

# *Structural analysis of the cervical epithelial tissue*

## *How to run the pipeline*

### 1. Input directory

This is the main directory containing the images to be analyzed grouped by SNIs. In Fig. 1, we can see the input directory *images* and three subfolders. Each subfolder corresponds to one SNI and each SNI folder contains the individual fluorescent channels saved as *tif* files (16-bit uncompressed). **All SNIs should be saved according to this template and the names of the corresponding SNIs subfolders must start with SNI of Neg SNI.**

In Fig 2, we can see the *scripts* folder containing all the necessary scripts to run the automated analysis of the cervical epithelial tissue, which will be detailed in the next sections.

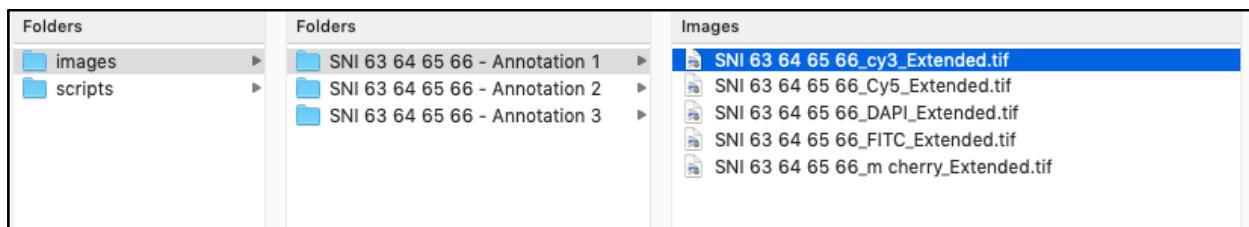


Fig. 1. Visualization of the *input* directory.

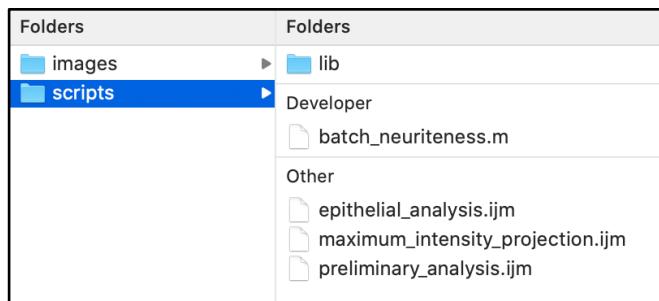


Fig. 2. Visualization of the *scripts* directory.

### 2. Requirements

The software listed below should be installed before running the scripts. Follow the instructions in each link on how to download and install them.

- a) **Fiji:** <https://fiji.sc>
- b) **MorphoLibJ plugin:** <https://imagej.net/MorphoLibJ>. For this specific Fiji plugin, open *Fiji*, go to *Plugins* menu and check if you can find *MorphoLibJ* among the available plugins. If not, then go to *Help / Update... / Manage update sites*. Select *IJPB-plugins* as shown in Fig. 3. Click on *Close* and then *Apply changes*. You will be requested to restart Fiji. This plugin must be installed in Fiji only once.
- c) **Matlab:** <https://se.mathworks.com/products/matlab.html>
- d) **Neuriteness library:** provided by *Obara et al.*

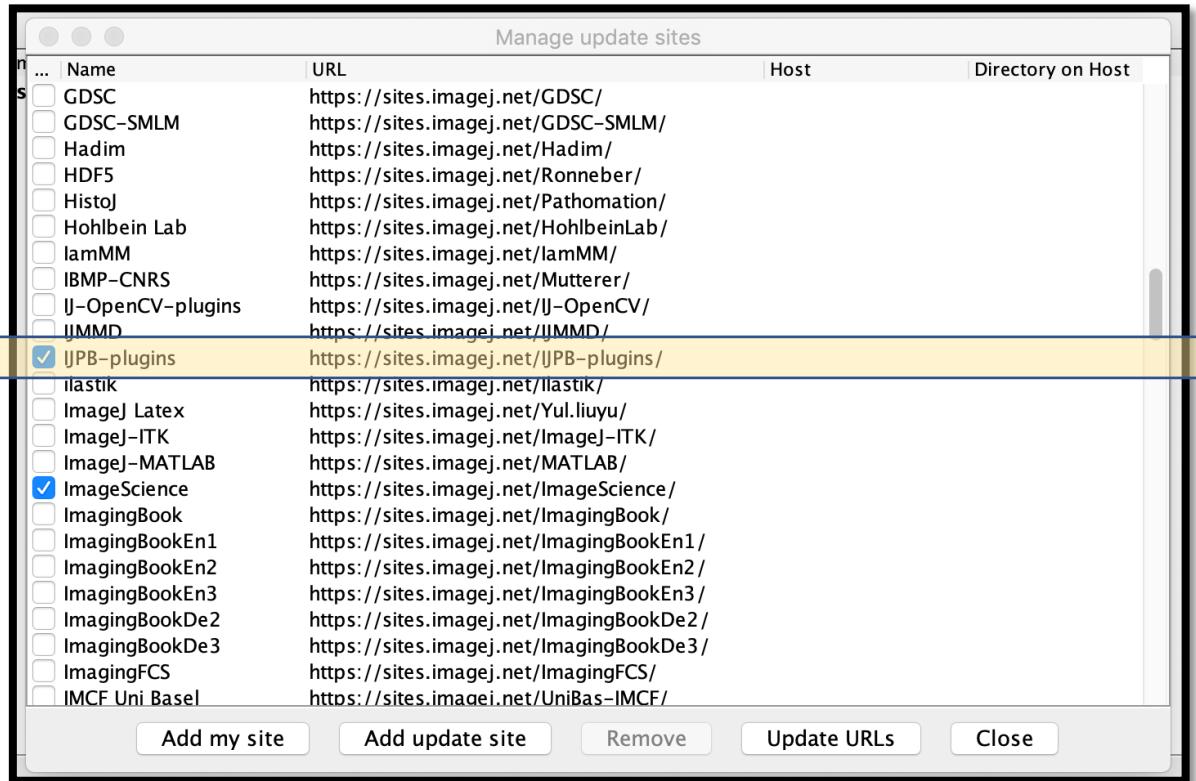


Fig. 3. Manage update sites interface.

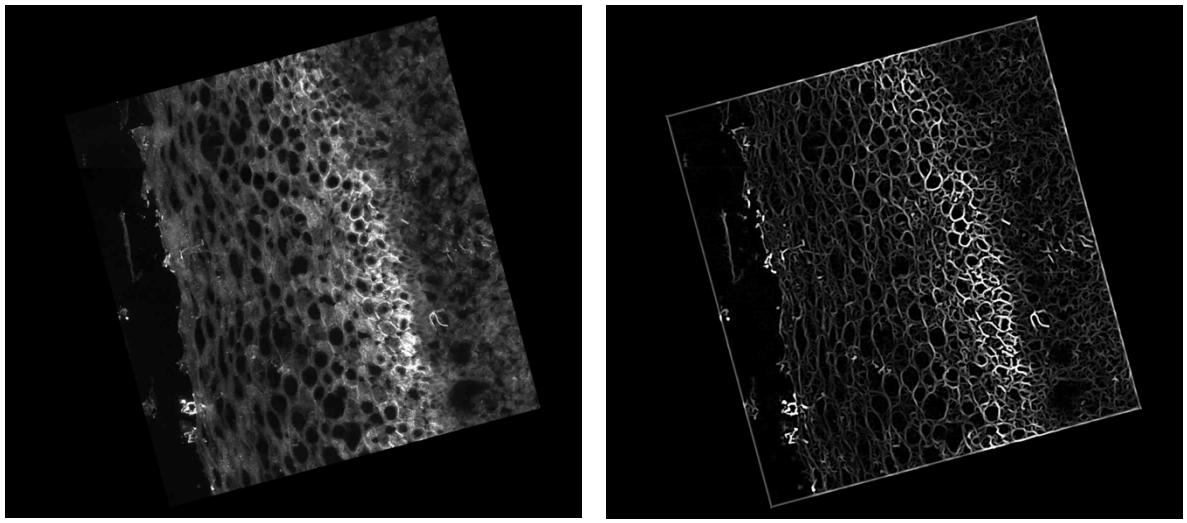
### 3. *Neuriteness* script - Matlab

The Matlab script *batch\_neuriteness.m* will generate the *neuriteness* network using the method proposed in: Obara, Boguslaw, et al. "Contrast-independent curvilinear structure detection in biomedical images." **IEEE Transactions on Image Processing** 21.5 (2012): 2572-2581. Fig. 4 shows an example of a *neuriteness* image, obtained from a sample image.

