

FISH-quant v3

Generate FQ outlines from automated cell segmentation

MUELLER Florian¹ muellerf.research@gmail.com

SAMACOITS Aubin¹ aubin.samacoits@gmail.com

WALTER Thomas² thomas.walter@curie.fr

¹Institut Pasteur Paris, *Computational Imaging & Modeling Unit*

² Institut Curie Paris

1. Introduction	3
Workflow	3
Image-acquisition	4
2. Z-projections	5
3. Workflow for CellCognition	7
Segmentation	7
Loading of images	7
Naming convention of CellCognition: position, well, subwell	8
Cell segmentation	8
Saving the results	9
Analysis and organization of results files	10
Generate FQ outlines files from CellCognition results	10
Specify experimental parameters	11
Define channel and segmentation method identifier	11
Specify folder where the results will be saved	11
Define folder	11
Create outlines	11
4. Workflow for CellProfiler	12
Segmentation	12
Analyze Images	12
Test mode	13
Generate FQ outlines files from CellProfiler results	13
Specify experimental parameters	13
Define naming scheme of original images and segmentation results	14
[Optional] Generating outline files for a second color	14
Specify folder where the results will be saved	14
Create outlines	14
5. Inspect outline files	15
6. References	16

1. Introduction

Cell segmentation can be performed with many specialized software packages.

- We use *CellCognition*¹, which is described as a “computational framework dedicated to the automatic analysis of live cell imaging data in the context of High-Content Screening (HCS)”.
- An alternative is *CellProfiler*² which is described as a “free open-source software designed to enable biologists without training in computer vision or programming to quantitatively measure phenotypes from thousands of images automatically”.

Either package allows the automated detection (segmentation) of cells and nuclei. This can then be used to automatically generate outline files that can be analyzed in FISH-quant.

FISH-quant is a freely available Matlab package to automatically analyze smFISH images in 3D,

CellCognition is available at <http://www.cellcognition.org>

CellProfiler is available at <http://www.cellprofiler.org/>

FISH-quant is available at <https://code.google.com/p/fish-quant/>

Installation instruction for the Matlab tool *FQ_seg* Matlab tool:

- **Copy the Matlab code** from dropbox on your computer. Do not let it in the “Download” folder, but copy it dedicated folder for Matlab source code.
- **Set the Matlab path.** The Matlab path is a set of folder that Matlab uses to look for each of functions you run. If you try to run a function which is not present in one of the folder on the Matlab path, you will have a “function not found” error message. You can change the path Matlab in the menu, selecting *File > Set Path*. You need now to add the folder which contains the code to the path with the button ‘Add with subfolders’. Before closing this interface, click on ‘save’ to save the updated path settings.
- You can now type *FQ_seg* in the command window of Matlab to launch the user-interface.

Workflow

Workflow requires *CellProfiler*, *FISH-quant* (Matlab), and a special Matlab user-interface (*FQ_seg*) that allows controlling many of the steps described below.

1. Image acquisition in 3D.
2. Generate z-projections to generate 2D images
3. Segmentation of 2D images in *CellCognition* or *CellProfiler*
4. Conversion of segmentation results in *FQ* outline files.
5. Check quality of segmentation results

Image-acquisition

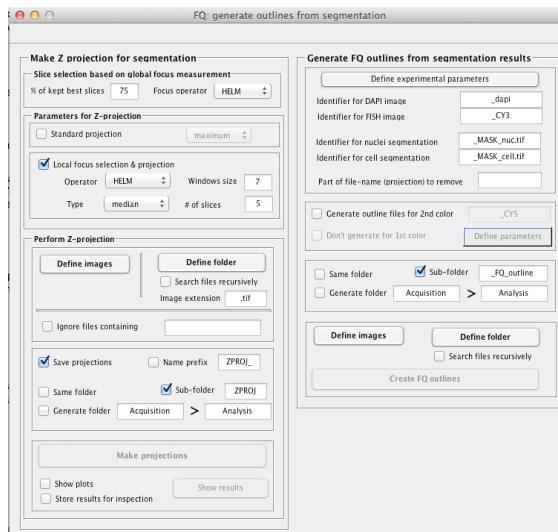
- Images have to be 3D and saved as a separate stacked TIF for each channel.
- Images have to follow a strict naming convention and start with a unique name, e.g. *img_001* where **001** is a running index. Subsequent identifiers specify the image type, e.g. **_DAPI** for DAPI images and **_CY3** for FISH. The complete file-name for different FISH image would be *img_001_CY3.tif*, *img_002_CY3.tif*, ... and similar for DAPI *img_001_DAPI.tif*, *img_002_DAPI.tif*,

