



This is a repository copy of *Exploring the impact of variability in cell segmentation and tracking approaches*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/219787/>

Version: Published Version

---

**Article:**

Wiggins, L. [orcid.org/0000-0003-4615-2379](https://orcid.org/0000-0003-4615-2379), O'Toole, P.J., Brackenbury, W.J. et al. (1 more author) (2025) Exploring the impact of variability in cell segmentation and tracking approaches. *Microscopy Research and Technique*, 88 (3). pp. 716-731. ISSN 1059-910X

<https://doi.org/10.1002/jemt.24715>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

## RESEARCH ARTICLE OPEN ACCESS

# Exploring the Impact of Variability in Cell Segmentation and Tracking Approaches

Laura Wiggins<sup>1,2</sup>  | Peter J. O'Toole<sup>1,2</sup> | William J. Brackenbury<sup>1,2</sup> | Julie Wilson<sup>3</sup><sup>1</sup>York Biomedical Research Institute, University of York, York, UK | <sup>2</sup>Department of Biology, University of York, York, UK | <sup>3</sup>Department of Mathematics, University of York, York, UK**Correspondence:** Laura Wiggins ([l.wiggins@sheffield.ac.uk](mailto:l.wiggins@sheffield.ac.uk))**Received:** 21 March 2024 | **Revised:** 15 August 2024 | **Accepted:** 4 October 2024**Review Editor:** Alberto Diaspro**Funding:** This work was supported by Biotechnology and Biological Sciences Research Council (BBSRC) grant number (B/S507416/1).**Keywords:** automation | cell phenotyping | image analysis | manual | segmentation | tracking

## ABSTRACT

Segmentation and tracking are essential preliminary steps in the analysis of almost all live cell imaging applications. Although the number of open-source software systems that facilitate automated segmentation and tracking continue to evolve, many researchers continue to opt for manual alternatives for samples that are not easily auto-segmented, tracing cell boundaries by hand and reidentifying cells on consecutive frames by eye. Such methods are subject to inter-user variability, introducing idiosyncrasies into the results of downstream analysis that are a result of subjectivity and individual expertise. The methods are also susceptible to intra-user variability, meaning findings are challenging to reproduce. In this pilot study, we demonstrate and quantify the degree of intra- and inter-user variability in manual cell segmentation and tracking by comparing the phenotypic metrics extracted from cells segmented and tracked by different members of our research team. Furthermore, we compare the segmentation results for a ptychographic cell image obtained using different automated software and demonstrate the high dependence of performance on the imaging modality they were developed to handle. Our results show that choice of segmentation and tracking methods should be considered carefully in order to enhance the quality and reproducibility of results.

## 1 | Introduction

With microscopic imaging becoming an essential step in almost all areas of biological research, there is an increased demand for image analysis tools that provide reliable phenotype quantification (Chen et al. 2017). A particular challenge in the face of image analysis is *segmentation*, a means of distinguishing objects from background. Objects of interest vary depending on application, with examples including single-cells within 2D microscopy images (Al-Kofahi et al. 2018), tumors within magnetic resonance images (Zhang et al. 2021) or tissue from histopathological whole-slide images (Bándi et al. 2017). At present, many studies make use of manual segmentation to identify regions of interest (ROIs), though recent research has highlighted

the issue of both inter- and intra-operator variability in manual segmentation and the impact this has upon reproducibility and biological conclusions (Dionisio et al. 2021; Covert et al. 2022; Joskowicz et al. 2019; Bø et al. 2017). The subjectivity of manual segmentation and the time required to identify ROIs by hand has motivated a shift to automated approaches that are standardized, delineated and high throughput (Sharma, Aggarwal, et al. 2010). This has had a particular impact in clinical settings where automated approaches are now used to make diagnoses through computer-aided diagnosis (CAD) systems (Chan, Hadjiiski, and Samala 2020).

Although providing benefits in terms of reproducibility, automated segmentation approaches still involve some degree of

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Microscopy Research and Technique* published by Wiley Periodicals LLC.

## Summary

- Manual segmentation and tracking of cells in microscopy time-lapses are subject to both inter- and intra-researcher variability, with results heavily dependent on researcher subjectivity and expertise.
- Variability in manual cell segmentation and tracking can lead to visible idiosyncrasies within extracted phenotypic metrics.
- Open-source software for automated cell segmentation and tracking produce more consistent results, although they too are subject to inter-software variability.

user interactivity, with manual segmentation often considered the gold standard against which to benchmark the results of automated segmentation (Morey et al. 2009; Verma, Kumar, and Paydarfar 2022). Furthermore, model-based segmentation approaches, such as U-Nets, make use of manually segmented ROIs for ground truth training sets (Walsh et al. 2022). Automated approaches are also often specific to certain imaging modalities, with models learning characteristics of such images and using this knowledge to accurately identify ROIs whilst excluding background and image artifacts. Enhancing the accuracy of these methods can require additional steps, including feature engineering, data augmentation, and model refinement. Consequently, automated approaches themselves can be variable in performance, requiring careful consideration and optimization to achieve reliable results across different datasets and imaging conditions. For all of these reasons, choice of segmentation approach prior to quantification from microscopy images is a hugely important step that should not be overlooked and should instead be carefully considered.

There exist several established, open-source software packages for the task of segmenting cells from 2D microscopy images, each with their own strengths and limitations (Wiesmann et al. 2015). Studies have been carried out to benchmark such software against one another based on accuracy of segmentation and ease of usability, though notably fluorescence images have been used to trial the majority of these software. With the emergence of label-free imaging modalities, like phase contrast, bright-field and quantitative phase imaging (QPI), efforts are now being made to form ground truth data sets for training of segmentation models specific to those label-free imaging techniques (Edlund et al. 2021). More recent cell segmentation software such as CellPose (Stringer and Pachitariu 2022) emphasize their focus on developing models that are adaptable to new imaging techniques. Here, we demonstrate the direct impact that manual segmentation can have on downstream analysis by quantifying the inter- and intra-user variability of phenotypic metrics extracted from manually obtained ROIs. Moreover, we show that manual cell tracking also results in large variability in the extraction of motility metrics due to subjectivity in the choice of cell centroid position as well as reidentification of cells on consecutive frames. By comparing the results of automated and manual cell segmentation and tracking, we demonstrate greater reproducibility for a panel of automated packages, but large

inter-software variability arising from differences between imaging techniques used for development.

## 2 | Methods

### 2.1 | Cell Lines

The MDA-MB-231 breast cancer cell line used within this study was a gift from M. Djamgoz, Imperial College, London. The molecular identity of this cell line was verified by short tandem repeat analysis (Masters et al. 2001). Thawed cells were subcultured one to two times prior to discarding and thawing a new stock to ensure that the molecular identity of cells was retained throughout. Cells were cultured in Dulbecco's modified eagle medium supplemented with 5% fetal bovine serum (FBS) and 4mM L-glutamine (Yang et al. 2012). FBS was filtered using a 0.22  $\mu\text{m}$  syringe filter prior to use to reduce artifacts when imaging. Cells were incubated at 37°C in plastic filter-cap T-25 flasks and were split at a 1:6 ratio when passaged. No antibiotics were added to cell culture medium.

### 2.2 | Image Acquisition

All microscopy images within this work are label-free ptychographic images acquired using the Phasefocus Liveocyte 2 microscope (Phasefocus Limited, Sheffield, UK). Cells were placed onto the microscope to incubate for 30 min prior to image acquisition to allow for temperature equilibration. For each frame, one 500  $\mu\text{m}$   $\times$  500  $\mu\text{m}$  field of view per well was imaged to capture as many cells, and therefore data observations, as possible. Note that cells were imaged at 20 $\times$  magnification, and the time interval between images within time-lapses was 6 min.

### 2.3 | Statistical Analyses

All tests of statistical significance within this study were performed using GraphPad Prism 9.1.0 (GraphPad Software, San Diego, CA). Data were tested for normality using the D'Agostino and Pearson test. Parametric tests (*t*-tests and *F*-tests) were used where suitable with nonparametric Mann-Whitney *U*-tests in place of *t*-tests where data did not follow a normal distribution. For comparison of three or more populations, ANOVA was used when data followed a normal distribution or Kruskal-Wallis tests if not. Tukey's post hoc test and Dunn's multiple comparisons were performed following ANOVA and Kruskal-Wallis respectively. Results were considered significant if  $p < 0.05$ . Levels of significance used: \*  $< 0.05$ , \*\*  $< 0.01$ , \*\*\*  $< 0.001$ , \*\*\*\*  $< 0.0001$ . Full details of statistical tests used for each analysis are provided in the figure legend for the corresponding figure. Sørensen-Dice coefficients to quantify the similarity of segmented regions were computed using MATLAB's `dice()` function.

### 2.4 | Automated Segmentation and Tracking

All automated segmentation and tracking was performed by one of the authors of the current study, having extensive

experience in the analysis of time-course images from ptychography and the MDA-MB-231 cell line. For comparison with manual results, the Phasefocus CATbox system was used for image segmentation with each cell image independently processed five times. Replicated automated tracking results were obtained with the Phasefocus CATbox by tracking the cells after three independent image processing pipelines were performed.

## 2.5 | ImageJ Segmentation

ImageJ (Schneider, Rasband, and Eliceiri 2012) is an open-source software for processing of biological microscopy images. It offers accessible implementation of a range of traditional image processing algorithms, with the user setting subjective thresholds throughout the course of a segmentation workflow. Initially a rolling ball algorithm with radius size 30px was applied to the image for the removal of background noise and imaging artifacts. The image was then smoothed using Gaussian blurring with a standard deviation of 3 to further denoise the image. The image was then converted to a binary mask where each pixel was classified as either background (black within the binary mask) or cell (white within the binary mask) and the watershed method used to separate connected cells ahead of final segmentation by using the “create selection” feature in ImageJ for cell boundary detection. A number of alternative workflows were assessed in ImageJ such as the use of “find edges” and thresholding prior to binary mask conversion but these methods proved unsuccessful, the details included here describe the workflow identified as being most successful for the segmentation of our images.