To run the *batch\_neuriteness.m* script, first open Matlab and then click on the *Open* menu in the main interface to browse the corresponding script file (Fig. 5). This file is located in the *scripts* folder, as highlighted in Fig. 2. Then, update the path in the *addpath* command (line 4). This path corresponds to the *neuriteness* library provided by Obara et al. (see *lib* folder in Fig. 2). **Notice that, in Matlab, Windows users should use the backslash to represent the library path, whereas Mac users should use the normal slash. Here is an example:**

- Library path in Mac: /Users/gisele.miranda/scripts/lib
- Library path in Windows: C:\Users\gisele.miranda\scripts\lib

After updating the library path, press the *Run* button, also highlighted in Fig. 5. A file browser window will appear and the *images* folder should be selected. After selecting it, wait until the processing is finished. The output of the *batch\_neuriteness.m* script can be seen in Fig. 6. An image named *Ne\_XXX.... .tif* is generated for each SNI channel.



(a) Original

(b) Neuriteness

Fig. 4. Generation of the *neuriteness* network.

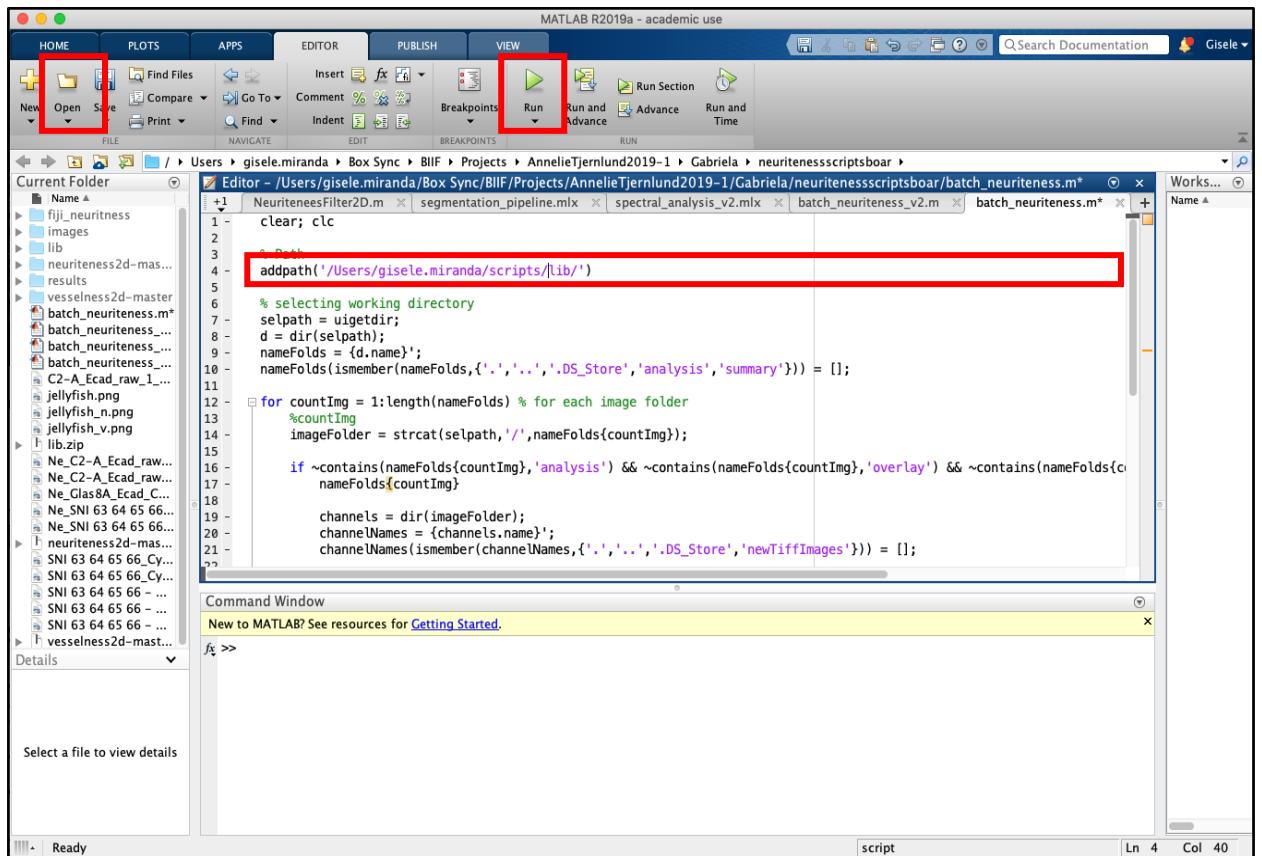


Fig. 5. Matlab interface displaying the script *batch\_neuriteness.m*.

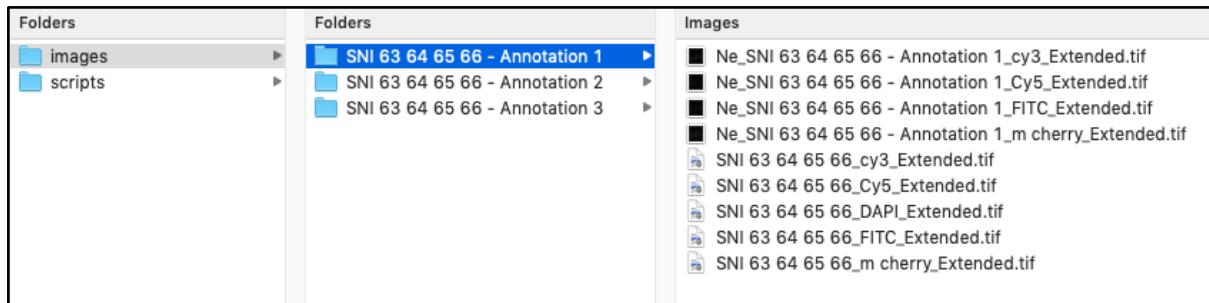


Fig. 6. Neuriteness files created for each SNI channel.

#### 4. Maximum intensity projection pipeline for layer annotation

The Fiji pipeline *maximum\_intensity\_projection.ijm* generates a maximum intensity projection for each SNI, given all the corresponding fluorescent channels. The four channels (cy3, Cy5, FITC and m cherry) are concatenated and then projected in a single image. The projected image will be used for manual annotation of the apical and basal layers in Fiji, since it provides a higher contrast image. This step is carried out only to facilitate the annotation of the layers.

To run the *maximum\_intensity\_projection.ijm* script, open Fiji and go to *Plugins – Macros – Edit...* and browse the corresponding file from the *scripts* folder (Fig. 2). The file extension must be *ijm* (ImageJ macro). Then, the interface showed in Fig. 7 will appear. In line 2, there is a parameter named *useNucleiChannel*, which can be set as *true* or *false*. If *true*, then the DAPI channel will be included in the maximum intensity projection.