2. Z-projections

CellProfiler and *CellCognition* work only on 2D images. Frequently, for the segmentation a **maximum intensity projection** of each channel is used. However, we found that better segmentation results can be obtained by performing a more advanced projection method. Specifically, we use focus measurements to obtain ‘shaper’ images to increase the segmentation accuracy. This is achieved by a 2-step process:

1. Calculation of the total focus of each z-slice. With this measurement, blurry slices, which have mainly out-of-focus intensity, can be removed.
2. Local focus projection. Here the focus is measured locally for a user defined window size. Then the z-projection is performed for each pixel over the best-identified slices.

These projections can be obtained with a Matlab tool which can openend by typing **FQ_seq** in the command window. This will open the user interface shown below.



Slice selection

In the first panel you can select the focus projection method (See ³ for details, default ‘HELM’). You can then select which percentage of slices will be kept (default 75%). To use all slices, defined a value of 100%

Z-projection

You can either perform either a **standard global projection** based on a mean, median, or maximum projection or projection based on local focus measurements.

For the local focus, you can again select the focus metric. In addition, you can define the window size that will be use to calculate the local focus. You can specify how many slices with the respective best focus values will be projected. Finally, the same projection method as for the global approach are available.

Define images / Define folder

In this panel, you can specify the images that will be analyzed. You can define manually images in a given folder with the button ‘Define images’. Alternatively, you can specify a folder in which all images will be analyzed. Here you have to define in addition the image extension, e.g. *.tif*. The image extension is case-sensitive. You can also define this recursively, i.e. images will be searched in the defined folder and all subfolders.

You can also specify that certain files will not be processed. Here you have to specify parts of the file-name that uniquely specifies them, e.g. ‘CY3’.

Save images

The generated images can then be saved. By default, they are saved under the same file-name as the original image in a subfolder of this imaged called 'ZPROJ'. There are different possibilities to change this behavior.

- You can add a pre-fix to the file-name.
- You can save the files in the same folder as the original files. **ATTENTION:** if you don't add a prefix the original files WILL be deleted.
- You can save the files in a different folder. This option allows you to replace parts of the folder-name by another string. For instance, files are saved in `/Users/FM/img/raw`. You can now say that 'raw' should be replaced by 'ZPROJ'. The images will then be saved in the folder `/Users/FM/img/ZPROJ`

Note: it is possible to combine saving in a subfolder and replace parts of a name of the folder.

Additionally to the images, the program will also save a small text file with the settings used to create the z-projections.

Show plots

If you enable this option, you will obtain a plot for each image that summarizes the results for each analyzed image. We recommend doing this only for a small number of images, to avoid having too many open images.

Show results

With this option enable, the program will save the standard maximum intensity projection and the advanced projection for each image. When the processing is done, you can then inspect these in a separate user interface by opening 'Show results'. Enable this option when processing only for a few images, otherwise the resulting data might be so large that Matlab crashes.

Perform z-projection

Pressing the button 'Make Projections' will then apply the Z-projection to all specified images. Images with fewer than 3 Z slices will not be considered in the analysis.

3. Workflow for CellCognition

Segmentation

CellCognition recursively scans user provided folders for images, performs the segmentation, and saves the segmentation results as masks. Such workflows are saved as **settings files**, which are provided with this toolbox (files ending with .conf).