## 2.6 | Livecyte Segmentation

Phase Focus' CATbox software facilitates automated cell segmentation by allowing the user to create a unique image processing recipe for each experiment. This recipe is produced through user-imposed thresholds on a range of image processing techniques allowing for live inspection of segmentation results. A rolling ball algorithm was applied to the image with radius  $29.24\mu\text{m}$  and a height of 2, to remove background noise and imaging artifacts. The image was then smoothed with a parameter of  $19.11\mu\text{m}$  ahead of seedpoint detection. Local pixel intensity maxima identified from the smoothed image were used as seedpoints with an intensity threshold of 0.8. Nearby seedpoints were then consolidated by combining two seedpoints if there was a path between them that did not change in phase by more than a threshold of 1.09. Thresholds were then applied to distinguish between the pixel intensities of cell and background. A background threshold of 0.05 was used, any pixel with intensity below this threshold was classified as background. A feature threshold (for detection of cells) of 0.6 was used, any pixel above this threshold was classified as belonging to a cell. Pixels that did not meet these requirements were then scaled to a range of 0–1 and formed a “gray area” where pixel classification was uncertain. A biased fuzzy distance transform (Saha, Wehrli, and Gomberg 2002) was then

used to classify “gray area” pixels with a fuzzy distance threshold of  $0.2\mu\text{m}$  and a phase weighting of 6. Final segmentation results were then exported as .roi files. All thresholds were selected subjectively by assessing live segmentation and revising thresholds to yield optimum results.

## 2.7 | Icy Segmentation

Icy (De Chaumont et al. 2012) is a free, open-source image processing software. A key component of Icy is its public repository for the sharing of analysis plug-ins and workflows. The “HK-means” plug-in was utilized with default settings for initial segmentation of the cells. This plug-in performs *K*-means clustering of the image pixel intensity histogram to distinguish between background and cells, and then uses size filtering to split undersegmented cells. A Gaussian prefilter was applied with standard deviation set equal to 5 for image smoothing prior to cell edge detection. *K* was set to 2 to establish two classes of pixel intensity: background and cell. Area size filters were then imposed with a minimum threshold of 100px and a maximum of 5000px, detected objects with areas outside of this range were discarded. The “active contours” plug-in was then used with default parameter settings for closer sculpting of ROIs to cell edges. Finally, the “split ROI” plug-in was used to separate cells that were originally undersegmented. This plug-in is semi-automatic and relies on the user to enter how many cells are incorrectly clustered within a group before selecting the group that needs separating. This method was applied to all instances of undersegmentation but some cells were unable to be corrected and were therefore left as is in the final segmentation.

## 2.8 | CellProfiler Segmentation

CellProfiler (McQuin et al. 2018) is an open-source software for the analysis and measurement of cell images in which pipelines can be constructed by the user for tailored cell segmentation and tracking. Cells were detected using the “identify primary objects” module with a minimum expected object diameter of 10px and a maximum of 50px. Objects detected outside of this range and objects that were touching the borders were discarded. The default image processing settings associated with this module were used, hence the image was globally thresholded prior to segmentation. A minimum cross-entropy thresholding method (Li and Lee 1993) was used to minimize the average error in pixel classification and Gaussian smoothing with a standard deviation of 1 was performed prior to final segmentation. CellProfiler allows the user to select which method should be used to distinguish and separate clumped objects, in this case, intensity was selected as opposed to the use of shape. After running the completed pipeline, segmentation masks exported as .tiff files. To convert image masks into ImageJ ROIs, the masks were read into ImageJ and an ROI of the full mask created using Edit > Selection > Create Selection. This selection was then added to the ROI manager where the “Split” function was used to create individual ROIs from the full mask selection.

## 2.9 | CellPose Segmentation

CellPose (Stringer et al. 2021) is an open-source, deep learning-based segmentation tool together with pretrained models for the segmentation of cells, nuclei, and tissue sections without the need for parameter fine-tuning. CellPose was utilized for cell segmentation using ImageJ's CellPose-TrackMate plugin. The pre-trained cytoplasm model was used for detection of cell boundaries with expected cell diameter of 10  $\mu\text{m}$ .

## 2.10 | Manual Segmentation and Tracking

We consider manual segmentation to involve users having to manually trace the outline of a cell as oppose to simply setting certain parameters to automate segmentation. Here manual segmentation was performed using ImageJ's freehand selection tool to trace around individual cells and selections were added to ImageJ's ROI manager using Edit > Selection > "Add to Manager." Morphological features were then extracted from each ROI using ImageJ's Measure function, where all features were selected from the available list provided in Analyze > Set Measurements. A full list of extracted features together with definitions is provided within the ImageJ [documentation](#).

ImageJ's MTrackJ plugin was used for manual tracking of cells. Each cell was tracked for 50 frames by clicking the center of the cell on consecutive frames, with  $X$  and  $Y$  coordinates from all final trajectories exported for feature extraction. R's trajr (McLean and Skowron Volponi 2018) package was used for the extraction of several features to characterize cell trajectories. The list of extracted metrics comprises: mean and standard deviation of cell speed, sinuosity, trajectory straightness index, mean, and standard deviation of track length, trajectory distance and mean trajectory turning angle.

To investigate inter-user variability, members of our research group with different research expertise performed the same segmentation and tracking tasks. Researchers 1, 2, and 5 are researchers who have received in-depth microscopy training, including the acquisition and interpretation of microscopy images but have no experience with the imaged cell-line. In comparison, Researchers 3 and 4 are not experienced microscopists but have extensive experience working with the MDA-MB-231 cell-line imaged in the study. Throughout the paper, we refer to the results obtained by these different research team members by a number so that the effect of differences in research experience can be considered. Researchers used a Windows Desktop computer and mouse for image annotation but, as no calibration of computer monitors was performed and the appearance of the image could differ with the researchers' computer monitor and display preferences, we provided a contrast-enhanced cell image in which cell protrusions not clearly visible within the original image could be observed (Figure 1a). The extent of protrusions that could be observed by researchers would likely be impacted by their monitor and display preferences. No training was provided as part of the study and there was no wash-out period as, with no feedback, researchers had no opportunity to learn and improve.

Researchers were set the task of manually segmenting three specified cells from a single microscopy image of MDA-MB-231

cells, the three cells are highlighted in Figure 1a. The segmentation of the three cells was repeated five times in order to calculate intra-user variability. As well as segmenting each of the three cells 5 times, the researchers tracked the three cells by selecting the cell's center on 50 consecutive frames. This was repeated three times for each cell to assess inter-user variability in cell tracking. Note that the researcher previously referred to as Researcher 3 did not take part in this cell tracking study and was replaced with a different researcher referred to as Researcher 6, with similar expertise to Researcher 3.

In order to compare the results from automated analyses with manual segmentation, a cell image was manually segmented by the author with extensive experience in the analysis of time-course images from ptychography and the MDA-MB-231 cell line. Although variability between users means that this manually segmented cell image cannot be considered "ground truth," it does provide a standard against which the accuracy of different automated software can be judged.

## 3 | Results

### 3.1 | Inter-User Variability in Manual Cell Segmentation

Inter-user variability in manual segmentation was observed from each researcher's identified ROIs (Figure 1b). Researchers 1, 2, and 5 (in-depth microscopy training) tended to closely sculpt the brightest intensity pixels, treating the cell boundary as the region around the cell in which bright intensity pixels are neighbored with dark intensity background pixels. Researchers 3 and 4 (MDA-MB-231 cell-line experience), on the other hand, tended to include a large number of dark intensity pixels within the areas they determined to be the cell interior, despite the resulting ROIs differing in morphology to that expected of an MDA-MB-231 cell. It is possible that Researchers 3 and 5 were considering the brightest intensity pixels as cell nuclei, and darker intensity pixels surrounding these areas as cytoplasm. This could be due to their experience with nucleic stains such as DAPI rather than label-free imaging in which whole cells can be visualized.

Sørensen–Dice coefficients confirmed the presence of inter-user variability in segmentation across all three cells, with higher coefficients indicative of significant overlap in segmentation results and lower coefficients representing manual segmentation that differs (Figure 1c). The Dice coefficients show the impact of research expertise on manual cell segmentation results, with researchers with similar expertise (Researchers 1, 2, and 5 or Researchers 3 and 4) obtaining high pair-wise coefficients in comparison to the coefficients obtained for researchers with differing expertise.

For each cell, bar charts showing the distribution of Dice coefficients for each researcher's segmentation data are plotted in Figure 1d to aid visualization. Two-way ANOVA revealed a statistically significant relationship between Dice coefficients and cell number, indicating greater inter-user variability in manual segmentation for certain cells. In comparison, no relationship was found between Dice coefficients and researcher