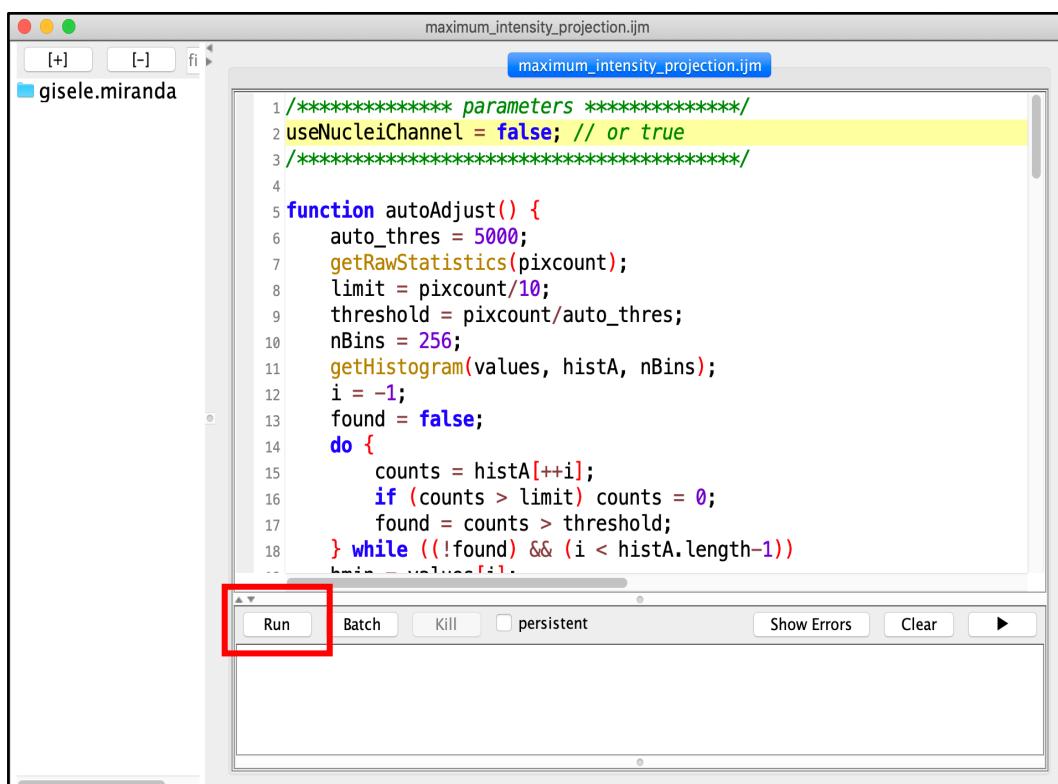


Fig. 7. Interface showing the *maximum\_intensity\_projection.ijm* pipeline.

After the execution of this script, a folder named *images\_MAX\_projection* will be created which contains the projected image of each SNI in *tiff*format, as shown in Fig. 8. Note that, the name of the input directory, *images*, is arbitrary. **Different names can be assigned to the input directory. However, the folder that contains the maximum intensity projections will be named with the same name of the input directory + *MAX\_projection*.**

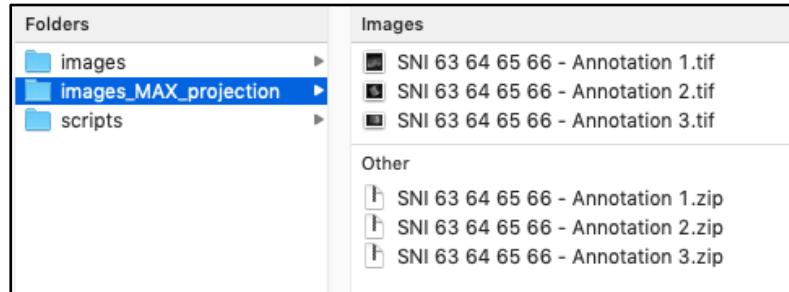


Fig. 8. Generation of the *MAX\_projection* folder.

The files with *.zip* extension in Fig. 8 correspond to the manual annotations of the basal and apical layers of the epithelial tissue. They can be opened in Fiji using the *Roi Manager* tool. The *maximum\_intensity\_projection.ijm* script also generates a new folder inside each SNI folder named *newTiffImages* (see Fig. 9), which saves the input channels in a *non-tile tiff* format.

**Notice that the scripts *batch\_neuriteness.m* and *maximum\_intensity\_projection.ijm* can be run in any order. The corresponding inputs and outputs do not depend on each other.**

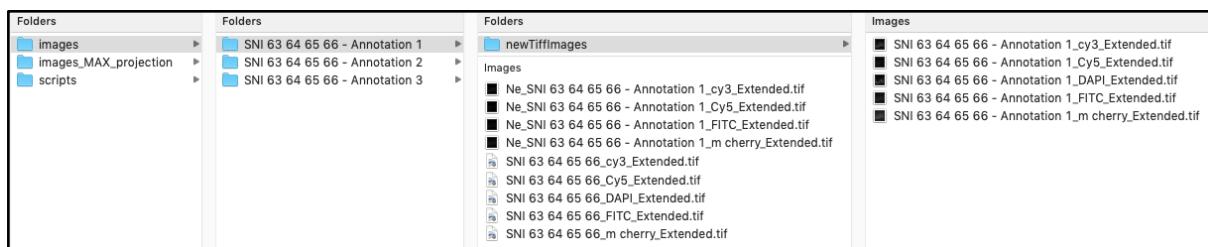


Fig. 9. *newTiffImages* foder containing the SNI channels saved in Tif files.

## 5. Basic statistics: preliminary analysis

The script *premilinary\_analysis.ijm* extracts basic statistics from the epithelium region. The input parameters are:

- **path**: path of the folder containing the SNI subfolders;
- **maxProjPath**: path containing the maximum intensity projection of the SNIs ;
- **nbins**: number of bins of the pixel intensity histogram
- **increment**: size of the histogram bin
- **channel**: fluorescent channel that will be analyzed (Ex.: FITC, Cy5, etc);

This script generates two output files:

- 1) **preliminary\_statistics\_FITC\_channel.csv**: stores the area, mean, standard deviation, maximum, minimum and median values for each input SNI, and
- 2) **preliminary\_statistics\_histogram\_FITC\_channel.csv** - stores the histogram of the pixel values for each input SNI

The histogram saved in file number 2) corresponds to the number of pixels that have values between a certain interval, e.g., the first column of this file is [0,500[, therefore, this column shows the number of pixels for which the pixel intensities are between 0 and 500, the next column shows the same counting for intensities between [500,1000[, and so on. Each line corresponds to the histogram of a single SNI. The number of histogram bins and the bin size (increment) can be modified through the input parameters. Since the input images are 16-bit, a possible combination is 70 bins with intervals of size 500.

## 6. Quantitative analysis of the cervical epithelium

The script *epithelial\_analysis.ijm* performs the automatic analysis of the SNI images. The workflow implemented in this script can be summarized in the following steps:

- a) Segmentation of the epithelium;
- b) Calculation of the epithelium height based on EDT (Euclidean distance transform);
- c) Segmentation of the *neuriteness* network for each fluorescent marker;
- d) Calculation of the relative area of the segmented *neuriteness*;
- e) Calculation of the distance between the segmented fluorescent marker to the apical and basal layers

### 6.1 Segmentation of the epithelium

The segmentation of the epithelium region is performed using the manual annotations of the basal and apical layers. These two layers are connected in order to obtain a polygon, which corresponds to the binary mask of the epithelium (Fig. 10).

