075 (2014).
19. Ostrander, J. H. et al. Optical redox ratio differentiates breast cancer cell lines based on estrogen receptor status. *Cancer Res.* **70**, 4759–4766 (2010).
20. Kalinina, S. et al. Bioenergetic Alterations of Metabolic Redox Coenzymes as NADH, FAD and FMN by means of fluorescence lifetime imaging techniques. *Int. J. Mol. Sci.* **22**, 5952 (2021).
21. Benninger, R. K. P. & Piston, D. W. Two-Photon Excitation Microscopy for the Study of Living Cells and Tissues. *Curr Protoc. Cell. Biol. Editor Board. Juan Bonifacino AI 0*, 4. Unit-4.1124, (2013).
22. Federau, C. et al. Improved segmentation and detection sensitivity of diffusion-weighted stroke lesions with synthetically enhanced deep learning. *Radiol. Artif. Intell.* **2**, e190217 (2020).
23. Song, H. et al. Establishment of patient-derived gastric Cancer Organoid Model from tissue obtained by endoscopic biopsies. *J. Korean Med. Sci.* **37**, e220 (2022).
24. Vlachogiannis, G. et al. Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science* **359**, 920–926 (2018).
25. van de Wetering, M. et al. Prospective derivation of a living Organoid Biobank of colorectal cancer patients. *Cell* **161**, 933–945 (2015).
26. Webb, J. M. et al. Comparing deep learning-based automatic segmentation of breast masses to expert interobserver variability in ultrasound imaging. *Comput. Biol. Med.* **139**, 104966 (2021).
27. 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).
28. Kratz, J. D. et al. Integrating Subclonal Response Heterogeneity to Define Cancer Organoid Therapeutic Sensitivity. 10.15.464556. Preprint at <https://doi.org/10.1101/2021.10.15.464556> (2021).
29. Mazzucchelli, S. et al. Establishment and morphological characterization of patient-derived organoids from breast Cancer. *Biol. Proced. Online* **21**, 12 (2019).
30. Rosenbluth, J. M. et al. Organoid cultures from normal and cancer-prone human breast tissues preserve complex epithelial lineages. *Nat. Commun.* **11**, 1711 (2020).
31. de Souza, N., Zhao, S. & Bodenmiller, B. Multiplex protein imaging in tumour biology. *Nat. Rev. Cancer* **24**, 171–191 (2024).
32. Van Doorselaer, L., Verboven, P. & Nicolai, B. Automatic 3D cell segmentation of fruit parenchyma tissue from X-ray micro CT images using deep learning. *Plant. Methods* **20**, 12 (2024).
33. Shah, A. T., Diggins, K. E., Walsh, A. J., Irish, J. M. & Skala, M. C. *Vivo* Autofluorescence Imaging of Tumor Heterogeneities post treatment. *Neoplasia* **17**, 862–870 (2015).
34. Gil, D. A., Deming, D. & Skala, M. C. Patient-derived cancer organoid tracking with wide-field one-photon redox imaging to assess treatment response. *J. Biomed. Opt.* **26**, 036005 (2021).
35. Samimi, K. et al. Light-sheet autofluorescence lifetime imaging with a single-photon avalanche diode array. *J. Biomed. Opt.* **28**, 066502 (2023).
36. Autofluorescence Imaging to Evaluate Cellular Metabolism (Video) | JoVE | Protocol. <https://www.jove.com/t/63282/autofluorescence-imaging-to-evaluate-cellular-metabolism-video-jove>
37. Stevens, M. et al. StarDist Image Segmentation improves circulating Tumor Cell detection. *Cancers* **14**, 2916 (2022).
38. Berg, S. et al. Ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* **16**, 1226–1232 (2019).
39. McQuinn, C. et al. CellProfiler 3.0: next-generation image processing for biology. *PLoS Biol.* **16**, e2005970 (2018).
40. Aguilar-Arnal, L. et al. Spatial dynamics of SIRT1 and the subnuclear distribution of NADH species. *Proc. Natl. Acad. Sci. U S A* **113**, 12715–12720 (2016).
41. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. Methods* **19**, 1634–1641 (2022).
42. Winfree, S. User-accessible machine learning approaches for cell segmentation and analysis in tissue. *Front Physiol* **13**, (2022).
43. Eschweiler, D., Rethwisch, M., Jarchow, M., Koppers, S. & Stegmaier J. 3D fluorescence microscopy data synthesis for segmentation and benchmarking. *PLoS ONE* **16**, e0260509 (2021).
44. Eschweiler, D., Smith, R. S. & Stegmaier, J. Robust 3d Cell Segmentation: Extending The View Of Cellpose. in *IEEE International Conference on Image Processing (ICIP)* 191–195. doi: <https://doi.org/10.1109/ICIP46576.2022.9897942> (2022).
45. Stringer, C. & Pachitariu, M. Cellpose3: one-click image restoration for improved cellular segmentation. 02.10.579780. Preprint at <https://doi.org/10.1101/2024.02.10.579780> (2024).
46. Ziyue, W. et al. Semi-Supervised Cell Instance Segmentation for Multi-Modality Microscope Images. (2022).
47. Pasch, C. A. et al. Patient-derived Cancer organoid cultures to predict sensitivity to Chemotherapy and Radiation. *Clin. Cancer Res.* **25**, 5376–5387 (2019).
48. datetime. — Basic date and time types. *Python documentation* <https://docs.python.org/3/library/datetime.html>
49. Cell Analysis Tools. — cell_analysis_tools documentation. <https://cell-analysis-tools.readthedocs.io/en/latest/>
50. Kaplan, R. M., Chambers, D. A. & Glasgow, R. E. Big Data and large sample size: a cautionary note on the potential for Bias. *Clin. Transl. Sci.* **7**, 342–346 (2014).
51. Sawilowsky, S. New effect size rules of Thumb. *J. Mod. Appl. Stat. Methods* **8**, (2009).
52. GLASS, G. V. & Primary Secondary, and Meta-Analysis of Research. *Educ. Res.* **5**, 3–8 (1976).

Acknowledgements

We thank Kratz et al. for providing the organoid FLIM images as well as their manually segmented nuclear masks. Additionally, we would like to thank the current members of the Skala and Deming labs for the discussions that led to the improvements in our computational methods and the refinement of our analytical techniques. We also thank graphic designer, Matthew Stefely, for assistance in finalizing our figures.

Author contributions

JMR: design, analysis, interpretation, drafted the work; AAG: conception, design, interpretation; ECG: analysis; MCC: creation of new software used in the work; AK: acquisition, analysis; SU and AS: interpretation; DD: conception, design, interpretation of data; BC: design, interpretation of data; MCS: conception, design, interpretation of data, drafted the work. All authors have read and approved the final manuscript.

Funding

P41 GM135019, R37 CA226526, R01 CA272855, R01 CA278051, R01 HL165726, P30 CA014520 (Core Grant, University of Wisconsin Carbone Cancer Center), ACI/Schwenn Family Professorship, JD Fluno Family Colorectal Cancer Precision Medicine Program, Carol Skornicka Chair, Morgridge Institute for Research.

Declarations

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/s41598-024-82639-6>.

Correspondence and requests for materials should be addressed to M.C.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024