ColocAnalyzer

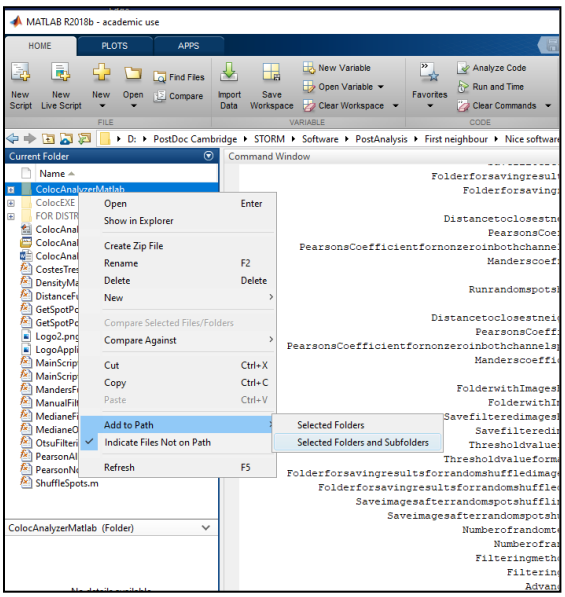
Manual

I. Installation

There are two ways of using ColocAnalyzer – either as MatLab function or as an independent .exe application. In general, there are almost no differences between these ways, except the log messages that can be viewed only when program is used as MatLab function. In this case, all messages will appear in a workspace. ColocAnalyzer should work for both – Windows and Macintosh operating system, but as it was made in Windows some problem may arise while using it in Macinstosh.

Installation of MatLab function

For this option you must have **MatLab R2018b** (or higher version) installed on your PC. Please also make sure that you have **Image processing Toolbox**. If it is already installed follow next steps:



1) Upload folder "ColocAnalyzer - master" from GitHub LAG/MNG account (https://github.com/LAG-MNG-CambridgeUniversity). Uncompress and save on your PC folder "ColocAnalyzer MatLab"
2) Open MatLab and find the path of the folder where saved folder "ColocAnalyzer MatLab" is located
3) With a right click on a folder "ColocAnalyzer MatLab": Add to Path -> Selected folder and subfolders
4) To run a program type *ColocAnalyzer* in a workspace and press Enter.

Installation of .exe application

For this option you **don't need** to have MatLab, but you have to install MatLab Runtime, that is completely free.

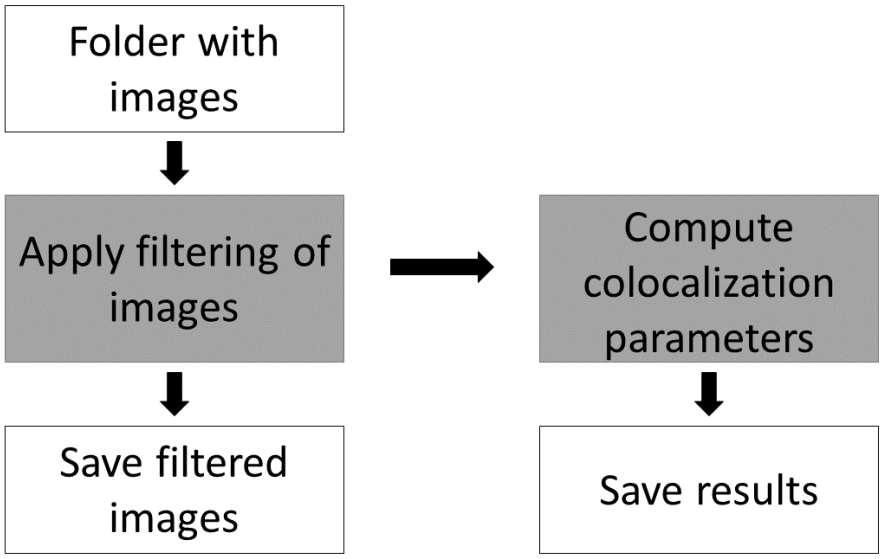
1) Go to https://uk.mathworks.com/products/compiler/matlab-runtime.html, and download MatLab Runtime with version 9.5 or higher.
2) Uncompress folder with "MatLab Runtime" and run file "setup.exe", follow instructions to install Runtime
3) Upload folder "ColocAnalyzer - master" from GitHub LAG/MNG account (https://github.com/LAG-MNG-CambridgeUniversity). Find in the folder file "ColocAnalyzer_installer.exe" and run it
4) Follow installation instructions. After program is installed you can find it in "Programs" and run it.

