FISH-quant v3

Co-localization analysis from dual/multi-color smFISH images

MUELLER Florian[1] muellerf.research@gmail.com

[1]Institut Pasteur Paris, *Computational Imaging & Modeling Unit, France*

1. Introduction

This document describes the workflow for how to perform a colocalization analysis in FISH-quant[1] for two or more channels. The workflow requires *FISH-quant* (MATLAB) with a dedicated user-interface (`FQ_DualColor`).

Co-localization between the two channels is calculated as a linear assignment problem (LAP) solved with the Hungarian algorithm. We use the Matlab function `hungarianlinker`[2] and `munkres`[3] for this purpose.

Image names have to have a common part, and a unique part, which is indicative of the channel. In this example we assume that the channel indicator is *_color1* and *_color2*. For instance, images could be called *Test_pos1_color1.tif* and *Test_pos1_color2.tif*

2. Generating outline files

It is important that the outline files for the different colors are identical with respect to defined cellular outlines and the names these cells have. This is necessary since the Matlab functions calculating co-localization will automatically match cells with identical names to perform the analysis.

There are several ways to generate these outlines files. These steps are necessary since a given outline is also associated with a specific image. So each time you load an outline, FQ will automatically load the associated image.

Image-by-Image

You can define in FISH-quant the outlines for one given color for one or multiple images and save each of them. Then you can open images in the other color one-by-one and select from the menu *[FQ] Main* > *Load* > *Outline from another color*. This will overlay the selected outline on the currently opened image. Then save the outline which will now generate an outline file with the open image and the loaded cellular outlines.

Batch-mode

In the module FQ_DualColor, you can create automatically outlines in a different color. You first have to specify unique strings for each color, e.g. *CY3* and *CY5*. By replacing the string of the first color in the file-name of the outlines/images by the string of the second color, you have to obtain the file-name of the second color. For instance, let's assume the outlines in the first color are called *Test_pXY_CY3_outline.txt*, where *pXY* are the different images (*p01*, *p02*, ..); and the outlines of the second color are called *Test_pXY_CY5_outline.txt*. Then replacing CY3 by CY5 will yield - as required - for each outline of the first color the outline of the second color. The same has to hold true for the image names that are referenced within these outlines files.

Then you have to press on Create. Here you have to select all outlines in the first color that you want to convert. FQ will then open each outline and will (a) change the name of the referenced image in the outline (b) save this outline file under the new file-name. These files will be stored in a subfolder with the name of the unique string of the second color, e.g. CY5 in our example.

3. Spot detection

After creating the outline files for each color, the actual spot detection has to be performed in FISH-quant. The details depend on your actual data, so you can either detect spot for one color in all images with the batch detection, or if your data is more heterogeneous you can analysis each image separately. In either case, the results of the spot detection have to be saved for each of the images (e.g. for each of the analyzed outline files). For more details on the different steps in the spot detection, we refer to the FISH-quant manual.

What is important that the same correspondence between the file-names with the spot detection results. So again, we have to be able to got from one color to another by simply replacing a string. For instance, results of the first color could be called *Test_pXY_CY3_spots_160101.txt*, while the results of the second color are called *Test_pXY_CY5_spots.txt*. Here the unique string of the first color is *CY3_spots_160101,* while for the second color it is *CY5_spots.*

4. Perform co-localization analysis

This analysis is performed in the module *FQ_DualColor*.

Used algorithm

FISH-quant performs the co-localization analysis in each cell separately. It treats this task as a Linear Assignment Problem (LAP). In short, the two spot detection results are considered as 3D point clouds. The Hungarian algorithm solving the LAP finds the best possible global assignment between these two points-clouds such that for each point in the first channel the closest point in the second channel is found. The LAP has the important property that assignment is exclusive, one point from the first channel can be linked to at most one point from the other channel, and conversely. The linking is also globally optimal – the sum of the squared distance is minimized, contrary to at the naïve nearest neighbor approach. We implement this with the Matlab function hungarianlinker[2] and munkres[3] which are available from Matlab FileExchange.

Define unique strings for each color

The first step is again to define the unique strings that define the outlines in each color as explained at the end of the preceding section.

Define files that should be analyzed

By pressing on the button ‘Specify Files’, you can define the files that will be analyzed. This will open a sequence of 4 dialog boxes where you have to specify

1. All result files of the first color that should be analyzed
2. Folder with the results files of the first color
3. Folder with the image files of the first color
4. Folder with the image files of the second color/

Pressing *Cancel* in either of this dialog boxes will cancel the definition.

Options for co-localization analysis

You can then define the maximum allowed distance between two spots to be considered co-localized. This distance is expressed in nm and is a 3D distance. After the analysis is done, FQ will show a plot with the number of co-localized spots as a function of different distances, which go up to the defined thresholds. In an ideal case, this number shows a plateau indicating that increasing the allowed distance is not changing the number of co-localized spots.

You can also define a maximum number of spots a cell can have to be considered in the analysis. With this option, highly expressing cells can be excluded from the analysis. These cells can pose problems since co-localization can occur by chance.

You can calculate the average drift in your data. Here the program will determine for each pair of co-localized spots their distance independently in X,Y, Z. After all data is analyzed a histogram for each axis is shown. You can then correct for this average drift in a second analysis run. Here FQ will subtract the average drift from the detected position in the second channel, before performing the co-localization. If you keep the option to calculate the drift enabled, FQ will show you the resulting final drift.

You can **save images summarizing the co-localization results**.

These files will be saved in a sub-folder of the results file of the first channel.

- *Save analysis details*. FQ will save the displayed plots summarizing the number of co-localized spots and (if available) the drift correction plot.

- *Save images of each cell*. FQ will save an image, where it shows in the upper row the two channels with all detected spots for each of the images. In the lower row, it shows the co-localization results, where co-localized spots are shown in blue. Red spots correspond to detected spots in the respective image that were not co-localized. Yellow circles are spots that have been detected in the other channel that didn't co-localize.

- *Same images of individual spots.* FQ will save an image of each spot. If will generate a folder **+** for co-localized spots and a **–** for not co-localized spots. The file-name indicates the name of the results file, name of the cell, and the index of the spot. In the image the cropped image around the detection in the first channel is shown on the left, for the image of the second channel on the right. Detected spot positions are shown as circle. On the right, information about the spots and their estimated distance is shown. In the case of spots that don't have a co-localization, FQ only looks at the channel with fewer detections. Here there the same plot is shown, expect that FQ will look for all detections in the other channel that are within the displayed window and will show (if available) the distance to the closest spot.

Saving results

After your satisfied with the results, you can save the results of the co-localization in tab-delimited text files.

- Summary file. FQ will save for each cell how many spots were detected and how many co-localize

- Detailed results. FQ will save each co-localized spots from either channel with the estimated positions, intensity, fitted amplitude and sigma, estimated integrated intensity, and maximum pixel intensity in both raw and filtered images. This file can be used for a more detailed analysis to infer if co-localized particles have certain characteristics.

5. References

1. Mueller, F. *et al.* FISH-quant: automatic counting of transcripts in 3D FISH images. *Nat. Methods* **10,** 277–278 (2013).

2. Hungarian based particle linking - File Exchange - MATLAB Central. at <http://fr.mathworks.com/matlabcentral/fileexchange/33968-hungarian-based-particle-linking>

3. Hungarian Algorithm for Linear Assignment Problems (V2.3) - File Exchange - MATLAB Central. at <http://fr.mathworks.com/matlabcentral/fileexchange/20652-hungarian-algorithm-for-linear-assignment-problems--v2-3->