

# Journal Pre-proof

ShapeMetrics: a userfriendly pipeline for 3D cell segmentation and spatial tissue analysis

Heli Takko, Ceren Pajanoja, Kristen Kurtzeborn, Jenny Hsin, Satu Kuure, Laura Kerosuo



PII: S0012-1606(20)30046-4

DOI: <https://doi.org/10.1016/j.ydbio.2020.02.003>

Reference: YDBIO 8207

To appear in: *Developmental Biology*

Received Date: 13 August 2019

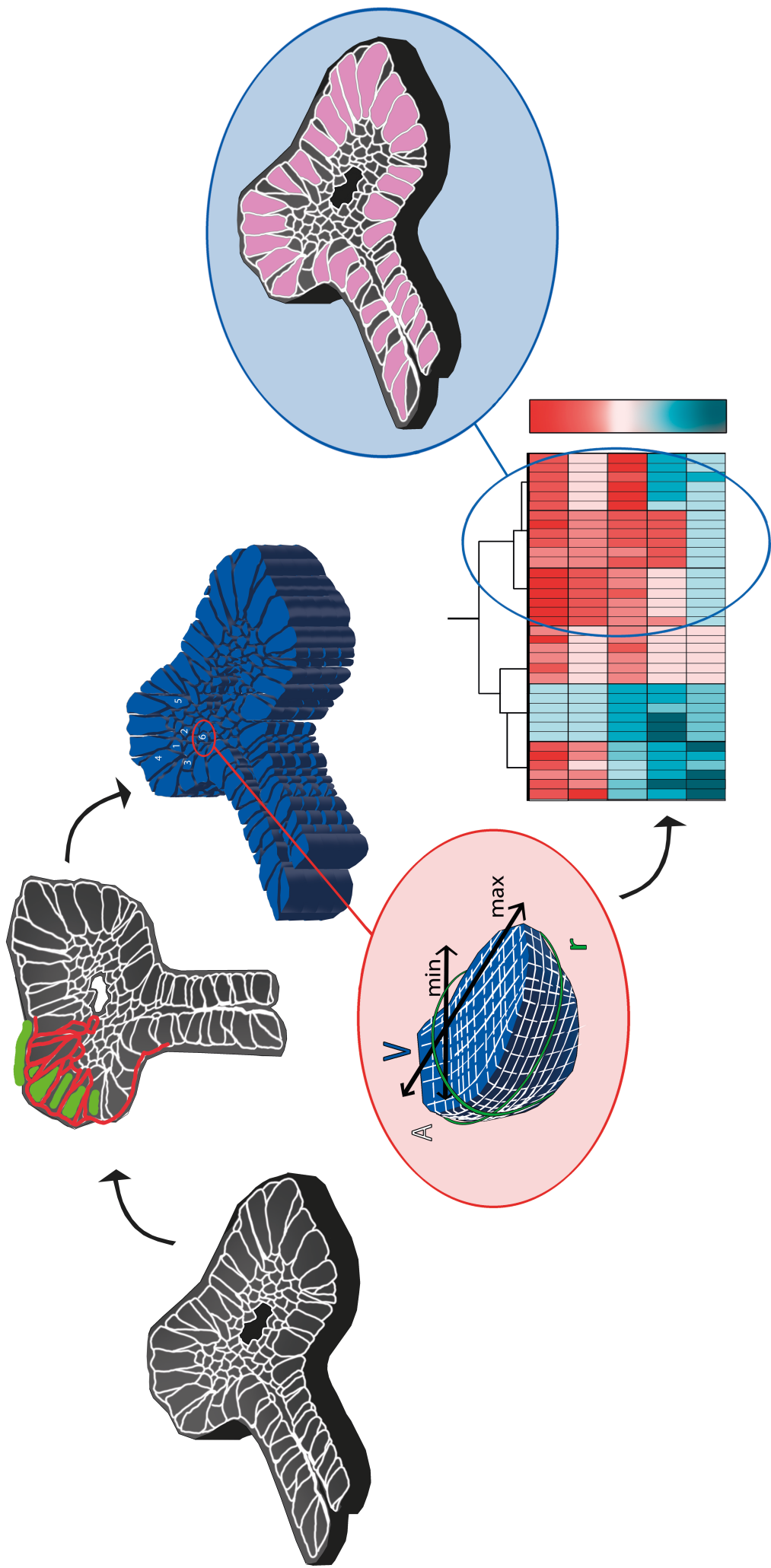
Revised Date: 29 January 2020

Accepted Date: 1 February 2020

Please cite this article as: Takko, H., Pajanoja, C., Kurtzeborn, K., Hsin, J., Kuure, S., Kerosuo, L., ShapeMetrics: a userfriendly pipeline for 3D cell segmentation and spatial tissue analysis, *Developmental Biology*, <https://doi.org/10.1016/j.ydbio.2020.02.003>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Published by Elsevier Inc.



## ShapeMetrics: a userfriendly pipeline for 3D cell segmentation and spatial tissue analysis

Heli Takko<sup>1</sup>, Ceren Pajanoja<sup>1</sup>, Kristen Kurtzeborn<sup>1,2</sup>, Jenny Hsin<sup>3</sup>, Satu Kuure<sup>1,2,4</sup> and Laura Kerosuo<sup>1,3</sup>

1. Department of Biochemistry and Developmental Biology, Biomedicum, University of Helsinki, Finland
2. Stem Cells and Metabolism Research Program, Faculty of Medicine, University of Helsinki, Finland
3. National Institute of Dental and Craniofacial Research, National Institutes of Health, Neural Crest Development and Disease Unit, National Institute of Dental and Craniofacial Research, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA
4. GM-unit, Laboratory Animal Centre, Helsinki Institute of Life Science, University of Helsinki

**Running title:** 3D cell morphology analysis pipeline

**Corresponding author:** [laura.kerosuo@nih.gov](mailto:laura.kerosuo@nih.gov)

**tel:** +1 (301) 827-5105

**Abstract**

The demand for single-cell level data is constantly increasing within life sciences. In order to meet this demand, robust cell segmentation methods that can tackle challenging *in vivo* tissues with complex morphology are required. However, currently available cell segmentation and volumetric analysis methods perform poorly on 3D images. Here, we generated ShapeMetrics, a MATLAB-based script that segments cells in 3D and, by performing unbiased clustering using a heatmap, separates the cells into subgroups according to their volumetric and morphological differences. The cells can be accurately segregated according to different biologically meaningful features such as cell ellipticity, longest axis, cell elongation, or the ratio between cell volume and surface area. Our machine learning based script enables dissection of a large amount of novel data from microscope images in addition to the traditional information based on fluorescent biomarkers. Furthermore, the cells in different subgroups can be spatially mapped back to their original locations in the tissue image to help elucidate their roles in their respective morphological contexts. In order to facilitate the transition from bulk analysis to single-cell level accuracy, we emphasize the user-friendliness of our method by providing detailed step-by-step instructions through the pipeline hence aiming to reach users with less experience in computational biology.

## Introduction

Within the past decade, the criteria for high quality data on *in vitro* cells, and increasingly on intact tissues, are shifting towards a demand for single-cell level data analysis. Acquisition of data at this resolution creates pressure to include computational methods to complement basic biological studies (Yanai and Chmielnicki 2017). To separate cells in 3D, usage of cell segmentation methods is a necessity in order to achieve robust, reliable results. Yet, although computational methods for cell segmentation have been available for decades (Hodneland et al., 2013, Meijering, 2012), many biologists still choose to use at least partially manual approaches, even for complicated quantitative analysis purposes (Ihermann-Hella et al., 2018, Li et al., 2019, Gordon et al., 2018), and thus remain vulnerable to inaccurate measurements and restricted to only eliciting a small proportion of the information held within the data acquired by advanced microscopy. One of the reasons for this gap has been the simple lack of scientists with overlapping skills in computer science and biology, which has rapidly improved during the past decade with the new generation of scientists having better computational knowledge. Given that most computational methods to accomplish cell segmentation were originally developed for applications in other fields (computer vision, robotics, materials science, medical imaging) and are often tailored to meet the needs of the particular tissue in question (Meijering, 2012), unavailability of user-friendly and cost-effective “general” analysis software created specifically for complex biological tissues has also played a role. Even today, when image analysis techniques are constantly improving in order to solve the problems of imaging data complexity and dimensionality, these segmentation applications are not trivial and still fail to catch everything a human eye can see.

The majority of the information obtained from advanced microscopy images in biomedicine as well as cell and developmental biology today is based on visualization of particular molecules in their biological milieu, such as specific transcripts or proteins, by using fluorescent dyes. Less effort has been put on the analysis of morphological changes in a given tissue and how they correlate with a certain phenotype or cellular function. Here, we provide a user-friendly, straightforward pipeline for machine learning-based cell segmentation, based on our previous pipeline for Spatial Genomic Analysis, a method that groups individual cells based on transcriptional profile similarity in their original tissues of origin (Lignell et al., 2017). In our

modified version here, we generated detailed instructions for unbiased grouping of the cells, via creation of a heatmap, based on novel morphological features that we created in MATLAB. Notably, unlike several other current software programs that segment cells based on nuclear recognition, ShapeMetrics is based on staining of the plasma membrane, which enables volumetric measurements of the individual cells. We tested our script on different sample types including mouse embryonic kidney, chick neural tube, and large human epithelial *in vitro* spheroids. The results show clear and statistically significant separation of cells within each tissue according to differences in parameters like longest axis of the cell, elongation, ellipticity, or ratio of cell volume to area. In order to emphasize the user-friendliness and truly make the script approachable for researchers with minimal experience in coding, we provide a thorough readme-file with step-by-step instructions. We believe this software will be a useful tool for a wide research community using image analysis generated data to gain additional information regarding variance of cell morphology in any tissue of interest.

## Methods

### *Overview of the Pipeline*

Our ShapeMetrics pipeline consists of three parts of which the first two are used for the 3D cell segmentation (both modified from Lignell et al., 2017). The first part takes advantage of a pre-existing segmentation software (Ilastik). This machine learning based software provides a superior and unbiased segmentation algorithm applicable to immunostained images. Briefly, Ilastik is used to create a prediction map that serves as a seed (a matrix) for MATLAB to further process the image, which significantly improves the segmentation accuracy (Fig 1A, part 1). The second part of the script (Fig. 1A, part 2) uses the MATLAB-based watershed segmentation for further image processing, which MATLAB as a coding language is specialized in. The possible gaps in cell borders are connected, and thresholding is used to clean the signal and to set the size of the cells to finalize the segmentation. Third (Fig. 1A, part 3), segmented cells in the image are collected into a single unbiased heatmap and divided into subgroups according to cell features that we have created here, such as volumetric and morphological measurements in the post-segmentation analysis. As the final step, the spatial location of cell groups with a certain morphology can be visualized by “mapping them back” to their spatial locations in the original tissue image by using a modification of our previous application (Lignell et al., 2017).

Instructions for the entire pipeline (as well as for the 2D versions presented in this work), including the creation of the prediction map together with our well-commented script for segmentation and further analysis as well as an example sample can all be found in GitHub: (<https://github.com/KerosuoLab/ShapeMetrics>).

#### *Immunostaining of cell membranes*

We used z-stacked images from multiple different sample types to test our script, which is based on cell segmentation from fluorescent staining of cell membranes. A strong staining with high signal to background ratio is a prerequisite for successful segmentation (Fig. 1B). Basically, as long as the membrane staining is readable with Ilastik, our code can be applied to any tissue sample. For the immunostaining, the following standard protocols were used.

Mouse husbandry and procedures were approved by EU legislation and Finnish Animal Care and Use Committee. The developing kidney samples from E12.5 mouse embryos were cultured for 4 h to allow tissue attachment to a nitrocellulose membrane (Millipore), as described previously (Ihermman-Hella et al. 2014). Samples on membranes were subsequently fixed for 10 min in ice-cold methanol, washed 3x 15 min with PBS-0.3% Triton X-100, incubated twice overnight at +4°C with primary antibody against E-Cadherin (BD Transduction Laboratories) to stain epithelial tissue, and N-Cadherin (MA1-91128, ThermoFisher) for the mesenchymal part. The samples were washed 2x 1 hr, then once overnight with PBS-0.3% Triton X-100, and incubated overnight with the respective AlexaFluor647 and AlexaFluor488 secondary antibodies (ThermoFisher).