II. Description of software

This software uses microscopy **tif** images with two channels (colors) to perform a filtering and then **compute colocalization parameters** between specified channels and/or run a **random shuffling test** for existing microscopy images.

Computing colocalization parameters

Software uses the folder with **tif** microscopy images, apply one of the chosen filters to remove low intensity spots and then compute chosen colocalization coefficients. The diagram of computing colocalization parameters part of ColocAnalyzer software:



Most of colocalization prameters and filtering methods are well described in *Dunn et al. A practical guide to evaluating colocalization in biological microscopy, Am J Physiol Cell Physiol 300, 2011.* Short description of parameters is below:

$$\mathrm{PCC} = \frac{\sum_i \left(R_i - \bar{R}\right) \times \left(G_i - \bar{G}\right)}{\sqrt{\sum_i \left(R_i - \bar{R}\right)^2 \times \sum_i \left(G_i - \bar{G}\right)^2}}$$

1) **Pearson's coefficient**:
   Where R_i, G_i – pixel values in Channel 1 and Channel 2 (*) respectively, $\bar{R}, \bar{G}$ – are averaged values in each channel respectively.
   PCC values ranges from 1 (perfect linearly colocalized channels) to -1 (perfect inverse colocalization – the pixel with lowest value in Channel 1 corresponds to the pixel with highest intensity in Channel 2). Due to extraction of average values, PCC is unsensitive and independent of signal bias (background).
   PCC assumes on the linear relationship between two channels, and may be inaccurate in case of non-linear dependence.
2) **Pearson's coefficient for "non-zero" pixels in both channels pixels**:
   computed with the formula above but only for those pixels which has intensity above zero in both channels simultaneously.
3) **Distance to closest neighbour**: computes distance from every spot in Channel 1 to the closest spot from Channel 2
4) **Manders coefficients**:

$$\mathrm{MOC} = \frac{\sum_i \left(R_i \times G_i\right)}{\sqrt{\sum_i R_i^2 \times \sum_i G_i^2}} \qquad M_1 = \frac{\sum_i R_{i,\mathrm{colocal}}}{\sum_i R_i} \qquad M_2 = \frac{\sum_i G_{i,\mathrm{colocal}}}{\sum_i G_i}$$

(*) – here and below values marked with blue color are the parameters that can be adjusted in section **Advanced parameters**

Where $R_{i,colocal}$, $G_{i,colocal}$ values of pixels in Channel 1 and Channel 2 respectively, where both channels has values above zero.
Manders Overlap Coefficient (MOC) values are ranged from 0 (when two channels are completely mutually exclusive) to 1 (when channels are identical). It is almost independent of signal proportionality and gives a big weight to colocalized pixels (the ones with positive values in both channels). So it might be not a good choice for the samples where only few pixels overlapped.
M1 and M2 are fractional overlap coefficients. They provide metric for measuring normalized levels of co-occurrence between two channels. M1 shows weighted fraction of colocalized pixels in Channel 1.

5) **Spearman Rank correlation coefficient**

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}.$$

Where n is total number of pixels, and d_i is difference in a pixel's values rank:

$$d_i = rank(R_i) - rank(G_i)$$

Rank is computed, such the highest pixel value will have the rank 1, second highest – 2, etc. Rank is computed for each channel separately, so pixels with highest intensity will have same rank 1 even if their intensity value differs. Spearman rank coefficients does not rely on specific (e.g. linear) dependence on two channels, but rather on rank correspondence between two channels. As ranks are made separately in each channel, this correlation coefficient is unsensitive to difference in channel average intensity. So it should work fine even if one channel is much brighter than other.