We can load the settings file from within *CellCognition* using *File>Open Settings*. Once the settings file is loaded, we just have to set the input and output folder and the segmentation is ready to start.

Two different settings file are provided:

- “Global method setting” which combine a global thresholding with a watershed segmentation on the image.
- “3 otsu method setting” where a 3 classes thresholding is performed with the possibility to assign the middle class to the background or the foreground combined to a watershed on the image. We assign the middle class to the foreground for our cell segmentation.

→ In our experience, the “3 otsu method settings” gives better results for cell segmentation.

Loading of images

CellCognition needs an **input folder**, which is scanned at the beginning of the analysis to detect all the images. It does a recursive search of the folder, looking into every subfolder of the input folder for images.

An important option is ‘**multiple plates**’, if when enabled indicates that the input folder contains subfolder that corresponds to different plate. If the option is enabled, the organization of the results will be organized in a per plate basis (more details below).

The import of images is based on a **regular expression** that allows extracting basic information about the images from their file names (well, time, position etc....). These regular expressions are stored in the “naming_schemas.ini” file, which can be found with *CellCognition* code package (On a Mac, right-click on the *CellCognition* icon and chose “Show package contents”) at “content\Resources\resource\naming_schemas.ini”. We provide an updated file, with a naming scheme called *FISH_HCS* adapted to the file-names with the naming convention (for more details see dedicated section in the FQ_HCS document)

well_cell_gene_pos_channel.tif.

For instance, *CellCognition* will interpret w3_HeLa_DYNC1H1_P4_DAPI.tif as

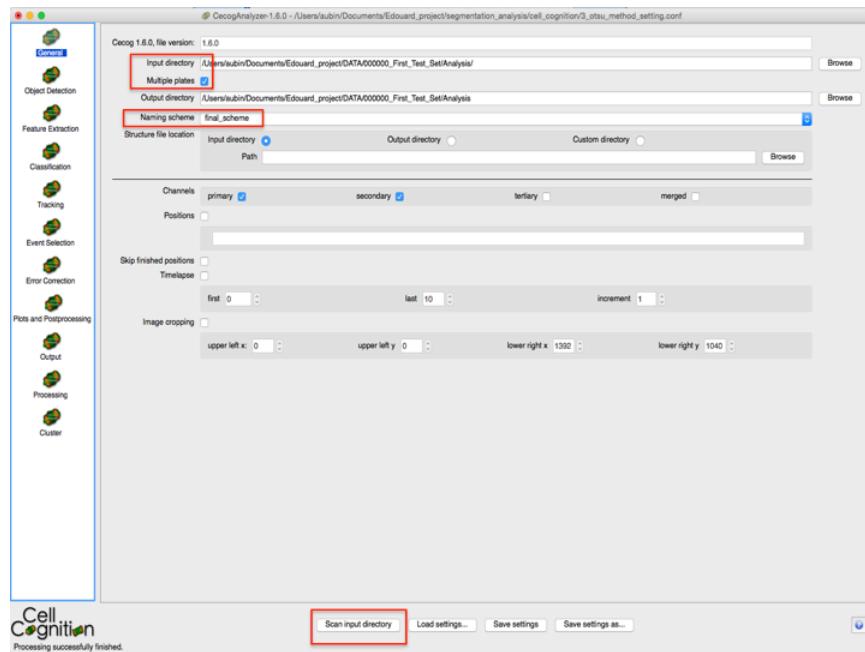
Plate :	Determined by the image folder
Well :	3
Subwell :	4
Channel :	DAPI

→ Position (as defined by *CellCognition*): 3_4

Once the input folder is defined and the correct naming schema is selected, you can click on “**scan input directory**” and *CellCognition* displays how many plates/images have been found.