The chick embryos (stage HH9) were fixed in 4% PFA overnight and washed five times with PBS-0.2% Tween for 45 minutes. The embryos were embedded in OCT embedding compound (TissueTek) , snap frozen in liquid N<sub>2</sub>, sectioned into 20 µm thick cryosections, and immunostained under “para-film coverslips” overnight at +4°C as previously described (Kerosuo et al. 2016). A cocktail of E-Cadherin and Beta-Catenin primary antibodies was used to stain the membranes (BD Transduction 610181 and Abcam ab6301). Similarly, for the chick mesenchyme we used a cocktail of Beta-Catenin and N-Cadherin (MNCD2, Developmental Studies

Hybridoma Bank). The primary antibodies were followed by AlexaFluor647 secondary antibody against E-Cadherin, and AlexaFluor488 against Beta-Catenin and N-Cadherin.

The human epithelial spheres were created by using the embryonic stem cell line H1, which were induced into neuroepithelial spheroids by using a previously described protocol (Bajpai et al., Nature 2010). The spheroids were fixed in 4% PFA at RT for 30 min. The same E-cadherin and Beta-catenin primary antibodies as listed above for the chick tissue were used. PBS-0.15% Triton X-100 was used for permeabilization. Spheres were washed 3x 10 min in PBS-Triton. Blocking was done with PBS-0.15% Triton X-100 and 5% donkey and goat serum overnight at +4°C. Primary and secondary antibodies were incubated for 3 days each at +4°C.

#### *Image acquisition*

The images were captured using confocal microscopes. Notably, epifluorescent microscopes can also be successfully used to create high quality images suitable for the segmentation pipeline (not shown here, see Lignell et al, 2017). All embryonic mouse kidney samples were mounted in 99.5% glycerol on glass slides with coverslip spacers (Invitrogen) and imaged using a Leica TCS SP8 X confocal microscope (63x oil-immersion objective, NA 1.4). The chick neural tube images were acquired by using the Zeiss Axio imager epifluorescence microscope (40x oil-immersion objective, NA 1.3, for figure 1B), and the Andor Dragonfly Spinning disc confocal microscope (63x water immersion NA 1.7; Figs 4 and 5). The epithelial spheroids were imaged by using a Nikon A1R+ confocal microscope (40x oil-immersion objective, NA 1.3.)

#### *Readme-file*

As common practice when sharing a code, this part is a stepwise instruction file to guide the user through the code using an example sample and explaining every step of the code. It thus explicitly explains what to do during each command and how to load in new samples. It should be open on the side as the code is in use. Some of the principles behind the algorithms and basic steps are explained separately below. The script is divided into sections, and we also refer to some of the sections here by using (Sect.).

#### *Ilastik machine learning: creating a prediction map*



Briefly, Ilastik software offers a Pixel Classification workflow, which labels pixels according to the user's pixel annotations. This part outputs a seed that has separated the raw image into values that represent a range of colors, which ultimately will provide MATLAB with instructions on how to interpret the image for further processing, as previously described (Lignell et al., 2017).

The input files (z-stacks) used for the segmentation need to be in .tif format. In order to define cell borders, we used images with strong membrane-specific immunostaining. The Pixel Classification workflow allows users to select the scale of pixel color/intensity, edge filtering, and texture. We have chosen all these features for our trainings.

The next step in this machine learning workflow is to train the random forest classifier to detect the membrane and cell interiors uniquely for each image. In order to start the training, the user needs to add two labels from the "Training" section. These labels correspond to the cell membrane and cell interior/background, respectively. After the selection of one label, one needs to mark the pixels that belong to that label by drawing over those areas. The predictions can be viewed by pressing the "Live Update" button. These predictions are displayed with the label colors on top of the original image. In addition, Ilastik software provides an uncertainty map where it indicates the unsure parts of the prediction results (displayed in turquoise). Users can draw more label annotations to improve the precision of the machine learning. After the user is satisfied with the prediction results, the prediction map can be saved to the disk by selecting "Prediction Export" and "Export" (or "Export all" for multiple images). The probability map file is saved in .h5 format (Fig. 1A, Part 1).

#### *Segmentation in MATLAB by using watershed algorithm*

The next part of the pipeline is to read the input files (user's image) into MATLAB. In the first section of the script, the prediction map is read in by using the MATLAB function `h5read` (Sect. 1.2) and converted into a form that is usable in MATLAB. Similarly, the second input file (the original z-stack image (Sect. 2.2)) is read in by using a function `h5read`.

The next step thresholds the pixel values provided by the Ilastik prediction map. The prediction map in MATLAB is a greyscale image with any pixel values between 0 (black) and 1 (white),

where 0 corresponds to the Ilastik pixel label for cell membrane and 1 for cell interior/background. We have pre-set four different threshold values ( $>0.7$  -  $>0.95$ ) and their visualizations (sect. 4.1) for the user to test. This step is followed by another threshold section where the minimum and maximum limits for cell size are set (sect. 4.3). The selected size threshold values are displayed for all six pixel-value thresholds, and the user should choose the best pixel value from these (sect. 5). Finally, the success of the segmentation can be evaluated by comparing the 3D rendering to the original image and also by confirming that the cell volumes are distributed into multiple peaks between the range of  $0-10^4 \mu\text{m}^3$  (the majority of cells in a given tissue will be in between  $50$  and  $1000 \mu\text{m}^3$ , see Fig. S1A for guidance on good vs bad histograms).

A seed for the watershed algorithm is created with MATLAB function `imimposemin` by using the chosen pixel value threshold together with the original image (sect. 6). Finally, a label matrix that assigns an individual label for each cell is created with the seed and the MATLAB function `watershed` (Fig. 1A, Part 2).

#### *Morphological features and their visualization*

In order to extract the spatial information from each of the label matrix cells, a MATLAB function `regionprops3` (Sect. 9) is used. This function provides volumetric measurements for 3D images. In our script, we utilize some of these built-in parameters, like number of cells and cell volume, while we also created new parameters by providing additional mathematical calculations to better meet the requirements for this type of image analysis on intact tissues. These include longest axis, cell elongation, cell ellipticity, and the cell volume to surface ratio (Fig. 1C). The cells are hierarchically clustered according to the parameters described above, and heatmaps are generated based on their z-scored values. The script provides multiple different combinations of parameters and their heatmaps. Importantly, we have preselected these five parameters that we use for our example images in this manuscript because we feel they provide the most useful information in a biological cell/tissue context. However, the `regionprops3` provides several other parameters, which can be equally utilized by ShapeMetrics if desired (Supplemental Table 1). For comparison we also listed additional parameters from other

software and conclude that `regionprops3` provides the largest selection (Supplemental Table 1).

In section 11 of the script, users select which heatmap to visualize and identify the cells of interest by typing in the branches that correspond to a certain group of cells. The script cuts off the labels of these selected cells from the original label matrix and assigns color to it to create pseudo-colored cells. Section 12.7 provides color choices for this step.

This sub-label is automatically saved to the disk in .tif format, and it can be displayed as a z-stack in Fiji (ImageJ). In Fiji, the original image and the sub-label can be merged together to display the spatial location of the sub-label as z-projection.

## Results

### *Testing of ShapeMetrics in a biologically relevant context*

We have created a new MATLAB-based method, ShapeMetrics, to segment cells and analyze their volumetric details in 3D. According to the pipeline, which is described in the Methods section, we tested ShapeMetrics on multiple samples from different species and tissue types to evaluate the performance of the code in a variety of biological example settings as shown in the results below.

### *Results interpretation: 3D rendition and histograms*

When a sample image consisting of multiple z-stacks, (as shown with maximum projections of our example tissues consisting of a ureteric bud of a developing kidney in Figure 2A, a neural tube in 3A, and a spheroid in 3F) is run through the ShapeMetrics, the result is a 3D label matrix visualized as a 3D rendition (Figs. 2B, 3B, and 3G) showing the volumetric appearance of the tissue with cell borders. The segmentation for these samples resulted in 328, 824, and 3690 individual cells, respectively, and the volumetric measurements for all the respective cells were calculated. As one of the most important features, the volume of each cell is presented separately in the histogram in voxels, the 3D equivalent of pixels, which we converted to  $\mu\text{m}^3$  to provide the results in a biologically meaningful setting (Figs. 2C, 6K, S1B). These histograms show the number of cells with certain volumes in the segmented sample and thereby demonstrate the success of segmentation. Based on the histogram, the majority of cells have volumes roughly within the range of  $10^2$ - $10^4 \mu\text{m}^3$ . As long as the histogram shows (one or several) Gaussian

distributions for the voxel /  $\mu\text{m}^3$  sizes, the segmentation can be considered successful (please see Fig S1A for potential pitfalls).

#### *Results interpretation: heatmaps and spatial visualization*

All calculated volumetric measurement values for each individual cell are presented by using unbiased hierarchical clustering in a heatmap (Figs. 2E, 3D, and 3I). In addition to generating heatmaps from single images, Shapemetrics allows acquiring data from bigger data sets as cells from multiple images can also be clustered into a single heatmap (Fig. 4A).

In the heatmap, individual cells are clustered on the top row (x-axis) while the morphological features are presented on the y-axis (Figs. 2E, 3D, and 3I). The heatmap clusters the individual cells into subgroups (branches) according to their similarity regarding the five volumetric cellular features, which thus present groups with high or low values in certain parameters in correlation with other parametric values. The subgroups of interest can be labeled with a color. The simplified heat map cartoons, typically presented on the right side of the heatmap (Figs. 2E-H and 3D, I) demonstrate the representation of each parameter as high/low values in each color-selected subgroup. These color-selected subgroups can then be “mapped back” and visualized in their original spatial context either as dots representing the centroid of each cell (on the upper row) or with pseudo-colored cells in 3D (on the lower row, Figs. 2F, 3E, 3J and S1B-C). By assigning colors to the chosen groups of cells in the sample with corresponding branch colors, the method thus allows combining the spatial information of individual cells with their morphological information. In addition, the visualization of volumetrically separated cell groups shows the accuracy of the method, since the visualization results can be compared to what is seen in the heatmap to verify that they match (e.g. Figs. 2F, S3C).

#### *Cellular characterization of ureteric bud in embryonic kidney*

We utilized ShapeMetrics for cellular characterization of multiple distinct tissues (ureteric bud and metanephric mesenchyme of mouse embryonic kidney, chicken neural tube, and human epithelial spheroids). Our results in the ureteric bud example show that these epithelial cells clustered into several subgroups (Figs 2A-G). A cartoon of T-bud stage that demonstrates known regional identities and segments of developing ureteric bud is shown in figure 2D (Kurtzeborn et

al., 2018). Highlighting of a few interesting clusters reveals, as expected, that none of the volumetric categories *per se* represent an anatomically defined structure. However, the analysis clearly demonstrates that the vast majority of the cells throughout the ureteric bud epithelium consist of large and elliptical cells (shown in blue in Figs. 2E-G) together with another population of cells that are smaller in volume with equally long axes and are thus elongated (shown in yellow in Figs. 2E-G). According to their small volume to surface area ratio, these yellow cells are less round in shape and may thus represent either cuboidal or polyhedral cells with more surface irregularities such as cellular projections. Finally, small and round cells (shown in green in Figs. 2E-G) are located mainly in the epithelial lumen, and another group of small but elongated and elliptical cells (shown in pink in Figs. 2F-G) are found on the outer edges of the lumen and the epithelium.

The code also allows users to select the number of features investigated. As an example of fewer features, grouping the cells shown in Figure 2A-C but based on only their volume and elongation, we identified a group of large, non-elongated cells (yellow), cells that are small and not elongated (turquoise), as well as a group of small elongated cells colored with pink (Figs 2H-I).

#### *Developing neural tube*

As our second example we used chick neural tube (Figs. 3A-B). The simplified structure of the tissue is demonstrated in a cartoon (Fig. 3C). Similar to our findings in the kidney sample, even though none of the volumetric clusters selected by the heatmap would fully represent any of the anatomically known cell types, clear conclusions could be drawn. For example, we identified the cells with the longest axis and biggest cell volume that are elongated and elliptical and marked them with blue (Figs 3D and E). The vast majority of these cells are found in the epithelium of the developing neural tube and not in the dorsal neural crest domain. On the other hand, many of the neural crest cells are found in the orange population of small round cells with longest axis located around the dorsal midline (Fig 3E). The third interesting cell group identified (highlighted in pink in Figs. 3D and E) are small in size, and although elongated and elliptical, they are less round in shape (polyhedral or otherwise more irregular in their membrane shape). Majority of these cells are found in the neural epithelium on the inner edge that faces the lumen.

