FISH QUANT v0I

Matlab package to analyze and visualize data of mRNA FISH experiments

***Florian Mueller**, muellerf.resarch@gmail.com*

Institut Pasteur Paris, Computational Imaging & Modeling Group
ENS Paris, Functional Imaging and Transcription Group

Overview

Matlab package to quantify mRNA FISH data. There are two main parts

- Single mRNA detection in the cell. Allows estimating the number of transcripts in a cell.
- Quantification of number of transcripts at the transcription site. (not implemented)

These functions can either be performed cell-by-cell or automatically for a larger population of cells.

The main program is FISH_QUANT. This allows accessing and controlling all other functions.

Typical workflow

- Finding optimized conditions for mRNA detection by analyzing a small number of cells.
- Run the automatic quantification.
- Quality check by manually looking at some of the analyzed cells.
- Analyzing the transcription site.

General use to detect single particles

FISH-QUANT was designed to analyze mRNA FISH data but it can also be used for any kind of single particle detection. Defining the outline of a 'cell' allows a restriction to certain areas of the image and the definition of a 'transcription site' allows the exclusion of certain areas of this image.

As an example, we use FISH-QUANT to analyze the PSF generated with fluorescent beads.

Installation

Requirements

Functions were programmed and tested for Matlab 7.11. We use several function of *the Piotr's Image & Video Toolbox for Matlab* [1]. The following Matlab toolboxes are needed additionally

- *Optimization toolbox*
- *Statistics toolbox*
- *Image Processing Toolbox*
- (Optional) *Parallel Computing Toolbox*

Matlab code

1. Unzip archive in user folder of Matlab. Under windows the Matlab work folder is usually called something like `C:\Users\muellerf\Documents\MATLAB`, where `muellerf` is the user name. In Matlab the user path can be found with the command `userpath`. After extracting there should now be a folder `C:\Users\muellerf\Documents\MATLAB\FISH_QUANT`.
2. **Update Matlab path definition**. This can be done with a few simple steps in Matlab
   a. In the Matlab menu select *File* > *Set Path*
   b. This will open a dialog box. In this box select *Add with subfolders …*
   c. This will open another dialog. Here select the folder of FISH_Quant from step 1, e.g. `C:\Users\muellerf\Documents\MATLAB\FISH_QUANT`. Click *OK*.
   d. To save this settings press on *Save*. Depending on the settings of the installation of Matlab this might results in a warning saying that the changes to path cannot be saved. Matlab proposes to save the path-definition file *pathdef.m* to another location. Click *Yes*. Select a directory of choice, e.g. the Matlab work directory of the user.

IMPORTANT: if an older version of FISH-QUANT is already installed on the computer, either delete this folder, move it to a different location, or remove it from the path-definition of Matlab.

Update path for ImageJ

FISH-QUANT provides the possibility to use ImageJ to visualize some of the analysis results. To use this functionality a few simple initialization steps have to be performed once.

1. ImageJ has to be installed properly according to the instructions on the website [2].
2. The ImageJ application path has to be set in FISH-QUANT. This can be done in the text file `FISH-QUANT_def.txt` which can be found in the installation directory of FISH-QUANT (see also Appendix 5). This files contains several important start-up parameters. The relevant path of ImageJ can be changed in the line `path_imagej=/Users/fmueller/`. In this line, change the red text to the installation path on your system. ***Important***: don't change the name of the file. Don't add any space in between the equal sign and the text!

Start FISH-QUANT

FISH-QUANT uses two different java-libraries for the read-in and display of images (MIJ, and BioFormats, details see below). To guarantee that these packages are successfully installed FISH-QUANT has to be called correctly. Several options are possible. Open Matlab and type in the command window and depending on installation type one of the following commands

- **`FISH_QUANT`**
  If the installation was performed as outlined above, FISH-QUANT can be called without any call-options. The necessary java-libraries are provided with FISH-QUANT and are references accordingly.

- **`FISH_QUANT('root')`**
  FISH-QUANT can be called with key-word `'root'` if the Java-files of MIJ and Bio-Formats were saved in the java sub-directory of the Matlab root directory.

- **`FISH_QUANT('dir','path')`**
  FISH-QUANT can be called with the key-word `'dir'` if the Java-files of MIJ and Bio-Formates were saved in a different directory. `'path'` indicates this directory. For example, if the files were saved in directory C:\java, then the function call will be `FISH_QUANT('Dir',' C:\java')`

Note: the main user-interface of FISH-QUANT should always be started first. This initiates MIJ and Bio-Formats properly.

MIJ – Interface between Matlab and ImageJ

FISH QUANT uses the software package MIJ to visualize results in ImageJ[3]. The package will be automatically installed during the installation process. The package requires a proper installation of ImageJ[2]. More detailed description can be found on the website http://bigwww.epfl.ch/sage/soft/mij/.

If large images will be used the (Java heap) memory has to be adjusted. Remember that increasing the size of the Java memory may correspondingly decrease the amount of space available for MATLAB arrays. There are two different ways to achieve this. Check the Matlab page for details.

- With more recent versions of Matlab, you can increase the memory by using File > Preferences > General > Java Heap Memory.

- It this is not possible a file called *java.opts* file can be either generated or changed. This file can be found $MATLABROOT/bin/$ARCH directory. $MATLABROOT is the installation directory of Matlab, /$ARCH is the architecture of the system, i.e. maci64 for a 64bit Mac OS. To increase the memory to 512 MB add the following line to the text file "-Xmx512m". If you don't have write access to this file you can save a local copy in the user path (see above for how to find the user path).

Bio-Formats

FISH QUANT uses the Matlab function `bfopen` to read-in image data [4]. This function software package utilizes the Bio-Formats Java tools developed for ImageJ. This guarantees that a wide range of imaging data can be read.

Single mRNA detection

The following section describes the different ways to analyze data in FISH-QUANT

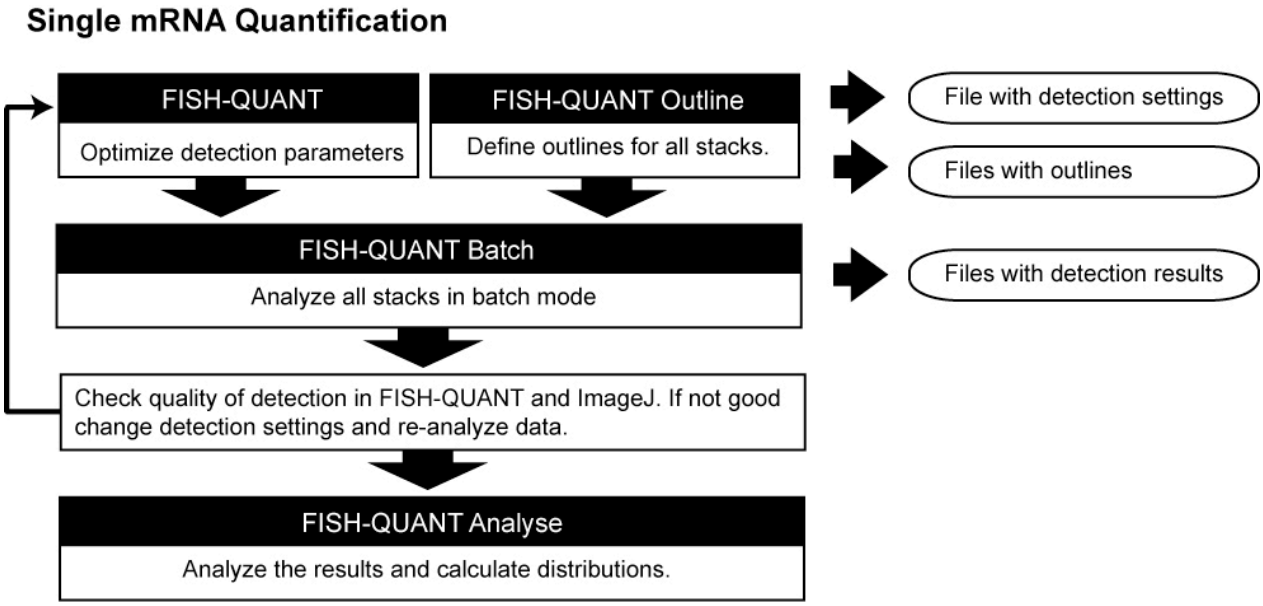


Fig. Workflow of automated single mRNA detection in FISH-QUANT. Black boxes indicated different graphical user interfaces.