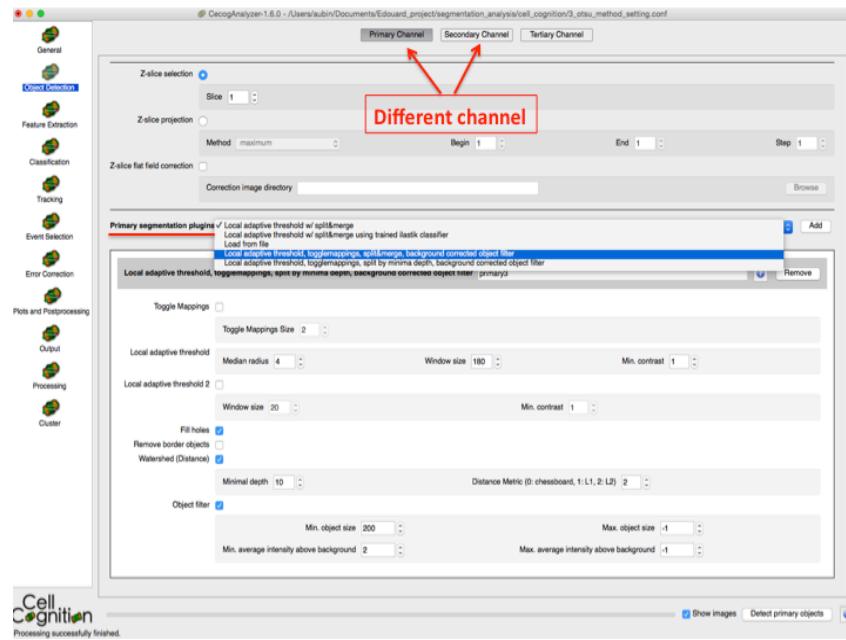
Naming convention of CellCognition: position, well, subwell

In *CellCognition*, the position of an image is a unique value for each plate; therefore it is NOT possible to have two identical positions for two wells belonging to the same plate. Alternatively, an individual image can be assigned to a well and a subwell. The position is then the combination of the well and the subwell. In a given plate, the position will be for example « 001_04 », where 001 is the well and 04 is the subwell.

