COLOCALIZR - USER MANUAL

ALLAN SAUVAT
allan.sauvat@gustaveroussy.com

MARION LEDUC
marion.leduc@gustaveroussy.com

September 28, 2017

1 Introduction

ColocalizR is an R-based image-analysis application developed for the quantification of co-localization in high-throughput. ColocalizR is aimed to assess co-distribution at the organelle level, thus can be used with widefield microscopes. It does not require any programming skill.

NB: ColocalizR has been optimized to be used on images acquired with Molecular Devices MetaXpress software (nomenclature and SQL database architecture), but can be used on any set of images.

2 Requirements

ColocalizR has been developed for Windows, Linux and MacOS X operating systems and has been tested on 64-bit systems. We recommend a computer with a minimum of 8 GB of RAM and equipped with a quad-core (or higher) central processing unit (CPU).

During installation an internet connection is required for downloading necessary R packages.

3 Installation

For installation, please refer to the "README.md" file on https://github.com/kroemerlab/ColocalizR.

4 File type and nomenclature

Images must be in tag image file format (TIFF), and have to be annotated according the nomenclature detailed as follows:

One plate, one timepoint. In that case the images must be stored in a folder, which name should correspond to an integer defining *Plate ID*. Each image should be named according to its acquisition parameters; it must contain the experiment name, the wellID (from A01 to P24), the siteID (when several images are acquired in the same well), and the acquisition channel (1,2, or 3) : *ExpName_Well_si_wi*.

```
Plate ID
├── MyExp_A01_s1_w1.tif
├── MyExp_A01_s1_w2.tif
├── MyExp_A01_s1_w3.tif
├── MyExp_A02_s1_w1.tif
└── ...
```

One plate, several timepoints. In that case, a subfolder must be created for each timepoint of the experiment : *TimePoint_i*

```
Plate ID
├── TimePoint_1
│   ├── MyExp_A01_s1_w1.tif
│   ├── MyExp_A01_s1_w2.tif
│   └── MyExp_A01_s1_w3.tif
└── TimePoint_2
    └── ...
```

Several plates, one timepoint. In that case, a root folder should be used, inside of which a subfolder will be created for each plate.

```
Plate Root
├── Plate_1
│   ├── MyExp_A01_s1_w1.tif
│   ├── MyExp_A01_s1_w2.tif
│   └── MyExp_A01_s1_w3.tif
└── Plate_2
    ├── MyExp_A01_s1_w1.tif
    ├── MyExp_A01_s1_w2.tif
    └── MyExp_A01_s1_w3.tif
```

Several plates, several timepoints. The application was not designed to run that kind of analysis. Still, this should work by mixing guidelines described above.

NB: Images are automatically annotated and stored as described above if they were acquired using the Molecular Devices MetaXpress software. In that case, no further file formatting is necessary to perform the analysis with ColocalizR.

5 Analysis workflow

5.1 Launching the app

After installation, the application can be launched by these commands (in a R or RStudio console) :

```
require(ColocalizR)
Launcher()
```

This will open a browser window with the main panel, as displayed in figure 1

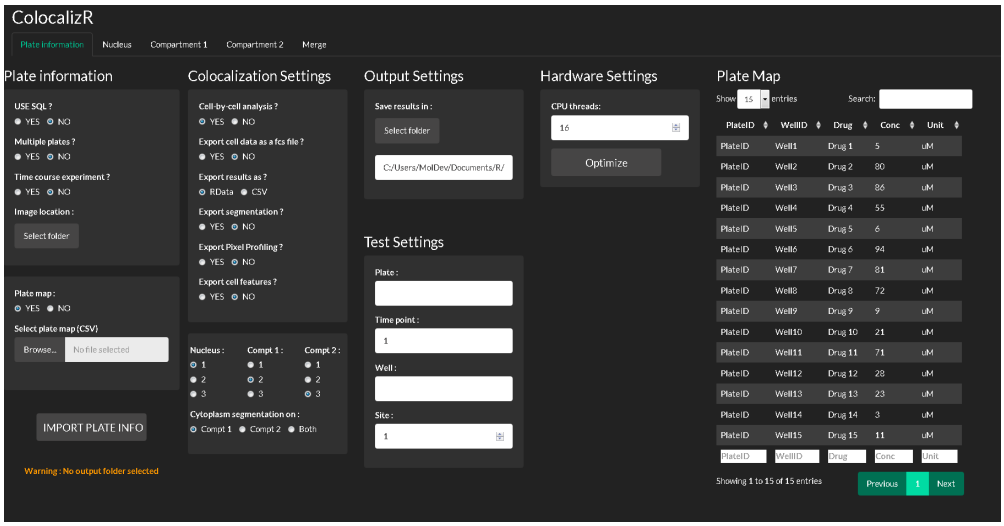

Figure 1: Main page of ColocalizR GUI.