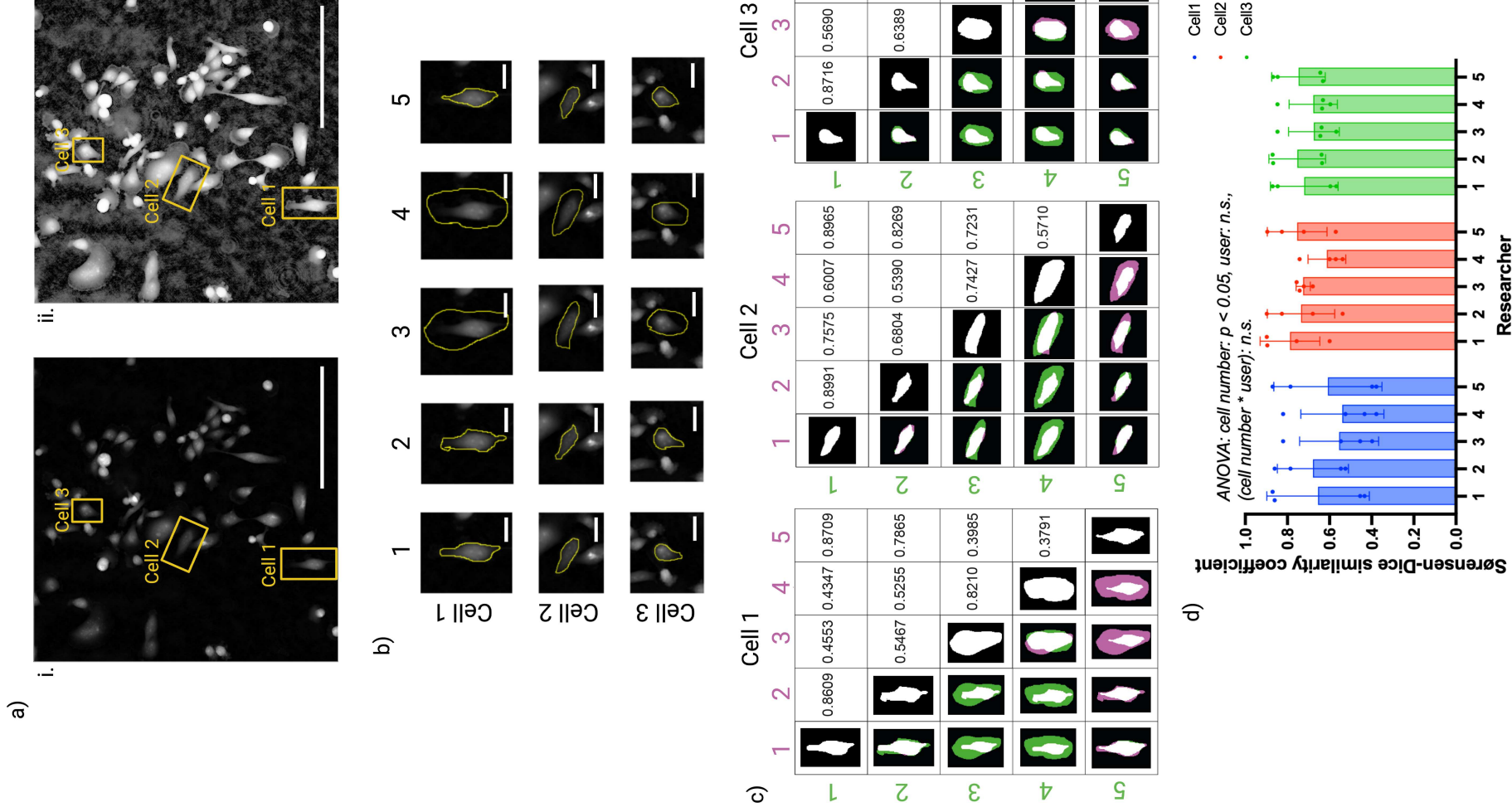
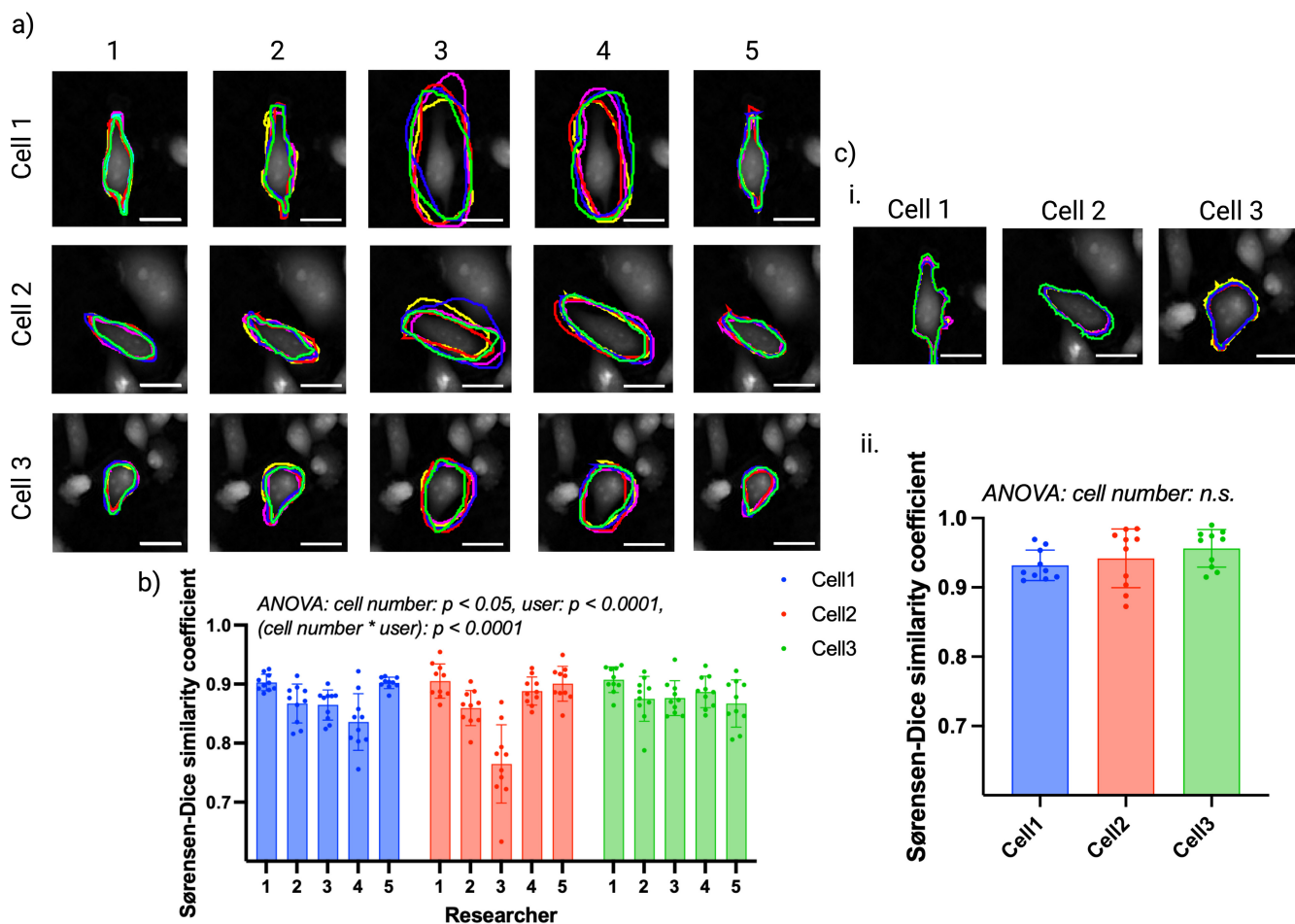


FIGURE 1 | Legend on next page.

number, suggesting that inter-user variability affected all researchers as opposed to one particular researcher. No statistical significance was found between Dice coefficients and the interaction between cell number and researcher, suggesting that inter-user variability is not related to particular cells.

### 3.2 | Intra-User Variability in Cell Segmentation

The segmentation of the three cells was repeated five times and the intra-user variability of each researcher's five attempts is shown in Figure 2a. Visual inspection of each researchers'



**FIGURE 2** | Intra-user variability in manual cell segmentation. (a) Ptychographic images overlaid with manual segmentations of each cell performed by each researcher, each color represents one of the five segmentation attempts performed. Intra-user variability can be observed from differences in outlined ROIs. Scale bar = 25  $\mu\text{m}$ . (b) Sørensen–Dice coefficients, grouped by researcher and colored by cell number. Bars show group means with error bars representing standard deviations. Two-way ANOVA showed that cell number, researcher and their interaction all have a statistically significant relationship with intra-user Dice coefficients, contributing to 3%, 25%, and 28% of the variance prevalent within the dataset, respectively. (c [i]) Automated segmentation of each cell performed using Phasefocus CATbox. Each color represents one of the five segmentation results obtained from the same cell image processed independently. Scale bar = 25  $\mu\text{m}$ . (c [ii]). Sørensen–Dice coefficients, colored by cell number. ANOVA showed no statistically significant relationship between Dice coefficients and cell number, indicating consistency in segmentation results obtained from Phasefocus CATbox irrespective of cell number.

**FIGURE 1** | Inter-user variability in manual cell segmentation. (a [i]). Ptychographic image of MDA-MB-231 cells with the three cells to be manually segmented highlighted and labeled. (a [ii]). The cell image with contrast enhanced to display cell protrusions that are not clearly visible within the original cell image. The three cells to segment are highlighted and labeled. Scale bar = 200  $\mu\text{m}$ . (b) Manual segmentation of each cell performed by each researcher. Inter-user variability can be observed from differences in outlined ROIs for each cell. Scale bar = 25  $\mu\text{m}$ . (c) Sørensen–Dice coefficients to quantify the degree of similarity of pair-wise manual segmentation. Images show segmentation maps where black represents background, white represents segmentation overlap, and green and pink represent differences in pair-wise segmentation. Areas in green appear in the segmentation map from the researcher listed in the row position, but not the researcher listed in the column position and vice-versa for areas in pink. (d) Sørensen–Dice coefficients, grouped by researcher and colored by cell number. Bars are representative of group means and error bars are representative of standard deviations to display the spread of coefficients. A two-way ANOVA was performed, obtaining  $p$  values of 0.0498, 0.3558, and 0.9944 for cell number, researcher, and their interaction, respectively.

segmentation attempts showed that some were more consistent in their segmentation approach, repeatedly outlining similar ROIs. This was the case for Researchers 1, 2, and 5 who all aimed to closely sculpt the bright intensity pixels that neighbor dark intensity background pixels. In comparison, Researchers 3 and 4 tended to show greater variance in their segmentation attempts where ROI selection, and therefore expected cell shape, was not guided as much by the bright intensity pixels.

Intra-user Dice coefficients were calculated and the results for each researcher and each cell are displayed in Figure 2b. Two-way ANOVA showed a highly significant relationship between intra-user Dice coefficients and researcher number, confirming greater consistency in manual segmentation for some researchers. Differences between researchers accounted for 25% of the total variance present within the data set. A statistically significant relationship was identified between intra-user Dice coefficients and cell number, indicating that intra-user variability was more prevalent for certain cells in comparison to others, though cell number only accounted for 3% of the total variance within the data set. High statistical significance was also found between intra-user Dice coefficients and the interaction between cell number and researchers, showing that maintaining consistency in segmentation proved more challenging with some cells than others for certain researchers. The interaction between cell number and researcher accounted for 28% of the total variance present within the data set.