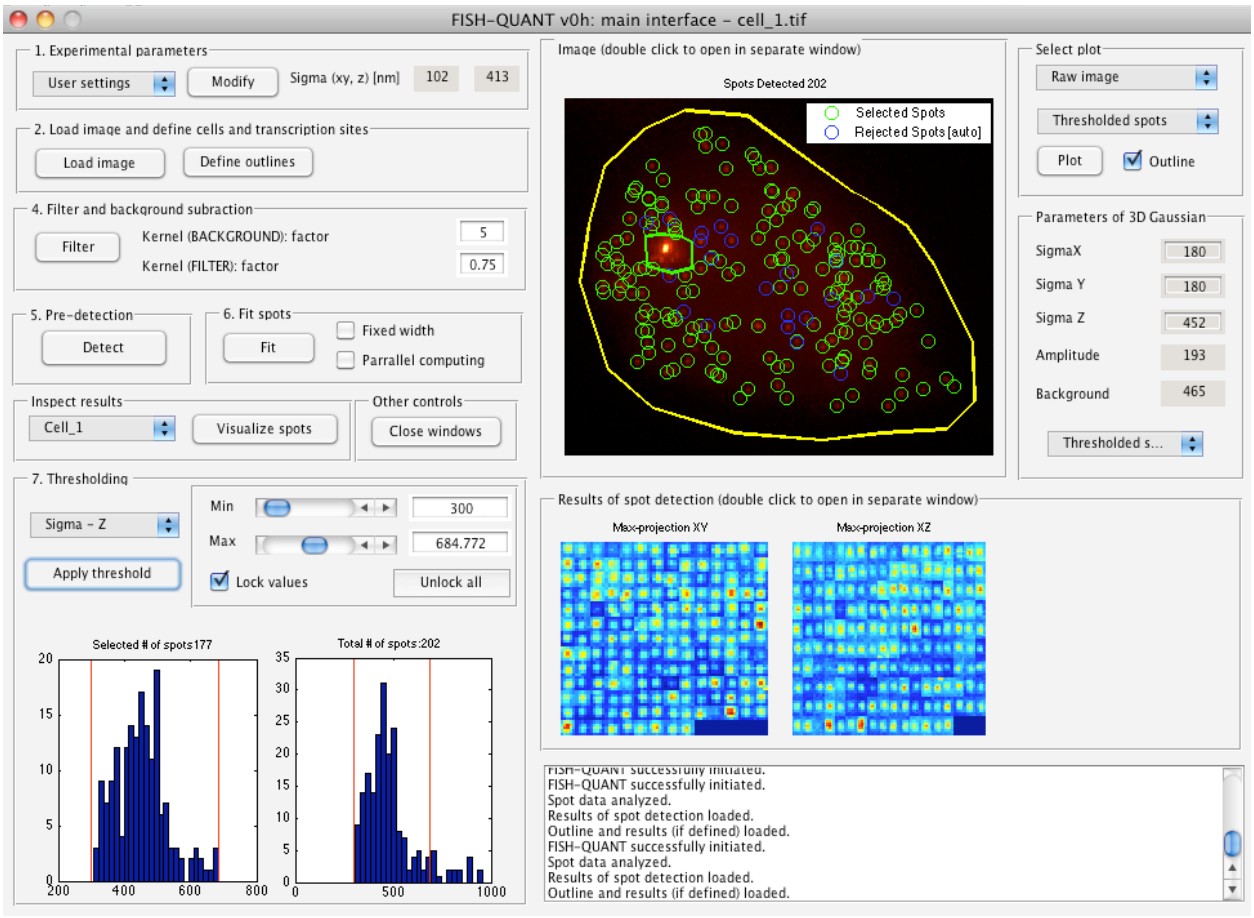
Example data-sets

Test-data can be downloaded the same web-site than FISH-QUANT. In the archive Test_data.zip you can find images which were already processed as well as images of the PSF.

Basic workflow

This section demonstrates the actual workflow which is then also implemented in the automated analysis of the data. The GUI allows optimizing the various parameters. Then this optimized set of parameters can be used in the automated detection.

The screen-shot below shows the main interface of FISH-QUANT with the results of the analysis of a typical experiment. In the next sections the main step will be explained.



1. Define experimental parameters

In this panel the parameters of the experiment can be specified. The user can choose from a number of presets that can be modified. Values needed are

- Pixel-size in XY in nanometer.
- Pixel-size in Z in nanometer.
- Refractive index of medium.
- Excitation wavelength of fluorophore.
- Emission wavelength of fluorophore.
- Type of microscope. Currently 'confocal', 'nipkow' and widefield are supported.

These values are used to calculate the theoretical PSF in XY and Z[5] (shown in the panel 'Experimental parameters'). These values are used for the filtering step and also as initial starting points for the fit with the 3D Gaussian. See Appendix 1 for more details.

2. Load image

There are two different options (a) load an image by pressing the button 'Load' or (b) load an already defined outline together with an image (see next section).

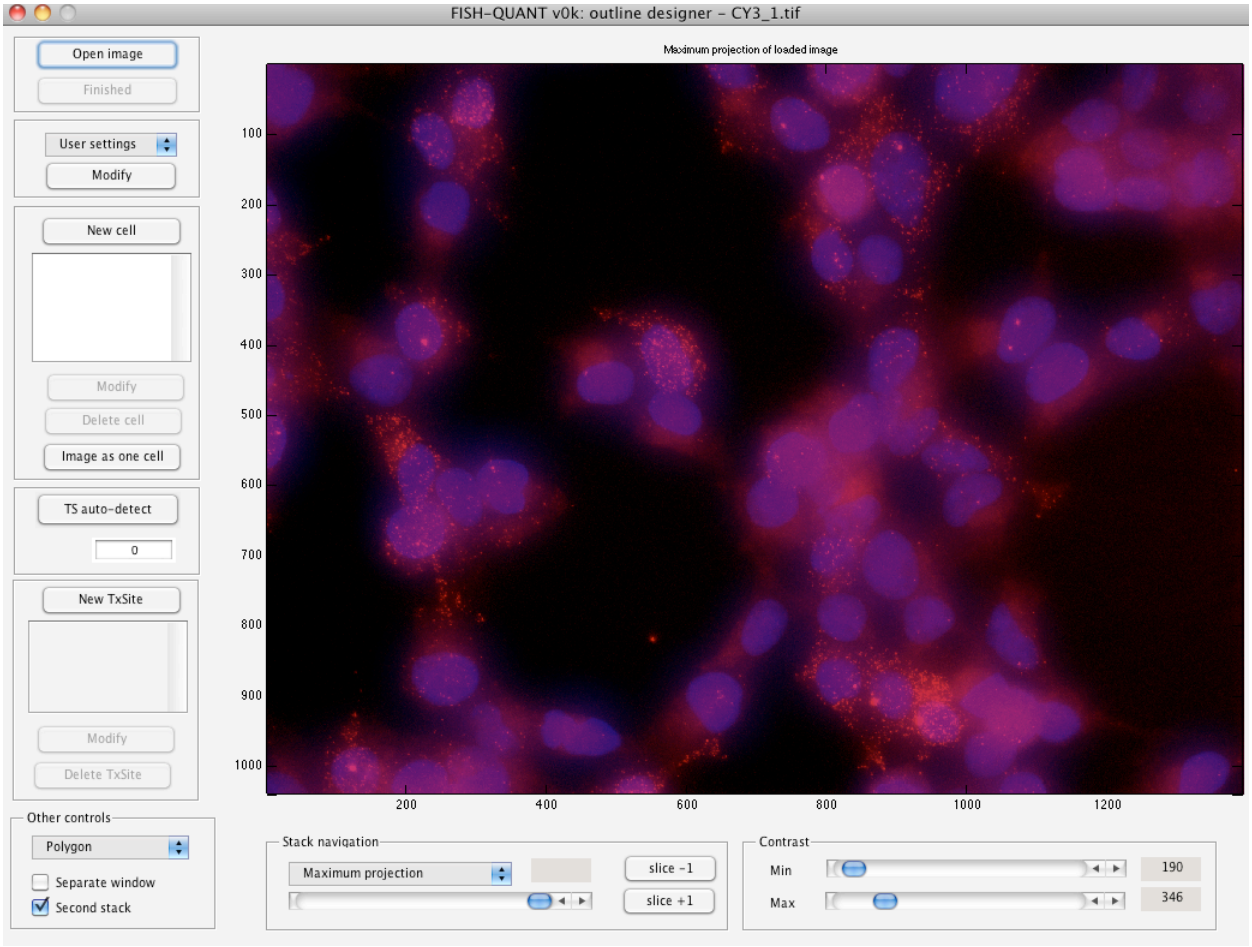
Either opens user dialog to specify the file name. Image files are read-in Matlab with the `bfopen` which utilized the BioFormats Java-library for reading life sciences image file formats [4]. For a list of supported format consult the Bio-Format web-site. In FISH-QUANT currently only single color stacks are supported.