5.2 Main tab : selecting images

Please refer to section 4 for image annotation. First, select if you want to load images by using a SQL connection (this option needs a SQL database architecture), and define your "Multiple plates?" and "TimeCourse experiment?" options. In the case you use a SQL connection, you have to indicate, if you have several databases for images, in which database your set of images is located. Otherwise, press "Select folder" to choose your image location.

A menu called "Plate selection" will appear. The selection of plates for analysis has to be confirmed by "IMPORT PLATE INFO". Following the information will appear in the "Test Settings" menu as shown in figure 2.

Several options can be selected to export data resulting from analysis. This will be detailed in section 5.4

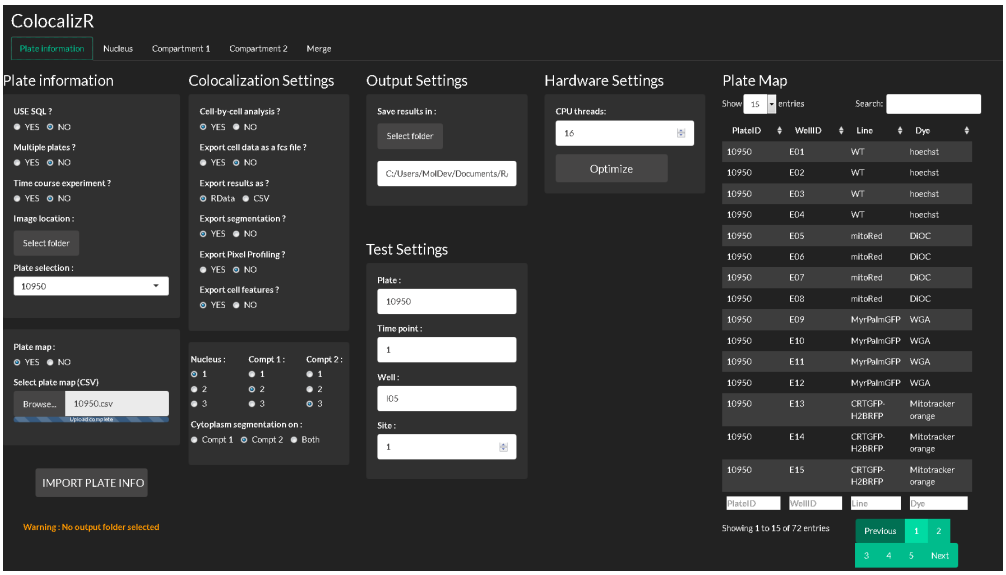

Figure 2: After plate Selection.

5.3 Adjusting segmentation settings

5.3.1 Before starting

Select images for testing. For navigating through the selected images, it is advisable to import a plate map ("Select platemap"), which has to be in comma-delimited csv format. It should be named with its corresponding Plate ID. Before adjusting the settings, the plate name, timepoint, well and site for testing have to be selected.

Define image channels. Colocalization quantification requires obviously at least two channels. For performing cell-by-cell segmentation, a supplemental nuclear stain is mandatory. You must define what channel corresponds to either "Nucleus", "Compartment 1" or "Compartment 2", as indicated in image name (w1, w2 or w3, see 4). If no nuclear stain is present, it has to be defined as channel 3.

Choose cytoplasm segmentation method. You can choose if the cell cytoplasm will be segmented according to compartment 1, compartment 2, or both. In all cases, this segmentation will be automatically conducted, but it is possible to adjust the segmentation with the adjustement parameter("Adjustement for cytoplasm segmentation" on figure 4) which ranges between 0 and 2 (the higher, the more stringent).

Define hardware settings. You can decide how many cores will be dedicated to the analysis. Press "Optimize" to calculate the optimal number of cores to be used according to your available RAM.

5.3.2 "Nucleus" tab

If no nuclear staining is present, leave "Remove nucleus from mask" on "YES". Otherwise, choose the nuclear region to be excluded from your colocalization assay (eg. if the colocalization compartments are exclusively cytoplasmic).

Here you must define the offset for segmenting the nucleus, which ranges from 0 to 1 (the higher, the more stringent). "Extrema smoothing" will blur the signal and can help segmenting heterogeneous stain. You can change the segmentation method from "Fast" to "Robust" and add a step of image denoising : in most cases, "Fast" method without denoising will provide satisfying results ; if cells are too confluent or with an heterogeneous staining, "Robust" method and/or image denoising can

give better results (yet increasing analysis time). Different combinations can be tested by clicking "TEST SETTINGS" and observing the segmentation results. The "Adjust image" slider adjusts the signal range for better visibility.