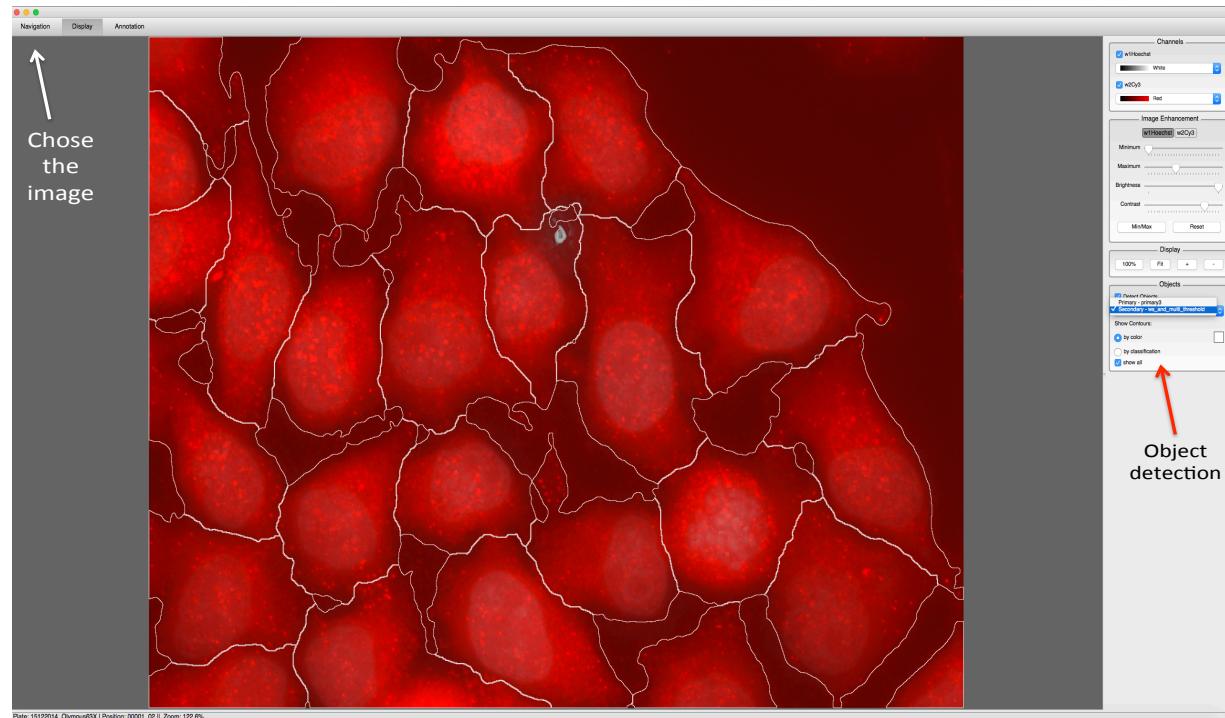


Cell segmentation

The segmentation of cells is a two steps process. First, nuclei will be segmented with a global thresholding strategy. Second, each nuclei will serve as a ‘seed’ for the segmentation of the cells. These two detection steps have to be defined by the user. Once in the “object detection” tab, it is possible to navigate between the different channels, and, for each one, to select a primary/secondary plugin in the list box and add it. The set of parameters are now visible and can be adjusted.



The easiest way to adjust the parameter is to use the browser (**CTRL + B**). It is possible with the browser to select an image in the list, to adjust the contrast of the several channel, and to perform and visualized the segmentation with the settings specified in the main interface of *CellCognition*.



Saving the results

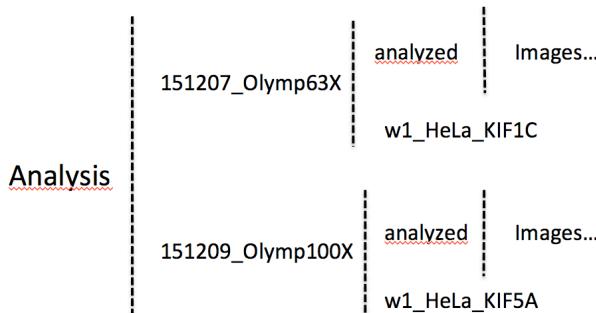
In the **output panel**, we can specify what segmentation result we want to export and save

- “**Contour images**” are the original images with the outline of cells segmented; it allows to visualize the segmentation results compared to the real images.
- “**Label images**” are simple images, with pixel value of 0 in the background, and a different value for each cells; these are the results we use to create FQ outlines.

Analysis and organization of results files

To launch the analysis on the whole image data set, use the tab “processing”. Press the “start processing” button to start the analysis.

The results will be saved in the output folder defined at the beginning of the analysis. CellCognition automatically creates 4 folders called ‘*hdf5*’, ‘*log*’, ‘*plots*’ and ‘*analyzed*’. If the *multiple plates* option is not checked, the set of 4 folders is created directly in the output folder defined at the beginning of the analysis. Otherwise, the 4 folders will be created in each subfolder corresponding to a plate.



The folders 151207_Olymp63X corresponds to a plate, containing different subfolders such as w1_HeLa_KIF1C which corresponds to wells (more information can be found in the dedicated FQ_HCS document).

The segmentation results can be found in:

`analyzed/images/labels/channel/segmentation_method/`

The label images are named based on the information extracted with the regular expression:

`P01_04_T00.tif`

Where *P* is the position as explained above and *T* the time (always 1 in our case since we don't have time-resolved data).

Generate FQ outlines files from CellCognition results

The segmentation masks generated by *CellProfiler* can be used to automatically create FQ outline files with the middle part of the user interface.

Specify experimental parameters

First, you have to define the '**Experimental parameters**'. These are the same parameters that have to be defined for the images in the main FISH-quant interface when pressing the 'Modify' button in the panel 'Experimental parameters'. Simply adjust them to your data by changing the corresponding values.

Note 1: You have to ALWAYS define the parameters – even if the default parameters are good. Only then the button to generate the outlines will be enabled.