3. Outline of cell and transcription site

Opens a new user interface that allows defining the outline of the cell(s) and the transcription-site(s) to be analyzed. In the analysis mRNA will only be detected inside each cell and the transcription sites will be ignored.

In this interface the position of each region can be specified and if needed also changed. Outlines can either be defined as polygons, rectangles, or ellipses. These geometries can be mixed in the same image.

The definition is either done on a XY-maximum projection or on the z-stack where a certain plane can be selected. The defined regions extend through the entire z-stack. The **contrast** of the displayed image can be changed with the sliders at the bottom part of the interface.



The FISH image can be overlaid with a second stack (usually a DAPI) stain. This stack can be loaded from the menu ‘Second stack’. The contrast of this stack and its transparency can be change with ‘Options’. In the lower left corner you can find a checkbox that allows enabling or disabling the display of this stack when loaded.

To **define a new cell** press ‘New cell’. Cells are defined with either polygons, rectangles, or ellipses and double-clicking on the first point closes the polygon. To *modify a cell*, select it in the list and press ‘Modify’; to delete press ‘Delete cell’. Transcription sites can be defined in a similar fashion. When finished, press button ‘Finished’ to return to main interface.

It is possible to zoom and pan in the window to make the selection of individual cells easier.

By selecting the ‘Image in separate window’ option the image will be shown in a separate figure. It is possible to **zoom into specific regions** for a more accurate selection. After selecting ‘New cell’ or ‘New TxSite’ press on the zoom icon and the cursor will turn into a magnification glass. Then enlarge the region of interest. Revert the cursor to the cross for outline selection by pressing on the zoom icon again. When the selection is finished the main interface for the outline selection will be shown. If a new cell or transcription site is defined, the same image as before (with the same zoom settings will be shown).

Transcription sites can be automatically detected. FISH-QUANT identifies regions that are above the specified threshold and assigns them to the corresponding cells. This function is implemented with the Matlab function bwlabeln, i.e. transcription sites are detected as connected components. The detection options can be changed in the Menu under options. The two parameters are the three-dimensional connectivity and the minimum distance between the centers of two detected transcription sites (in pixel). The detected sites can then be modified or deleted as described above.

Images can be either shown in a maximum intensity projection or as a z-stack.

Important points

When defining a new cell or a new transcription site the cursor will change from an arrow to a cross. Depending on the computer used this might take a few seconds; please wait until the cursor changes appearance before continuing.

Other features

- Each transcription site is associated to a cell.
- Transcription sites cannot be defined outside of cells.
- Interface can be called again to change the already defined geometry.
- This step can also be omitted and the detection will be performed on the entire image.
- The entire image can be selected as one cell with the option ‘Whole image as cell’. Then transcription sites can be defined. This is a useful feature to process for instance images of fluorescent beads. The ‘transcription sites’ could be region which are excluded from the analysis.
- The outline can also be saved in a text file, or already saved outline can be loaded. See next section “Saving and loading data’ for more details. The file-name of the image and also (if saved) the filtered image are stored in the outline definition file.
- Interface can also be loaded without a pre-loaded image from the Menu *Tools* > *Outline designer*. This is convenient for defining outlines for multiple images which can then be used in the batch-processing tool.

Note

- Filtering step will be performed for entire image. Pre-detection and fitting, however, only for the currently selected cell in the 'Define outline' panel.

4. Filtering

This step comes before the pre-detection. Filtering with a Gaussian Kernel will smooth the image. Two filtering steps are performed in 3D. First, the image is convolved with a larger Kernel to obtain a background image. This background image is then subtracted from the original image. Second, the background-subtracted image is smoothed with a smaller Kernel, which enhances the single particles.

The size of these two kernels can be determined by the two multiplicative factors in the filtering panel.

The filtered image can also be save (Menu: Load/Save > Save > Filtered Image). The name of the filtered image will be stored in the outline definition file and also in the results file (Appendix 2 and 3). It can also be used in the batch processing mode to avoid filtering images multiple times.

5. Pre-detection

Pre-detection is performed on the filtered image in a separate graphical interface (see Fig XYZ). Pre-detection involves two main steps. First, possible candidates are identified base on a non-maximal suppression (`nonMaxSupr` function of Piotr's Matlab toolbox[1]). Second, a quality score is calculated for each candidate and is used as a thresholding parameter. Pre-detection will be performed on the cell that is currently shown in the main interface.

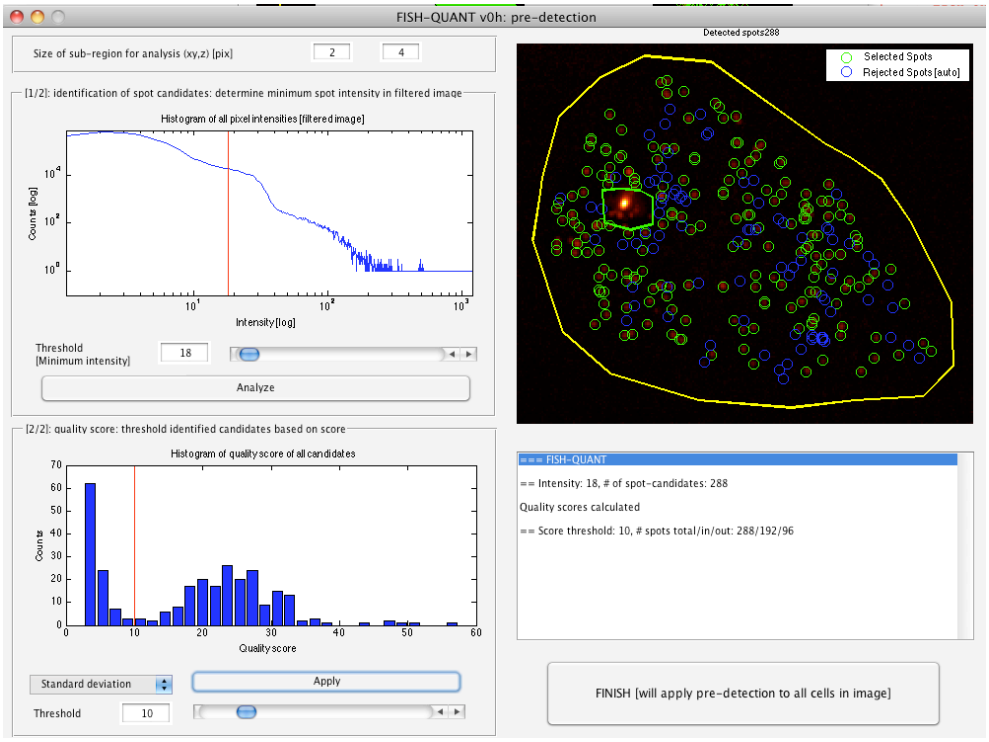


Fig ???. User-interface for the pre-detection.

The **sub-region** describes the size of a sub-region around each identified spot which will be considered in the following analysis steps. For each candidate spot a region +/- the specified number of pixel values with respect to the identified spot location will be extracted. This region is then used for the calculation of the quality scores and also for the fitting of the 3D Gaussian. By default only spots for which this sub-region can be extracted will be considered. This excludes spots that are close to the first or last focal plane. Selecting the 'Allow smaller region for detection in Z' option deactivates this setting. Now all spots independently of their z-position are considered. If spot are only partly in the image FISH-QUANT still attempts to fit them with the 3D Gaussian function.

The non-maximal suppression identifies pixel location and values of local maximums - that is a location is returned only if it has a value greater or equal to all pixels in the surrounding window. In the first panel the **minimum intensity** of such a spot can be set. This value has to be adjusted to the actual image data. The plot shows the histogram of all pixels in log-log. By moving the slider (or editing the text box) the minimum intensity can be set. The value should be set such that all spots are considered without considering too much background. To obtain a better feeling for the actual range of intensities in the filtered image open the **filtered** image in the main FISH-QUANT in a separate window (Select 'Filtered image' in the select plot panel, deselect all other options, press plot and then double click on the figure in the main user interface to open it in a separate window). Activate the data cursor from the figure toolbar and inspect the intensities of typical spots and background. When the filtering parameters are set well these regions should be well separated. Set the threshold to a value close to the background or slightly above. This value can be set not too stringent since there is a subsequent thresholding step based on a quality score. To

In the next step for each spot a **quality score** is calculated. Currently two scores are implemented

- ***Standard deviation***. Standard deviation of all pixels in the sub-region. Works robust also for dimmer spots
- ***Curvature***. 3D curvature [6]. Works best for bright spots.

