SquasshAnalyst manual

Aurélien Rizk

SquasshAnalyst graphical user interface is designed for intuitive use. For the shortest introduction to Squassh and SquasshAnalyst follow the installation instructions section 2 and the toy example analysis section 4.1. You will then be able to analyze your own data or further look into the provided datasets showcasing SquasshAnalyst features. The reference section lists and describes SquasshAnalyst functionalities.

1 Squassh and SquasshAnalyst software files

Squassh and SquasshAnalyst are available as additional file with SquasshAnalyst paper and for download at:
`http://www.psi.ch/lbr/SquasshAnalystEN/SquasshAnalyst.zip`
The .zip archive contains the following files :

- `Lif_extractor.ijm` is an ImageJ macro enabling extraction of '.lif' files form Leica's microscopes to a series of '.tiff' images.
- `Squassh_.jar` contains the Squassh plugin [1] for segmenting subcellular structures.
- `SquasshAnalyst.zip` contains the source code of SquasshAnalyst.

2 Installation and launch

2.1 Squassh

Install Squassh in ImageJ by drag and dropping the `Squassh_.jar` file on ImageJ main window. For more information on Squassh and its use refer to [1]. ImageJ can be downloaded from `http://imagej.nih.gov/ij/`.

2.2 Lif extractor

The Lif extractor is an ImageJ macro working in combination with Bio-Formats an ImageJ plugin allowing to open a large number of life sciences image file formats. Bio-Formats is required for the Lif extractor to function and can be obtained from `http://downloads.openmicroscopy.org/bio-formats/5.1.1/`.

Use 'Plugins→Macros→Run' in ImageJ menus and select the `Lif_extractor.ijm` file for a single use of the macro.

Alternatively, for a permanent install of the Lif extractor macro, copy the content of the `Lif_extractor.ijm` file to the end of 'StartupMacros.txt' contained in ImageJ's macros folder. This can be done by opening the `Lif_extractor.ijm` file in a text editor and copy and pasting its content. ImageJ's macro folder is located in ImageJ's installation folder, next to ImageJ's application binary. The macro is then available in Plugins→Macros ImageJ menu.

2.3 SquasshAnalyst

SquasshAnalyst is an R program running as an interactive web application using the web application framework for R Shiny [2]. It works with any modern browser and requires the free statistical software R to be installed on your system. Follow instructions at `http://cran.r-project.org/` to install the latest R version.

Shiny installation
Before first launch of SquasshAnalyst it is necessary to install Shiny in R with the following R command:

```
install.packages("shiny")
```

SquasshAnalyst launch
SquasshAnalyst is launched from R with:

```
shiny::runGitHub("SquasshAnalyst", "a-rizk")
```

Note that this requires an internet connection.

Alternative launch method
It is possible to launch SquasshAnalyst with no internet connection by decompressing 'SquasshAnalyst.zip', changing R working directory to the parent folder and running SquasshAnalyst with:

```
shiny::runApp("SquasshAnalyst")
```

3 Example data

We provide example datasets online on a cloud file sharing service.

- *Toy example.* This small toy example dataset allows to test the whole analysis workflow from the .lif extraction, Squassh segmentation to SquasshAnalyst analysis in less than fifteen minutes. The set provides two conditions with five 2D one channel fluorescence images in each condition. The images show COS cells 30 minutes after EGF stimulation and before EGF stimulation. The data is given as .lif files as would be obtained when exported from a Leica microscope.

- *EGF RABx colocalization.* This set investigates the colocalization of EGF with RAB5a and RAB11b in COS cells before and 30 minutes after EGF stimulations. Each condition contains five 2D images with three fluorescence channels.

- *RABx localization.* This set investigates the colocalization of four RAB GTPases with four subcellular markers in HEK cells. It contains 3D images with two fluorescence channels to examine the colocalization of RAB4a, RAB5a, RAB7a and RABB11b with subcellular markers for early endosomes (EEA1), late endosomes (LBPA), endoplasmic reticulum (PDI) and lysosomes (LBPA). The first channel is the subcellular marker while the second channel is the RAB. There are thus 4×4 conditions in this set. Each condition contains 20 to 25 images with a total of 330 images. Segmenting this whole set in Squassh takes approximately 24 hours on a quad core computer. We provide the raw data as well as already segmented Squassh output.

- *colocalization of VEGF121a with RAB7a and RAB11b in live cell imaging.* This example contains a single movie obtained after stimulation with VEGF121a. The three fluorescence channels correspond to RAB7a, RAB11b and VEGF121a.

The toy dataset (427KB) is available at:
`http://www.psi.ch/lbr/SquasshAnalystEN/Toy_example.zip`
It contains raw data only.