As here all pixels are taken to account (even with zero values), they may give a big impact to colocalization coefficient. So, if the most of pixel are dark, it might be not a good idea to use this coefficient as it will show high level of colocalization due to good correspondence in rank of pixels with zero intensity. Using all pixels from image and sorting them result in quite long computational time. In average it takes 2 minutes to compute Spearman rank coefficient for one image, with size 512 x 512 px. The time is growing proportional to a number of pixels, so, for example one image with 1024 x 1024 px will be analyzed for approximately $2*2^2 = 8$ min etc

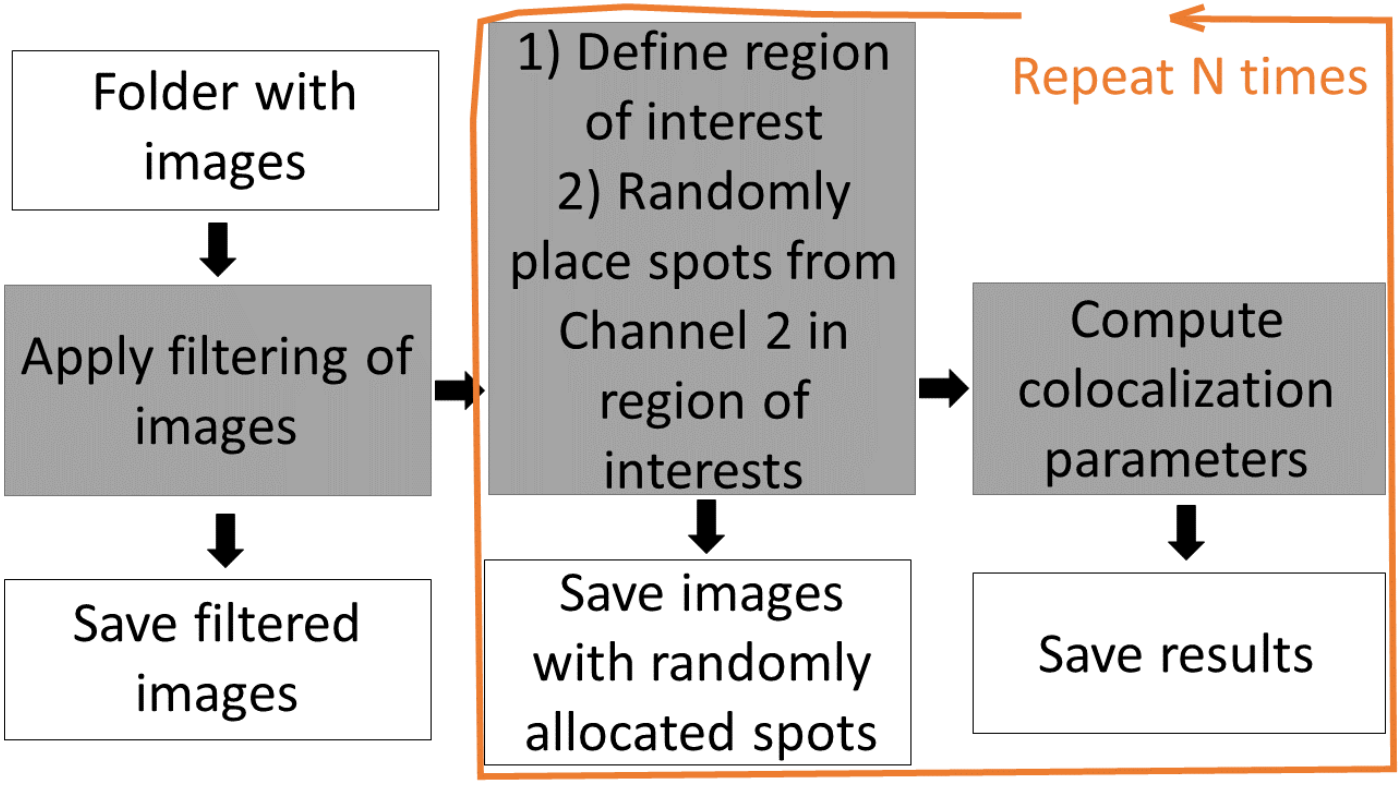
6) **Spearman Rank correlation coefficient for "non-zero" in both channels pixels**
Computed for the pixels with positive values of intensity in both channels. This method might be used in case of big fraction of black pixels (pixels with zero intensity values) to remove their influence on the correlation coefficient. Also this will greatly reduce the computational time, as only some portion of pixels is used. It

can be indeed biased in case of low number of “non-zero” pixels, so the information about number of pixels that were analyzed is also saved in .csv and .mat files.