When pre-detection is finished a **histogram of the quality score** is shown and the user has to specify a cut-off. In the title of the histogram the total number of spots and the currently consider number of spots are shown. The degree of thresholding depends on the quality of the data on how well the above parameters are adjusted. Good spots populate the right part of the histogram, whereas bad spots can be found on the right. If the threshold for the minimum intensity has been set optimally and the image data is not noise only one distribution can be seen (Fig A and B). Here the cut-off should be set that the entire distribution is considered.

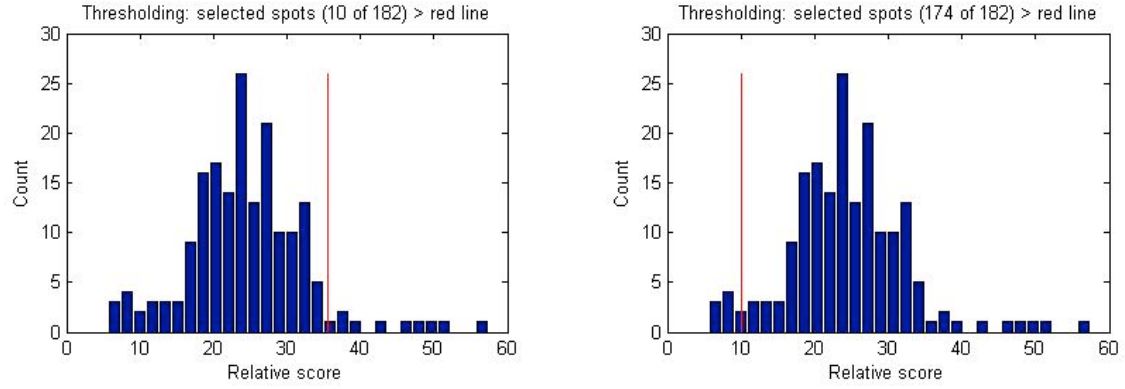


Fig A (left). Initial thresholding parameter for intensity is 30. Thresholding parameter for quality score is set to

high and too many spots are rejected. Fig B (right). As in figure A but thresholding parameter for quality score is set correctly.

Often, the histogram looks bi-modal, and then the cut-off should be chosen such that the two populations are well separated. If there is only one distribution, the pre-detection worked either very well or the initial cut-off was already too stringent.

Press 'Finish' to finish the pre-detection. The determined pre-detection settings will then be applied to all other cells in the image.

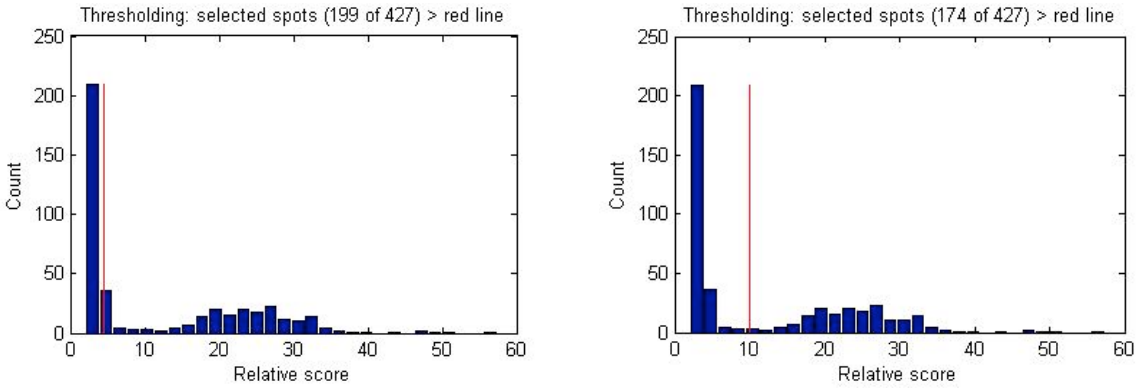


Fig C (left). Initial thresholding parameter for intensity is 15. Therefore more spots are considered. Ad in Fig A. Thresholding parameter for quality score is set to low and too many spots are considered. The spots identified in Figure A and B are on the right side of the histogram. The false-positives are on the left side (with smaller quality scores) . Fig D (right). As in figure C but thresholding parameter for quality score is set correctly.

Note: for the first step a relatively modest thresholding strategy can be applied since a second round of thresholding will be performed on the fitted spots.

Note: if multiple cells are processed the threshold will be determined for the first cell and then used for all other cells.

6. Fit spots

All pre-detected spots will be fit with a 3D Gaussian considering the same sub-region used in the pre-detection. In the first round of fitting all fitting parameters are free, i.e. for each spot the

- width of the Gaussian in XY,
- width of the Gaussian in Z,
- position in X,Y,Z,
- amplitude, and
- local background

are calculated. As soon as the fitting process is finished various plots are shown. See next section 'thresholding' and the chapter on visualization for more details. The averaged values for the estimated parameters are shown in the panel 'Experimental PSF'.

The option '**Fixed width'** allows fitting all spots with fixed values for the width of the Gaussian in XY and Z. For this fit the values shown in the experimental PSF are used. Here different sets of parameters can be chosen, see 'Thresholding' and 'Averaging of spots' for more details.

The results of the spot detection can either be inspected in the GUI directly or by using the ImageJ visualization functionality (see 'Visualization' for more detail).

Note: this step supports parallel computing. This option can be enable by checking the respective option in the lower part of the GUI.

7. Thresholding

In this panel the detected spots can be thresholded with the different fitting parameters. The parameter can be selected in the pop-up menu. This changes the histogram to this parameter. Minimum and maximum values for this parameter can be specified. These values are indicated in the histograms by a red and a green line. To activate this threshold the button 'Apply threshold has to be pressed'. The histogram on the right shows always the entire histogram of the selected parameter, the histogram on the left after thresholding.

The option '**Lock values'** allows saving these thresholding parameters. Then the next parameter can be inspected. The option 'Unlock all' removes all applied thresholds. The left histogram reflects locked parameters, i.e. if one parameter is locked the left histogram of another parameter will already show the thresholded parameters. This option is very useful to check how thresholding of one parameter affects other parameters. Thresholded parameters will be applied for all analyzed cells.

The results of the spot detection can either be inspected in the GUI directly or by using the ImageJ visualization functionality (see ''Visualization' for more detail).

In the panel 'Averaged spots' the averaged values for each of the estimated parameters is shown. The pull-down menu allows selecting either the average of all spots, all thresholded spots, or (if calculated) the results of fitting the averaged spot. For each new selection in the pull-down menu the results will also be shown in the workspace together with the standard deviation. This is useful if fitting results from different images are compared.

8. Inspection of detected spots

FISH_QUANT allows a detailed inspection of the detected and thresholded spots in a separate user-interface (see Fig ???). **Image data** can either be shown in a maximum intensity projection or as a z-stack. Further either the raw data or the filtered data can be shown. The **detected spots** can be inspected individually with 'Data cursor'. When selecting one spots the sub-region of this spots as used in the pre-detection and fitting steps of FISH-QUANT will be shown in the windows on the right side of the interface, further the corresponding fitting parameters will be shown on the left side. This allows a better adjustment of the thresholding parameters. The tool '**Region inspector'** opens the matlab region inspector which allows a more detailed inspection of the image data. **Zoom In/Out** allows zooming the image. This can either be done with the mouse wheel or by pressing the left mouse button to zoom-in and shift+mouse to zoom out. **Pan** activates the pan mode that allows moving view of a (zoomed) graph up and down as well as left and right.

This tool can be accessed in two ways. First, by pressing the 'Visualize' button in the 'Visualize results' panel. Here the currently processed data-set in FISH-QUANT will be automatically loaded into the spot-inspector tool. The tool can be closed by the 'Close Window' button in the upper left corner. Alternatively the tool can be called from the Tools Menu. Here already saved data with results of spot detection can be loaded and saved from the 'File' menu. When called with this option it is possible to manually remove spots from the data-set.