Finally, the green population of small, non-elliptical, non-elongated cells with irregular non-round shapes seem randomly distributed across the developing neural tube (Figs 3D-E).

#### *Neuroepithelial spheroids*

The human ES-cell derived neuroepithelial spheroid is large in size with almost 3700 cells expanding through 120  $\mu\text{m}$  in diameter and serves as an example of our method being applicable to whole organoids (Fig 3F-J, S1B-C). Although artificial in nature, these spheres spontaneously form rosettes that mimic neural tube-like structures with small lumens inside the spheres (Fig 3H). Due to the large size of the spheroid and in order to get a better visualization regarding the anatomy, the heatmap and pseudo-coloring shown in figure 3 only represents half of the sample. Of note, the characterization results for all 3700 cells were very similar (not shown).

Characterization of the cell features in the epithelial spheroid (Figs I-J) shows that elongated, large elliptical cells (blue cells) are concentrated around these epithelial rosettes, whereas equally long and elongated cells that have a smaller volume (yellow) are mostly found inside the lumen. On the other hand, the pink group that consist of smaller, elongated cells that are not round and may be cuboidal or otherwise irregular in their shape due to cellular projections are predominantly seen outside the rosettes. Finally, small round cells (green population) are seen throughout the sphere but seem to be predominantly focused around the rosettes (Figs 3I-J).

#### *Statistical differences between samples, between clusters, and between developmental stages*

The results above from three different sample types show that our code works well in all the tissues we tested. Although all our samples were chosen purely as examples with no intention to generate biologically meaningful data comparable to a research project, we were able to reveal trends in cell shapes in certain spatial locations. In order to ultimately apply our method, ShapeMetrics, to acquisition of actual biologically relevant data, next we wanted to evaluate a) to what extent do we detect sample to sample variation within the defined clusters in a heatmap, and b) whether the differences in volumetric parameters between cell clusters are statistically significant.

First, we wanted to confirm that cells from parallel samples (developing kidney ureteric buds from three different E11.5 embryos) within a certain cluster are not statistically different from each other. For this, we clustered all segmented cells from all three samples into one heatmap and, for this example, chose two clusters that were very different from each other (Fig S2A, light blue; small and round and yellow; big and elongated clusters). We plotted the cells of each embryo separately according to the five used parameters. The results, displayed as box plots, show the cells from parallel samples within one cluster are very similar to each other (Fig. S2A). Next, we tested whether the differences in the highlighted clusters, the light blue and yellow clusters, shown in the heatmap (Figs. 4A,B; S2A) are statistically significant. Since biological replicates often are presented as individual data points, we compared average values ( $n=3$  embryos) from both groups by plotting them according to all five parameters used in the heatmap. The results showed the differences in cellular characteristics identified by the ShapeMetrics and indicated by different coloring on the heatmap indeed are statistically significant ( $t$ -test  $p<0.01$ ) for all parameters except V/SA, which is in line with the visual color coding in the heatmap (Fig. 4C). However, box plots also allow the possibility of analyzing the two groups without taking average values by pooling all individual yellow cells into a single group ( $n=646$ ) and comparing them to the group of cyan cells ( $n=654$ ). This way, the high amount of data points radically diminished the  $p$ -values, which was significant ( $p<0.01$ ) also for V/SA, although the value was much higher than for the other parameters (Fig. 4D).

Finally, we wanted to showcase an example of using ShapeMetrics for monitoring of a certain morphologically distinct subgroup of cells during development. We chose two timepoints during kidney development (E11.5 and E12.5 ureteric bud samples, respectively). As described above, we clustered all segmented cells from each developmental time point into a single heatmap, and chose two similar clusters from the two respective heatmaps for comparison: pink cells that are big, elliptical and have a high V/SA ratio and yellow cells that are elongated, elliptical and have the longest axis but a smaller volume than other cells included in the comparison (Fig 5A-B, S3). For the cells in the pink clusters, when average values of cells representing each biological sample are compared (E11.5  $n=3$ ; E12.5  $n=3$ ), the box plots show similarity ( $t$ -test  $p>0.01$ ) between the two developmental stages for all five parameters both in the pink (Fig. 5C) and yellow groups (Fig. 5D).

### *Equal segmentation success in epithelial and mesenchymal tissue*

Another possible bias we wanted to rule out was potential tissue type-dependent unequal segmentation performance. For this, we immunostained mesenchymal and epithelial tissues separately within the same developing kidney sample and evaluated their segmentation success. As expected, the 3D rendering and histograms for cell volumes looked reliable and we did not find any reason to suspect systematic bias in the success rate between epithelium and mesenchyme (Figs. 6A-D). We also successfully segmented mesenchymal cells from the developing HH9 chicken embryo (Fig. S2B). We conclude that as long as the membrane staining is strong and the signal to noise ratio is high, the tissue type is not a contributing factor for the success of the segmentation.

### *Combining volumetric data with traditional biomarker results*

Finally, we asked whether we can, as an additional feature of ShapeMetrics, combine volumetric information with other biomarker expression data such as protein expression. For this, we used an image of the developing neural tube that was immunostained with an antibody to Sox9 that marks the neural crest cells in the dorsal part (Kerosuo and Bronner-Fraser, 2012; Fig.6E). The Sox9+ nuclei were segmented in order to define them as objects inside the segmented cells, taking advantage of the fact that all segmented cells in our code are assigned an identification number (Fig. 6E-I). Of all the 824 cells, 150 (18%) were Sox9-positive (Fig 6 J-L). Next we clustered all the cells and displayed the different subgroups in a heatmap accordingly (Figure 6M that is also previously shown in 3D). In order to find out more information about the volumetric properties of the Sox9+ cells, we tested which clusters contained most of them. Indeed, the majority of the Sox9+ cells (137 cells, 91% ) were found in one of the two main branches of the heatmap that consisted of all the large cells with the longest axis (divided into three subclusters shown in yellow, pink and green) and also were non-elongated with high volume to membrane surface ratio (Figs 6M-O).

### *ShapeMetrics segmentation capacity is excellent as compared to other existing methods*



Next, in order to further evaluate the segmentation performance of our ShapeMetrics code, we compared it to other existing free software such as CellProfiler<sup>TM</sup> (McQuin et al., 2018), ImageJ, and CellSegm as well as the commercially available image processing program Imaris (Fig 7). Analysis of the same sample with all four methods revealed that while our software performed equally well or better in 2D as CellProfiler<sup>TM</sup> or ImarisCell (the module designed specifically for 2D and 3D cell analysis), which were the best of the pre-existing methods, our ShapeMetrics performed extremely well in 3D segmentation and resulted in similar cell count and volume numbers as the commercial ImarisCell. ShapeMetrics outcompeted all others in recapitulating the shape of the cells in the 3D rendering as compared to the original image. Notably, none of the pre-existing non-commercial methods or “basic” Imaris were able to produce meaningful 3D volumetric data or cell counts based on membrane staining (Fig. 7, S4 and Table 1). Furthermore, user-friendliness was not emphasized in any of the pre-existing methods including ImarisCell. To note, even though CellProfiler<sup>TM</sup> provides a package for Linux platform, its usage in our hands was problematic and seemed poorly compatible. Furthermore, we also tested CellSegm (Hodneland et al., 2013), yet another published approach for cell segmentation. However, rather than being a complete pipeline like the other software we used for comparison, CellSegm is more of a display of a selection of individual MATLAB functions (some of which share similar features that we used in our ShapeMetrics code) available to be included in the end user’s script if one were to create a new code for individual novel needs.

We also performed nuclear segmentation comparisons. In 3D, our method, ShapeMetrics, and ImarisCell were superior in their performance, whereas in 2D CellProfiler<sup>TM</sup> was also excellent (Fig S4, Supplemental Table S2, and Table 1). However, we analyzed several additional samples which showed that if the nuclei were closely packed to each other, ShapeMetrics was not optimal in separating them as individual objects (not shown). We don’t thus recommend ShapeMetrics for nuclear segmentation for densely packed cells. Taken together, these results suggest that our method developed here, ShapeMetrics, is excellent and accurate for cell membrane staining-based segmentation and volumetric analysis. Most importantly, none of the other methods provide further analysis on cell shape in a single cell level, nor do they provide information on the spatial location in the tissue.

## **Discussion and Conclusions**

The techniques for cell and nuclear segmentation have been available for decades (Meijering 2012), but methods that actually perform accurate 3D segmentation and volumetric analysis at the cellular level are still missing. Depending on the particular needs of given experiments, multiple possibilities for executing the required tasks have been used and recently improved (Boutin et al., 2018, Caicedo et al., 2017, Hodneland et al., 2013, Lignell et al., 2017, Lou et al., 2014, Molnar et al., 2016). Although several methods partially take advantage of model-based choosing of the algorithms and manual optimization of parameters using histogram-based thresholding, these approaches do not provide tools for addressing hurdles that are faced with robust high-throughput 3D segmentation of complex tissue samples (Caicedo et al., 2017, Xinghua et al., 2014). Here, we developed ShapeMetrics, a new cell analysis pipeline capable of performing unbiased segmentation of membrane-stained cells in 3D according to their volumetric and morphological features. ShapeMetrics accurately segregates tissue-residing cells based on their biologically meaningful features and allows mapping them back to the original organ structures. By these means, ShapeMetrics provides a new, user-friendly method with unique features for single-cell level data analysis.

ShapeMetrics is specialized in separation of individual cells based on their morphological features in 3D. Based on the volumetric differences in morphology we clustered cells into subgroups in three different example tissues and visualized (pseudo-colored) them in their original images. Our examples show that cells from parallel samples clustered in the same population are similar with each other, while differences between cell features in different clusters are statistically significant. Likewise, volumetric values of cells clustered in similar subgroups in two different developmental stages were shown to be statistically similar. To note, an alternative option we did not showcase here, for comparing similarities between biological timepoints, is to combine all segmented cells from several developmental stages into one single heatmap and then analyze and compare the spatial localization of the cells that are clustered together.

With these results we conclude ShapeMetrics can be applied to generate biologically meaningful data and to perform comparisons between spatially, developmentally or phenotypically different subgroups. In sum, we hope to provide a useful tool for computational single-cell analysis for

routine phenotype and tissue characterization. Although our goal here is to provide a method for volumetric tissue analysis, our results in the example tissues already allow us to also draw preliminary biological conclusions. None of the selected subpopulations in our examples, however, *per se* fully corresponded to known anatomically defined domains, which likely reflects the fact that structural domains are complex and constructed based on several variables with cell volumetrics only being one of them. Due to lack of userfriendly and accurate volumetric analysis softwares like Shapemetrics, we currently know very little about the impact of volumetric properties of cells on defining developmental stages or pathological states. We believe usage of our method will thus not only complement but also extend the information that is routinely acquired from imaging data to help us get started in understanding the role of cell shapes and sizes in tissue architecture on single cell level.

ShapeMetrics identifies individual cells based on approximately twenty previously identified MATLAB parameters. Of these we preselected the ones we think provide the most valuable information in a cell biological context such as e.g. principle axis length as well as the new parameters cell elongation, longest axis, ellipticity, or the prediction of round vs. irregular or more pointy shape. Some similar features are provided in a recent analysis on spheroids, where the approach combines primary detection of 2D and 3D nuclear clusters followed by a robust watershed-based segmentation. However, while this method allows performance of volumetric calculations for the spheroids and to some extent provides details about the nuclei shapes it does not allow volumetric analysis of individual cells (Boutin et al., 2018). Similarly, another MATLAB implemented approach, SpheroidSizer, specializes in volumetric measurements of spheroids but not individual cells (Chen et al., 2014).

The performance of our new MATLAB-based pipeline, ShapeMetrics, was compared to multiple existing free programs including ImageJ/FIJI, CellProfiler<sup>TM</sup>, and CellSegm as well as the commercially available Imaris and its application ImarisCell specifically developed for this type of 3D and 2D analysis of cells. We conclude that in 2D analysis CellProfiler<sup>TM</sup> and ImarisCell performed equally well as our method ShapeMetrics. However, ShapeMetrics and ImarisCell were superior to all the other methods for 3D analysis, and ShapeMetrics in fact turned out to be the only one capable of providing additional information about cell shapes and sizes.