Random spot shuffling tests

Every image have random noise and artefacts, that may influence values of computed colocalizations. Random spot shuffling test is used in order to check “randomness” of obtained colocalizations. The scheme of the test is similar to **computing colocalization parameters** part (see below) with one additional step. After images are filtered, before computing colocalization parameters program performs next steps:

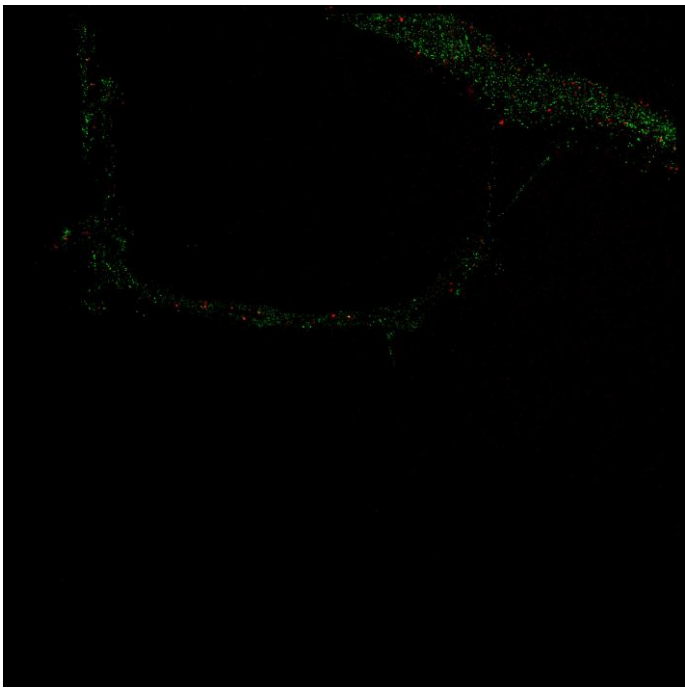
1) Define regions of interest (usually approximate area occupied by the cell) by defining regions with high densities of the Channel to determine cell boundary
2) Random positioning of spots from Channel 2 in the regions of interest while leaving unchanged Channel 1 in regions of interest. Concentration (number) of randomly allocated spots from Channel 2 is regulated by Density correction factor



Random shuffling tests can be run N times, that is defined with Number of random tests. Then, program will produce N random versions of every image with randomly distributed spots from Channel 2 and compute chosen colocalization parameters for every image.

III. How to use software

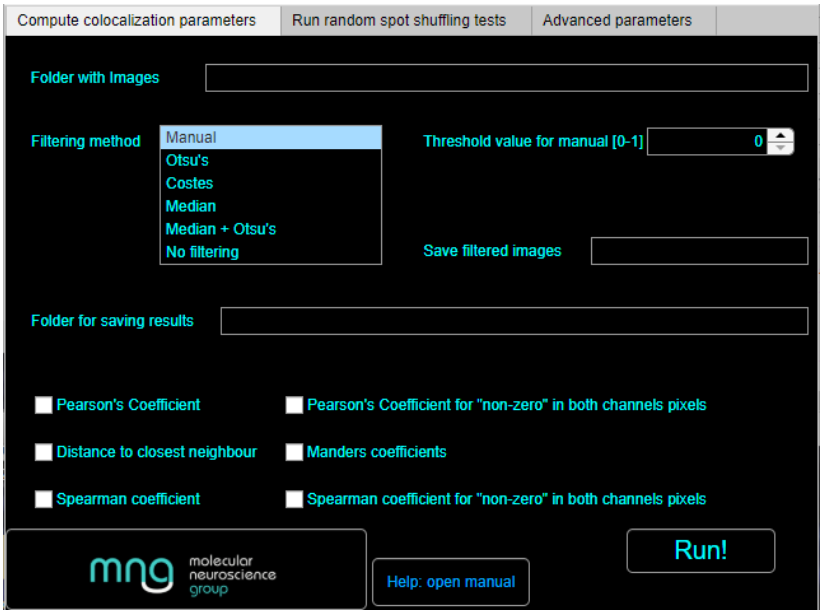
Preparing images



Images should be RGB where channels should be represented in one of 3 main colors – red, green or blue. It can be done in **FiJi → Image → Color → Merge Channels** . So image should have at least 2 colors. You should specify which colors you have chosen for colocalization in **Advanced Parameters .** For example image (shown at left) Channel 1 could be either 1 (red) either 2 (green), and respectively other value for Channel 2