The three other datasets (637MB) are packed in a single .zip file. The file contains the segmentation results from Squassh for the three datasets. It is therefore possible to use them directly in SquasshAnalyst without segmenting the raw data in Squassh. It also contains the raw data of *VEGF121a* and *EGF RABx* datasets. In order to reduce file size, raw data (1.86Gb) for the RABx localization dataset is available in a separate file. Both are available at `http://tinyurl.com/squasshanalystdata`.

4 Tutorial

4.1 Toy example

Data preparation

- Download the toy example dataset at:
  `http://www.psi.ch/lbr/SquasshAnalystEN/Toy_example.zip`
- Unzip the file. The obtained 'toy_example' folder contains two .lif files. Each .lif file contains the images corresponding to one experimental condition.
- Extract the .lif files using the .lif extractor macro. in ImageJ choose 'Plugins→Macros→Run' and select the
  `Lif_extractor.ijm` file to launch the macro. Select the 'toy_example' folder. The Lif extractor will search recursively in the directory for .lif files and will create one new subdirectory for each .lif file containing the extracted .tiff images. Each .tiff image contains one image (with all stacks and channels if present). The toy example contains only one channel and one layer per image.

Important note : when analyzing your own dataset the images should be organized in the same way as this toy example. They should either be extracted by the lif extractor with one .lif file per experimental condition or organized as a series of .tiff images with all .tiff images in a given folder corresponding to an experimental condition. Analyzing all experimental conditions in SquasshAnalyst is then done by selecting the parent folder in SquasshAnalyst data input tab.

Image analysis with Squassh

- Launch Squassh from ImageJ Plugins→Squassh
- Click on 'Select File/Folder' and select the 'toy_example' folder.
- In background subtraction options activate 'Remove background' and set the rolling ball window size to 25. In the segmentation parameters panel set the regularization to 0.125 and minimum object intensity for channel one to 0.28. Check that 'intensity rescaling' is active, 'subpixel segmentation' inactive and 'intensity estimation' is set to automatic.

These settings can then be used as default settings for analyzing your own dataset. Adapt regularization and minimum intensity thresholds first to obtain a good segmentation. Refer to [1] for a description of further Squassh settings.

- Unselect all options in the cell masks and visualization options panels. Click 'Ok' on the main Squassh window to launch the segmentation. Squassh will recursively analyze all the images in the folder.

Data analysis with SquasshAnalyst

- Open the statistical analysis software R and launch SquasshAnalyst with:
  `shiny::runGitHub("SquasshAnalyst", "a-rizk")`
- Under the data input tab select the 'toy_example' folder.
- Click on the 'Segmentation overview tab' and select a condition for an overview of the segmentation made in Squassh.
- Select the 'Analysis' tab and click on 'Select all' under the experimental conditions selection field.
- Select 'Total object signal' in the 'Subcellular object/Image attribute field'. This will display a comparison of the amount of internalized EGF before and after stimulation.

4.2 RABx subcellular localization

Data selection

- From SquasshAnalyst 'Data input' tab click 'Select Directory' and select 'RABx colocalization' directory.
- Rename channel one and two to 'marker' and 'RABx'.

Colocalization of EEA1 with RABx

- Under 'Analysis' tab in 'compare experimental conditions' field select the four conditions containing EEA1.
- Select 'Colocalization (size)' in 'Subcellular object / image attribute' field.
- Select 'Percentage of channel.. ' Marker 'colocalizing' with channel 'RABx' to display the amount of EEA1 colocalizing with the four RABs.
- Sort the four RABs with 'sort conditions'. RAB4a and RAB5a colocalize the most with EEA1 (early endosmes).
- EEA1 fluorescence channel is obtained with immunostaining. In the 'Subcellular objects and image filters' panel increase the intensity threshold of the Marker channel to 0.3. This will remove low intensity objects of this channel from the analysis. Notice how it increases the colocalization scores of RAB4a and RAB5a with EEA1. This can be explained by the fact that in immunostaining low intensity objects can correspond to non specific staining.
- Under 'Most representative image' select 'RAB5_EEA1' condition to obtain the most representative image of this condition. A preview of the image is displayed as well as the file name of the raw image.
- Statistical difference between conditions is given at the bottom of the page.

Colocalization of RAB5 with subcellular markers EEA1, LBPA, H4B4 and PDI

- Select all conditions containing RAB5a and repeat above given steps to obtain the colocalization of RAB5a with the four subcellular markers.

4.3 EGF RABx colocalization

Data selection

- From SquasshAnalyst 'Data input' tab click 'Select Directory' and select 'EGF RABx colocalization' directory. Rename channel one to three respectively to RAB5, RAB11 and EGF.

Segmentation overview

- In 'segmentation overview tab' select condition 'R5_R11_EGF_T30' to overview segmentation of three channel images. Color legend of preview images is given on top : red for RAB5a, green for RAB11b and blue for EGF. Original image is on left side. Segmentation is on right side.