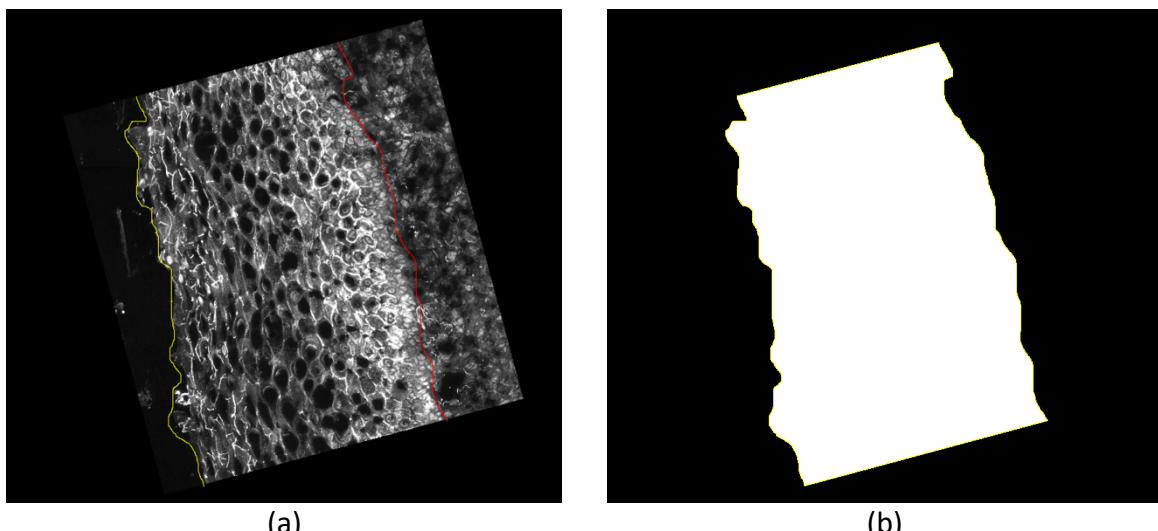


Fig. 10. (a) Original image with manual annotations for the apical and basal layers. (b) Binary mask corresponding to the epithelium region.

## 6.2 Calculation of the epithelium height based on EDT

The calculation of the epithelium height is performed using the Euclidean distance transform (EDT) of the basal and apical layers (see Fig. 11). The EDT is a distance map which assigns a distance value for each pixel of the image or selected region. In Fig. 11 (a), the EDT is calculated using the basal layer as reference. Therefore, the pixels that are farther from the basal layer will be assigned higher distance values (brighter regions). Finally, in the same image we can observe the overlay of the apical line (in yellow) on the distance map. The pixels along this line will be taken into account to calculate the height of the epithelium from the basal layer towards the apical layer. In Fig. 11 (b), we can see a similar map, but now taking the apical layer as reference for the EDT.

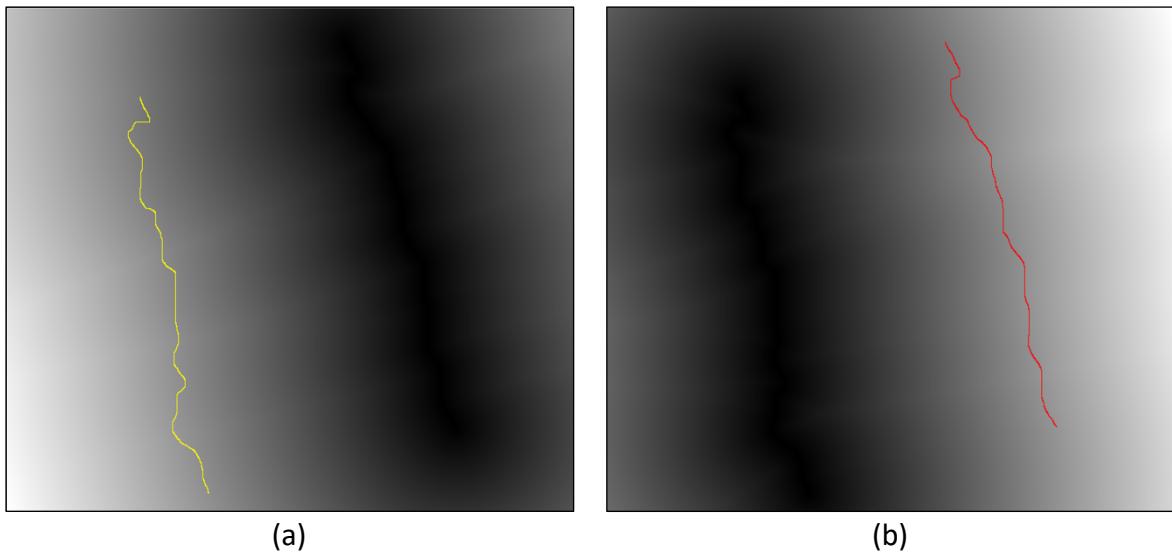


Fig. 11. Euclidean distance transform - EDT. (a) EDT of the basal layer and overlay of the apical layer in yellow, where the height is measured. (b). EDT of the apical layer and overlay of the basal layer in red, where the second height is measured.

## 6.3 Segmentation of the *neuriteness* image

The Otsu threshold is used for the segmentation of the *neuriteness* image. In addition, it is also possible to define a correction factor which will be multiplied by the threshold value given by the Ostu method. This allows some adjustments in the final segmentation result. In Fig. 12, we can observe the effect of different values of the correction factor.

After choosing the correction factor, the next step is the generation of the EDT map for the thresholded *neuriteness*, as shown in Fig. 13. In Fig. 13 (a) we have the inverted binary mask of the thresholded *neuriteness* and in Fig. 13 (b) we have the corresponding distance map. Finally, we can define a distance threshold in order to obtain the binary mask shown in Fig. 13 (c). Different distance thresholds lead to different binary masks. In Fig. 14, we can see the effect of different values of the distance threshold.

Notice that, the EDT transform is now applied on the thresholded *neuriteness*. The same transform is applied on the apical and basal layers to calculate the epithelium height, as explained in section 6.3.