Visual inspection of the ROIs from replicate segmentations of the same three cells using the Phasefocus CATbox showed greater consistency in comparison to the intra-user results of the researchers, across all three cells (Figure 2c[i]), with differing image processing pipelines resulting in subtle differences in segmentation. Dice coefficients were calculated to assess the similarity of ROIs obtained from the five processed images, the results are displayed in Figure 2c[ii]). One-way ANOVA revealed no statistically significant difference in Dice coefficients as a result of cell number ( $p$  value = 0.2407), indicating greater consistency in segmentation results irrespective of cell number for Phasefocus segmentation.

### 3.3 | The Impact of Cell Segmentation on Extracted Metrics

As the morphological metrics that can be extracted from cell images rely on initial segmentation of individual cells, accuracy and consistency in identification of ROIs is vital for meaningful downstream analyses. We hypothesized that the inter- and intra-user variability present in manual segmentation would also be prevalent in the morphological metrics extracted from the ROIs. We therefore extracted several morphological metrics describing cell shape, size, and texture from the ROIs in Figure 2a, as well as the Phasefocus CATbox segmentation results (described from now on as Livecyte) in Figure 2c, and assessed their variability, both inter- and intra-user. Cell area, circularity, and mean gray value were chosen as representative descriptors of cell size, shape, and texture,

respectively and their varying values can be seen in Figure 3a. ANOVA of all three metrics identified statistically significant differences in size, shape, and texture of cells as a result of inter-user manual segmentation with  $p$  values < 0.0001 for all three metrics. Post hoc analysis was performed using Tukey's multiple comparisons test and the full results for cell area, circularity, and mean gray value are provided in Tables S1 and S2, respectively. Notably, Researchers 1, 2, and 5 frequently displayed no significant differences in extracted metrics when compared with one another (with the exception of Researchers 2 and 5 scoring a  $p$  value of 0.0356 for circularity), but did consistently differ significantly from Researchers 3 and 4 across all metrics. Furthermore, Researchers 3 and 4 frequently displayed no significant differences between one another. Differences in research expertise between Researchers 1, 2, and 5, and Researchers 3 and 4 highlights the influence of research background on the results of downstream analyses. Researchers 1, 2, and 5 showed no statistical significance from Livecyte segmented cells in terms of cell area, but were statistically significant for both circularity and mean gray value. In contrast, Researchers 3 and 4 showed no statistical significance from Livecyte segmented cells in terms of circularity, but were statistically significant for both area and mean gray value. The impact of inter- and intra-user variability in extracted metrics was further investigated using PCA. Scores plots for Cells 1, 2, and 3 are provided in Figure 3b(i–iii), respectively. Researcher-specific clusters can be identified within the scores plots demonstrating idiosyncrasies in cell segmentation and their impact on the resulting extracted metrics. Clusters from Researchers 1, 2, and 5 tended to overlap with one another as did the clusters from Researchers 3 and 4, again suggesting a relationship between research experience and the results obtained. Intra-user variability is visible within all three scores plots, with larger ellipses representative of greater variability. Tighter clusters were frequently obtained with Livecyte, even in cases where all researchers displayed extensive variability. This is the case for Cell 3, where greater consistency in results is obtained with Livecyte across all three cells.

Additionally, intra-user variability in extracted metrics was quantified through the calculation of Euclidean distances to measure the pair-wise similarity of data vectors obtained by each researcher, the results are shown in Figure 3c. ANOVA identified statistically significant differences as a result of cell number ( $p$  value = 0.0273), indicating that certain cells invoked greater variability in extracted metrics in comparison to others. A statistically significant difference was also identified between researchers, demonstrating that some were more consistent in their segmentation approach and this was reflected in the consistency of the extracted metrics. The interaction between cell number and researcher was also highly statistically significant, indicating that consistency posed a greater challenge for some cells for certain researchers, resulting in greater variability in extracted metrics for some cells. Notably, the distances calculated for Livecyte segmentation remained consistent across all three cells, with the distances themselves minimized in comparison to those obtained by other researchers indicating greater consistency in extracted metrics as a result of Livecyte segmentation.



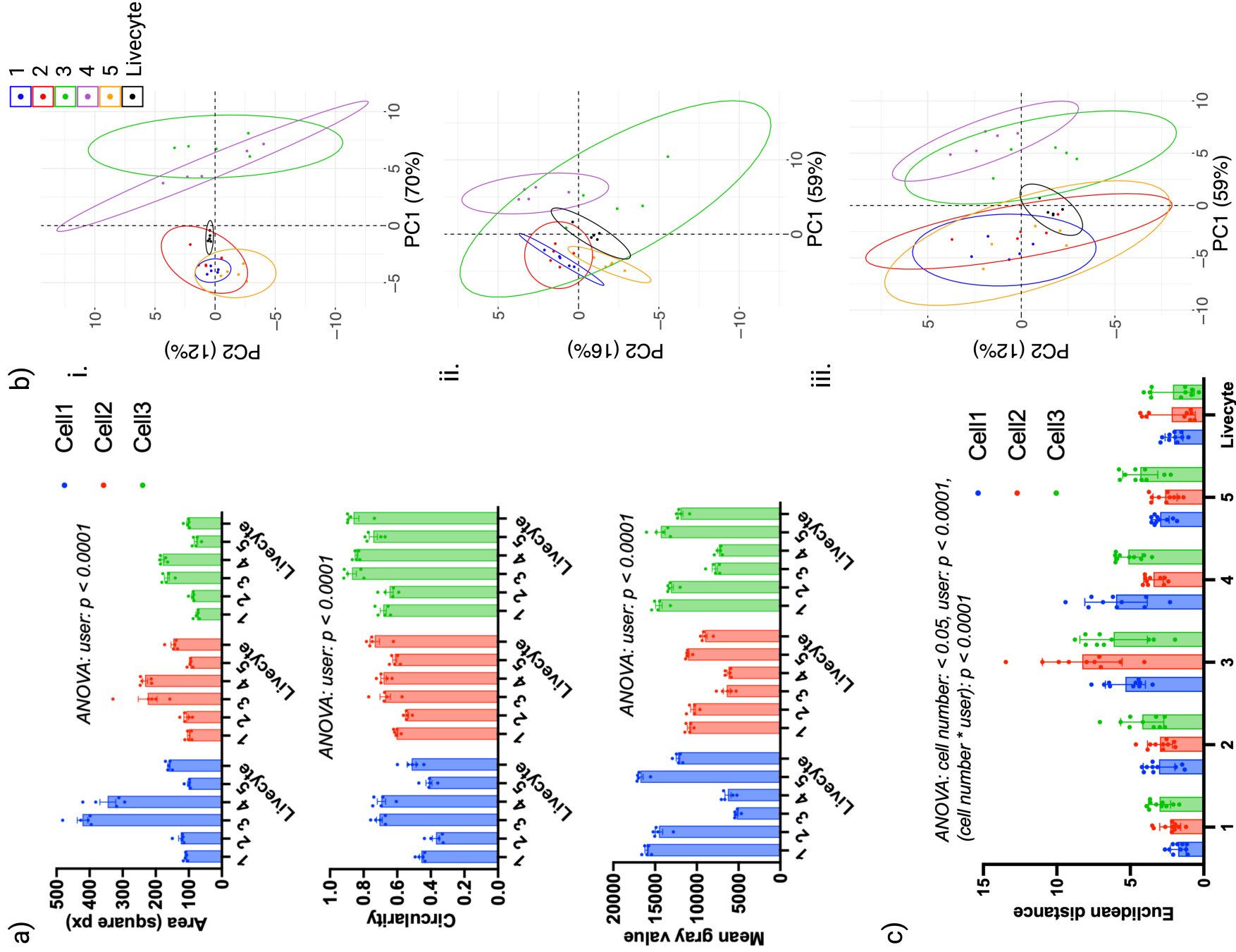
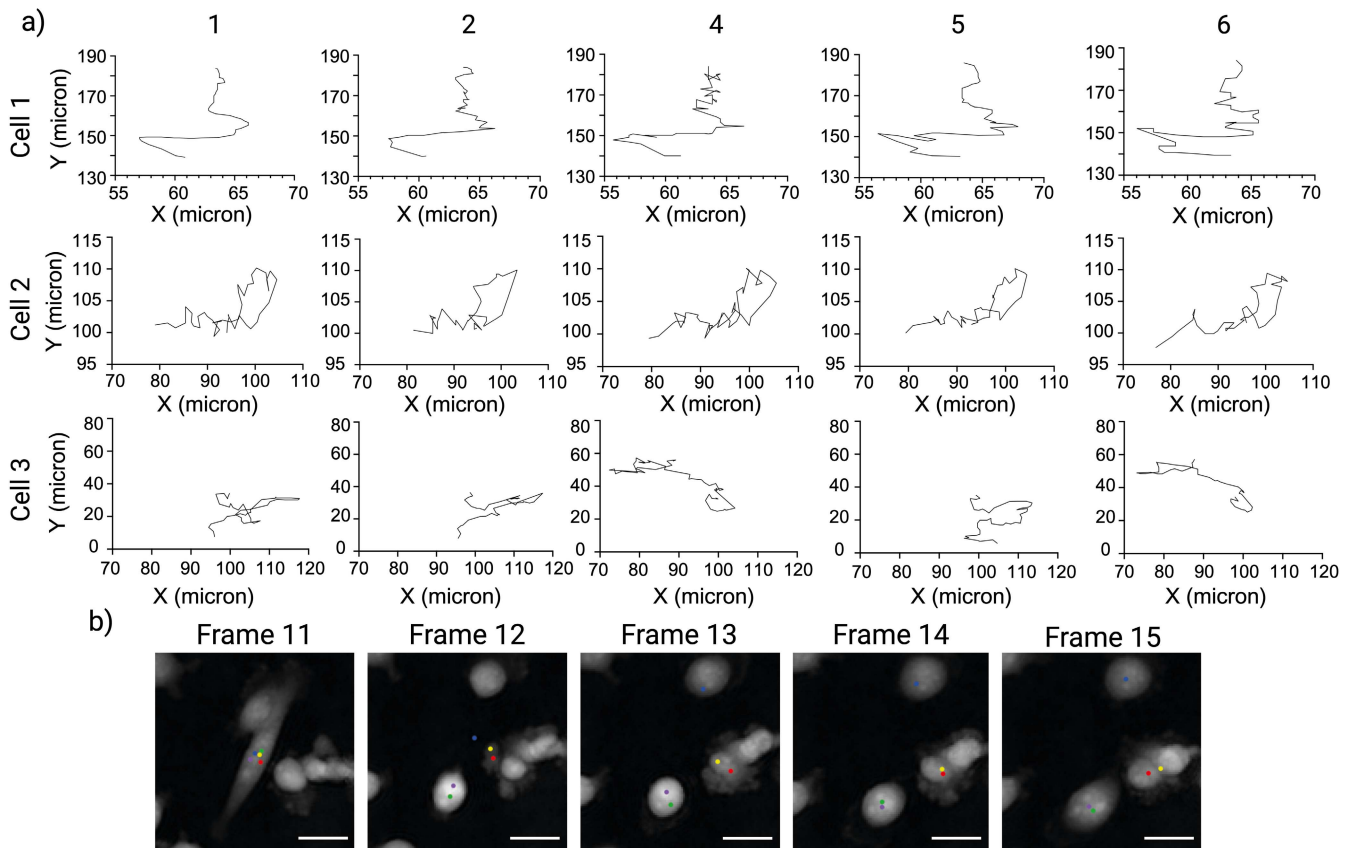