A fairly recent paper nicely brings together a selection, which they named CellSegm, of 13 previously established MATLAB-based applications/scripts that perform 3D cell segmentation based on both membrane and nuclear staining and provides steps for illumination correction and image smoothing, although none of these scripts *per se* contain full segmentation pipelines ready for usage (Hodneland et al., 2013). An obstacle not addressed in our code, segmentation of overlapping cells, is overcome in a 2D set up of circular cells by Molnar and colleagues (Molnar et al., 2016). However, our method uses watershed segmentation combined with Ilastik machine learning that greatly improves the accuracy of the 3D segmentation of cells intermingled in intact *in vivo* tissues. This is similar to CellProfiler<sup>TM</sup>, another open source software that uses similar advanced algorithms for precise and sophisticated 2D cell segmentation but lacks the ability to perform true 3D analysis (McQuin et al., 2018).

Furthermore, all these previously existing applications mentioned above lack the fundamental feature of spatial mapping, which is a unique feature of ShapeMetrics and our previously published tool SGA that served as the base for the cell segmentation part in our method described here (Lignell et al., 2017, Lignell and Kerosuo 2019). ShapeMetrics works equally well on images from any tissue type as long as the membrane staining is successful and outputs a high signal to noise ratio. Finally, we provide an option of combining the volumetric analysis with traditional biomarker expression information to further explore the identity of the segmented cells. These are essential features required for the detailed cellular characterization of complex tissues and organs *in vivo*, which together with user-friendliness makes our script applicable for wide range of users, regardless of discipline and computational skills.

In order to reach a wide range of potential users, including those with limited coding experience, we have emphasized on making the usage of our script as user-friendly as possible by providing thorough commenting and additional step-by-step instructions in the readme-file, unlike some of the existing methods we tested for comparison. Several people trained in biology with moderate coding skills successfully tested our code, which gives us confidence to describe ShapeMetrics as a user-friendly tool. Furthermore, it's worth noting that even though our example samples in this work are all based on confocal microscopy images, our previous work shows that the cell

segmentation part is also applicable for samples imaged with an epifluorescent microscope (Lignell et al, 2017), which we hope will make the use of ShapeMetrics applicable for as many researchers as possible. In sum, our goal is that our script and others will push the trend and demands of our field forward so that the next decade will implement usage of single-cell level quantitative analysis as a standard approach to be routinely included in biomedical and developmental biology research.

### **Acknowledgements**

This research was supported in part by the NIDCR imaging core ZIC DE000750-01 and we thank Dr. Duy Tran for technical assistance. We also thank the Biomedicum Imaging Unit at the University of Helsinki for assistance with imaging as well as Laboratory Animal Centre at the University of Helsinki for housing the mice. We thank Dr. Antti Lignell for conversations during the initiation phase of the project.

Work in the laboratories at the University of Helsinki was supported by grants from the Academy of Finland (to LK and SK, respectively), Sigrid Juselius Foundation (to LK), The Finnish Cultural Foundation (to SK), Väre Foundation (to CP), The Finnish Kidney Foundation (KK), Research Program Unit, Faculty of Medicine, University of Helsinki (SK), and in part by the Division of Intramural Research of the National Institute of Dental and Craniofacial Research at the National Institutes of Health, Department of Health and Human Services (LK).

## Figure Legends

### Figure 1. Pipeline for the segmentation and volumetric analysis.

(A) **Part 1: Ilastik machine learning.** The original z-stack image of epithelial ureteric bud in developing kidney (.tif) used as an input file. *Pixel annotations* are made by manual labeling of the membrane (red) and the background (green). *Prediction overlay* based on the labels that have been used for the training, colors indicate the predicted pixel values: red = 0, green = 1. *Uncertain pixels* are marked with cyan to indicate the pixel value 0.5. *Prediction map* is the final file obtained after full training; the file consists of a matrix array of pixel values between 0 and 1, with one value corresponding to each pixel in the original input image. **Part 2: Segmentation in MATLAB.** *Prediction map* serves as an input file for the MATLAB segmentation script. *Thresholding* the prediction map selects all the pixels with a value greater than 0.95. *Substraction* sets the size threshold for the pixel clusters. This allows getting rid of the white background (that survived from the pixel value threshold) and only the cells are left. *The (watershed) segmentation* is done by using the thresholded prediction map together with the original z-stack image as the seed for the watershed segmentation algorithm. *3D rendering* of the final watershed label matrix of segmented cells. The *histogram* shows the distributions of the cell volumes. **Part 3: Extraction of spatial parameters in MATLAB.** The *heatmap* visualizes the hierarchical clustering of spatial parameter values calculated individually for each cell in the segmented label matrix. Each row corresponds to one parameter whereas each column corresponds to each individual cell. Red color points to high representation of a certain parameter and blue to low, respectively. *Spatial localization* is based on the heat map clusters where each clustered group of cells is mapped back to their spatial location and visualized by pseudo-coloring the cells using the same, respective color-code. (B) A strong and specific immunostaining is a requirement for successful cell segmentation as shown in single plane images of our example tissues, which are from the top: ureteric bud in E13.5 mouse kidney, embryonic HH9 chicken neural tube, and a human neuroepithelial spheroid. Scale bar 40  $\mu\text{m}$ .

(C) Presentation of the volumetric parameters used in the ShapeMetrics script. *Number of cells:* Number of cells is calculated from the final segmented label matrix as a number of distinguishable (separable) group of voxels. *Cell volume:* Calculated as the number of voxels in each cell. *Longest Axis:* Length of the longest axis out of three principal axes from parameter

“Principal axis length” extracted from Matlab function `regionprops3`. *Cell Elongation*: Longest axis length divided by the average length of intermediate and minor axis. High values of elongation are presented in red and small values in blue. *Cell ellipticity*: As demonstrated in the figure, ellipticity is calculated by dividing the subtraction of the longest and minor axis lengths by the longest axis length. *Volume-to-surface area ratio*: This parameter is calculated by dividing the volume with surface area. It shows the difference between round and platonic cells (polyhedronic or irregular spiky membranes). A ball will display the darkest intensity of red.

**Figure 2. Analysis of the segmented ureteric bud in developing kidney: hierarchical clustering and visualizations.** (A) A maximum projection of the ureteric bud raw image. Scale bar 40  $\mu\text{m}$ . (B) 3D rendition. (C) Histogram showing the distribution of the segmented cell volumes in cubic microns. (D) Schematic illustration of the ureteric bud in developing mouse kidney. The ureteric bud is an epithelial tube with lumen (pink), and it is compartmentalized into tip (aqua) and trunk (light blue) regions. (E) Hierarchical clustering of the distinct cells based on their differences identified by the script according to the five used volumetric features. The individual clusters with interesting features, chosen as examples here, are color-labelled. The same clusters are also illustrated in the simplified table on the right. For example, the cells in the green population are round and small, and they are not elongated. (F) Visualization of the selected subgroups by mapping them back to the original image, as shown by single color images. The upper panel shows centroids of the segmented cells marked on top of the maximum projection image. The lower panel shows the visualization of selected clusters by filling the respective cells from a selected cluster with a pseudo-color. (G) Magnification and overlay of two-colored cell clusters, respectively, and all clusters together as indicated by the dashed box. (H) Hierarchical clustering of the ureteric bud in developing kidney by using only cell elongation and volume as parameters. The heatmap shows the existence of subgroups, among other subgroups not color-selected here, that represent either small elongated cells (pink), large non-elongated cells (yellow), or small non-elongated cells visualized (I) separately or as an overlay.

**Figure 3. Analysis of the segmented neural tube and spheroid: hierarchical clustering and visualizations.** (A) A maximum projection of the neural tube raw image. Scale bar 40  $\mu\text{m}$ . (B) 3D rendition. (C) Schematic illustration of the developing Hamburger-Hamilton Stage 9 chick

neural tube. **(D)** Hierarchical clustering of the cells based on their differences identified by the script according to the five used volumetric features. The individual clusters of interest are color-labelled and also illustrated in the simplified table on the right. For example, the cells in the pink population are small, elongated and elliptical. **(E)** Visualization of the selected subgroups by mapping them back to the original image, as shown by single color images as well as magnification and overlay of two-colored cell clusters, respectively, and all clusters together as indicated by the dashed boxes. **(F)** A maximum projection of the raw image of an epithelial spheroid. Scale bar 40  $\mu\text{m}$ . **(G)** 3D rendition. **(H)** Schematic illustration of the neuroepithelial spheroid/organoid. **(I)** Hierarchical clustering of the cells based on the differences identified by the script according to the five used volumetric features. The individual clusters of interest are color-labelled and also illustrated in the simplified table on the right. For example, the blue cells are big, have the longest axis and are elongated. **(J)** Visualization of the selected subgroups by mapping them back to the original image, as shown by single color images as well as magnification and overlay of two-colored cell clusters, respectively, and all clusters together as indicated by the dashed boxes.

**Figure 4: Differences between hierarchical clusters are statistically significant.**

**(A)** All 5464 segmented cells from three developing kidney samples from E11.5 were pooled into one heatmap. Two subgroups that, according to the heatmap, represent different values for elongation, ellipticity and longest axis were selected for statistical comparison (high in yellow vs low in the cyan group, as indicated by colored arrows). **(B)** Maximum projections of each sample image with the number of segmented cells in the respected embryo displayed in white. Visualization of the individual cells in the selected clusters in their respective original sample images are shown in the lower row with the number of segmented cells belonging to the respected colored group in the heatmap displayed. Scale bars 50  $\mu\text{m}$ . **(C)** Comparison of the two selected groups for each of the five parameters displayed as box plots of average values ( $n=3$  for both yellow and cyan groups, respectively) shows statistically significant ( $t$ -test  $p<0.01$ ) differences for all parameters except volume to surface area ratio ( $V/SA$ ), which is in accordance with the information provided by the heatmap. The blue dots represent the individual data points. **(D)** An alternative way of presenting the same data is to pool values of the individual cells into the box plot. Although the results provide the same information than shown in Fig. 4C, the high



increase in data points changes the p-values to be tremendously smaller, and the p-value for V/SA is now smaller than 0.01. The line in the middle represents the median value, and the first and second quartiles are defined inside the box. The whiskers mark the calculated maximum and minimum value within the data points counted as results, outliers are shown in red.

**Figure 5:** Comparison of ureteric buds at two different developmental stages. **(A)** All segmented cells from three developing kidney samples from embryonic day E11.5 and E12.5, respectively, were pooled into two heatmaps. Two colored clusters with distinct profiles were chosen from both heat maps: pink cells representing large, elliptical nonelongated cells while the yellow cells are small but elongated, elliptical and have the longest axis. The groups were chosen based on their similar profiles in both heat maps. **(B)** Maximum projections of the original images of each sample together with visualization of the spatial localization of the individual selected cells in the original images. Scale bars 50  $\mu\text{m}$ . **(C)** The box plots show comparison of the differences in the five individual parameters within cells in the pink groups between E11.5 and E12.5, shown as average values from three different embryos from each age group ( $n=3$  per stage). As expected, there is no statistically significant difference between any of the parameters (ttest  $p>0.01$ ) in the pink group **(D)** or the yellow group. The blue dots represent the individual data points.

**Figure 6: ShapeMetrics can be combined with traditional biomarker analysis.** **(A)** ShapeMetrics shows no bias between different tissue types as shown by segmentation of mesenchyme and epithelium from the same embryonic kidney sample stained with two different cadherins (E-cadherin in red and N-cadherin in green). Scale bar 50  $\mu\text{m}$ . **(B)** Maximum projections of the distinct channels, **(C)** 3D renderings and **(D)** histograms showing distribution of cell volumes in a realistic biological range (epithelium in the upper row and mesenchyme in the lower row). **(E)** Single stack of the original neural tube image used for the analysis and **(F)** the respective 3D rendition and **(G)** corresponding histogram of cell volume distributions. **(H)** Sox9 immunostaining for the same sample and **(I)** the resulting 3D rendition of the Sox9-positive nuclei. **(J)** 3D rendition of the cells that express Sox9 (i.e. share the same voxel coordinates with the Sox9 rendition). **(K)** The histogram of volume distributions as well as **(L)** the spatial localization of these cells labelled with pseudo-coloring in the original image. **(M)** Heatmap of all the cells segmented from the sample. One of the two main branches that contains the large

cells is selected in red. **(N)** Visualization of the spatial localization of the cells included in the red cluster from the heat map. **(O)** Visualization of only the Sox9 positive cells in the red subgroup presented by pseudo-coloring on the original image.