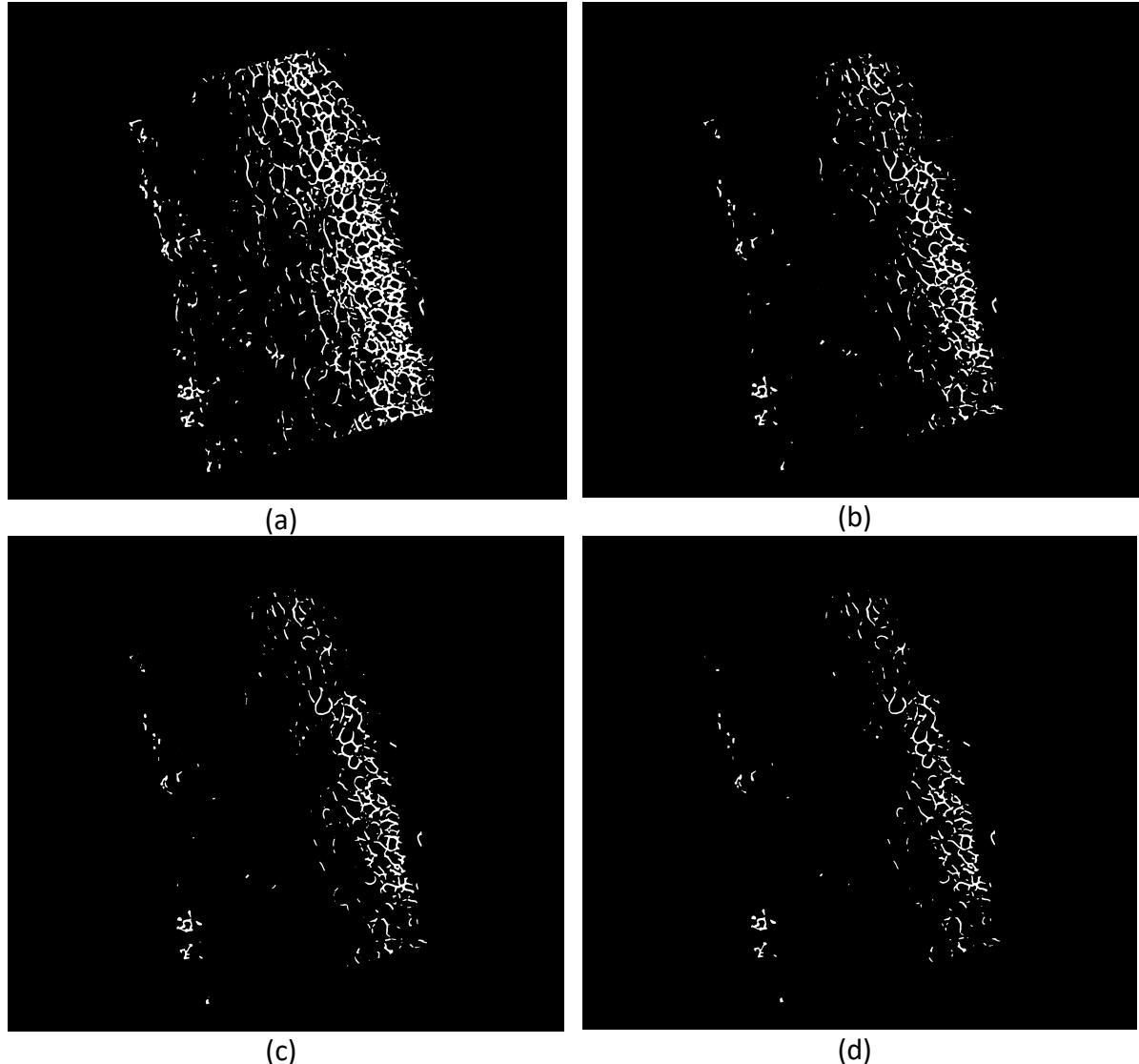


Fig. 12. Threshold of the *neuriteness* image. (a) Otsu threshold (without correction factor). (b) Otsu threshold + corrections factor 1.5. (c) Otsu threshold + corrections factor 1.8 and (d) Otsu threshold + corrections factor 2.

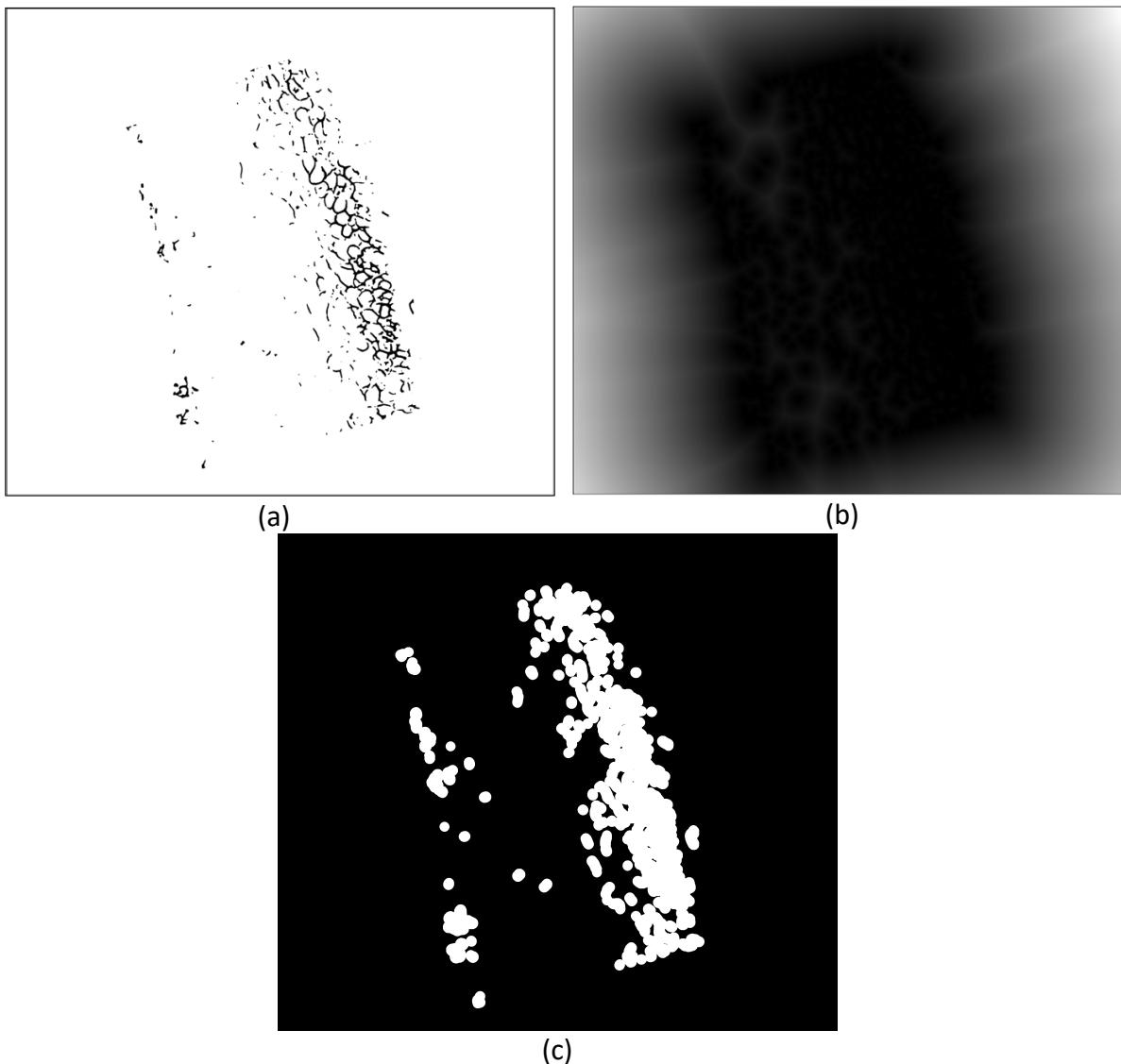


Fig. 13. Generation of the distance map from the thresholded neuriteness. (a) Thresholded neuriteness (inverted). (b) Distance transform of (a), darker values indicate shorter distances. (c) Threshold of the distance map shown in (b).