Computing colocalization parameters

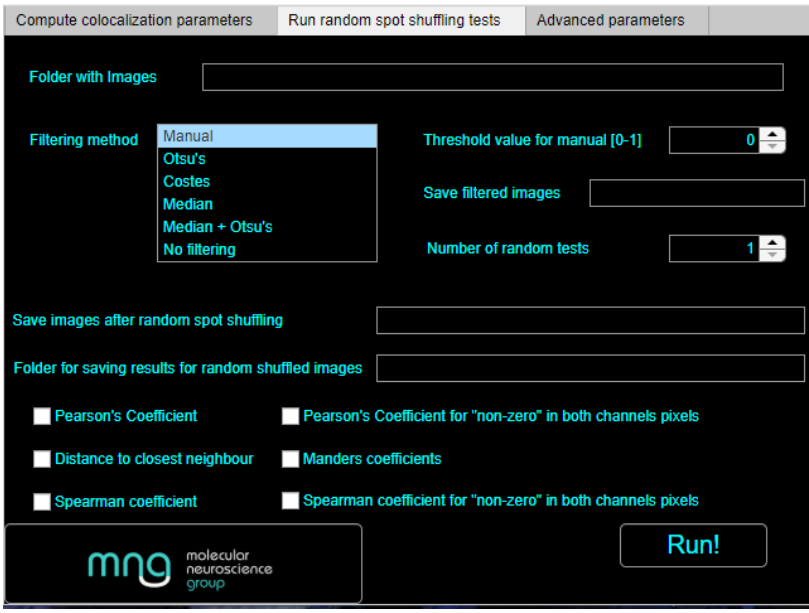
1) Put **tif** microscopy images for analyzing in one folder. Provide path to this folder in Folder with Images filed
2) Choose a filtering method. If you have chosen Manual filtering you will need to specify relative threshold value from 0 to 1 at Threshold value for manual.
3) (Optional) Specify the path where images after filtering would be saved in the field Save filtered images. If this field is empty, filtered images will not be saved but program will be still running.
4) Specify path for the folder where results will be saved in the field Folder for saving results. If the path field will be empty or invalid program will not compute the results and will only



perform image filtering. Program saves all data together in .mat file; each parameter results in a separate .csv file and plots

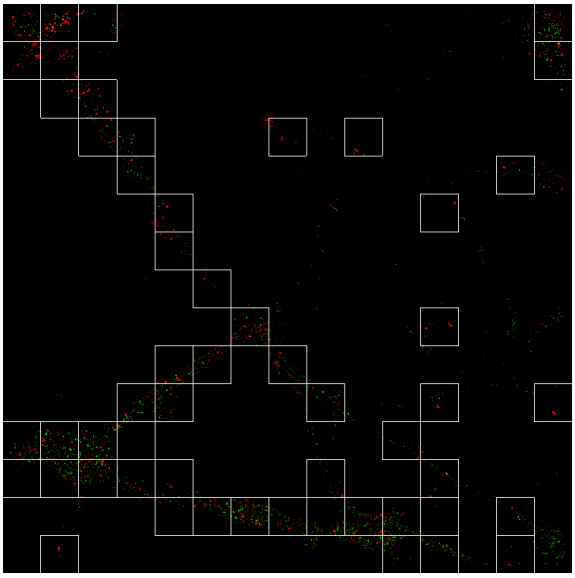
5) Choose by clicking on a checkbox which parameters you would like to be computed (Pearson's Coefficient, Person's Coefficient for "non-zero" in both channels pixels, Distance to closest neighbour, Manders coefficients, Speraman coefficient and Spearman coefficient for "non-zero" in both channels pixel).
6) Go to **Advanced parameters** tab to set 4 top parameters.
   Channel 1 and Channel 2 defines pair of colors that will be used to compute colocalization parameters. Please note, that only three colors can be used: red (value = 1), green (value = 2) and blue (3).
   Size of the averaging area median filtering is used only when Median or Median+Otsu's filtering was chosen. This is a length of the box side around chosen pixel over which median value of intensity is computed. Then, this median value is substracted from pixel intensity, so usually larger than one spot, but smaller than the distance between two spots. In practice this value usually kept constant and doesn't have a huge influence on final colocalization parameters computed.
   Pixel size – length of one pixel side in nanometers.
7) Press Run! Button to start the program

Random spot shuffling tests



1) Put **tif** microscopy images for analyzing in one folder. Provide path to this folder in Folder with Images filed
2) Choose a filtering method. If you have chosen Manual filtering you will need to specify relative threshold value from 0 to 1 at Threshold value for manual.
3) (Optional) Specify the path where images after filtering would be