**Figure 7: Comparison of 2D and 3D segmentations performed with ShapeMetrics and other pre-existing segmentation software.** **(A)** The original image used for all softwares. **(B) FIJI:** visualization of 2D (left) and 3D (right) segmentation results of the binary watershed segmentation option. **(C) CellProfiler:** Visualization of the 2D (top) and 3D (below) monolayer segmentation results using IdentifyObjects and Measure modules. **(D) ImarisCell:** Visualization of the 2D (left) and 3D (right) segmentation results. **(E) ShapeMetrics:** visualization of the 2D (left) and 3D (right) watershed segmentation results.

**Table1:** The quantitative results of the comparison show that our code ShapeMetrics outputs very similar cell numbers and volumes as ImarisCell. Fiji, on the other hand, outputs a similar number of cells but the mean volume of the cells as well as the visualization in figure 7 reveal that the segmentation is not successful. CellProfiler is not applicable for 3D segmentation of cells.

## References

- Bajpai, R., Chen, D. A., Rada-Iglesias, A., Zhang, J., Xiong, Y., Helms, J., Chang, C. P., Zhao, Y., Swigut, T., Wysocka, J., 2010. CHD7 cooperates with PBAF to control multipotent neural crest formation, *Nature*. 463, 958-62.
- Boutin, M. E., Voss, T. C., Titus, S. A., Cruz-Gutierrez, K., Michael, S., Ferrer, M., 2018. A high-throughput imaging and nuclear segmentation analysis protocol for cleared 3D culture models, *Scientific reports*. 8(1), 11135.
- Caicedo, J. C., Cooper, S., Heigwer, F., Warchal, S., Qiu, P., Molnar, C., ..., Wawer, M., 2017. Data-analysis strategies for image-based cell profiling, *Nature Methods*. 14(9), 849.

- Chen, W., Wong, C., Vosburgh, E., Levine, A. J., Foran, D. J., Xu, E. Y., 2014. High-throughput image analysis of tumor spheroids: a user-friendly software application to measure the size of spheroids automatically and accurately, *JoVe (Journal of Visualized Experiments)*. (89), e51639.
- Gordon, H. B., Lusk, S., Carney, K. R., Wirick, E. O., Murray, B. F., Kwan, K. M., 2018. Hedgehog signaling regulates cell motility and optic fissure and stalk formation during vertebrate eye morphogenesis, *Development*. 145(22), dev16506811(4).
- Hodneland, E., Kögel, T., Frei, D. M., Gerdes, H. H., Lundervold, A., 2013. CellSegm - a MATLAB toolbox for high-throughput 3D cell segmentation, *Source code for biology and medicine*. 8(1), 16.
- Ihmann-Hella, A., Lume, M., Miinalainen, I.J., Pirttiniemi, A., Gui, Y., Peränen, J., Charron J., Saarna, M., Costantini, F., Kuure, S., 2014. Mitogen-activated protein kinase (MAPK) pathway regulates branching by remodeling epithelial cell adhesion, *PLOS Genetics*. 10(3), e1004193.
- Ihmann-Hella, A., Hirashima, T., Kupari, J., Kurtzeborn, K., Li, H., Kwon, H.N., Cebrian, C., Soofi, A., Dapkunas, A., Miinalainen, I. and Dressler, G.R., 2018. Dynamic MAPK/ERK activity sustains nephron progenitors through niche regulation and primes precursors for differentiation, *Stem cell reports*, 11(4), 912-928.
- Kerosuo, L., Bronner, M. E., 2016. cMyc Regulates the Size of the Premigratory Neural Crest Stem Cell Pool, *Cell Reports*. 17(10), 2648-2659.
- Kerosuo L, Bronner-Fraser M.; 2012. What is bad in cancer is good in the embryo: importance of EMT in neural crest development. *Semin Cell Dev Biol*. 2012 May;23(3):320-32. Review.
- Kurtzeborn K, Cebrian C, Kuure S., 2018. Regulation of Renal Differentiation by Trophic Factors. *Front Physiol*. 12;9:1588. eCollection. Review.

- Li, H., Jakobson, M., Ola, R., Gui, Y., Kumar, A., Sipilä, P., Sariola, H., Kuure, S., Andressoo, J. O., 2019. Development of the urogenital system is regulated via 3'UTR of GDNF, *Scientific reports*. 9(1), 5302.
- Lignell, A., Kerosuo L., 2019. Spatial Genomic Analysis: A Multiplexed Transcriptional Profiling Method that Reveals Subpopulations of Cells Within Intact Tissues, *Methods Mol Biol*. 2002, 151-163.
- Lignell, A., Kerosuo L., Streichan, S. J., Cai, L., Bronner, M. E., 2017. Identification of a neural crest stem cell niche by Spatial Genomic Analysis, *Nature communications*. 8(1), 1830.
- Lou, X., Kang, M., Xenopoulos, P., Munoz-Descalzo, S., & Hadjantonakis, A. K., 2014. A rapid and efficient 2D/3D nuclear segmentation method for analysis of early mouse embryo and stem cell image data, *Stem cell reports*. 2(3), 382-397.
- McQuin, C., Goodman, A., Chernyshev, V., Kamensky, L., Cimini, B.A., Karhohs, K.W., Doan, M., Ding, L., Rafelski, S.M., Thirstrup, D. and Wiegraebe, W., 2018. CellProfiler 3.0: Next-generation image processing for biology, *PLOS Biology*. 16(17), p.e.2005970.
- Meijering, E., 2012. Cell segmentation: 50 years down the road [life sciences], *IEEE Signal Processing Magazine*. 29(5), 140-145.
- Molnar, C., Jermyn, I. H., Kato, Z., Rahkama, V., Östling, P., Mikkonen, P., ..., Horvath, P., 2016. Accurate morphology preserving segmentation of overlapping cells based on active contours, *Scientific reports*. 6, 32412.
- Yanai, I. and Chmielnicki, E., 2017. Computational biologists: moving to the driver's seat, *Genome biology*. 18(1), p.223.

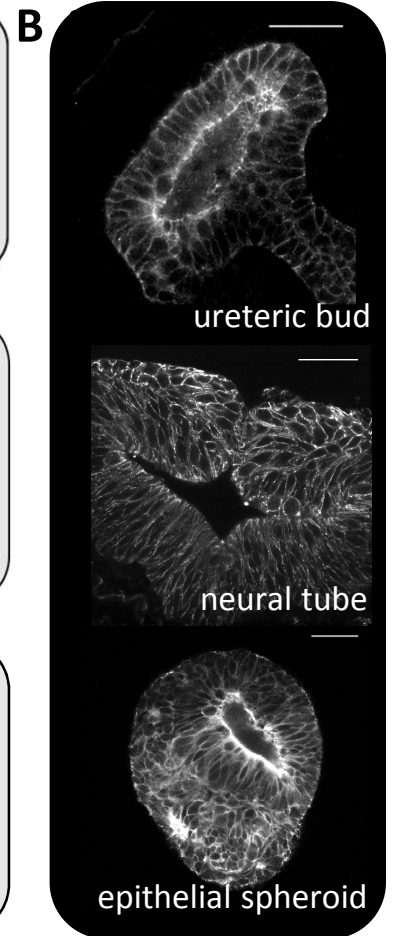
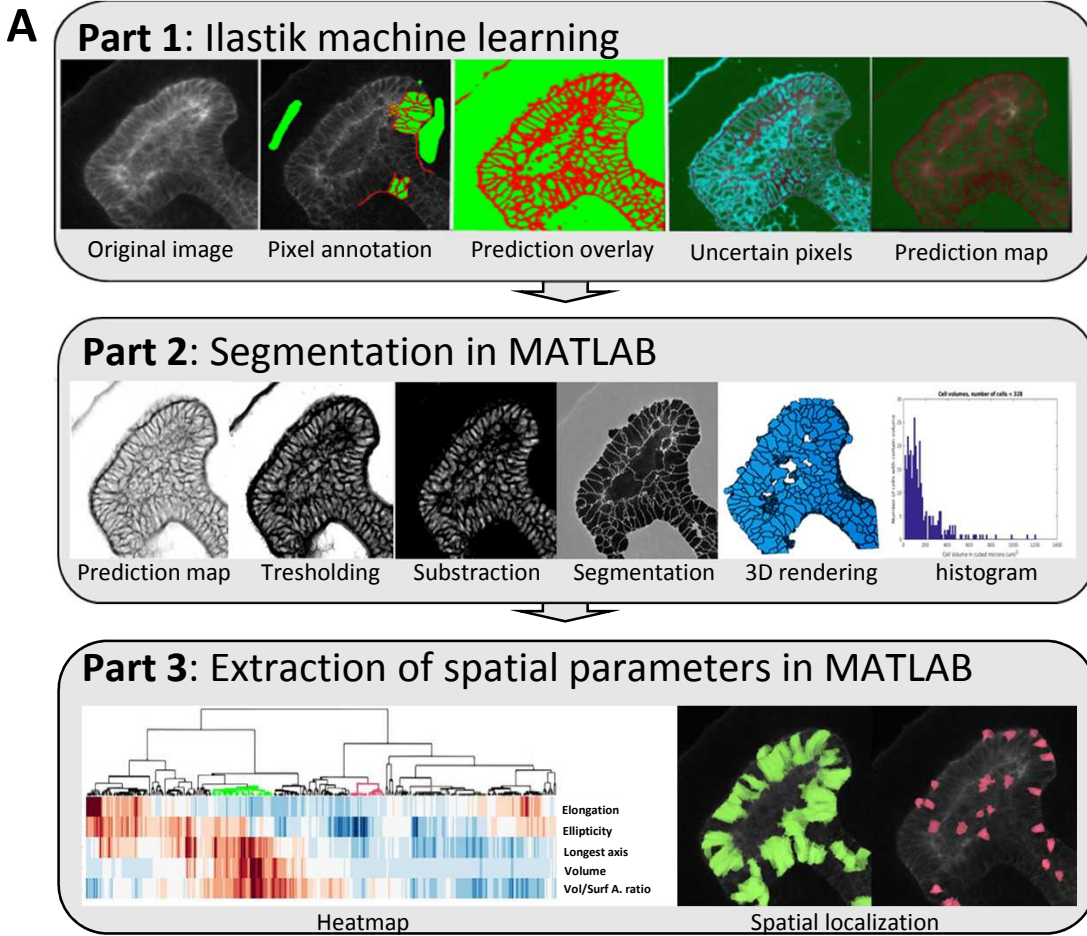
**KEY RESOURCES TABLE (Takko et al.)**

Reagent or resource	Source	Identifier
<b>Antibodies</b>		
Mouse anti E-cadherin	BD Transduction Laboratories	Cat # 610181
Mouse anti beta-catenin	Abcam	ab6301
Rat anti N-cadherin	Hybridoma Bank	MNCD2
Mouse anti N-Cadherin	Thermofisher	MA1-91128
Secondary antibodies	Molecular probes	Alexa fluor 647, 488
<b>Bacterial and Virus Strains</b>		
<b>Biological Samples</b>		
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
Human neural crest induction protocol	Bajpai et al, 2010	
Neurobasal Medium	Gibco	Cat # 21103049
DMEM/F12 w/ GlutaMax	Gibco	Cat # 10565018
EGF	Sigma Aldrich	Cat #E9644
FGF	Peprtech	Cat #100-18B
Insulin	Sigma Aldrich	Cat #11070-73-8
Matrigel HESC qualified	Cornig	Cat# 354277
mTeSR	Stem cell technologies	Cat #85850
<b>Critical Commercial Assays</b>		
<b>Deposited Data</b>		
<b>Experimental Models: Cell Lines</b>		

Human embryonic stem cell line H1	Commercial source	the Wisconsin stem cell bank
Experimental Models: Organisms/Strains		
Mouse embryos; C57BL/6J01aHsd	Commercial source	Envigo
Chicken embryos	Commercial source local farm	Pavelan siitoskanala, Finland
Oligonucleotides		
Recombinant DNA		
Software and Algorithms		
Cell segmentation based on volumetric features	This paper	<a href="https://github.com/KerosuoLab/ShapeMetrics">https://github.com/KerosuoLab/ShapeMetrics</a>
ilastik	free downloadable software	<a href="https://www.ilastik.org/">https://www.ilastik.org/</a>
MATLAB	Commercial source	<a href="https://www.mathworks.com">mathworks.com</a>
Other		

Figure 1

ShapeMetrics PIPELINE



**C**

**Number of cells**

**Cell Volume**

**Longest Axis:**  
Matlab parameter "Principal Axis Length" gives 3 major axis of ellipsoid (length in voxels),  
Major Axis = Longest Axis