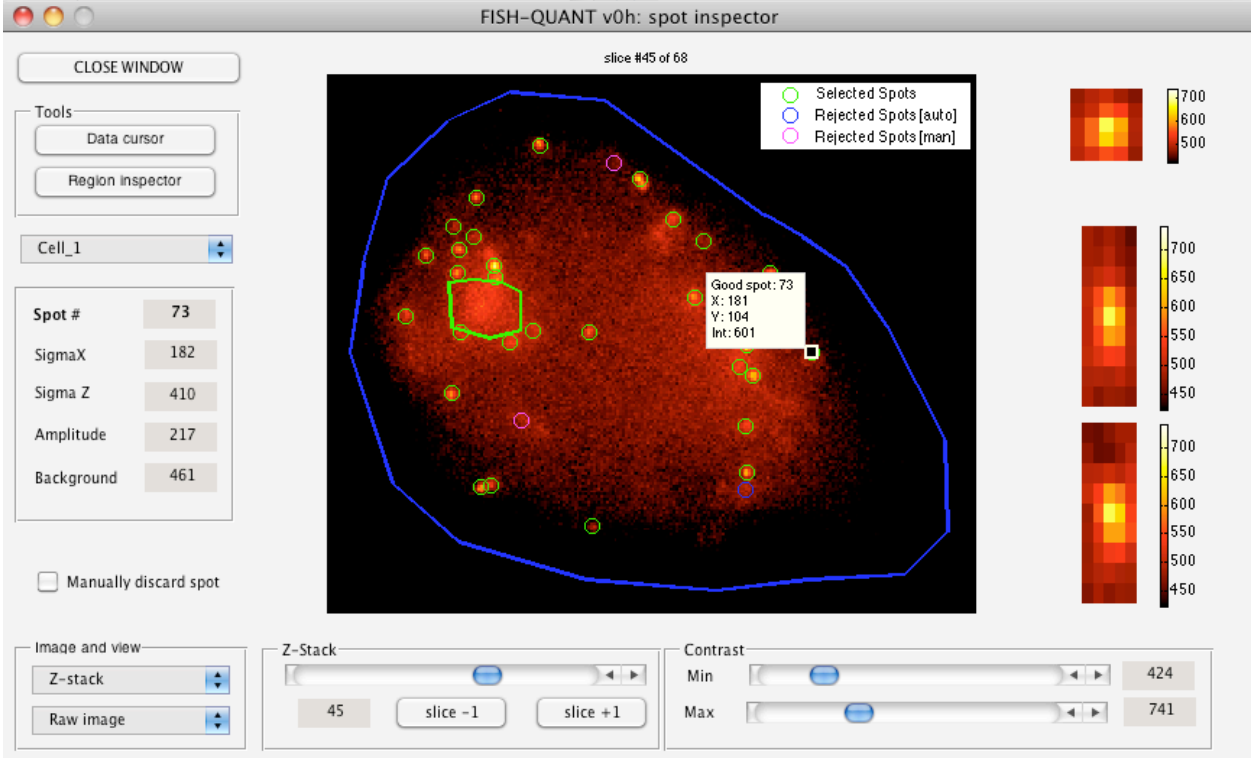


Fig C (left). User interface for the inspection of the detected and thresholded spots.

9. Averaging and fitting spots with fixed size

This step should only be performed if it is known that the individual spots are diffraction limited in size. It is then possible to average all spots together and fit the resulting spot with the 3D Gaussian (for more details see section 'Averaging of spots. This gives a very accurate estimate of the size of an individual spots, i.e. the width of the Gaussian in XY and Z. This value can then be fixed in the fitting process by checking the corresponding option in the 'Fit spots' panel and all spots will be fit again with the width of the Gaussian being fixed to this estimated value. The values displayed in the panel 'Averaged spots' will be used.

Note: alternatively the averaged values of all fits can be used. For this purpose the corresponding option in the panel 'Experimental PSF' can be selected.

10. Quantification of transcription site

Transcription sites can be quantified with the functions available in the *TxSite* Menu. The reconstruction method requires several additional parameters, which have to be defined first (see Appendix 8 for a description of the reconstruction method).

Point Spread Function

The PSF can be defined either by a 3D-image or by a parametric PSF model. When using **an image of the PSF**, two 3D stacks have to be specified: first, a stack describing the actual PSF (PSF_beads_3D_ns in the example data). Second, a stack describing the background of the PSF measurement (PSF_beads_3D_bgd_ns in the example data). These stacks are usually obtained from measuring and

averaging the images of fluorescent beads. Details can be found in Appendix 7 - 'Measuring the PSF'. When defining a parametric PSF the following parameters have to be specified: ….

Amplitudes of mRNA.

The amplitudes of individual mRNA's show a certain distribution. This distribution can either be obtained from the currently analyzed image or from a saved image with the results of a spot detection.

Other options

For the quantification a number of default settings are defined which can be reviewed and changed here. See Appendix 8 for details.

After defining of these parameters, the quantification can be started. The transcription sites in the currently selected cell will be analyzed. When the quantification is finished the results of the quantification can be inspected in ImageJ. Two different plots are shown. The first image shows the z-stack with the transcription site on the left and the best reconstruction on the right. The second image shows the residuals of this reconstruction. Positive residuals are shown in red, negative residuals are shown in green.

Note: a useful function in ImageJ is reslice (*Image* > *Stacks* > *Reslice [\]* …). This allows flipping the stacks, i.e. rather than having different xy slices, xz slices are shown.

Analysis in batch mode

The steps above allow optimizing the detection, fitting and thresholding steps involved in the mRNA quantification. These settings can then be saved in a text file (see Appendix 3 for details). FISH-QUANT then provides the possibility to use these settings in a batch-processing mode to analyze a large number of pictures automatically.

Usually the **outline of the cells and transcriptions sites** for each picture is defined with the outline tool. This tool can called be from the *Tools menu* in FISH-QUANT. See 'Outline of cell and transcription site' for more details. If called from the menu, images can be loaded with the 'Load Image' button and the outline for cells and transcription sites can be saved for each loaded file. This step can be omitted if no outline definition is necessary or the same outline will be used for all images (see below).

A large number of files can be processed with the **batch-processing tool.** The respective user-interface can be called from the menu Tools in FISH-QUANT. In this interface the various parameters of the batch-processing mode can be specified. First, the file describing the detection settings has to be loaded. This file is usually obtained by optimizing the settings for a small number of images with the main interface of FISH-QUANT. Second, the files that will be processed have to be specified. Here either the respective outline definition files or the image file can be added. If the image files are added the analysis is usually performed on the entire images. Alternatively a single outline can be specified that will be applied to all images (*Menu* > *Advanced* > *Same outline for all images*). Different pre-detection thresholds can be applied to the different images. The respective thresholds for the detection (minimum intensity of the spot and minimum quality score) have to be specified in a text file (see Appendix 7). This file can be loaded into FISH-QUANT from the menu *Settings* > *Detection* > *Load threshold for each image*. The detection settings will be used for all specified files; the default settings will be used for all other files. The batch-processing can be started by pressing on PROCESS.

When finished the data will be **thresholded** based on the parameters defined in the settings file. In the status window the number of all detected spots before and after thresholding are shown. Then the spots can be additionally thresholded. This is done exactly as described above for an individual cells but now for all detected spots. Only LOCKED thresholding parameters will be considered.

Detection results can be either **saved** for each processed image separately or in one large file that summarizes all detected spots. Further it is possible to save all detected spots or only the spots that remain after thresholding. The user can also choose from two different ways to label the rows in the summary file: either with the name of the file and the corresponding cell or with a short file identifier extracted from the file-name. By default the last four characters of the file-names are used (can be changed in the Advanced menu). Summary files for counting of mature and nascent mRNA in all cells in all files can be saved.

All files will be processed with the same settings!

- It is possible to **use already filtered images** in the batch processing mode. This avoids re-filtering images and saves some time. It is also possible to save filtered images which are generated during the batch processing mode and make them available for later use.

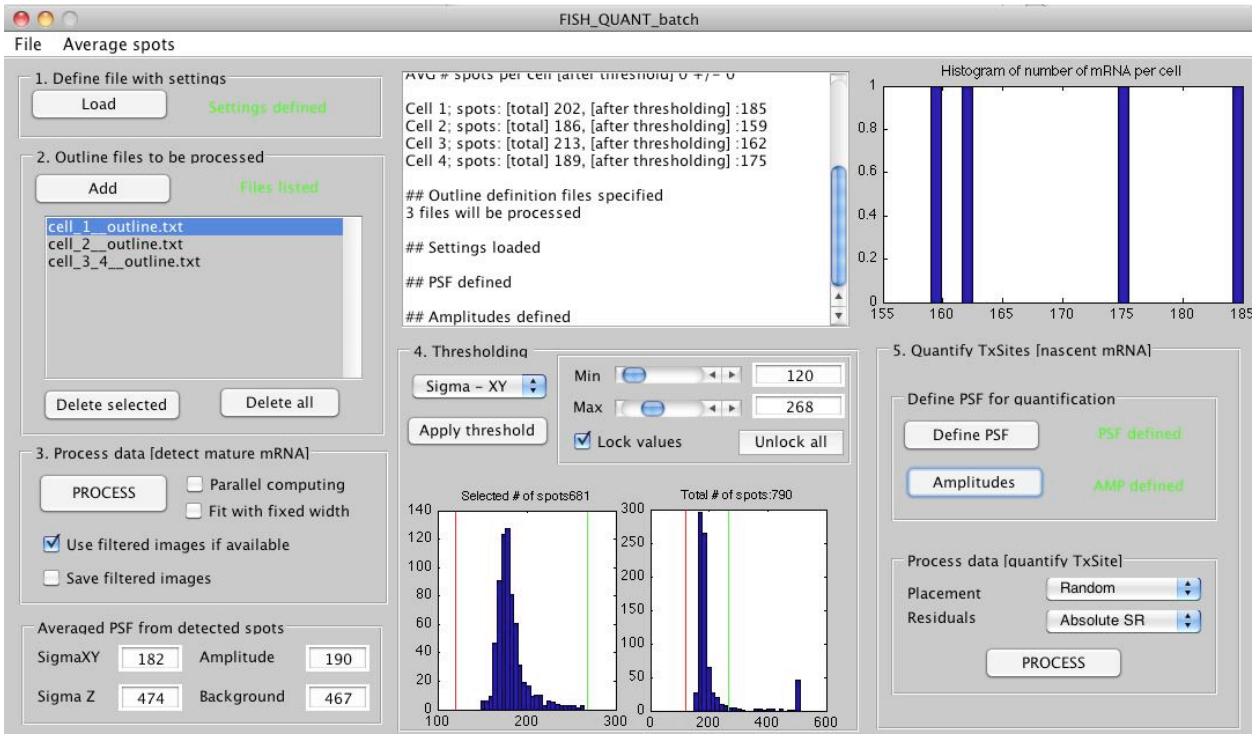
- It is possible to **re-fit the data** with fixed values for the size of the PSF, i.e. the sigma in XY and Z. The corresponding option can be check-marked. The values shown in the experimental PSF will be used for the fit.