FIGURE 3 | Legend on next page.

### 3.4 | Inter-User Variability in Cell Tracking

The inter-user variability in cell trajectories can be observed in Figure 4a, where, for each cell, the researchers' first tracking attempt is shown. By eye, the general shape of trajectories obtained for Cells 1 and 2 appear fairly consistent, with subtle differences along the trajectories induced by variability in click position of each researcher. The trajectory shape for Cell 3, shows a noticeable difference for Researchers 4 and 6 (relevant cell-line experience) in comparison to Researchers 1, 2, and 5 (in-depth microscopy training) although the trajectories produced by Researchers 4 and 6 are similar.

Further investigation showed that the difference is due to disagreement in cell identification from one frame to the next, as visualized in Figure 4b. All researchers correctly identify Cell 3 in Frame 11, as demonstrated by all 5 points within the cell interior. The cell then undergoes a notable change in morphology on Frame 12, making it correct identification of Cell 3 challenging on this frame. Note that the correct cell can be identified by considering the timelapse in full rather than frame-by-frame as the researchers did. Cell 3 is correctly identified by Researchers 4 and 6 (green and purple points) in Frame 12, whereas Researchers 2 and 5 (red and yellow points) misidentify a different cell as Cell 3. Researcher 1



**FIGURE 4** | Inter-user variability in cell tracking. (a) Trajectories from each researcher's first cell tracking of Cells 1, 2, and 3 from a ptychographic time-lapse. For Cells 1 and 2, subtle differences between trajectories can be observed, induced by variability in click position between researchers. More notable inter-user variability can be observed for Cell 3, where differences in cell identification between researchers resulted in the tracking of different cells that were misidentified as Cell 3. (b) Visualization of the misidentification of Cell 3 throughout the time-lapse. All researchers correctly identify Cell 3 in Frame 11, where points represent click position colored by the researcher. Cell 3 undergoes a change in morphology on Frame 12 which causes disagreement on which cell is actually Cell 3. Researchers 4 and 6 correctly identify Cell 3 on Frame 12, Researchers 2 and 5 misidentify a different cell as Cell 3 and Researcher 1 identifies background as Cell 3. Each researcher then continues to track their chosen cells for the remainder of the time-lapse, with Researcher 1 misidentifying a cell as Cell 3 in Frame 13. Scale bar = 25  $\mu$ m.

**FIGURE 3** | The impact of cell segmentation on extracted metrics. (a) Distributions for cell area (top), circularity (middle) and mean gray value (bottom) extracted from different manual segmentation results and Livecyte automated segmentation. Plots are grouped by researcher and colored by cell number. Bars show group means with error bars representing SEM. ANOVA results show high statistical significance across all three metrics, demonstrating the impact of researcher-specific segmentation on extracted morphological metrics. (b) PCA scores plots for (i). Cell 1, (ii) Cell 2, and (iii) Cell 3 obtained from morphological metrics extracted from each manual segmentation and Livecyte automated segmentation. Points are colored according to researcher and 95% concentration ellipses are displayed. Small and large ellipses are representative of low and high intra-user variability in extracted metrics, respectively. (c) Pair-wise Euclidean distances between extracted data vectors, grouped by researcher and colored by cell number. ANOVA revealed statistical significance differences in Euclidean distances as a result of cell number, researcher, and the interaction between the two. Bars show group means and error bars represent standard deviations to display the spread of coefficients.

(blue point) selects a position that is actually background as opposed to a cell interior on Frame 12. This could be due to an unexpected change in morphology, and therefore position, of Cell 3 between Frames 11 and 12. Researcher 1 does select a cell interior on Frame 13 but again this is a misidentification rather than Cell 3. All researchers then continue to track their selected cells for the remainder of the time-lapse, resulting in different trajectories with only Researchers 4 and 6 correctly tracking the original Cell 3.

### 3.5 | Intra-User Variability in Cell Tracking

As the main source of variation in cell tracking arises from inconsistencies in researcher click positions, we sought to explore the intra-user variability present across three independent cell tracking attempts for each cell. We quantified each researcher's click spread for each of the three cells and scatter plots displaying these values are provided in Figure 5a. Here, the origins are representative of the average click position from three tracking attempts, with the click spread of each point then calculated as its deviation from the average. Researchers who remain consistent in their click positions will have all points closely clustered and centered around the origin. Researchers with inconsistencies in click position will have deviated scatter plots with points positioned at varying distances from the origin. All researchers experienced intra-user variability in click position across all cells, with Researcher 6 providing the largest inconsistencies for Cells 1 and 2 as evidenced by greater deviation in personal click spread for these cells.

Click spread for Cell 3 was considered separately to Cells 1 and 2 due to the impact of frequent misidentification of Cell 3. Researchers 4 and 6, the two researchers who successfully identified Cell 3 in Figure 4b successfully identified Cell 3 in all three of their tracking attempts, resulting lower click spread for Cell 3 in comparison to all other researchers. All other researchers identified different cells as Cell 3 within their three tracking attempts, resulting in large click spreads for Researchers 1, 2, and 5 for Cell 3. Interestingly, these results suggest a relationship between cell tracking results and research expertise, with cell biologists 4 and 6 displaying greater accuracy in cell tracking in comparison to Microscopists 1, 2 and 5.

In the cases of Cells 1 and 2, the click spread deviation for each researcher is likely a result of differences in determination of the cell centroid. Researchers with low click spread potentially aim for the same subcellular region on each attempt, whereas high click spread indicates inconsistency in selection of subcellular region. The click spreads for Cell 3 are dominated by the challenge of identifying Cell 3 when it experiences a change in morphology on Frame 12 with the additional deviation as a result of variability in selection of a subcellular region.

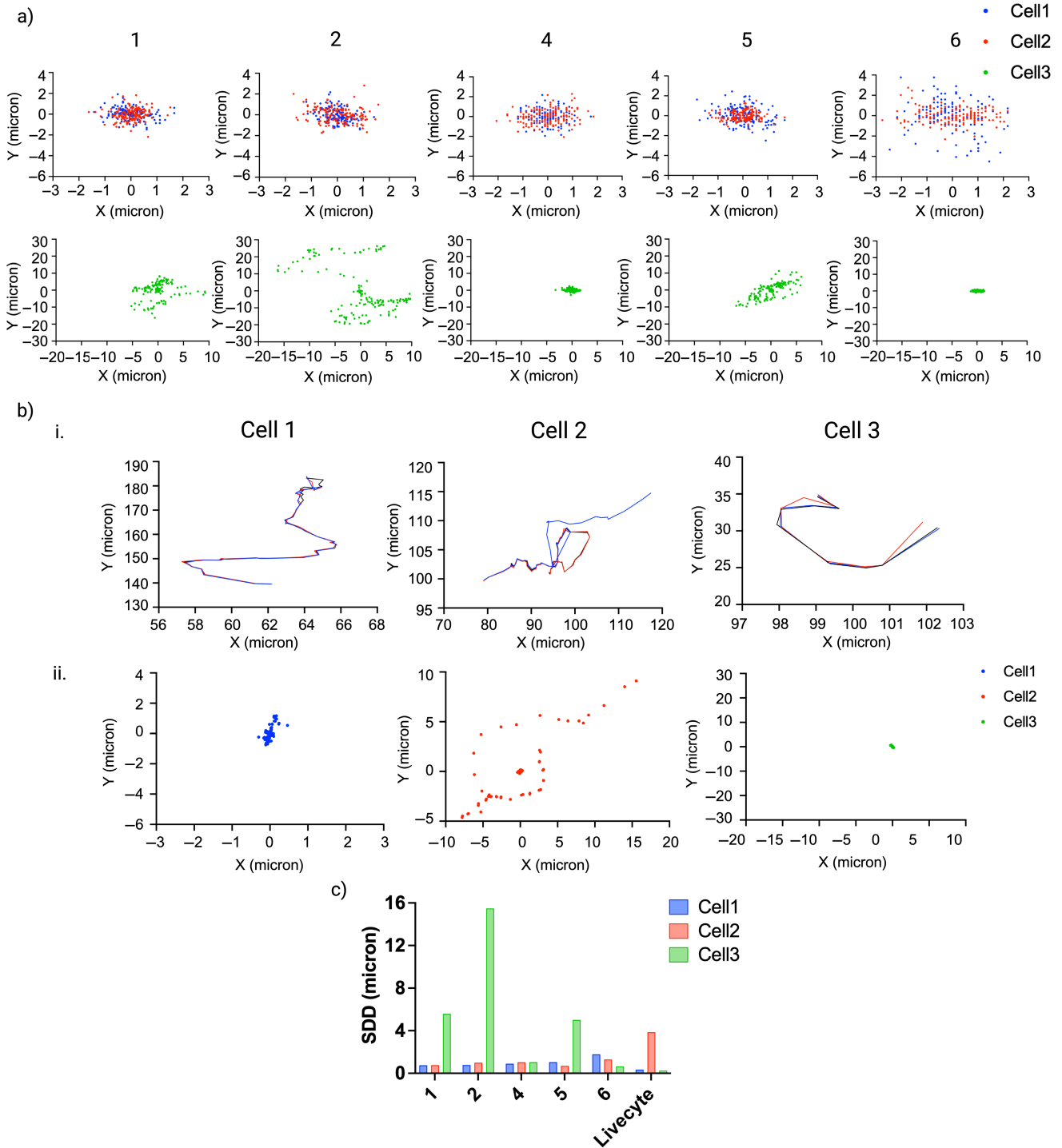
For comparison with manual tracking results, we considered the replicates obtained from automated tracking with the Phasefocus CATbox. Different processing methods were found to affect the determination of cell centroids. The