**Cell Ellipticity:**  
(measure of the compression of a circle or sphere along a diameter to form an ellipse or a spheroid respectively)

$$f = \frac{a - b}{a}$$

Longest axis - Minor axis / Longest Axis

**Cell Elongation:**  
Longest Axis / average of Intermediate and Minor axis

	Different long axis	Same long axis
Bigger value for elongation		
Smaller value for elongation		

High value (red) / Low value (blue)

**Cell Volume / Surface Area Ratio:**  
For the same volume;  
Surface area is  
-Smaller for rounder shapes  
-Bigger for platonic solids

Bigger value for V/SA:

Smaller value for V/SA:

Figure 2

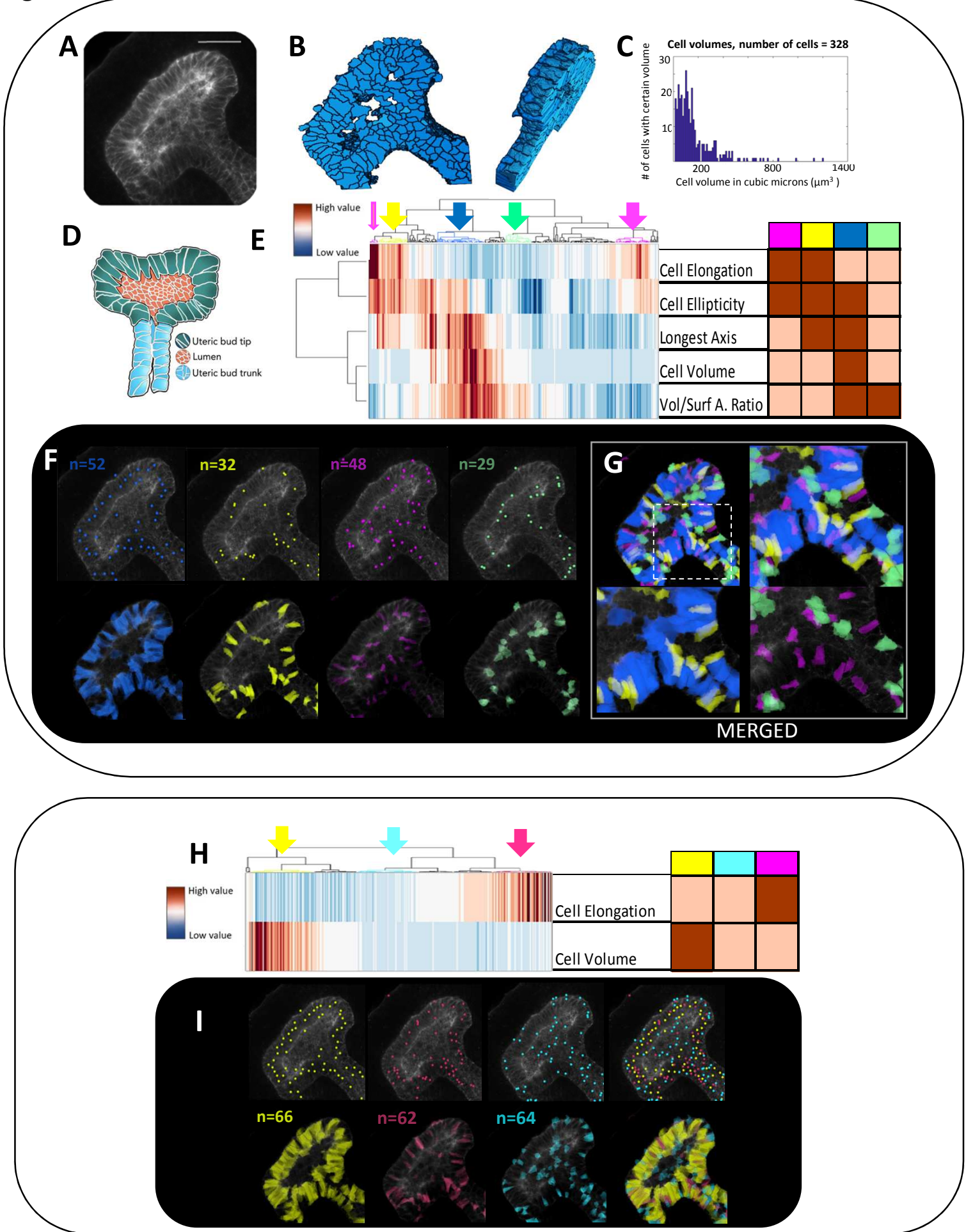




Figure 3

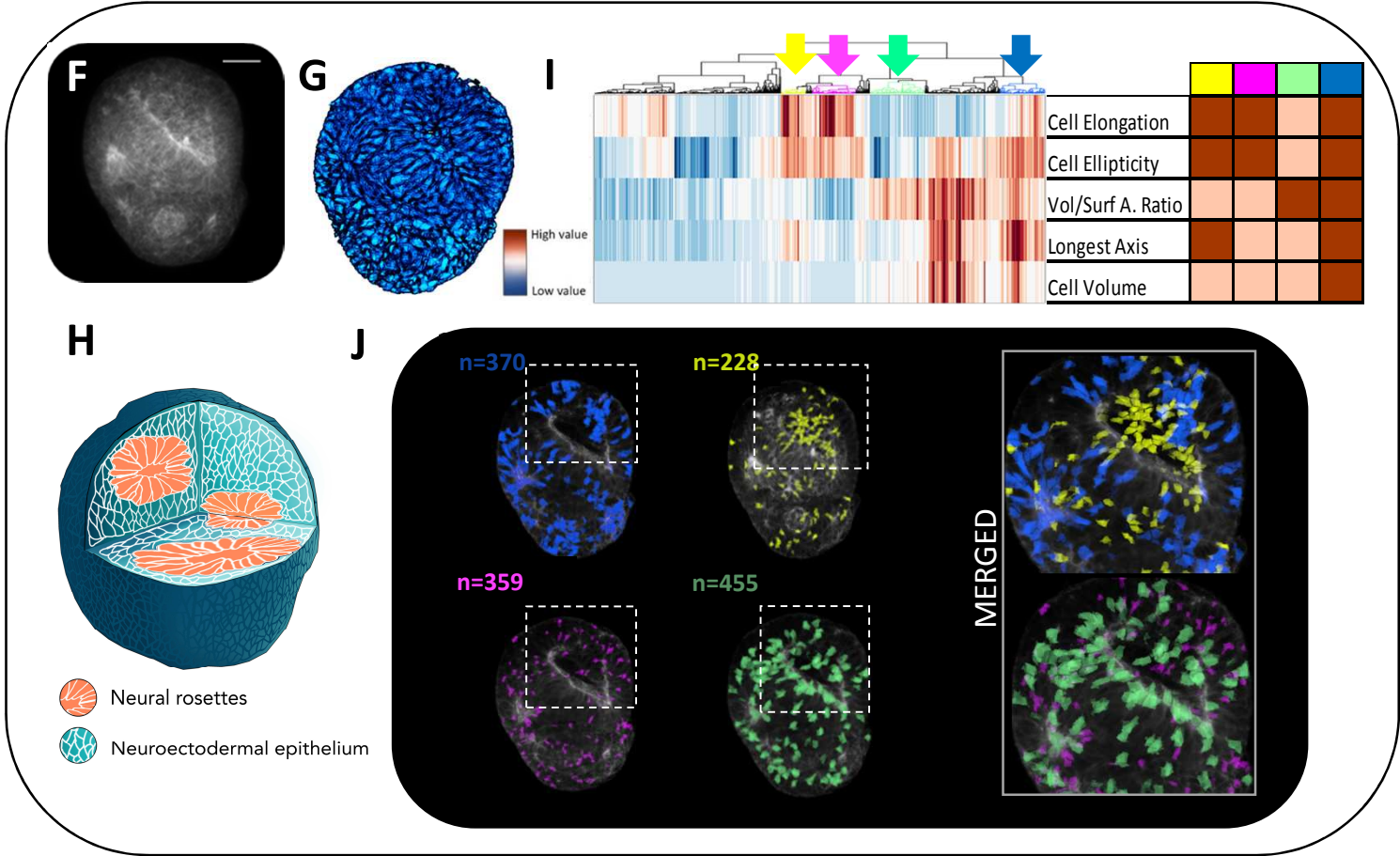
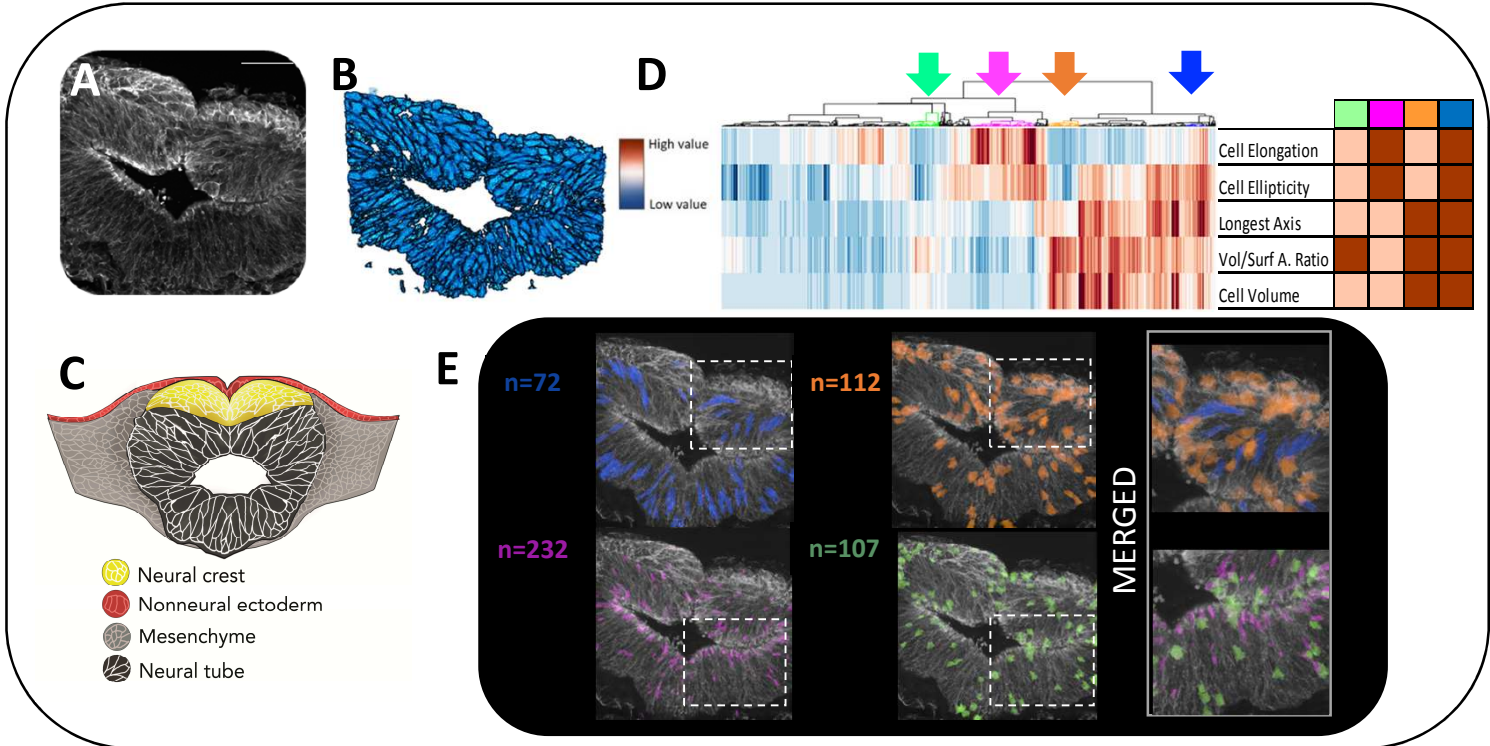
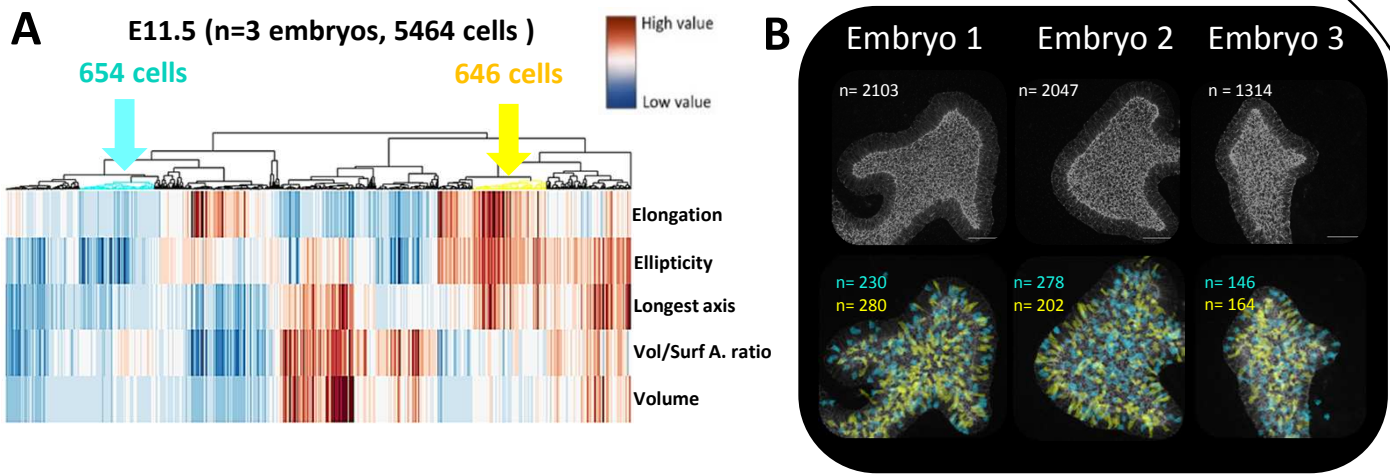
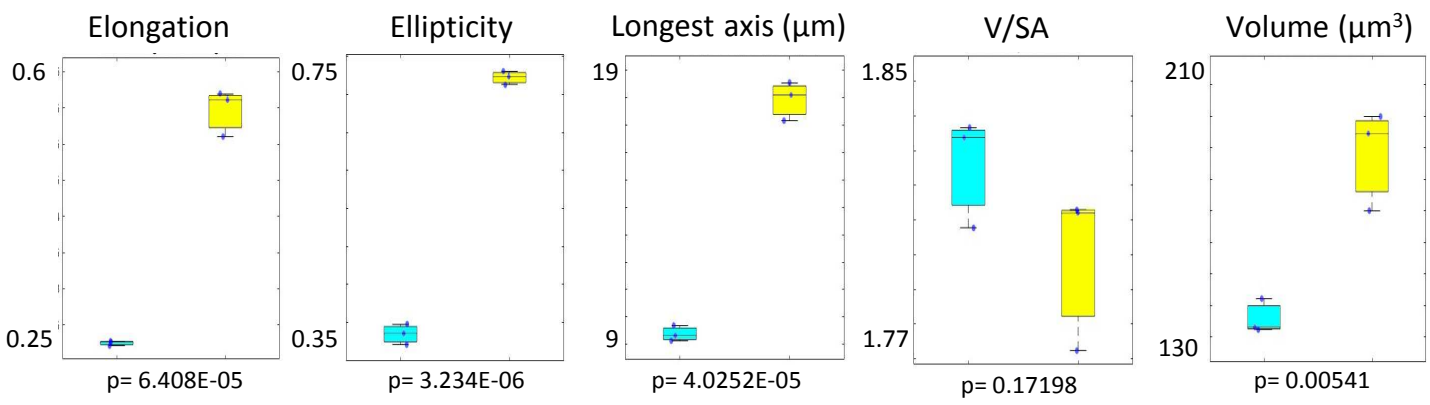