Note 2: you can change the default parameters in a text-file called 'FISH-QUANT_default_par.txt' which is located in the same directory as the *FQ_seg* Matlab file.

Define channel and segmentation method identifier

Second, you have to define the names of the two channels and the names of the segmentation method for the first and for the second channel. When saving the results, Cell Cognition creates a folder per channel, and a subfolder per segmentation method used to segment the channel. You have to indicate to the FQ_Seg which channel result you want to use, and from which segmentation method. The names have to correspond to these folders names.

Specify folder where the results will be saved

You only have the possibility to save the outlines file in a subfolder of the well folder of the projection. You can specify the name of the subfolder from the "Outline folder" of the GUI. Your outline will always be saved in "analysis→w1_HeLa_KIF1C→"Outline_folder".

Define folder

You have to define one or multiple folder that **corresponds to a plate**.

Create outlines

Lastly, press the button '**Create FQ outlines**'. The script will then automatically search for the files describing the segmentation of cells and nuclei. For each image an outline file with the reference to the original 3D image will be generated and nuclei assigned to their respective cells.

4. Workflow for CellProfiler

Segmentation

Below a typical workflow in CellProfiler for the segmentation of nuclei and cells is shown. The results are saved in masks that can be used to generate outline files for FISH-quant. Such workflows are saved in so-called pipelines, which are provided with this toolbox (files ending with `.cppipe`).

1. Open CellProfiler
2. Import Pipeline from Menu *file > Import Pipeline > from file*
3. Important first steps are the **properly load and assign the images of the different channels**. This is done in the 'Input model' and 'Select Images'. To provide images, you can simply drag & drop images (or entire folders) into the file list (select Images on the left).
4. The next step is extracting of meta-data. In the provided script this is done based on regular expressions analyzing the actual image-name. The image name follows the convention `cell_gene_pos_channel.tif`, where `cell` is the used cell line, `gene` the analyzed gene, `pos` the position of the image (i.e. a value that will be increased for each acquired multi-color stack), and `channel` the imaging channel. If your image names don't follow this convention, you have to adapt the regular expression (See *CellProfiler* for more help.)

To apply the settings to extract the Metadata press the button 'Update metadata'. This might then take a little while. If everything works fine, a table in the lower part of the panel should populate. Containing information (at least) about the channel and position. If that's the case you can continue

5. Go to the 'NamesAndType' module. Here simply click update. This should populate a table in the lower part, where the different images of the different channels are correctly assigned to each other. Please consult the CellProfiler help if this doesn't work.
6. You can then either simply apply the segmentation settings to all images (Button 'Analyze Images') or start a test mode (Button 'Start test mode').
7. We recommend saving the CellProfiler project in the folder with the Z-projections. This allows later to inspect the settings used for the segmentation.

Analyze Images

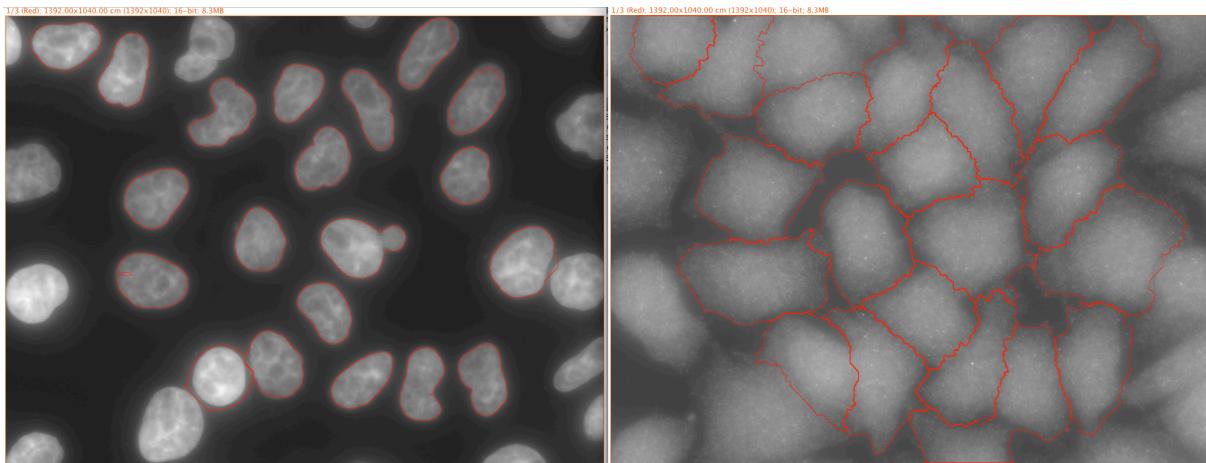
This will perform the automated segmentation of all images and save the results automatically.