obtained cell trajectories, together with centroid deviation, for each cell are provided in Figure 5b. The plots show that despite different image processing pipelines used, the tracking results are consistent for Cell 1 with negligible variation in the cell trajectories and very little deviation in centroid position with points in the scatter plot forming a tight cluster centered at the origin. The same can be seen for Cell 3 although each track ends at Frame 11, just before the cell changes morphology on Frame 12 where each tracking repeat generated a new identifier for the cell that was used for the remainder of the time-lapse. There was notable misidentification of Cell 2 in one tracking repeat as evidenced by a change in the cell trajectory and centroid deviation, shown for a subset of points within the centroid scatter plot.

Standard distance deviation (SDD) was used to quantify the overall deviation of points for each cell and each researcher (Figure 5c). This confirmed observations from the click spread plots with Researcher 6 displaying greater inconsistencies in click positions for Cells 1 and 2 in comparison to all other researchers. Variability in SDD was greatest for Cell 3 due to repeated misidentification of Cell 3 by multiple researchers. SDD was minimal for Livecyte tracking of Cells 1 and 3, but higher than all other researchers for Cell 2. These results highlight greater consistency and reproducibility of automated tracking in comparison to manual tracking but also highlight problems in automated cell tracking.

### 3.6 | The Impact of Cell Tracking on Extracted Metrics

As inter- and intra-user variability in manual segmentation was found to have a major impact on morphological metrics, we hypothesized that the same would be the case for extracted characteristic metrics of cell trajectories. Variability in cell trajectories, both between researchers and within each researcher's own repeats, occurs due to variability in click positions on consecutive frames. Certain dynamical features, such as speed of cells, are extremely sensitive to such deviations resulting in large variability in the time series obtained for these features (Figure 6a[i,ii]). The variability across the whole cell trajectory has a cumulative effect on summary statistics that are calculated to characterize cell behavior throughout the whole time-lapse (such as mean speed), with each researcher obtaining different values dependent on their own tracking approach. The mean speeds calculated from each researchers' Cell 1 trajectories are shown in Figure 6a(iii). ANOVA revealed statistically significant differences in mean speed between researchers performing the tracking ( $p$  value = 0.0017), highlighting the impact of subjectivity on quantification. Notably, the minimal standard error of the mean was achieved by automated Livecyte tracking (SE = 0.034) in comparison to manual tracking (SE = 0.057, 0.054, 0.088, 0.12 for Researchers 1, 2, 4, and 6, respectively). PCA was used to assess the cumulative effect of such variability on a combination of motility metrics. As for PCA of morphological metrics shown in Figure 3b, researcher-specific clusters can be identified within the scores plots demonstrating idiosyncrasies in cell tracking and the impact these have on the resulting extracted metrics. Intra-user variability is visible within all three scores plots, with larger ellipses representative of greater variability.



**FIGURE 5** | Intra-user variability in cell tracking. (a) Click spread plots for each researcher, colored by cell number. Origins represent the average click position across three tracking attempts, with the click spread of each point calculated as deviation from this average. Click spread plots for Cell 3 show greater deviance due to repeated misidentification of Cell 3 throughout the time-lapse resulting in variable cell trajectories. (b [i]) Cell trajectories for Cells 1, 2, and 3 obtained from automated tracking using Phasefocus CATbox. Each color represents one of three trajectories obtained from the same cell time-lapse independently processed to replicate inter-user variability. Misidentification of Cell 2 in one time-lapse results in an inaccurate trajectory (shown in blue). Trajectories for Cell 3 are shorter than those obtained by manual tracking due to the premature ending of tracking on Frame 11 when Cell 3 undergoes a morphology change, after which the true Cell 3 is given a new cell identifier for the remainder of the time-lapse. (ii) Centroid spread plots for each researcher, colored by cell number. Origins are represent the average centroid position across each of three tracking repeats, with the spread of each centroid calculated as deviation from this average. Deviation is minimal for Cells 1 and 3, but the misidentification of Cell 2 in one trajectory causes deviation in centroids. (c) Standard distance deviation (SDD) values to quantify the manual click spread from (a) and automated centroid spread from (b).

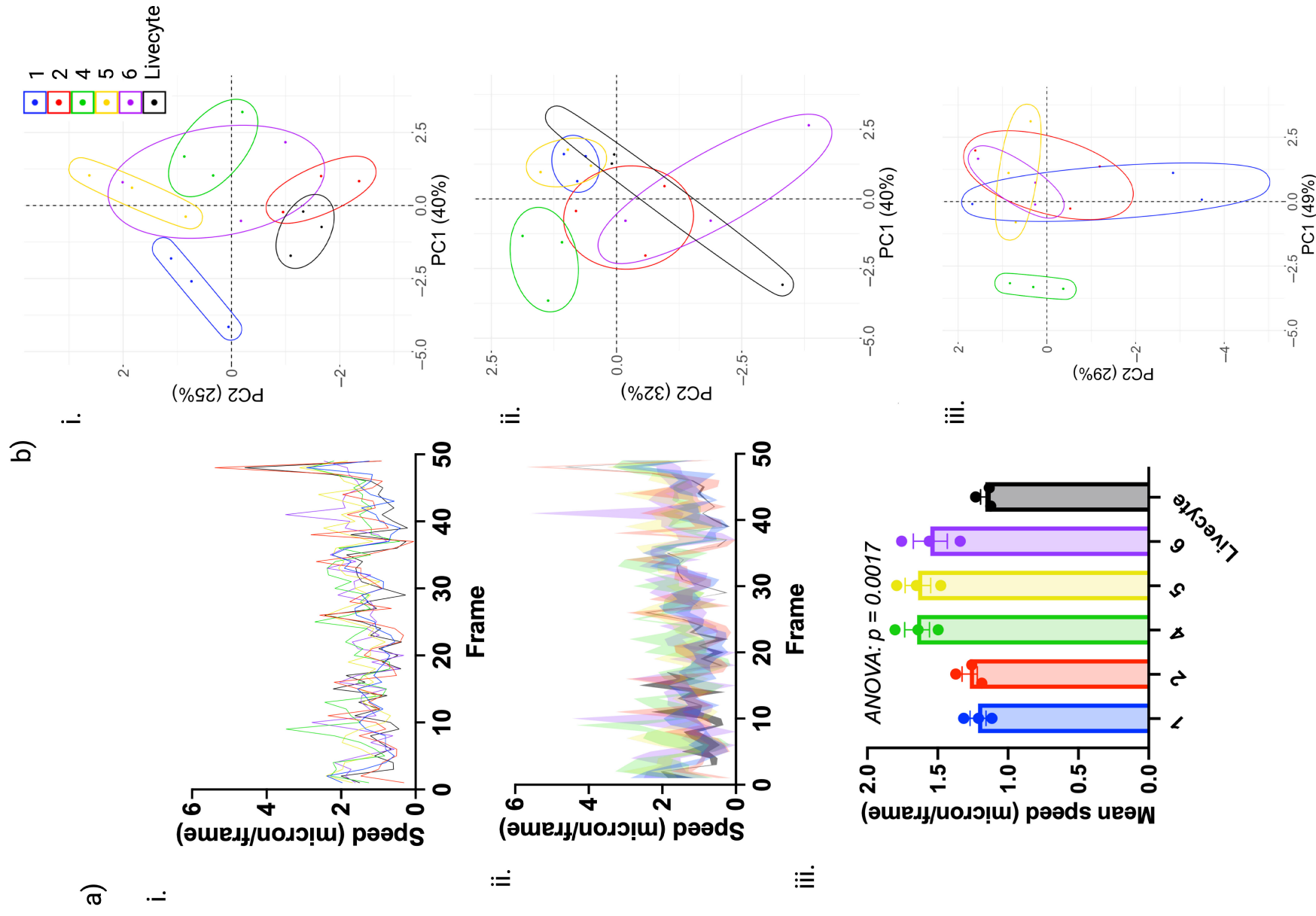


FIGURE 6 | Legend on next page.

Largest ellipses within scores plots for Cells 1 and 2 are observed for Researcher 6, found to have the largest click spread deviation for these cells. Automated Livecyte tracking resulted in the smallest ellipse, and therefore the least variability in motility metrics, for Cell 1 but the impact of misidentification of Cell 2 can be observed in the Cell 2 scores plot where a much larger Livecyte ellipse and more dispersed points can be observed. Note that Livecyte points are not included in the scores plot for Cell 3 due to the premature ending of all three Livecyte trajectories for this cell meaning metrics such as track length could not be compared with full-length manual trajectories.