- Finally it is possible to save the **results of the detection**. Only LOCKED thresholding parameters will be considered. The file-names will start with the name of the image followed by the suffix that can be specified in the interface. For each image a results file as described in Appendix 2 will be saved. The experimental parameters (pixel-size, wavelength, ..) from the detection setting file will be used rather than the ones save in the individual outline definition files. Additionally a summery file will be generated. Additionally, it is possible to save a summary of results of the spot detection as well. This file contains the number of transcripts before and after thresholding for each processed cell in a compact format (see Appendix 4). Additionally the detection-settings are saved as well (including all locked thresholding parameters).

- In the *Load/Save menu* the results of previous detections can be loaded and re-analyzed.

- The Average spots menu allows averaging the detected spots after thresholding. This can be used to average the images of beads to obtain a PSF with good signal-to-noise ratio.

It is also possible to quantify all transcription sites specified in the outline-definition files (for details see above ‘Quantification of transcription site’). First, the PSF has to be specified. This either can be a PSF which is saved as an image or a model PSF (not yet implemented). Next the distribution of estimated amplitudes has to be specified. This can be either from the current selections in interface or from a stored file of the spot detection. Then simply press process for the quantification. The results of this detection can then be saved in a summary file.


FISH_QUANT_batch
File
Average spots
1. Define file with settings
Load
Settings defined
2. Outline files to be processed
Add
Files listed
cell_1__outline.txt
cell_2__outline.txt
cell_3_4__outline.txt
Delete selected
Delete all
3. Process data [detect mature mRNA]
PROCESS
Parallel computing
Fit with fixed width
Use filtered images if available
Save filtered images
Averaged PSF from detected spots
SigmaXY
182
Amplitude
190
Sigma Z
474
Background
467
Cell 1; spots: [total] 202, [after thresholding] :185
Cell 2; spots: [total] 186, [after thresholding] :159
Cell 3; spots: [total] 213, [after thresholding] :162
Cell 4; spots: [total] 189, [after thresholding] :175
Outline definition files specified
3 files will be processed
Settings loaded
PSF defined
Amplitudes defined
Histogram of number of mRNA per cell
4. Thresholding
Sigma - XY
Apply threshold
Min
120
Max
268
Lock values
Unlock all
Selected # of spots681
Total # of spots:790
5. Quantify TxSites [nascent mRNA]
Define PSF for quantification
Define PSF
PSF defined
Amplitudes
AMP defined
Process data [quantify TxSite]
Placement
Random
Residuals
Absolute SR
PROCESS

Visualization

Double click on all plots will open them in a separate figure.

Histogram of estimated parameters

Two histograms are in the thresholding panel. They show the histogram of the selected fitting parameter. The right plot shows the histogram of all parameters, the left after thresholding.

Location of detected spots

In the main plot a maximum projection in XY of the image is displayed. After the spot detection the identified spots are shown. After thresholding the thresholded spots among the detected spots are shown. Next to the plot window is a panel which allows selecting several plot options

- **Image**: raw image or filtered image
- **Spots**: all detected spots, detected spots and thresholded spots, no spots
- **Outline of cell and transcriptions site**: on / off

Quality of detected spots

There are three smaller plots situated under the main window. The first two panels show a maximum projection of all detected spots in XY and XZ is shown. The last panel shows the maximum projection of the residuals in XY.

Visualize results

This opens a separate user-interface that allows a detailed inspection of the detected and thresholded spots. When called from the tools Menu spots can be manually thresholded and removed.

Saving and loading data

FISH QUANT allows loading and saving of different parameters. Can be accessed by the Load/Save menu.

Filtered image

The filtered image can be saved as a stacked TIF file. The file-name of this image can be stored in the outline definition file. This is useful if the same data-sets are analyzed multiple times since the filtering step can be omitted and the already filtered image can be used. It is, however, necessary to FIRST save the filtered image and then save the outline definition file. Alternatively the file-name of the filtered image can manually be added (see Appendix 1).

Outline of cell

The outline of the cell and the transcription site(s) can be stored in a simple text file (see Appendix 1). This files does, however, not only define the outlines but contains the file-name and information about the experimental details as well.

Results of spot detection

This options saves two files and the user is asked to specify two file-names

1. **Results of the actual spots detection** in a simple text file. This file contains all the estimated parameters for the detected spots of one cell. In addition it repeats the information contained in the outline-definition file (see Appendix 2).

2. The detection settings are saved as well (see next section).

Detection settings

In this text-file all relevant parameters of the detection process are stored (see Appendix 3).

Results of batch processing

This text file contains a brief summary of results of the batch-mode detection (see Appendix 4). For each cell the total number of transcripts before and after thresholding, and the number of transcripts at the transcription-site, are listed.

Other Features

There are a number of other features implemented which simply the workflow.

Double click opens in new window

Double click on any of the plots in the GUI will open them in a separate figure.

Close all windows

The button ‘Close all windows’ will close all windows except the different GUI’s of FISH-QUANT.

Reset GUI

Brings GUI back to default settings and deletes all data.

Parallel computing

This option enables parallel computing for the spot detection and spot fitting. The Parallel Computing Toolbox of Matlab is required.

Advanced features

Averaging of spots

FISH-QUANT allows averaging the identified spots. This is useful to obtain a PSF with improved signal-to-noise ratio. Some of the functionality is implemented in the GUI itself. More functions are implemented in the menu.

Visualization

The results of the various steps involved in the averaging process can be visualized in ImageJ. This plot functions can be found in the menu 'averaged spot'

Averaging

The user can average the images of the averaged spots. The respective functions can be found in the menu 'Averaged spots'. The first step is averaging all spots together. When selecting this function a user-dialog opens where several options for this averaging process can be specified.

- The **size of the resulting image.** This is the area around each spot that will be considered in the averaging process. If for one spot an area of this size can not be extracted, the spot will be ignored for the averaging process.
- During the averaging process the spots are aligned based on the estimated center of the PSF. Since the center is known with sub-pixel accuracy, the alignment can be done with sub-pixel accuracy as well. This is implemented in FISH-QUANT as **oversampling (OS)** by sub-dividing the pixels in smaller sub-pixels. The corresponding factors are specified in the fields 'Factor OS'. The default settings are 3 for xy and z. This means that each pixel is sub-divided in 3*3*3 = 27 sub-pixels. Each sub-pixel has the same intensity than the larger pixel but now spots can be aligned with higher precision.
- When starting the averaging process the user is asked if the **local background** of each spots should be subtracted before the averaging process. The local background for each spots is obtained from the fit with the 3D Gaussian. The results are shown in a separate Matlab figure.

Fitting

The resulting spot can be fit with the 3D Gaussian function. The size of the fitting region can be changed in the panel 'Predetection of spots'. The estimated numbers are shown in the panel 'Experimental PSF'. The width in xy and z can now also be used as fixed parameters to fit the individual mRNAs again.

Radial averaging

To reduce the noise levels the resulting image can further be radial averaged (Menu > Averaged spot > Radial average). The function determines the center of mass in XY and calculates for each z-plane the radial intensity distribution with respect to this center. This is done on the over-sampled data.