FOR FASTER PROCESSING DISABLE THE DISPLAY OF THE SEGMENTATION. This can be done by clicking on the icon with the eye next to the different processing steps.

The **segmentation results** are saved in the folder of the MIP

- a. Masks for cells and nuclei. These are images with the same name as the MIP, followed by '_MASK_cell' or '_MASK_nuc'. In these images each identified cell or nucleus has a different gray-scale value. They will then serve as an input to automatically generate FQ outline files.
- b. Images that allow to check how well the segmentation worked, they also have the same file-name as the MIP and end by '_CONTOUR_cell' or '_CONTOUR_nuc'

Below are an example for segmentation nuclei (left) and cells (right).



Test mode

This allows you to change settings and look at their impact at the individual steps of the pipeline. More information can be found in the *CellProfiler* help.

Below typical settings used for nuclear and cellular segmentation are shown.

<p>Nuclei</p> <p>Select the input image DAPI (from NamesAndTypes)</p> <p>Name the primary objects to be identified Nuclei</p> <p>Typical diameter of objects, in pixel units (Min,Max) 50 200</p> <p>Discard objects outside the diameter range? <input checked="" type="radio"/> Yes <input type="radio"/> No</p> <p>Discard objects touching the border of the image? <input checked="" type="radio"/> Yes <input type="radio"/> No</p> <p>Threshold strategy Global</p> <p>Thresholding method Otsu</p> <p>Two-class or three-class thresholding? Three classes</p> <p>Minimize the weighted variance or the entropy? Weighted variance</p> <p>Assign pixels in the middle intensity class to the foreground or the background? Background</p> <p>Select the smoothing method for thresholding Manual</p> <p>Threshold smoothing scale 4</p> <p>Threshold correction factor 1.1</p> <p>Lower and upper bounds on threshold 0 1</p> <p>Method to distinguish clumped objects Shape</p> <p>Method to draw dividing lines between clumped objects Intensity</p> <p>Automatically calculate size of smoothing filter for declumping? <input checked="" type="radio"/> Yes <input type="radio"/> No</p> <p>Automatically calculate minimum allowed distance between local maxima? <input checked="" type="radio"/> Yes <input type="radio"/> No</p> <p>Speed up by using lower-resolution image to find local maxima? <input checked="" type="radio"/> Yes <input type="radio"/> No</p> <p>Retain outlines of the identified objects? <input checked="" type="radio"/> Yes <input type="radio"/> No</p> <p>Name the outline image NucoOutlines</p> <p>Fill holes in identified objects? <input checked="" type="radio"/> Yes <input type="radio"/> No</p> <p>Handling of objects if excessive number of objects identified Continue</p>	<p>Cells</p> <p>Select the input image CY3 (from NamesAndTypes)</p> <p>Select the input objects Nuclei (from IdentifyPrimaryObjects #05)</p> <p>Name the objects to be identified Cells</p> <p>Select the method to identify the secondary objects Watershed - Image</p> <p>Threshold strategy Global</p> <p>Thresholding method Otsu</p> <p>Two-class or three-class thresholding? Three classes</p> <p>Minimize the weighted variance or the entropy? Weighted variance</p> <p>Assign pixels in the middle intensity class to the foreground or the background? Foreground</p> <p>Select the smoothing method for thresholding Manual</p> <p>Threshold smoothing scale 5</p> <p>Threshold correction factor 1.1</p> <p>Lower and upper bounds on threshold 0 1</p> <p>Fill holes in identified objects? <input checked="" type="radio"/> Yes <input type="radio"/> No</p> <p>Discard secondary objects touching the border of the image? <input checked="" type="radio"/> Yes <input type="radio"/> No</p> <p>Retain outlines of the identified secondary objects? <input checked="" type="radio"/> Yes <input type="radio"/> No</p> <p>Name the outline image CellOutlines</p>
--	--