### 3.7 | Variability in Extracted Metrics From Segmentation Software

Cell segmentation software were benchmarked on their ability to obtain accurate cell ROIs that reliably capture the true morphology of the cells within the images. Figure 7a shows the cell image manually segmented to provide a standard against which the accuracy of automated software could be assessed. From the automated segmentation results presented within Figure S1, it can be observed that morphology of ROIs from ImageJ and Icy, in particular, differ from those of the other three software. Notably, ImageJ and Icy tend to segment the cell body, excluding the lower intensity cell protrusions that are identified by Livecyte, CellProfiler, CellPose, and manual segmentation. PCA of all morphological metrics extracted from each software's obtained ROIs confirmed this to be true, with ImageJ and Icy ellipses providing smallest overlap with the manual segmentation ellipse and greatest distance between group means in the PCA scores plot (Figure 7b).

Further investigation into the morphological metrics impacted by this difference in segmentation showed that ImageJ and Icy segmentation result in cells that are significantly smaller in size, have increased circularity and have higher mean gray value compared with manually segmented cells (Figure 7c). In comparison, Livecyte, CellProfiler, and CellPose segmentation resulted in the extraction of morphological metrics that were not significantly different to those obtained by manual segmentation, the exception being circularity of CellPose ROIs ( $p$  value = 0.046).

## 4 | Discussion

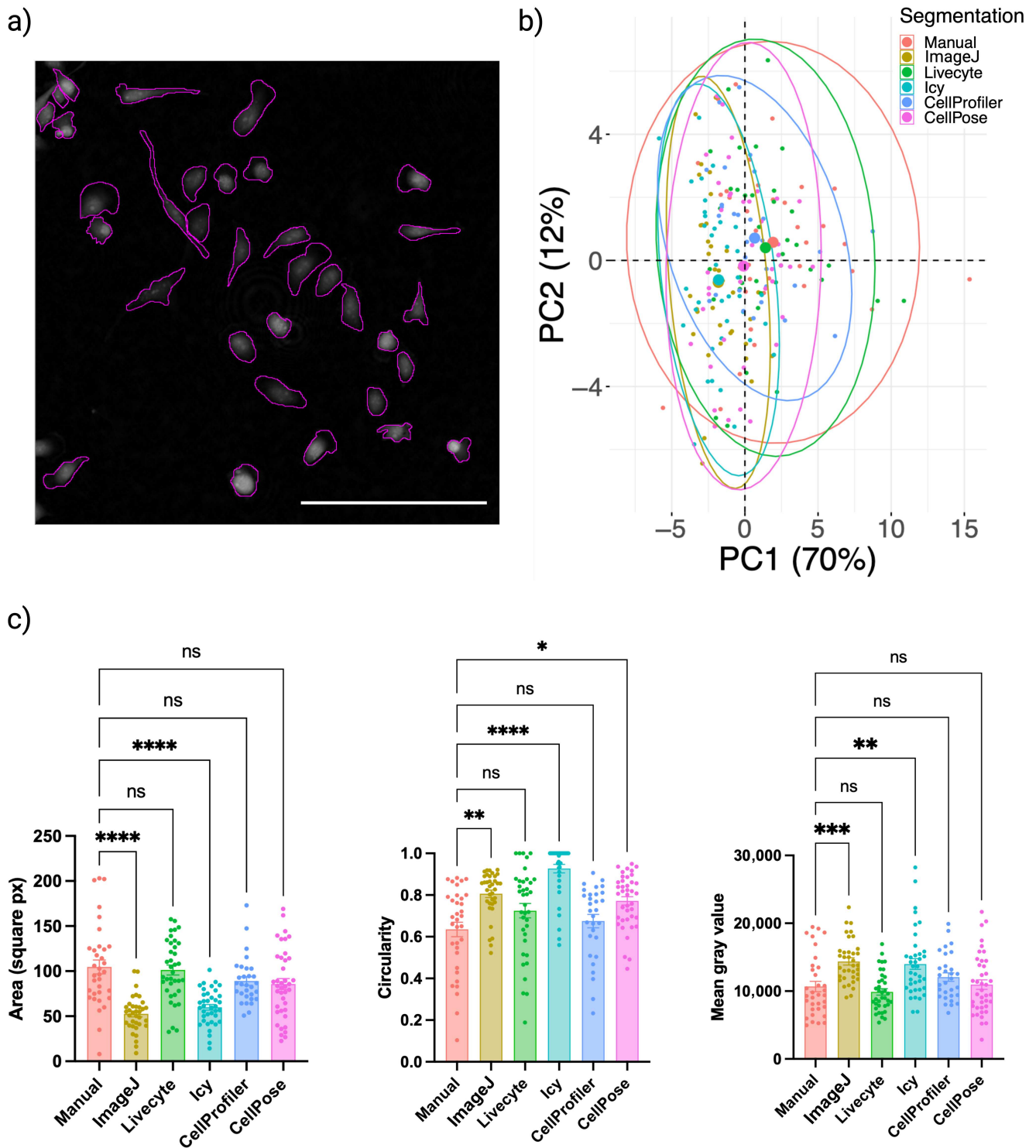
Whilst our limited pilot study cannot provide meaningful statistics, the results demonstrate the importance of choosing a segmentation and tracking approach that is optimized for the application in question. Most segmentation and tracking software rely on the contrast in intensity between cells and background to accurately detect cell boundaries. These software are often optimized for fluorescence images where fluorescent labels enhance the intensity of pixels within cell interiors. With pychography being a relatively

new imaging modality, pre-existing algorithms do not perform as well on these images due to differences in the distribution of cell pixel intensities within label-free images. The newer segmentation and tracking software, Livecyte, and CellPose-TrackMate, were found to be optimal for our images in comparison to the other, well-established software tested: ImageJ, Icy, and CellProfiler. Livecyte algorithms are optimized for pychographic imaging and this shows in its superiority for both segmentation and tracking. CellPose is the most recent software within the field and already offers several pre-built models that are optimized for imaging modalities, namely phase-contrast and brightfield, that often pose the greatest challenge for segmentation (Stringer and Pachitariu 2022). As label-free imaging becomes more established, it is likely that pre-existing segmentation and tracking software will further develop their current algorithms to handle such images in order to compete with newer approaches.

Our findings suggest that the results of manual segmentation and tracking should be used as a gold standard with caution. Deep learning algorithms frequently make use of manual results for training and testing performance, yet manual results vary significantly with experience and subjectivity. It is important that manual segmentation and tracking for such ground truth data sets is performed by an expert in the particular field in question, and that their own results provide minimal intra-user variability. The subjectivity of manual segmentation is of particular concern for medical applications in which segmentation results are used for diagnosis of diseases, such as the detection of breast tumors in mammograms. Furthermore, the morphological metrics extracted from segmented regions are used to assess response to treatment and disease progression (such as size of tumor following chemotherapy). This motivates the need for standardized, automated approaches to segmentation that eradicate the potential for variability within results.

The variability in segmentation and tracking results, both inter- and intra-user as well as inter-software support the ongoing replicability and reproducibility crisis within scientific studies. The results of this study showed that idiosyncrasies present within initial segmentation and tracking are further exacerbated by the extraction of metrics to quantify cell morphology and motility. This led to results from identical cell populations and even same cells being found statistically significantly different based on their initial segmentation and tracking. Automated software takes on an algorithmic approach to cell segmentation and tracking, following the same formulaic pipeline both within and between images. On the other hand, researchers performing manual segmentation and tracking often do not have a set of rules to follow in order to detect cell boundaries or centroids, as evidenced by intra-user variability in identified ROIs and researcher click spread, therefore adding cumulative inaccuracies to their results. Intra-software variability was minimal which motivates the

**FIGURE 6** | The impact of cell tracking on extracted metrics. (a [i]) Time series for mean speed obtained from three tracking attempts for Cell 1, colored by researcher. (ii) The same plot as in (i) but with the region between SEM error bars shaded to display intra-user inconsistencies. (iii) Separate scatter plots showing mean speeds calculated for each researcher and each tracking repeat of Cell 1. ANOVA revealed a statistically significant relationship between mean speed and the researcher performing tracking ( $p$  value = 0.0017). (b) The resulting PCA scores plots for (i) Cell 1, (ii) Cell 2, and (iii) Cell 3, performed with motility metrics extracted from the trajectories. Points are colored according to researcher and 95% confidence ellipses are displayed. Small and large ellipses are representative of low and high intra-user variability in extracted metrics, respectively. Note that Livecyte points are not included in the scores plot for Cell 3 due to the premature ending of all three Livecyte trajectories.



**FIGURE 7** | Variability in extracted metrics from segmentation software (a) Results of manual segmentation on the 0h ptychographic image, performed by an expert with extensive experience working with ptychographic images and the MDA-MB-231 cell line used in this study. Scale bar = 200  $\mu\text{m}$ . (b) PCA of morphological metrics extracted from cell ROIs, colored by segmentation approach. ImageJ and Icy display greatest variability to manual segmentation with their ellipses sharing the minimum overlap with the manual segmentation ellipse. Larger points are representative of group means. (c) Distributions for cell area, circularity and mean gray value extracted by manual segmentation and five different automated software packages. Bars show group means and error bars represent SEM. Results of Dunn's multiple comparisons test are displayed on plots, with greatest significance across all metrics obtained by ImageJ and Icy segmented cells when compared with manual segmentation.

use of automated approaches to improve the repeatability of results. However, though results may be more reproducible this does not ensure that they are accurate. It is therefore vital

that background research is done on the choice of automated segmentation and tracking software prior to downstream analysis.

The similarity of results from manual segmentation and tracking tend to be associated with the similarity of research background. Notably, the trained microscopists (Researchers 1, 2, and 5) produced manual segmentation results with greater similarity to each other than to cell biologists (Researchers 3 and 4). This was supported by the calculation of Dice coefficients for quantification of segmentation overlap. This suggests that research background is a latent variable in terms of quantification from cell segmentation and tracking, though this impact is often not assessed or accounted for in scientific research.