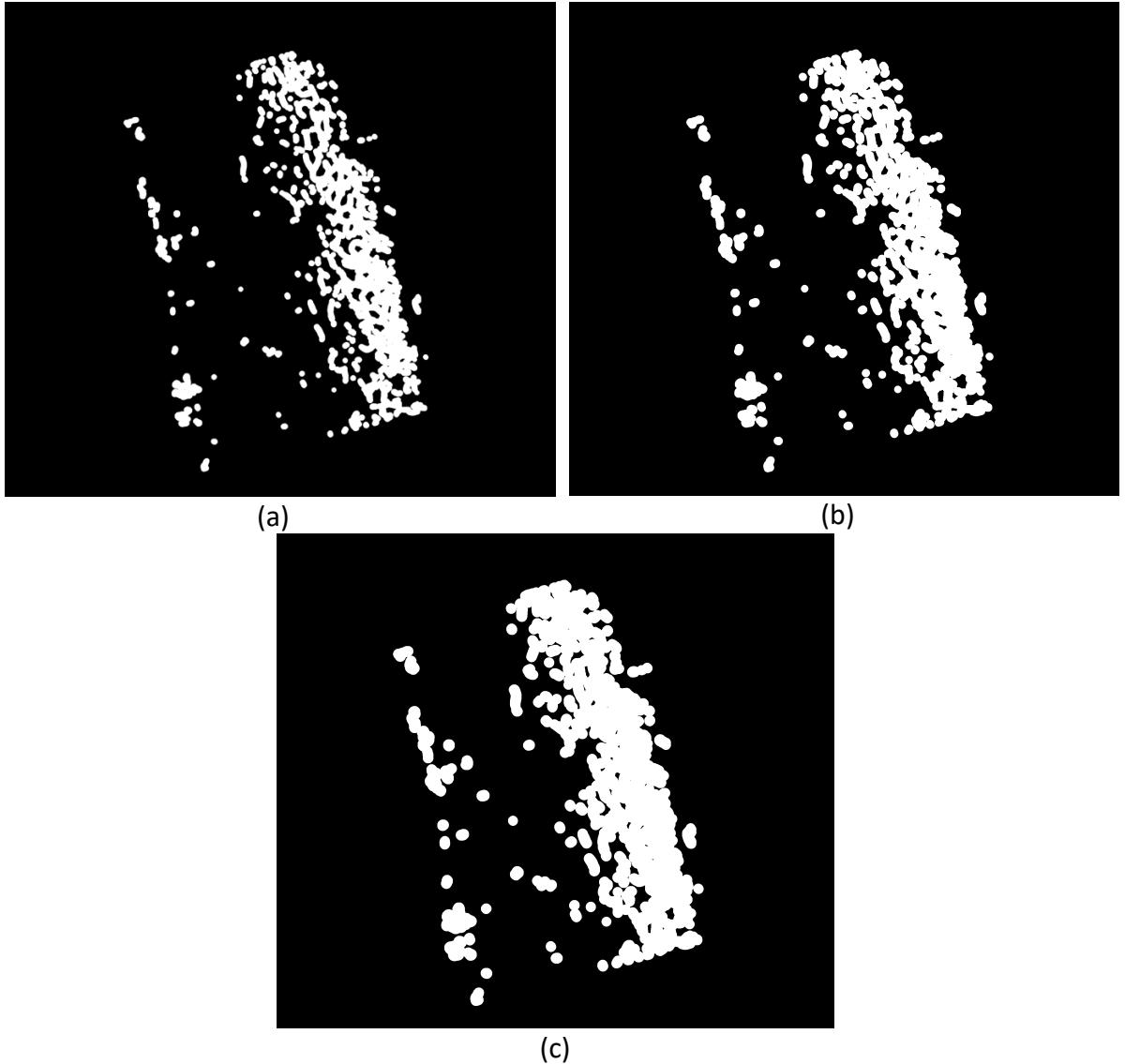


Fig. 14. Influence of the distance threshold. (a) 10, (b) 15 and (c) 20.

Instead of using the binary mask obtained only with the Otsu threshold, using the binary mask obtained via EDT over the thresholded neuriteness can help to overcome further issues. For instance, we can observe that in Fig. 13 (a), the segmented *neuriteness* is very fragmented, which leads to many small connected components with high variability in size. In this case, it will be difficult to define an area threshold to select the components of interest. Meanwhile, using the thresholded image calculated over the EDT of the segmented *neuriteness*, leads to bigger connected components encompassing all the brighter regions of the *neuriteness* image, which can be easily filtered by area. This additional step is used to filter regions of the *neuriteness* image. Fig. 15 (a) shows the binary mask of the FITC channel after the removal of small elements and the ones that are too close to the apical layer. Finally, Figs. 15 (b), (c) and (d) correspond to the overlay of the binary mask in (a) over the thresholded *neuriteness*, the *neuriteness* image and the original image, respectively.

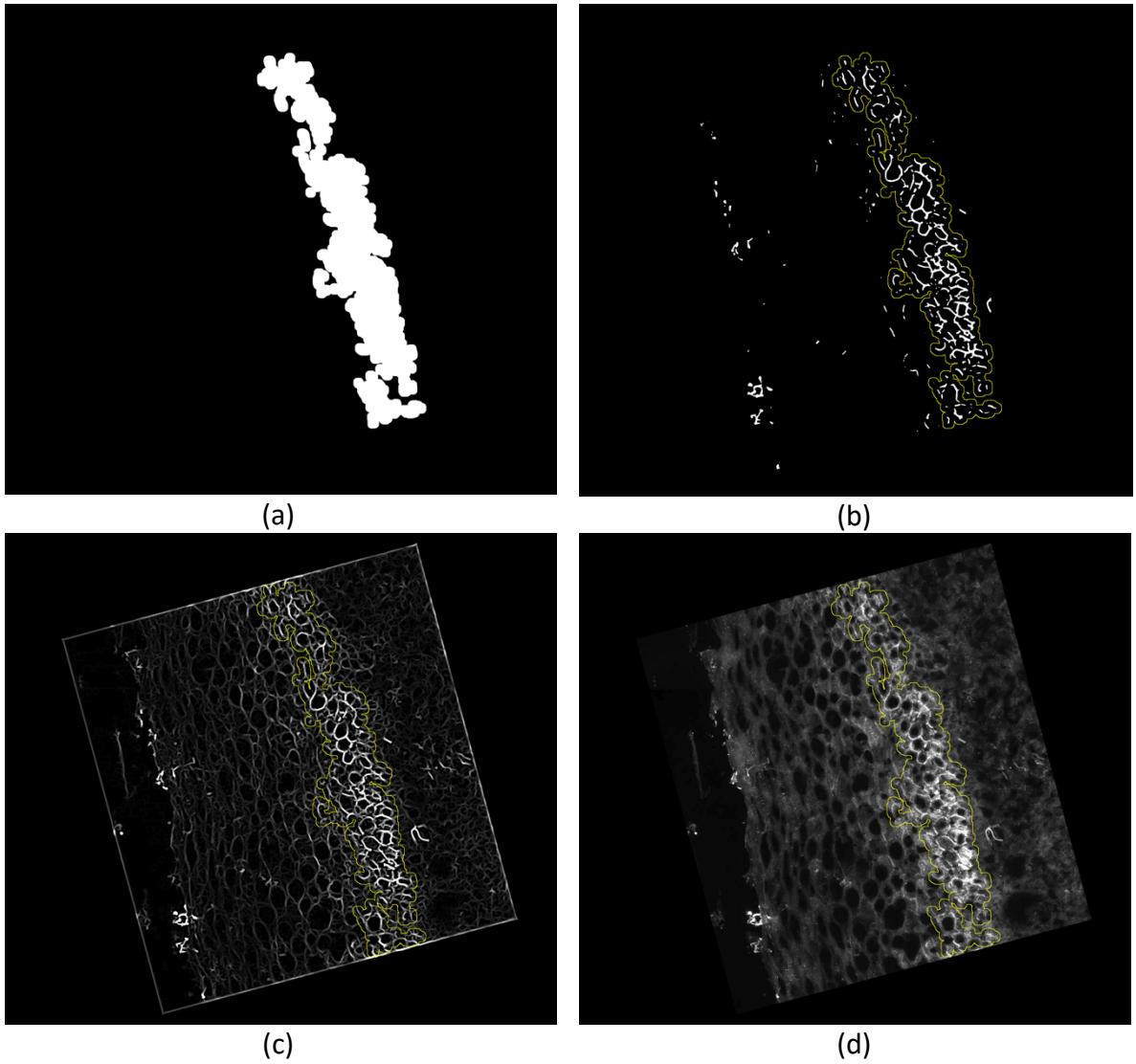


Fig. 15. (a) Binary mask of the segmented neuriteness image. Overlay of the binary mask in (a) over the thresholded *neuriteness* (b), the *neuriteness* image (c) and the original image (d)

#### 6.4 Calculation of the area of the segmented *neuriteness* and other layers

Finally, after obtaining the binary mask corresponding to the segmented fluorescent marker (Fig. 15 (a)) we can calculate the corresponding **area of the marker of interest** and the area of the *neuriteness* inside the binary mask (Fig. 15 (b)). In addition, the area of the following binary masks can also be obtained.