Generate FQ outlines files from CellProfiler results

The segmentation masks generated by *CellProfiler* can be used to automatically create FQ outline files with the middle part of the user interface.

Specify experimental parameters

First, you have to define the '**Experimental parameters**'. These are the same parameters that have to be defined for the images in the main FISH-quant interface when pressing the 'Modify' button in the panel 'Experimental parameters'. Simply adjust them to your data by changing the corresponding values.

Note 1: You have to ALWAYS define the parameters – even if the default parameters are good. Only then the button to generate the outlines will be enabled.

Note 2: you can change the default parameters in a text-file called ‘FISH-QUANT_default_par.txt’ which is located in the same directory as the *FQ_seg* Matlab file.

Define naming scheme of original images and segmentation results

Second, you have to define a few parameters regarding the naming convention of your files. You have to define the unique identifier for the FISH and DAPI images. These identifiers have to be defined in a way that when you take the full file-name of the FISH image and you replace the identifier for FISH, e.g. CY3, by the identifier of the DAPI, e.g. ‘DAPI’, you get the DAPI file-name. Then, you have to define the identifier of your segmentation results using the suffix that *CellProfiler* added in the name of your masks (MASK_nuc and MASK_cell for example).

[Optional] Generating outline files for a second color

This option allows to generate outline files for a second color, e.g. for a dual-color FISH experiment. The outlines for this color will be based on the segmentation results of the first color and the exact same cells will be used. This allows a simple comparison between the detection results.

As above, the identifier for the second color has to be specified, e.g. ‘CY5’. You also have to redefine the experimental parameters (most often to adjust the excitation and emission wavelength).

You can also choose that only the outlines of the second color are saved. This option can be useful if the first color doesn’t contain actual smFISH data but results of a dedicated cell segmentation stain.

Specify folder where the results will be saved

The same options as for the projections are available. More details can be found there.

Create outlines

Lastly, press the button ‘**Create FQ outlines**’. The script will then automatically search for the files describing the segmentation of cells and nuclei. For each image an outline file with the reference to the ORIGINAL 3D image will be generated and nuclei assigned to their respective cells.

5. Inspect outline files

The next few steps summarize a simple workflow for a manual inspection of the outlined cells

1. Open FQ and define the folders with images and the new outlines.
2. Open 'List directory content' from the tools menu. Here navigate to the folder with the outlines. Double-click of an outline will display it on the outline designer.
3. Click on the list with the different cells. You can use the arrows to navigate through this list (see shortcuts below).
4. Delete each cell that's not good with CTRL-X
5. Save outline with CTRL-S

The following short cuts help to speed up the analysis

CTRL-S	Save outline
CTRL-X	Delete currently selected cell <i>Note: works only when you clicked on the list with the cells before</i>
Arrow UP	Previous cell <i>Note: works only when you clicked on the list with the cells before</i>
Arrow DOWN	Next cell <i>Note: works only when you clicked on the list with the cells before</i>

6. References

1. Held, M. *et al.* CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. *Nat. Methods* **7**, 747–754 (2010).
2. Kamentsky, L. *et al.* Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics* **27**, 1179–1180 (2011).
3. Pertuz, S., Puig, D. & Garcia, M. A. Analysis of focus measure operators for shape-from-focus. *Pattern Recognition* **46**, 1415–1432 (2013).