EGF internalization before and after stimulation

- In 'Analysis' tab select all conditions with 'Select all'. Select 'Total object signal' in image attribute selection and EGF in the channel selection field. By default box plots then display the total EGF signal present in both conditions.
- There is as expected no EGF present at time = 0 min.

Colocalization of EGF with RAB5a and RAB11b

- Select only condition 'R5_R11_EGF_T30' in condition selection field.
- Select 'Total object Volume Venn diagram' in image attribute selection.
- The Venn diagram shows that in average over all images at time = 30 min objects with a total size of only 1.4 pixels are positive for RAB5a, RAB11b and EGF at the same time. EGF is colocalizing about twice as much with RAB5a than with RAB11b at this time point.

4.4 Live cell imaging

Data selection

- From SquasshAnalyst 'Data input' tab click 'Select Directory' and select 'R7_R11_VEGF121a_movie' directory. Rename channel one to three respectively to RAB7, RAB11 and VEGF121a.

Internalization of VEGF121a over time

- Under 'analysis' tab select condition 'R7_R11_VEGF121a'.
- Select 'Total object signal' in image attribute selection.
- The curve displays the total amount of VEGF121a in segmented objects. This reflects VEGF121a internalization after stimulation.
- Data coming from movie analysis is often displaying high variability with time (from experimental noise or from cell stochasticity). It is possible to average over time with rolling mean. The slider controls the length of the sliding window used to average over time. Set it to 10 to smooth the curve.

Colocalization of VEGF121a with RAB7 and RAB11

- Select 'Colocalization (size)' in image attribute selection and colocalization of VEGF121a with RAB7.
- Click 'Keep plot'. Now select colocalization of VEGF121a with RAB7. The graph will display an overlay of two curves for the colocalization of VEGF121a with RAB7 and RAB11. This allows to monitor colocalization of the VEGF ligand with different RABs over time.

5 Reference

5.1 Data input

Select a folder that contains Squassh analyzed fluorescence microscopy images.

Directory selection: When selecting a directory Squassh Analyst will recursively search for all '.csv' data files generated by Squassh in subdirectories and will display found conditions data. Each subdirectory should contain files corresponding to at most one conditon (contact sheet image files and '.csv' files). All conditions should contain the same type of data (2D/3D, one channel/two channels/three channels, fixed images/movie). When using Squassh Analyst on Windows computer the directory selection window opens in R.

Movie analysis: For live cell imaging analysis data should be analyzed by Squassh as one .tiff file for all channels, slices and time points.

Channel renaming: After folder selection and if Squassh data is found text fields for optional channel renaming appears.

5.2 Segmentation Overview

Survey Squassh segmentation and exclude images from analysis

Squassh parameters: After selecting a condition Squassh Analyst will display parameters used by Squassh for the segmentation.

Images and segmentation preview: Squassh Analyst displays contact sheet previews of original microscopy images maximum intensity z-projections alongside Squassh segmentation z-projection. Cell mask if present is displayed on both images as white outline.

Image features overview: Squassh Analyst displays next to contact sheets images a table containing mean number of objects in each channel and colocalization quantification (for two channel images).

Image removal: Exclude from further analysis bad images by ticking 'Exclude image box'.

5.3 Analysis

Compute image features and compare across experimental conditions

Subcellular object/image attribute: Select wich image feature to compute and display.

- *Mean object volume*: computes the mean volume of subcellular objects. Mean is computed for each image over all its subcellular objects. For 2D images this computes the surface of subcellular objects.

- *Mean object length:* computes for each image the mean length of its subcellular objects.

- *Mean object intensity:* computes for each image the mean intensity of its subcellular objects. Intensities are normalized in each image between 0 and 1.

- *Colocalization (number):* percentage of objects in one channel that overlap with objects from the other channel. An object is considered overlapping with the other channel if at least 50% of its volume is overlapping with objects form the other channel.

  More precisely the colocalization of channel one with channel two is defined as:

  $$C_{\text{number}}(O1_{O2+}/O1) = \frac{|\{o \in O1 : O_o > 0.5\}|}{|O1|}$$

  where $O1$ is the set of all objects detected in channel 1 and $O2$ the set of objects detected in channel 2. $|\cdot|$ denotes the total number of objects in the given set. O_o is the fraction of object o that overlaps with any other object from the other channel. The notation $O1_{O2+}$ intuitively means "objects in channel one that are positive for objects in channel two".

- *Colocalization (size):* percentage of the total size of objects in one channel overlapping with objects from the other channel.