- **Parabasal layer + marker of interest** (Fig. 16 (b)): obtained from segmentation of the DAPI channel + Fig. 15 (a);
- **Upper layer** (Fig. 16 (c)): obtained by the subtraction of the parabasal layer from the epithelium region;
- **Lower layer** (Fig. 16 (d)): obtained from the subtraction of the upper + Fig. 15 (a) from the epithelium region.

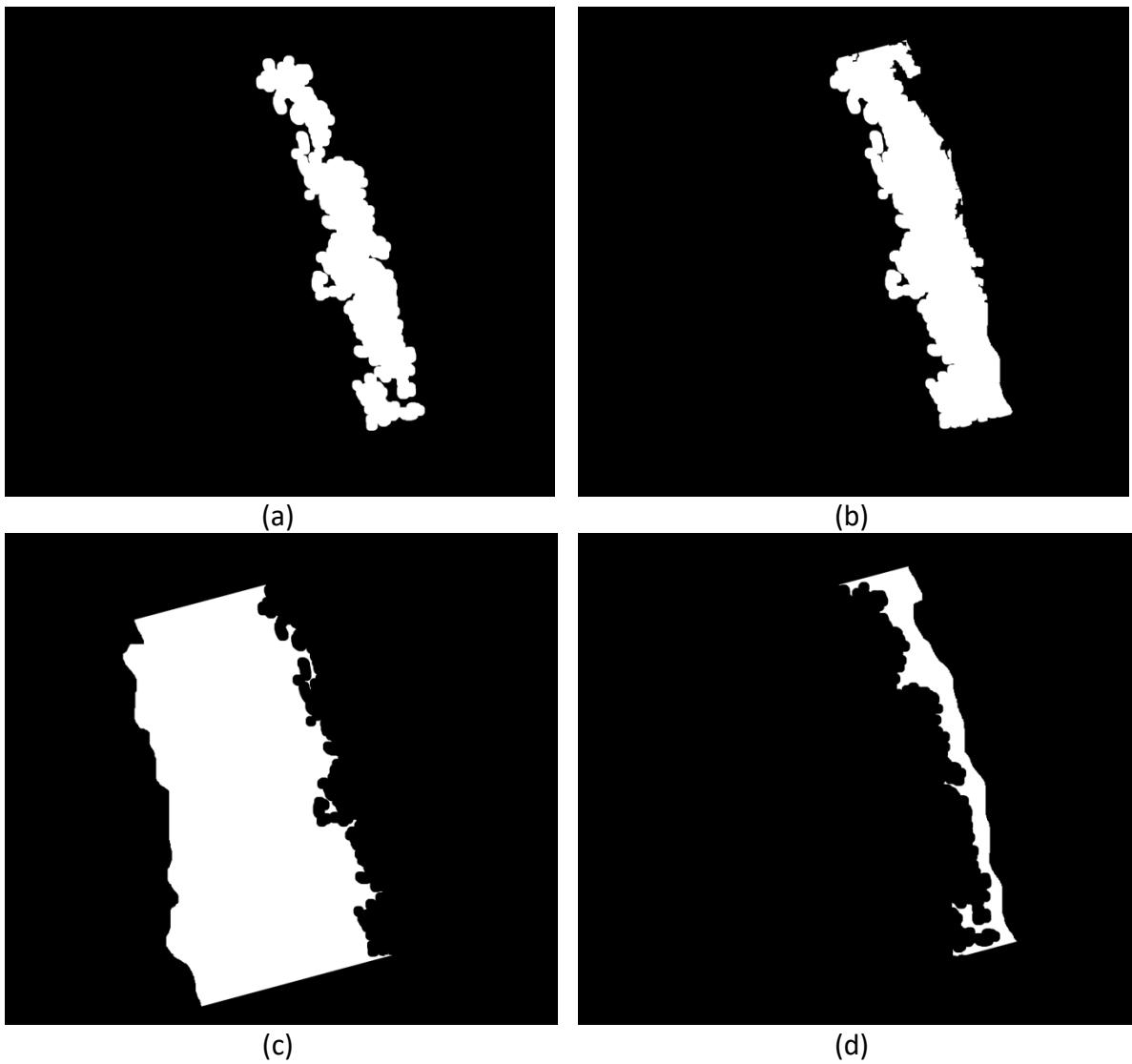


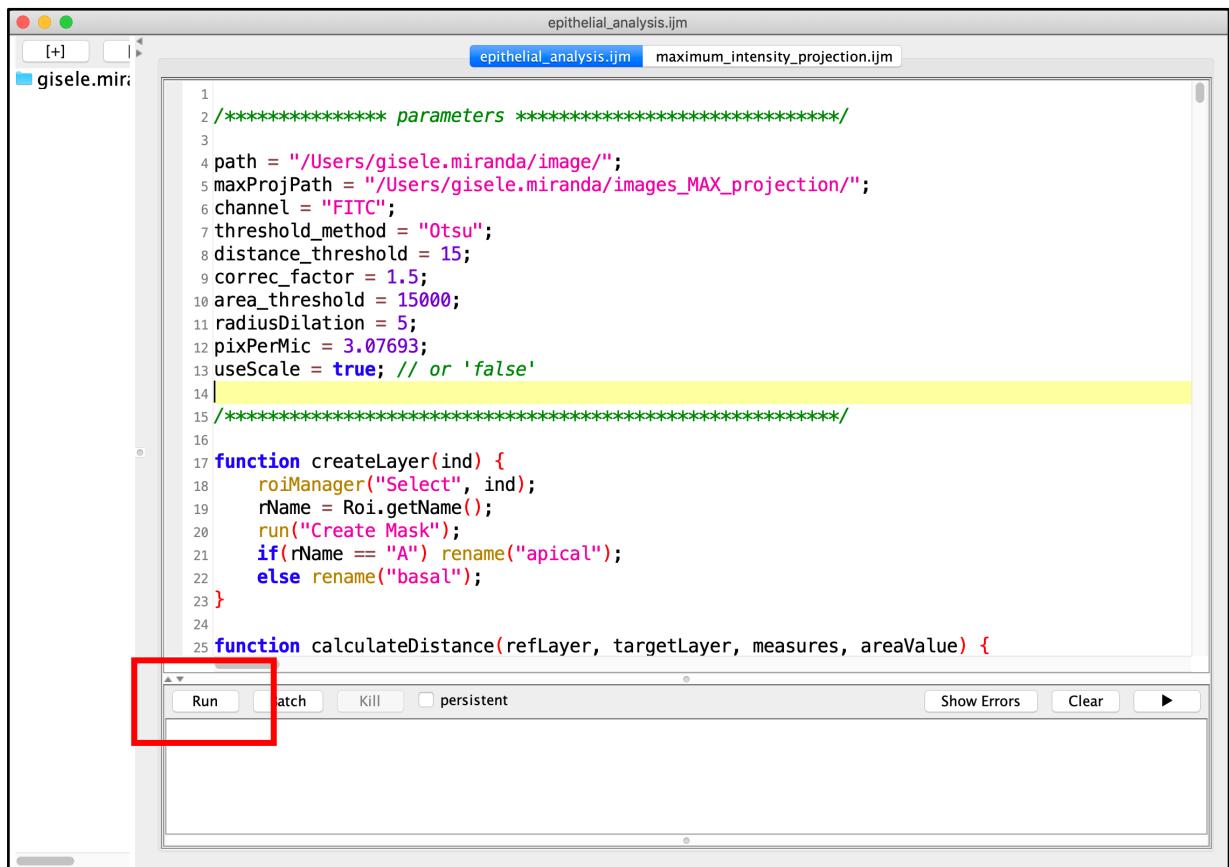
Fig. 16. (a) Binary mask of the segmented fluorescent marker, (b) parabasal layer + marker of interest, (c) upper layer and (d) lower layer.