Figure 4



**C** AVERAGE VALUES (n=3, n=3)



**D** INDIVIDUAL CELLS (n=654, n=646)

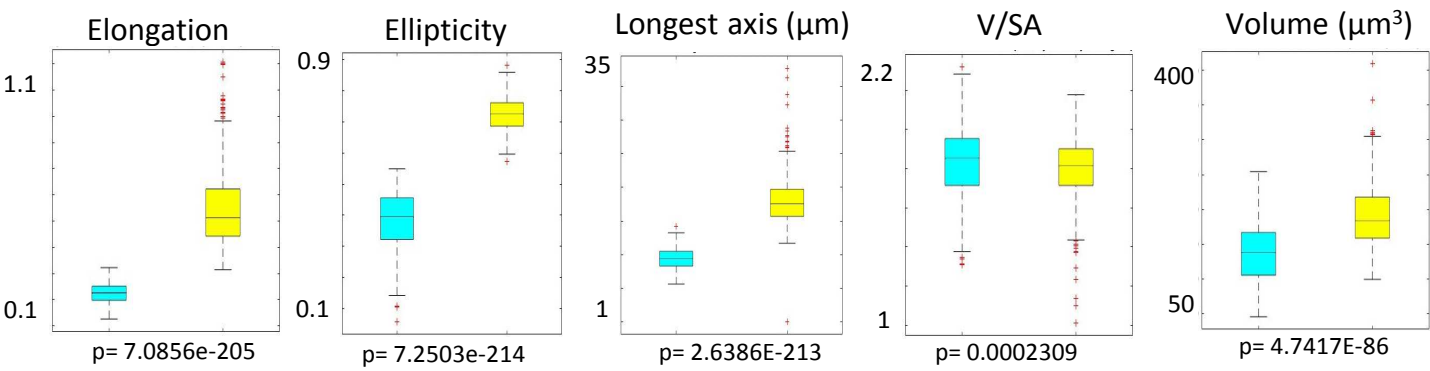


Figure 5

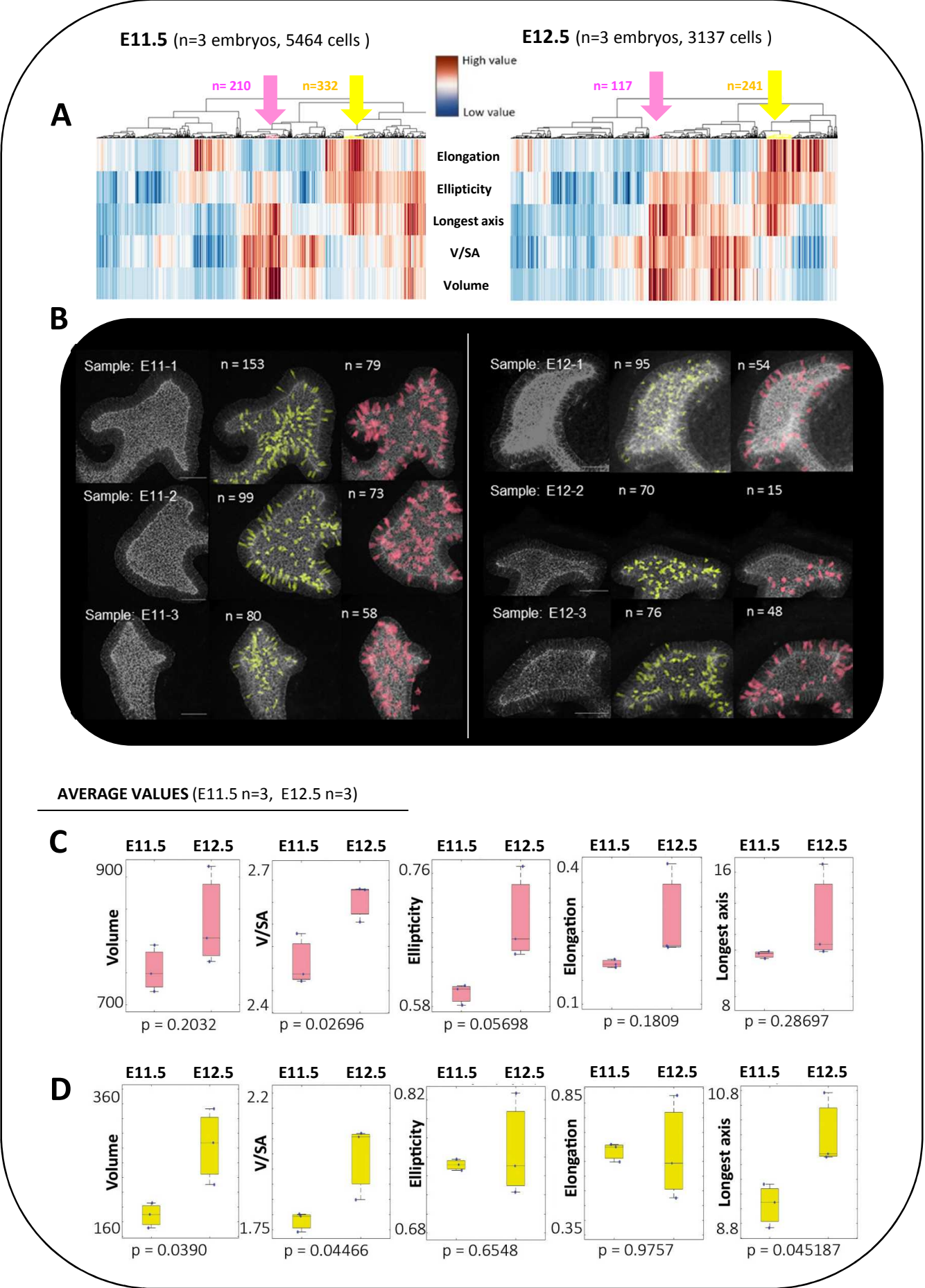


Figure 6

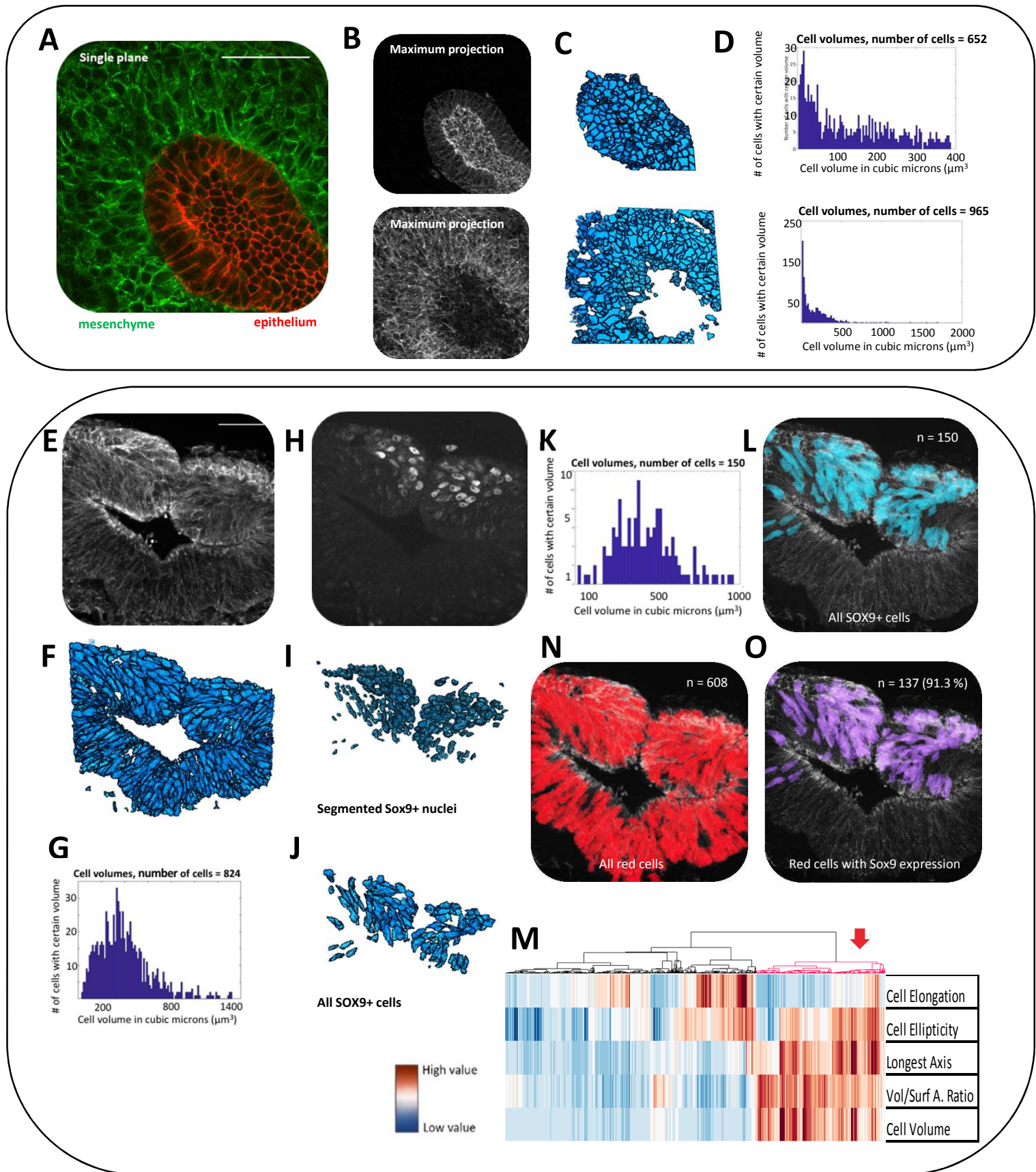


Figure 7

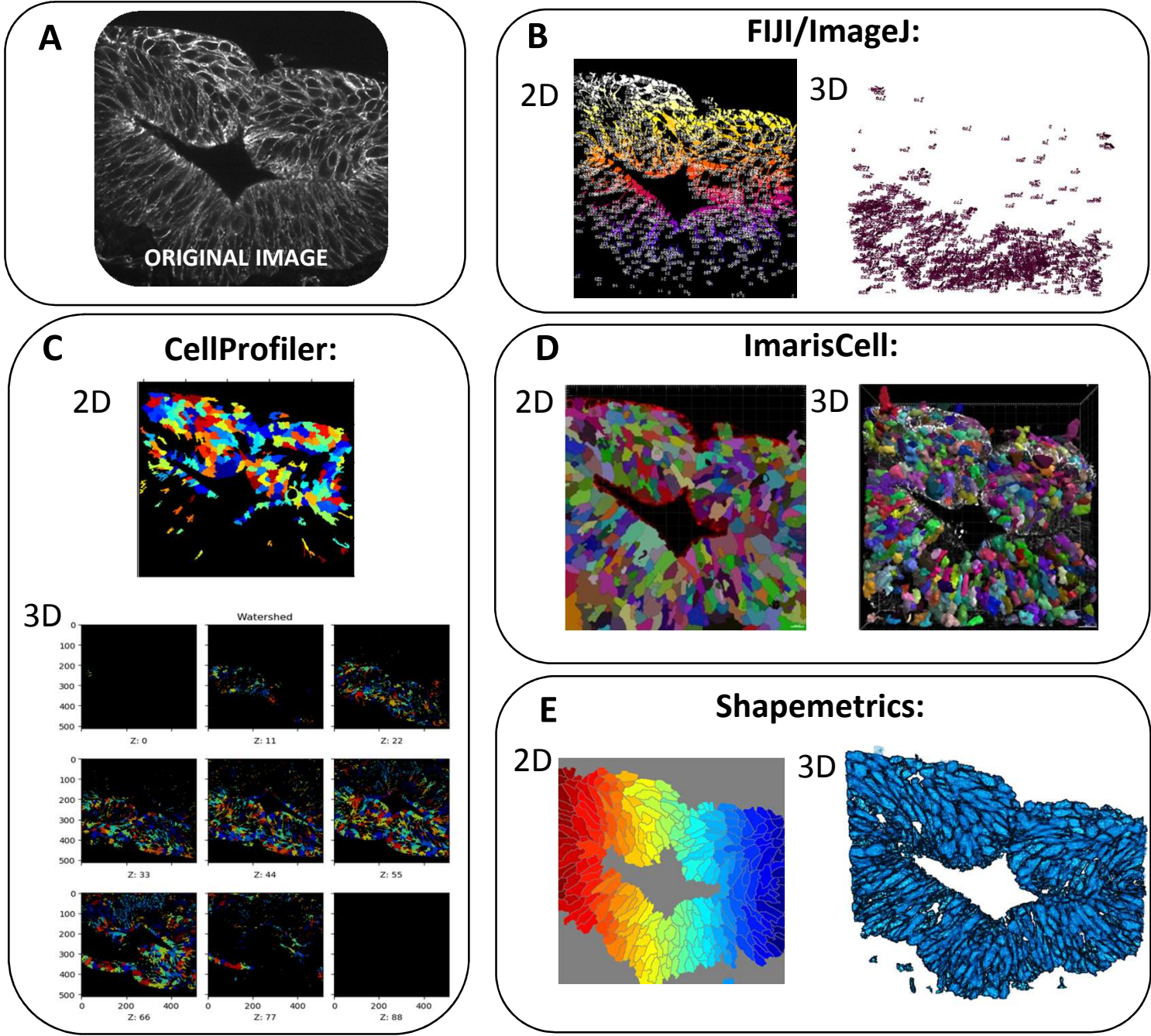


Table 1

<b>2D</b>	<b>CellProfiler</b>	<b>FIJI</b>	<b>ImarisCell</b>	<b>Shapemetrics</b>
Number of cells	261	811	393	271
Mean Surface Area ( $\mu\text{m}^2$ )	21.1	8.04	167.47	10.22
<b>3D</b>	<b>CellProfiler</b>	<b>FIJI</b>	<b>ImarisCell</b>	<b>Shapemetrics</b>
Number of cells	-	790	821	824
Mean volume ( $\mu\text{m}^3$ )	27.45	2.364	379.59	401.6

Figure 6

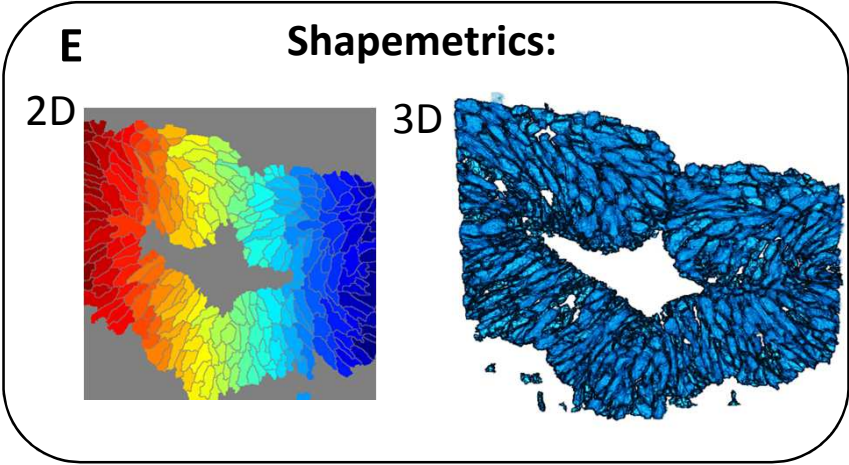
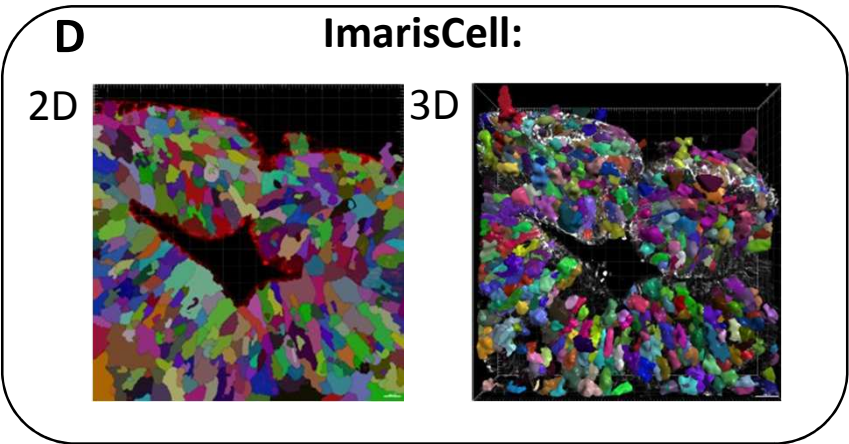
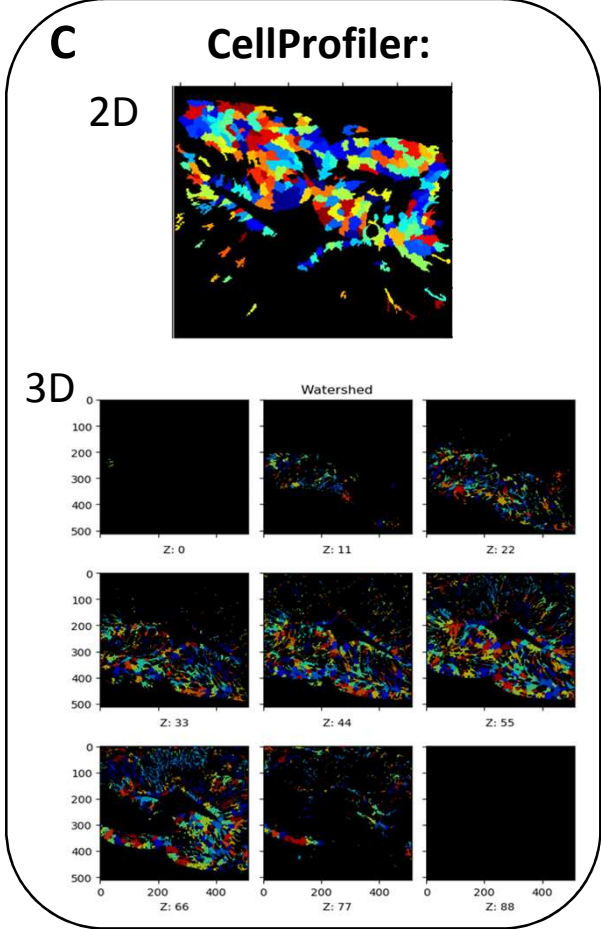
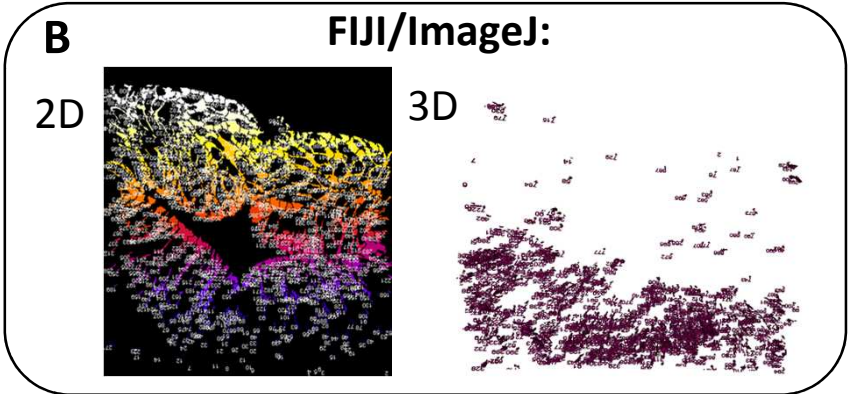
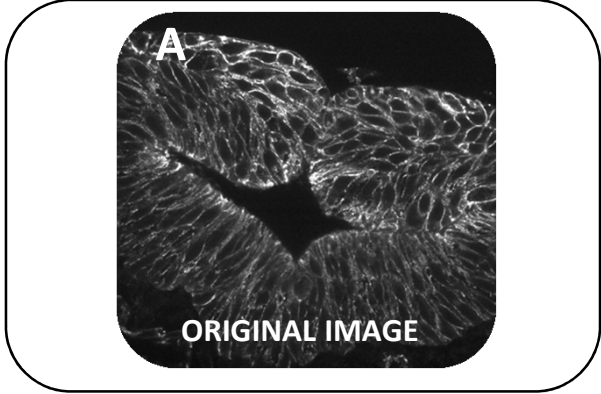


Table 1

<b>2D</b>	<b>CellProfiler</b>	<b>FIJI</b>	<b>ImarisCell</b>	<b>ShapeMetrics</b>
Number of cells	261	811	393	271
Mean Surface Area ( $\mu\text{m}^2$ )	21.1	8.04	167.47	10.22
<b>3D</b>	<b>CellProfiler</b>	<b>FIJI</b>	<b>ImarisCell</b>	<b>ShapeMetrics</b>
Number of cells	-	790	821	824
Mean volume ( $\mu\text{m}^3$ )	27.45	2.364	379.59	401.6

- A MATLAB based pipeline that subgroups cells in intact tissues according to volumetric, morphological features.
- A method aimed to complement tissue and phenotype analysis from 3D microscopy data in addition to traditional information based on usage of fluorescent biomarkers.
- Provides the option of visualization of the spatial location of selected subgroups or individual cells within the original tissue image.
- User-friendliness aims to provide a tool that meets modern needs for single cell level data acquisition for a broad readership.
- Step by step instructions provided for users with limited computational skills.

Journal Pre-proof