  The size-based colocalization coefficient of channel one with channel two is defined as:

  $$C_{\text{size}}(O1_{O2+}/O1) = \frac{\sum_{o \in O1} S_o O_o}{\sum_{o \in O1} S_o}$$

where S_o is the size of object o given by the number of pixels belonging to it. Hence, $\sum_{o \in O1} S_o O_o$ is the total size of the colocalizing regions and $\sum_{o \in O1} S_o$ is the total number of pixels covered by any object in that channel.

- *Colocalization (signal):* percentage of the total signal (=volume*intensity) of objects in one channel overlapping with objects from the other channel

  The intensity-based colocalization coefficients of channel one with channel two is:

$$C_{\text{signal}}(O1_{O2+}/O1) = \frac{\sum_{o \in O1} I_o S_o O_o}{\sum_{o \in O1} I_o S_o}$$

  where I_o is the estimated intensity of object o. Hence, $\sum_{o \in O1} I_o S_o O_o$ is the total signal in the colocalizing region in channel 1, while $\sum_{o \in O1} I_o S_o$ is the total signal present in channel 1 altogether.

- *Pearson correlation:* pearson correlation coefficient between two channels

- *Pearson correlation inside cell mask:* pearson correlation coefficient computed inside cell masks only.

- *Total object volume Venn diagram:* displays a Venn diagramm representing the overlap between the objects volumes in the different channels. This attribute can only be displayed for one experimental condition at a time. The first condition selected in 'Compare experimental conditionss' is used.

- *Total object signal:* total signal (=volume*intensity) of objects in an image. This quantitiy is extensive with the cell size : the larger a cell is, the more subcellular objects it can have.

- *Total object signal / Cell Size:* same as Total object signal but normalized with the cell size. Cell size is computed from the cell mask.

- *Total object volume:* total volume of objects in an image.

- *Total object volume / Cell Size:* same as total object volume but normalized with cell size.

- *Object number:* total number of objects in an image.
- *Object number / Cell Size:* same as Object number but normalized with cell size.
- *Channel selection:* Select the channel in which the image attribute is computed. For three channel images, when a Pearson correlation attribute is selected the field turns into a two channel selection form. For a colocalization attribute the selection is ordered to be able to select either channnel 1 colocalizing with channel 2 or channel 2 colocalizing with channel 1. In the case of three channel images and colocalization attribute it is possible to compute the colocalization of one channel with two others by selecting one channel in the first field and two in the second field.

Condition selection: Experimental conditions on which the image attribute will be computed. Each condition will be displayed in the chart as a separate bar or boxplot. The condition selection field is non reactive, use 'Apply conditions changes' for modifications to take effect. Use 'Select all' or 'Deselect all' as shortcuts to select/ deselect all conditions.

Excluded images: Images excluded in the 'segmentation overview' tab are given here as a reminder.

Chart type: Plot image attributes with Box plot, Strip chart, or Bar chart. Bar charts are displayed with mean values and standard error of the mean. Conditions can be sorted by increasing mean value by ticking 'Sort conditions'.

Movie analysis: In the case of movie analysis there is no chart type option. The plot displays the evolution over time of the selected image attribute.

Figure export: Use 'Export pdf' button to export current chart in .pdf format. For png format export, right click the chart and save as image. Export the data corresponding to the currently displayed chart in a spreadsheet file with 'Export .csv'.

Filters: Remove subcellular objects from the analysis by setting thresholds on their minimum volume, maximum volume or minimum intensities. Remove images from the analysis by setting a threshold on the minimum number of subcellular objects an image has to contain.

Most representative image: SuqasshAnalyst displays the image whose attribute value is closest to the median value of all images in the chosen condition. This thus provides the most representative image of the group for the calclated attribute. The file name of the image is given on top of its contact sheet preview.

Statistical analysis: When at least two experimental conditions are selected statistical significance tests results are performed in this section. Two statistical analysis are possible, parametric or non parametric. The first one is one way ANOVA followed by Tukey Honest Significance difference test for multiple comparisons. The second one is the Kruskal-Wallis test followed by Dunn's test adjusted with the Bonferonni correction for multiple comparisons. ANOVA and Tukey are parametric, that is they require that samples come from Gaussian distributions. The non parametric tests Kruskal-Wallis and Dunn do not require this assumption but only take into account the rankings of the values and have thus less statistical power.

Star ratings are given next to p values with `****` for p value < 0.0001, `***` for $0.0001 <$ p value < 0.001, `**` for $0.001 <$ p value < 0.01,`*` for $0.01 <$ p value < 0.05 and `ns` for p value > 0.05.

References

[1] Rizk et al. *Segmentation and quantification of subcellular structures in fluorescence microscopy images using Squassh.* Nat Protoc, 9 (3), 586-596 2014.

[2] Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie and Jonathan McPherson (2015). *shiny: Web Application Framework for R. R package version 0.11.1.* `http://CRAN.R-project.org/package=shiny`