Constructing an image from radial averaged data

The radially averaged data can then be used to construct and xyz image. The constructed image will be on normal sampling size, e.g. with the original pixel-size. Since the construction is based on the over-sampled data the constructed image can have off-set positions of the center with respect to xy. When

using this function (Menu > Averaged spot > Construct from radial average) the user will be asked for this offset in pixels.

Fit of constructed image

The constructed image can be fit with the 3D Gaussian function (Menu > Averaged spot > Fit constructed image). The size of the fitting region can be changed in the panel ‘Predetection of spots’.

Problems

Filtering takes a very long time

This filtering is performed on the entire image independently of the defined cells. This step can take a long time for large stacks. Cropping the images to contain only a single cells leads to a faster filtering. Filtered images can also be saved and then an updated outline-file can be saved that will reference to this save file so the filtering step doesn't have to be performed again.

Visualization in ImageJ doesn't work

This can either results either in an error message or Matlab will crash. This can usually be avoided by increasing the Java Heap Memory in Matlab. See section 'MIJ – Interface between Matlab and Imagej' in chapter about installation for more details.

Appendix

Appendix 1 – definition of files defining outlines

These files are defined for each recorded stack. They define the parameters of the experiment and for each cell in the field of view the outline of this cell and the location of the transcription site(s). **They have to be stored in the same directory than the image**. An example for such a file can be found below (grey box).

ROW	FILE
1	FISH-QUANT v0e
2	OUTLINE DEFINITION, 11-Mar-2011
3	COMMENT Outline definition performed in FISH-QUANT (Main program)
4	FILE FISH_test_data.tif
5	FILTERED FISH_test_data_filtered.tif
6	PARAMETERS
7	Pix-XY Pix-Z RI Ex Em NA Type
8	160 300 1.458 568 568 1.25 widefield
9	ANALYSIS-SETTINGS FISH_test_data__settings.txt
10	CELL Cell1
11	X_POS 60 5 40 238 230 100 END
12	Y_POS 7 90 197 195 104 11 END
13	TxSite TxSite_1
14	X_POS 43 45 75 72 END
	Y_POS 73 103 101 72 END

- **Line 1-2** Header row, specifying version number, type of file and date
- **Line 3** is reserved for a user-comment and starts with the key word **COMMENT**
- **Line 4** is reserved to save the file-name and starts with they key word **FILE**
- **Line 5** is reserved to save the file-name of the filtered image and starts with they key word **FILTERED**
- **Line 6** is the header to indicate that the experimental parameters will come next
- **Line 7** contains a description of the parameters in line 8.

Pix-XY	Pixel-size in XY in nanometer
Pix-Z	Pixel-size in Z in nanometer
RI	Refractive index of medium
Ex	Excitation wavelength of fluorophore
Em	Emission wavelength of fluorophore
Type	Type of microscope. Currently only widefield is supported.

- **Line 8** specifies the experimental parameters.
- **Line 9** contains the file-name of the corresponding analysis settings (if they were saved).
- **Line 10** specifies the first cell. Starts with keyword **CELL**, followed by a unique identifier of this cell.
- **Line 11** specifies the x-coordinates of the polygon describing the outline of the cell. Line is terminated with key-word **end**
- **Line 12** specifies the y-coordinates of the polygon describing the outline of the cell. Line is terminated with key-word **end**

- **Line 13** specifies a transcription site. Starts with keyword **TxSite**, followed by an identifier.
- **Line 14** specifies the x-coordinates of the polygon describing the outline of the cell. Line is terminated with key-word **end**
- **Line 15** specifies the y-coordinates of the polygon describing the outline of the cell. Line is terminated with key-word **end**

Notes

- It is not necessary to defined transcription-sites. This block can be omitted.
- Additional transcription-sites can be defined by repeating block from line 14-15 and changing the identifier in line 13.
- Multiple cells can be defined by repeating the block from line 9 to line 14 with unique cell label.
- All values are separated by tabs.
- No empty rows in between definitions.
- Files are saved as .txt files.

Appendix 2 – definition of files summarizing results of spot detection

These files are defined for each cell in the recorded stack. Relevant information from outline definition files is repeated (experimental parameters, outline of cell and transcription site. An example for such a file can be found below (grey box).

ROW	FILE
1	FISH-QUANT v0e
2	RESULTS OF SPOT DETECTION PERFORMED ON 11-Mar-2011
3	COMMENT User specified comment
4	FILE FISH_test_data.tif
5	FILTERED cell_1_filtered.tif
6	PARAMETERS
7	Pix-XY Pix-Z RI Ex Em NA Type
8	160 300 1.458 568 568 1.25 widefield
9	ANALYSIS-SETTINGS FISH_test_data__settings.txt
10	CELL Cell1
11	X_POS 60 5 40 238 230 100 END
12	Y_POS 7 90 197 195 104 11 END
13	TxSite TxSite_1
14	X_POS 43 45 75 72 END
15	Y_POS 73 103 101 72 END
16	SPOTS
17	Pos_Y Pos_X Pos_Z AMP BGD RES SigmaX SigmaY SigmaZ Cent_Y Cent_X Cent_Z MuY MuX MuZ ITERY_det Y_det X_det Z_det Y_min Y_max X_min X_max Z_min Z_max SC_det SC_det_norm TH_det TH_fit
18	1017.58 242.307 824.836 3455.96 444.688 1.82951e+07 120 120 320 180.819 175.36 406.952 177.583 182.307 444.836 8 17 190 2 14 20 1 193 1 3 268.56 0.280481 1 1
19	1019.35 246.091 346.884 3649.55 516.118 1.16962e+07 120 120 320 182.086 177.036 153.38 179.347 186.091 -33.1162 9 17 190 1 14 20 1 193 1 2 350.316 0.293912 1 1

- **Line 1-3** Header row as above, second row specifies that the results of spot detection are listed.
- **Line 4-15** repeat information about file-name, experimental parameters, cell position and transcription site location as specified above.
- **Line 16** Key word SPOTS indicates that results of spot detection start

- **Line 17** Header for spot detection

Pos_Y	Estimated position of center in image (Y, in nanometer)
Pos_X	Estimated position of center in image (X, in nanometer)
Pos_Z	Estimated position of center in image (Z, in nanometer)
AMP	Estimated amplitude
BGD	Estimated background
RES	Squared sum of residuals
SigmaX	Estimated width = sigma of Gaussian (X, in nanometer)
SigmaY	Estimated width = sigma of Gaussian (Y, in nanometer)
SigmaZ	Estimated width = sigma of Gaussian (Z, in nanometer)
Cent_Y	Centroid of sub-image (Y, in nanometer)
Cent_X	Centroid of sub-image (X, in nanometer)
Cent_Z	Centroid of sub-image (Z, in nanometer)
MuY	Position in sub-image (Y, in nanometer)
MuX	Position in sub-image (Y, in nanometer)
MuZ	Position in sub-image (Y, in nanometer)
ITER	Iterations used for fit
Y_det	Pre-detected position in image (Y, in pixel)
X_det	Pre-detected position in image (X, in pixel)
Z_det	Pre-detected position in image (Z, in pixel)
SC_det	Pre-detection score
SC_det_norm	Normalized pre-detection score (to maximum)
Y_min	Start of sub-region for detection in image (Y, in pixel)
Y_max	End of sub-region for detection in image (Y, in pixel)
X_min	Start of sub-region for detection in image (X, in pixel)
X_max	End of sub-region for detection in image (X, in pixel)
Z_min	Start of sub-region for detection in image (Z, in pixel)
Z_max	End of sub-region for detection in image (Z, in pixel)
TH_det	Thresholded after detection (1 = good, 0 = bad)
TH_fit	Thresholded after detection (1 = good, 0 = bad, -1 manually removed)

- **Line 18- end** Results

Appendix 3 – definition of files summarizing detection settings

These files define all parameters describing the processing and detection process. They can be read-in by the batch processing tool. Settings are stored in a simple format. Each property starts with the name followed by a '='and the actual value. There is NO space in between the equal sign and the identifier and the actual value. The row will be ignored if the identifier is not known or if there is no equal sign. So additional rows with comments can be introduced.

Below is such file (grey box). On the left side is an explanation of what the key-words (bold) mean.