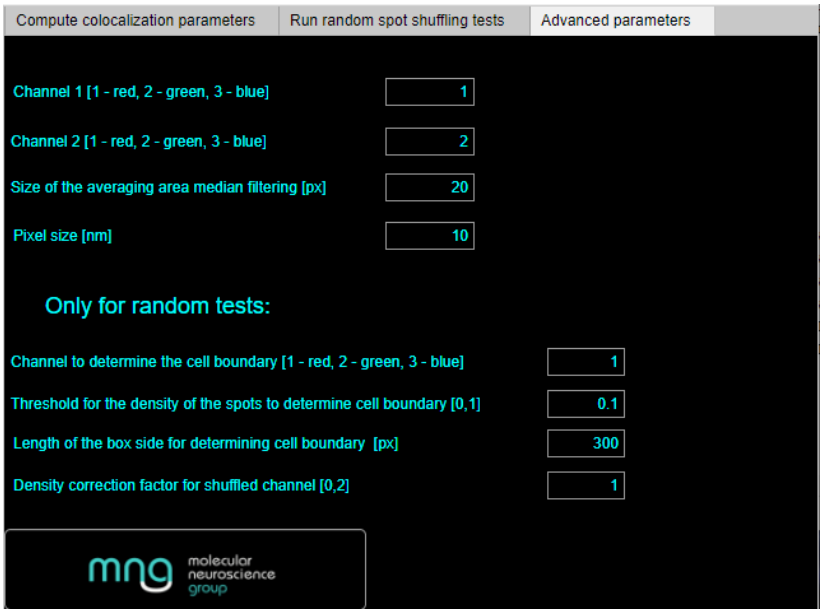
saved in the field Save filtered images. If this field is empty, filtered images will not be saved but program will be still running.



4) Choose Number of random tests as any integer number above 0
5) (Optional) Specify the path where images after filtering and random shuffling of the spots would be saved in the field Save images after random spot shuffling. If this field is empty, filtered and shuffled images will not be saved but program will be still running.
6) Specify path for the folder where results will be saved in the field Folder for saving results. If the path field will be empty or invalid program will not compute the results and will only perform image filtering and spot shuffling. Program saves all data together in .mat file; each parameter results in a separate .csv file and plots (mostly for
7) Choose by clicking on a checkbox which parameters you would like to be computed.
8) Go to **Advanced parameters** tab to set parameters. Top 4 parameters described above and bottom 4 parameters are used only in the spot shuffling:

Channel to determine cell boundary – the channel that used in order to measure the spot density and determine the regions with high density of the spots as region of interest (see image below – regions inside white squares). It can be one of the channels that are used in analysis or different one. As for Channel 1 and Channel 2 it should be one from the next colors – red (1), green (2) or blue (3).

Length of the box side for determining cell boundary – length of the side of one box used to determine field of interest. It shouldn't be neither too big (that may cause low resolution in the boundary defining) neither too small (may result in missing parts inside the cell). Playing with this parameter and Threshold for the density of the spots is an essential in defining approximate area covered by the cell as accurate as possible.



Threshold for the density of the spots to determine cell boundary defines relative value (from 0 to 1) for the spot density threshold inside a box. If value is lower – program will take to account larger area of interest

Density correction factor – user can change number of spots (so density of spots) in Channel 2 which will be randomly located from original one. Density correction factor then is defined as the ratio between the desired number of the shuffled spots to original amount of the spots in Channel 2:

$$DensFactor = N_{shuffled}(Channel2)/N_0(Channel2)$$

So if you want to place randomly only half of all spots from Channel 2 you have to set this value to 0.5, if you want to have twice more spots - to 2. $DensFactor$ could be any number from 0 to 2. Default value is 1 that corresponds to exactly the same number of shuffled spots as in original image in Channel 2

9) Press Run! Button

Please note that two parts of software - **Computing colocalization parameters** and **Random spot shuffling tests** should be runned independently, eventhough they are mostly identical. Also top 4 parameters in **Advanced parameters** tab works simultaneously for both of this parts, so it worth to check them every time before running program.

Stas Makarchuk, Molecular Neuroscience Group, Chemical Engineering and Biotechnology Department, University of Cambridge
2020

All questions, suggestions and critics can be send to sm2425@cam.ac.uk