### 6.1 Running the Fiji script

The script *epithelial\_analysis.ijm* implements all the steps described in this section. Lines 4 – 14 correspond to the parameters that can be modified.

- **path:** path of the folder containing the SNI subfolders;
- **maxProjPath:** path containing the maximum intensity projection of the SNIs in *path*;
- **channel:** fluorescent channel that will be analyzed (Ex.: FITC, Cy5, etc);
- **threshold\_method:** *Otsu* or other any other threshold method available in Fiji;
- **distance\_threshold:** see Fig. 14 for examples;
- **correc\_factor:** see Fig. 12 for examples;

- **area\_threshold**: minimum area of a connected component that should be included in the final binary mask;
- **radiusDilation**: radius of the dilation operation which is applied on the final binary mask;
- **nucleiSize**: size threshold for the segmented nuclei. It aims at excluding small segmented particles;
- **pixPerMic**: image scale – number of pixels per one micron (check image metadata);
- **useScale**: whether or not to use the provided scale. It can assume *true* or *false* values. If *false*, then the measures will be given in pixels.



```

1 //***** parameters *****/
2
3 path = "/Users/gisele.miranda/image/";
4 maxProjPath = "/Users/gisele.miranda/images_MAX_projection/";
5 channel = "FITC";
6 threshold_method = "Otsu";
7 distance_threshold = 15;
8 correc_factor = 1.5;
9 area_threshold = 15000;
10 radiusDilation = 5;
11 pixPerMic = 3.07693;
12 useScale = true; // or 'false'
13
14 //*****
15
16
17 function createLayer(ind) {
18   roiManager("Select", ind);
19   rName = Roi.getName();
20   run("Create Mask");
21   if(rName == "A") rename("apical");
22   else rename("basal");
23 }
24
25 function calculateDistance(refLayer, targetLayer, measures, areaValue) {

```

Fig. 16. *epithelial\_analysis.ijm* script.

After choosing the values of each parameter, press the *Run* button (highlighted in Fig. 16). Wait for the execution of the pipeline. The execution time will depend on the number of SNIs in the input directory. During the execution, the *Run* button will be disabled.

## 7. Output files

Finally, after execution of the *epithelial\_analysis.ijm*, the following data will be created:

- A summary file (“summary\_XXX.csv”) inside the input directory, as shown in Fig. 17. This file contains measures obtained for each image. These measures are detailed in section 7.1.
- Output folder: an output folder is created inside each SNI folder to save the images generated during the execution of the script (Fig. 18).
- A folder named “overlay\_XXX” is saved in the input directory containing a file of the thresholded *neuriteness* and its corresponding segmentation. The images in this folder were created to facilitate the visualization of the segmented regions. They can also be found in the output folders of each SNI.

Obs: the nomenclature of the output folders and files contains a combination of the selected parameter values. Therefore, if the pipeline is executed with different parameters, the output of each unique combination will be saved once.

Folders	Folders	Images
images	overlay_Otsu_correc-factor-1.5_dist-15_area-15000_radDilation-5	network_Ne_SNI 63 64 65 66 - Annotation 1_FITC_Extended.tif network_Ne_SNI 63 64 65 66 - Annotation 2_FITC_Extended.tif network_Ne_SNI 63 64 65 66 - Annotation 3_FITC_Extended.tif neuriteness_distance_transform_N... - Annotation 1_FITC_Extended.tif neuriteness_distance_transform_N... - Annotation 2_FITC_Extended.tif neuriteness_distance_transform_N... - Annotation 3_FITC_Extended.tif original_distance_transform_Ne_S...6 - Annotation 1_FITC_Extended.tif original_distance_transform_Ne_S...6 - Annotation 2_FITC_Extended.tif original_distance_transform_Ne_S...6 - Annotation 3_FITC_Extended.tif
images_MAX_projection	SNI 63 64 65 66 - Annotation 1 SNI 63 64 65 66 - Annotation 2 SNI 63 64 65 66 - Annotation 3	
scripts		

Fig. 17. Output files and folders

Folders	Folders	Folders	Images
images	overlay_Otsu_correc-f...-15000_radDilation-5	newTiffImages	lower_layer.tif mask_epithelium.tif mask_nuclei.tif Ne_SNI 63 64 6...elium_overlay.tif
images_MAX_projection	SNI 63 64 65 66 - Annotation 1	output_Otsu_correc-factor-1...5_area-15000_radDilation-5	network_Ne_SNI...FITC_Extended.tif neuriteness_dist...FITC_Extended.tif original_distan...FITC_Extended.tif parabasal_layer.tif segmented FITC channel.tif segmented_net...FITC_Extended.tif upper_layer.tif
scripts	SNI 63 64 65 66 - Annotation 2 SNI 63 64 65 66 - Annotation 3	Images	Ne_SNI 63 64 65 66 - Annotation 1_cy3_Extended.tif Ne_SNI 63 64 65 66 - Annotation 1_Cy5_Extended.tif Ne_SNI 63 64 65 66 - Annotation 1_FITC_Extended.tif Ne_SNI 63 64 65 66 - Annotation 1_m_cherry_Extended.tif SNI 63 64 65 66_cy3_Extended.tif SNI 63 64 65 66_Cy5_Extended.tif SNI 63 64 65 66_DAPI_Extended.tif SNI 63 64 65 66_FITC_Extended.tif SNI 63 64 65 66_m_cherry_Extended.tif

Fig. 18. Output folder inside each SNI folder.

## 7.1 Information stored in the summary file

Each column of the summary file corresponds to the following measures:

1. The corresponding SNI file;
2. Number of pixels in the basal layer;
3. Average height of the basal layer;
4. Standard deviation of the height calculated from the apical towards the basal layer;
5. Minimum height calculated from the apical towards the basal layer;
6. Maximum height calculated from the apical towards the basal layer;
7. Number of pixels in the apical layer;
8. Average height of the apical layer;
9. Standard deviation of the height calculated from the basal towards the apical layer;
10. Minimum height calculated from the basal towards the apical layer;
11. Maximum height calculated from the basal towards the apical layer;
12. Area of the whole epithelium;
13. Area of marker of interest;
14. Mean fluorescence intensity (MFI) of marker of interest;
15. MFI Standard deviation of marker of interest;
16. Area of the segmented *neuriteness* inside the area defined in 13;
17. Mean fluorescence intensity (MFI) of the segmented *neuriteness* inside the area defined in 13;
18. MFI Standard deviation of the segmented *neuriteness* inside the area defined in 13;
19. Relative area: 13 divided by 12;
20. Relative area: 14 divided by 12;
21. Relative area: 14 divided by 13;
22. Area of the parabasal layer + marker of interest;
23. Area of the upper layer;
24. Area of the lower layer;
25. Average height calculated between 13 and the apical layer;
26. Standard deviation of the height calculated between 13 and the apical layer;
27. Minimum height calculated between 13 and the apical layer;
28. Maximum height calculated between 13 and the apical layer;
29. Average height calculated between 13 and the basal layer;
30. Standard deviation of the height calculated between 13 and the basal layer;
31. Minimum height calculated between 13 and the basal layer;
32. Maximum height calculated between 13 and the basal layer.

Obs: Null values correspond to SNIs that were not segmented either because they were missing or the area of the segmented regions were below the defined threshold.