Explanation	**FILE**
	FISH_QUANT v0e ANALYSIS SETTINGS, 10-Mar-2011 # EXPERIMENTAL PARAMETERS
Emission wave-length	**lambda_EM**=568
Excitation wave-length	**lambda_Ex**=568
Numerical aperture	**NA**=1.25

Refractive index	**RI**=1.458
Type of microscope	**Microscope**=widefield
Pixel-size [XY] in nm	**Pixel_XY**=160
Pixel-size [Z] in nm	**Pixel_Z**=300
	# GENERAL PROPERTIES
Flag: parallel computing	**flag_parallel**=0
# Setting for filtering	# FILTERING
Factor for BGD-Kernel	**Kernel_bgd**=5
Factor for PSF-Kernel	**Kernel_psf**=0.75
# Pre-detection	# PRE-DETECTION
Size of region around each detected region in pixel	**Detect_Region_XY**=2
	Detect_Region_Z=4
Threshold for intensity	**Detect_Thresh_int**=30
Threshold for score	**Detect_Thresh_score**=0
Type of quality score	**Detect_Score**=Standard deviation
	# AVERAGING
Size of region used for averaging	**AVG_Region_XY**=13
	AVG_Region_Z=13
Factor for oversampling, i.e. division in how many sub-pixels	**AVG_OS_XY**=3
	AVG_OS_Z=3
	# THRESHOLDING OF DETECTED SPOTS
Parameters for thresholding of fitted spots *	**SPOTS_TH_sigmaXY_min**=148
	SPOTS_TH_sigmaXY_max=279

Thresholding can be done for each parameter. In order to activate thresholding the allowed minimum and maximum values for this parameter have to be specified. The always have the format **SPOTS_TH_PAR_min** and **SPOTS_TH_PAR_max** where PAR stands for the respective parameter. Supported are **sigmaXY** (Width of the Gaussian in xy), **sigmaZ** (Width of the Gaussian in Z), **amp** (Amplitude of the Gaussian), **bgd** (Local background of the Gaussian), **score** (Quality score of the pre-detected spot), **iter** (Number of iterations in the fit), **resNorm** (Residuals of the fit). Paramters which are not listed will not be thresholded.

Appendix 4 – summary of batch detection: mature mRNA

This file summarizes the results of batch detection for mature mRNA. For each analyzed cells the total number of detected transcripts in the cell are listed before and after thresholding.

Below is such file (grey box). On the left side is an explanation of what the key-words (bold) mean.

Row	**FILE**
1	FISH-QUANT v0f
2	RESULTS OF SPOT DETECTION PERFORMED IN BATCH MODE ON 16-Mar-2011
3	COMMENT Batch detection
4	ANALYSIS-SETTINGS Detection_settings.txt
5	FILE CELL N_total N_thres
6	cell_1.tif Cell_1 189 156
7	cell_2.tif Cell_2 179 129
8	cell_3_4.tif Cell_1 193 129
9	cell_3_4.tif Cell_2 177 148

- **Line 1-2** Header row, specifying version number, type of file and date
- **Line 3** is reserved for a user-comment and starts with the key word **COMMENT**
- **Line 4** contains the file-name of the corresponding analysis settings (if it was saved)
- **Line 5** contains a description of the parameters which are listed afterwards.

File	File-name of image
CELL	Identifier of the cells in the image
N_total	Total number of spots in cell before thresholding
N_thresh	Total number of spots in cell after thresholding

- **Line 6-end** contains the results of the analysis.

Appendix 5 – file with start-up parameters for FISH-QUANT

This file is called FISH-QUANT_def.txt and is located in the same folder than the main Matlab files of FISH-QUANT. It contains important parameters to initialize FISH-QUANT. Each parameter is specified with the corresponding key word followed by an equal sign (=) and the parameter value. NO spaces are allowed. They keyword should not be changed and NO lines should be removed.

ROW	FILE
1	## FISH-QUANT
2	path_imagej=/Users/fmueller/ImageJ
3	version=v0f
4	ij_macro_name=FQ_IJ_vis_v2.txt

- **Line 1** Header row, specifying version number, type of file and date
- **Line 2** Installation path of ImageJ (key word `path_imagej`)
- **Line 3** Version of FISH-QUANT (key word `version`)
- **Line 3** Name of Macro for ImageJ visualization (key word `ij_macro_name`)

Appendix 6 – summary of batch detection: nascent mRNA

This file summarizes the batch detection for nascent mRNA. For each analyzed transcription site the estimated number of transcript and the standard deviation is reported. As comparison the number of transcripts as obtained with the traditional method of quantification is given as well.

Below is such file (grey box). On the left side is an explanation of what the key-words (bold) mean.

Row	FILE
1	FISH-QUANT v0g
2	RESULTS TxSite quantification performed ON 04-May-2011
3	COMMENT Batch detection
4	ANALYSIS-SETTINGS FISH-QUANT__batch_settings_2011-05-04.txt
5	FILE CELL TS N_mean N_std N_trad
6	cell_1.tif Cell_1 TS_1 12.54 1.26507 9.06702
7	cell_2.tif Cell_2 TS_1 18.56 1.21487 12.6082
8	cell_3_4.tif Cell_1 TS_1 35.34 1.89101 17.6967
9	cell_3_4.tif Cell_1 TS_2 30.68 1.58359 18.0982
10	cell_3_4.tif Cell_2 TS_1 12.1 1.0351 9.08074

- **Line 1-2** Header row, specifying version number, type of file and date
- **Line 3** is reserved for a user-comment and starts with the key word **COMMENT**
- **Line 4** contains the file-name of the corresponding analysis settings (if it was saved)
- **Line 5** contains a description of the parameters which are listed afterwards.

File	File-name of image
CELL	Identifier of the cells in the image
TS	Identifier of the transcription sites in the cells
N_mean	Number of nascent mRNA at the TxSite [mean]
N_std	Standard deviation of the number of nascent mRNA at the TxSite
N_trad	Number of nascent mRNA at the TxSite with traditional method

- **Line 6-end** contains the results of the analysis.

Appendix 7 – specifying different detection thresholds in batch mode

This file can be used when different detection thresholds are applied to different files in the batch detection. Each file is specified with the file-name of the image (INCLUDING the extension), followed by thethreshold for the minimum intensity, and the minimum quality score (ORDER IS IMPORTANT, values are tab-delimited). The name has to match the name that you specify in the batch mode (either the outline definition file or the actual image file). FISH-QUANT will use the default settings if no exact match for the name can be found!

ROW	FILE
1	Name TH_predect TH_score
2	Img1.tif 490 50
3	Img2.tif 500 45

- **Line 1** Header row, specifying rows
- **Line 2-3** Detection thresholds for the different

Installation Appendix 8 – measuring the PSF

A precise knowledge of the PSF is necessary for an accurate quantification of the transcription site. We recommend measuring the PSF by imaging fluorescent beads. The resulting images can be analyzed in FISH-QUANT and images of multiple beads can be averaged together to increase the signal-to-noise ratio.

Image of PSF

If only the beads from one image will be averaged the main interface of FISH-QUANT can be used. If the results of multiple images will be averaged the batch mode of FISH-QUANT can be used. The processing steps are the same. First, the experimental parameters are defined. Second, the regions in the image what will be analyzed are specified. Depending on the density and quality of the image the entire image can be analyzed or only certain regions. For the latter the outline definition of FISH-QUANT can be used and one cell encompassing the relevant region of the image can be defined. Then the pre-detection and fitting is performed as described for mRNA. The spots can then be thresholded to select only the good ones. Then the thresholded spots can be averaged and the resulting imaged displayed in ImageJ and saved.

Image of background

We found that the best background correction for the PSF measurement can be obtained by collecting a z-stack for the background obtained under similar conditions than the PSF measurement.

The best way to achieve this is by collecting a second z-stack in a region where no beads are present. Then the identified locations of the beads from above can be used to calculate and averaged image of the background. If no separate image with only background was collected one option in the dialog for the averaging of the images can be used. Here an offset for the averaging can be specified. This means that not the identified positions for the PSF will be used but this offset (in pixels) will be added to the identified center. This works satisfyingly for images with only a few beads but is problematic if many beads are present in the image.

Appendix 9 – Quantification of transcription sites

- Description of the reconstruction method.
- Description of different parameters.

References

1. Piotr's Matlab Toolbox. at <http://vision.ucsd.edu/~pdollar/toolbox/doc/>

2. Sage, D., Prodanov, D. & Ortiz, C. MIJ - A Java package for bi-directional communication and data exchange from Matlab to ImageJ. at <http://bigwww.epfl.ch/sage/soft/mij/>

3. ImageJ. at <http://rsb.info.nih.gov/ij/>

4. Bio-Formats | LOCI. at <http://www.loci.wisc.edu/software/bio-formats>

5. Zhang, B., Zerubia, J. & Olivo-Marin, J. Gaussian approximations of fluorescence microscope point-spread function models. *Appl. Opt.* **46**, 1819-1829 (2007).

6. Thomann, D., Rines, D.R., Sorger, P.K. & Danuser, G. Automatic fluorescent tag detection in 3D with super-resolution: application to the analysis of chromosome movement. *J Microsc* **208**, 49-64 (2002).