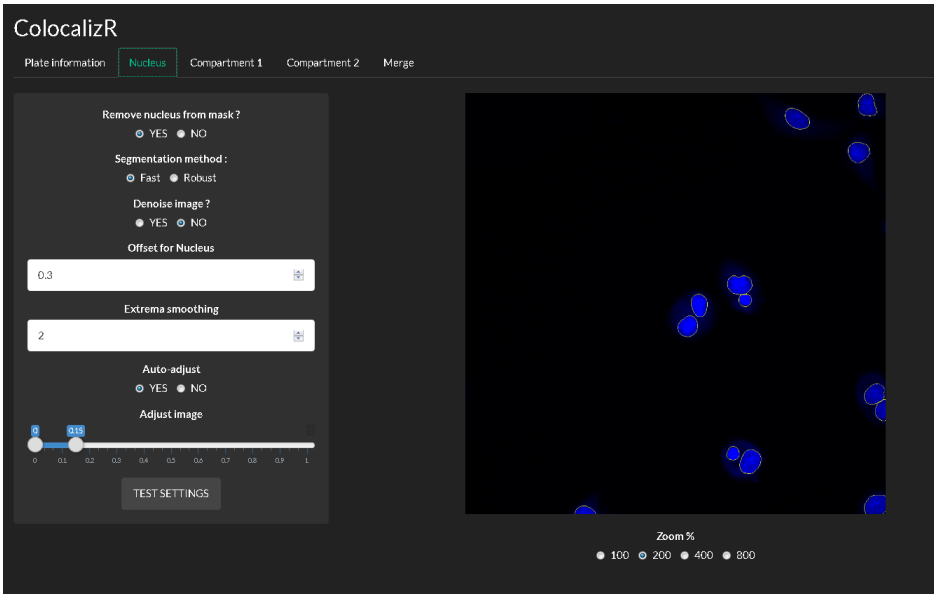

Figure 3: "Nucleus" settings tab.

5.3.3 "Compartment 1/2" tab

Here the choice between automated and manual segmentation has to be made. We recommend leaving this option set to "YES", to avoid any bias due to manual thresholding. Still, the offset can be set manually for refining the channel mask. This can be useful when working with surface overlap coefficient rather than Pearson coefficient.

"Extrema smoothing" is a parameter that removes pixel intensity extremes from the measured image. The bigger the entered integer , the more pixels will be removed and might finally influence calculated correlation coefficients.

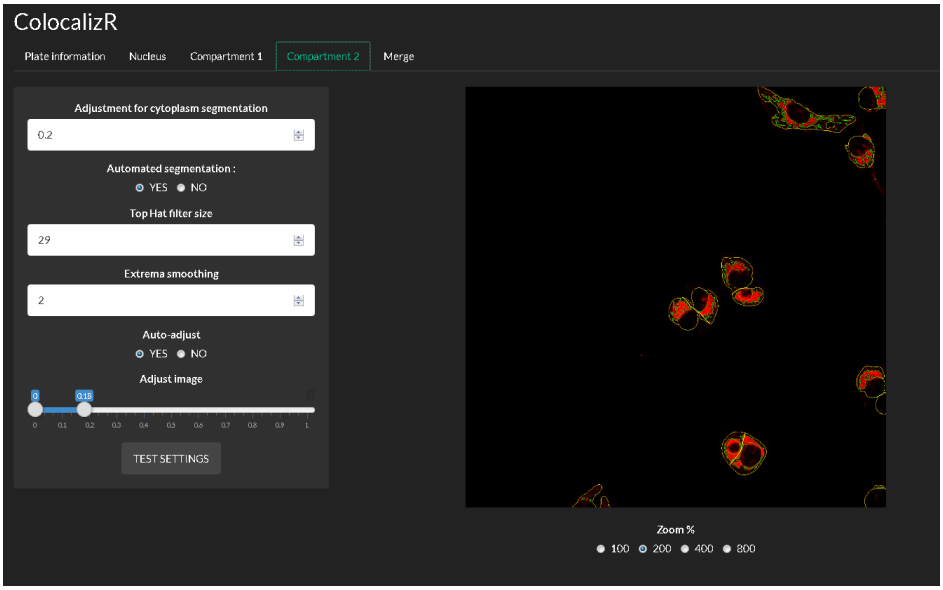

Figure 4: "Compartment2" settings tab.

5.3.4 "Merge" tab

Here the channels will be displayed as a RGB image, with the segmentation overlay. PCC and SOC calculated based on the selected settings are displayed, so that they can be refined if necessary.