The challenge caused by Cell 3, even for trained microscopists and automated software, emphasizes the importance of minimizing interval times between frames during image acquisition. Cells that undergo a drastic morphology change from one frame to another are difficult to identify in consecutive frames, and misidentification results in erroneous changes to trajectories and consequently the time series of extracted metrics. These results demonstrate that both misidentification and erroneous cell segmentation are visible within extracted morphology and motility metrics and suggest the possibility of automating detection of such instances to clean up datasets prior to downstream analyses. Shorter cell tracks would not cause a problem in terms of biological conclusions, as the remaining cell track is still represented within the data set but linked to a different cell identifier. To handle the longer, inaccurate cell tracks, machine learning models could be used to identify interruptions in cell time series induced by misidentification and exclude these cells from further analysis, a step that is included in the CellPhe toolkit for automated cell phenotyping (Wiggins et al. 2023).

Additional experimental approaches to enhance cell segmentation accuracy include optimizing cell seeding density to minimize cell occlusion and accounting for the proliferation rate throughout the time-lapse sequence. Furthermore, addressing microscopy-related factors, such as minimizing focal drift and ensuring a high signal-to-noise ratio, can significantly improve segmentation accuracy and mitigate potential issues for downstream analysis. It is possible that 3D imaging could prove superior to 2D imaging in terms of minimizing segmentation errors, as the ability to capture the full spatial context of cells may reduce the overlap or intersections that are common causes of mis-segmentation in 2D cultures. Software such as 3DeeCellTracker (Wen et al. 2021) and Imaris leverage 3D-specific spatial features to enhance their segmentation and have proved useful for applications in dynamic cell time-lapses (Liu et al. 2024).

## 5 | Conclusion

The manual segmentation and tracking study within this pilot study demonstrated inter-user variability in segmentation and tracking as a result of researcher subjectivity and experience. The study also demonstrated intra-user variability in manual segmentation and tracking, a potential source of irreproducible findings. Both inter- and intra-user variability in segmentation and tracking caused imprecision in the phenotypic metrics extracted from obtained ROIs and trajectories, with

metrics from the same cells being statistically significantly different based on initial segmentation and tracking. This motivated the benchmarking of a panel of automated segmentation and tracking software, whose performance proved variable on the representative ptychographic time-lapse images. The choice of segmentation approach is a hugely important step that should not be overlooked and should be carefully considered. We would recommend testing various options and optimizing the parameters at the start of a project. It was determined that Livecyte, CellProfiler and CellPose performed best for our application in terms of accuracy, when compared with manual segmentation performed by an expert in the field. Although this person had both ptychography expertise and experience with the cell line used in this study, we would suggest combining the results of several experts to provide ground truth data for a larger study.

## Author Contributions

**Laura Wiggins:** conceptualization, investigation, writing – original draft, methodology, validation, visualization, writing – review and editing, formal analysis, data curation. **Peter J. O’Toole:** conceptualization, funding acquisition, methodology, writing – review and editing, supervision, resources. **William J. Brackenbury:** resources, supervision, methodology, writing – review and editing, conceptualization, funding acquisition. **Julie Wilson:** conceptualization, funding acquisition, writing – original draft, methodology, writing – review and editing, supervision.

## Acknowledgments

This work was supported by the BBSRC (grant number B/S507416/1). The authors thank the following members of the research team for assisting with data collection: Dr. Serife Yerlikaya, Nattanan Sajjaboontawee, Jodie Malcolm, and Blythe Wright from the University of York, and Kundi Umar and Abdurrahman Alkham from Yobe State University.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## References

- Al-Kofahi, Y., A. Zaltsman, R. Graves, et al. 2018. “A Deep Learning-Based Algorithm for 2-D Cell Segmentation in Microscopy Images.” *BMC Bioinformatics* 19, no. 1: 1–11.
- Bándi, P., R. Van de Loo, M. Intezar, et al. 2017. “Comparison of Different Methods for Tissue Segmentation in Histopathological Whole-Slide Images.” In *2017 IEEE 14th International Symposium on Biomedical Imaging (Piscataway, New Jersey, USA, 2017)*, 591–595. IEEE.
- Bø, H. K., O. Solheim, A. S. Jakola, K. Kvistad, I. Reinertsen, and E. M. Berntsen. 2017. “Intra-Rater Variability in Low-Grade Glioma Segmentation.” *Journal of Neuro-Oncology* 131: 393–402.
- Chan, H., L. M. Hadjiiski, and R. K. Samala. 2020. “Computer-Aided Diagnosis in the Era of Deep Learning.” *Medical Physics* 47, no. 5: e218–e227.
- Chen, W., W. Li, X. Dong, et al. 2017. “A Review of Biological Image Analysis.” *Current Bioinformatics* 12: 337–343.



- Covert, E., K. Fitzpatrick, J. Mikell, et al. 2022. "Intra-and Inter-Operator Variability in MRI-Based Manual Segmentation of HCC Lesions and Its Impact on Dosimetry." *EJNMMI Physics* 9, no. 1: 1–16.
- De Chaumont, F., S. Dallongeville, N. Chenouard, et al. 2012. "Icy: An Open Bioimage Informatics Platform for Extended Reproducible Research." *Nature Methods* 9: 690–696.
- Dionisio, F., L. S. Oliveira, M. Hernandez, et al. 2021. "Manual Versus Semi-Automatic Segmentation of Soft-Tissue Sarcomas on Magnetic Resonance Imaging: Evaluation of Similarity and Comparison of Segmentation Times." *Radiologia Brasileira* 54: 155–164.
- Edlund, C., T. R. Jackson, N. Khalid, et al. 2021. "Livecell—A Large-Scale Dataset for Label-Free Live Cell Segmentation." *Nature Methods* 18, no. 9: 1038–1045.
- Joskowicz, L., D. Cohen, N. Caplan, and J. Sosna. 2019. "Inter-Observer Variability of Manual Contour Delineation of Structures in Ct." *European Radiology* 29: 1391–1399.
- Li, C. H., and C. K. Lee. 1993. "Minimum Cross Entropy Thresholding." *Pattern Recognition* 26: 617–625.
- Liu, B., Y. Zhu, Z. Yang, H. H. Yan, S. Y. Leung, and J. Shi. 2024. "Deep Learning-Based 3D Single-Cell Imaging Analysis Pipeline Enables Quantification of Cell-Cell Interaction Dynamics in the Tumor Microenvironment." *Cancer Research* 84, no. 4: 517–526.
- Masters, J. R., J. A. Thomson, B. Daly-Burns, et al. 2001. "Short Tandem Repeat Profiling Provides an International Reference Standard for Human Cell Lines." *Proceedings of the National Academy of Sciences of the United States* 98, no. 14: 8012–8017.
- McLean, D. J., and M. A. Skowron Volponi. 2018. "Trajr: An R Package for Characterisation of Animal Trajectories." *Ethology* 124, no. 6: 440–448.
- McQuin, C., A. Goodman, V. Chernyshev, et al. 2018. "Cellprofiler 3.0: Next-Generation Image Processing for Biology." *PLoS Biology* 16: e2005970.
- Morey, R. A., C. M. Petty, Y. Xu, et al. 2009. "A Comparison of Automated Segmentation and Manual Tracing for Quantifying Hippocampal and Amygdala Volumes." *NeuroImage* 45, no. 3: 855–866.
- Saha, P. K., F. W. Wehrli, and B. R. Gombert. 2002. "Fuzzy Distance Transform: Theory, Algorithms, and Applications." *Computer Vision and Image Understanding* 86: 171–190.
- Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. "NIH Image to ImageJ: 25 Years of Image Analysis." *Nature Methods* 9, no. 7: 671–675.
- Sharma, N., and L. Aggarwal. 2010. "Automated Medical Image Segmentation Techniques." *Journal of Medical Physics* 35, no. 1: 3.
- Stringer, C., and M. Pachitariu. 2022. "Cellpose 2.0: How to Train Your Own Model." *BioRxiv* 19, no. 12: 2022-04.
- Stringer, C., T. Wang, M. Michaelos, and M. Pachitariu. 2021. "Cellpose: A Generalist Algorithm for Cellular Segmentation." *Nature Methods* 18, no. 1: 100–106.
- Verma, K., S. Kumar, and D. Paydarfar. 2022. "Automatic Segmentation and Quantitative Assessment of Stroke Lesions on MR Images." *Diagnostics* 12, no. 9: 2055.
- Walsh, J., A. Othmani, M. Jain, and S. Dev. 2022. "Using U-Net Network for Efficient Brain Tumor Segmentation in MRI Images." *Healthcare Analytics* 2: 100098.
- Wen, C., T. Miura, V. Voleti, et al. 2021. "3DeeCellTracker, a Deep Learning-Based Pipeline for Segmenting and Tracking Cells in 3D Time Lapse Images." *eLife* 10: e59187.
- Wiesmann, V., D. Franz, C. Held, et al. 2015. "Review of Free Software Tools for Image Analysis of Fluorescence Cell Micrographs." *Journal of Microscopy* 251: 39–53.
- Wiggins, L., A. Lord, K. Murphy, et al. 2023. "The Cellphe Toolkit for Cell Phenotyping Using Time-Lapse Imaging and Pattern Recognition." *Nature Communications* 14, no. 1: 1854.
- Yang, M., D. J. Kozminski, L. A. Wold, et al. 2012. "Therapeutic Potential for Phenytoin: Targeting Nav1.5 Sodium Channels to Reduce Migration and Invasion in Metastatic Breast Cancer." *Breast Cancer Research and Treatment* 134, no. 2: 603–615.
- Zhang, Y., P. Zhong, D. Jie, et al. 2021. "Brain Tumor Segmentation From Multimodal MR Images via Ensembling UNets." *Frontiers in Radiology* 1: 704888.

### Supporting Information

Additional supporting information can be found online in the Supporting Information section.