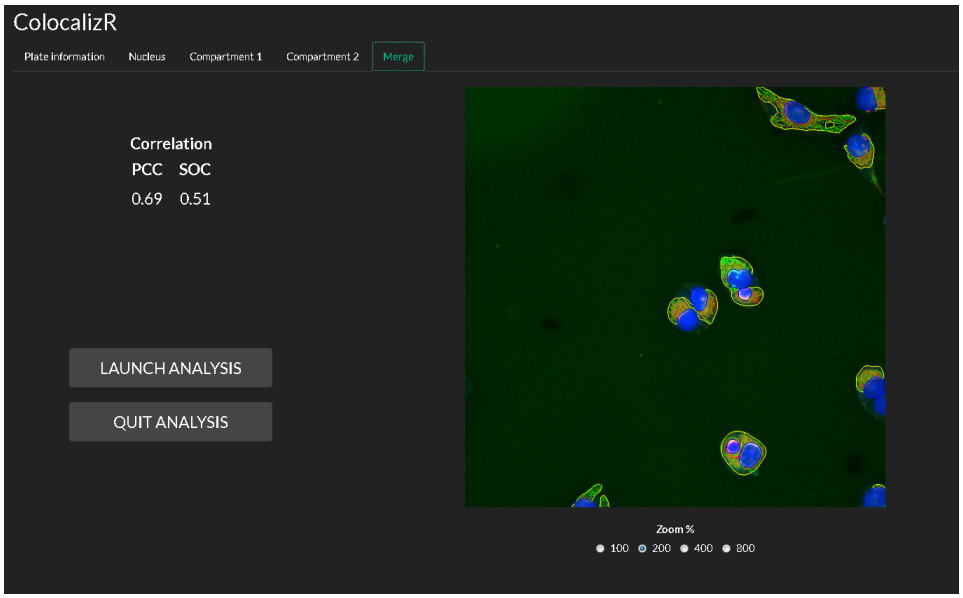

Figure 5: "Merge" settings tab.

5.4 Analyzing Results

5.4.1 Colocalization coefficients

Colocalization can be expressed by different means. Here is a quick summary of the correlation coefficients calculated by ColocalizR.

Pearson Correlation Coefficient. Pearson Correlation Coefficient (PCC) is the most robust and widely-used coefficient to quantify correlation. It is weakly sensitive to signal segmentation. It ranges from -1 (perfect anti-correlation) to 1 (perfect correlation), its formula is described as :

$$PCC = \frac{\sum (R_i - R_{av}).\sum (G_i - G_{av})}{\sqrt{\sum (R_i - R_{av})^2.\sum (G_i - G_{av})^2}}$$

Intensity Correlation Quotient. Intensity Correlation Quotient (ICQ) is derived from PCC. It ranges from -0.5 to 0.5, its formula is described as :

$$ICQ = \frac{\sum (R_i > R_{av}) = \sum (G_i > G_{av})}{N_{pix}} - 0.5$$

Manders Overlap Coefficient. Manders Overlap Coefficient (MOC) differs from PCC in the way that it's taking absolute fluorescence intensities, and not their departure from the mean. It ranges from 0 to 1, its formula is described as :

$$MOC = \frac{\sum (R_i).\sum (G_i)}{\sqrt{\sum (R_i)^2.\sum (G_i)^2}}$$

where R_i is the intensity of the first stained compartment (in red) in individual pixels and R_{Av} the arithmetic mean, whereas G_i and G_{Av} are the corresponding intensities for the second labeled compartment (in green) in the same pixels.

Surface Overlap Coefficient. Surface Overlap Coefficient (SOC) is different from the coefficients described above. Indeed, it does not take into account the pixel intensities, but only their co-occurence at the same position. By its nature, this method is very sensitive to compartments segmentation. It ranges from 0 to 1, and it can be calculated as :

$$SOC = \frac{\sum (P_G \cap P_R)}{\sum (P_G \cup P_R)}$$

where $P_G \cap P_R$ equals 0 when there is no co-occurence at a given pixel position (and 1 in the opposite case), and $P_G \cup P_R$ equals 0 in the absence of any signal at the same pixel position and 1 if at least one compartment is detected.

5.4.2 Interpreting data

In the main tab (see figure 1), a (1) cell-by-cell or (2) image-by-image based analysis has to be chosen. In the first case ColocalizR returns one set of coefficients per cell, which can be used for example for clustering subpopulations. This however requires more computing resources and increases the time for analysis. In the second case, the amount of data decreases (as compared to cell-by-cell) and consequently the speed of analysis increases. However, a much higher rate of false negative and positive data has to be expected.

Our recommendation is to perform the analysis on a cell-by-cell basis. For subsequent analysis the generated data can be exported as flow cytometry standard (fcs) files, which will allow using any flow-cytometry analysis software for graphical display. Data will also be exported as *.csv or *.Rdata (R environment) files, for analysis by other means. In addition to colocalization data and in the case of a cell-by-cell analysis, cell features (such as nuclear and cell area, fluorescence intensities in both channels) can also be calculated and exported. Besides, you can choose to export the segmentation as a TIFF file (one site per well), and the pixel profiling, i.e. scatterplot of pixel intensities distribution within the image.