# Documentation for Multi-colour molecular visualization of signaling proteins reveals how C-terminal Src kinase (Csk) nano-clusters regulate T-cell receptor activation.

Please refer to the original paper for details regarding the algorithm.

**Requirements:**

- MATLAB (http://uk.mathworks.com/products/matlab/?refresh=true) with version no earlier than 2014b.

- R (v3.1.2; https://cran.r-project.org/) and RStudio (v1.0.136; https://www.rstudio.com/)

- ImageJ (http://imagej.nih.gov/ij/), with the plugins Grid (http://rsb.info.nih.gov/ij/plugins/grid.html) and ThunderSTORM (https://code.google.com/p/thunder-storm/) downloaded and installed.

- Two additional R libraries, 'splancs' and 'igraph', which can be installed directly from the RStudio interface via 'Tools' 'Install Packages'

- Post processed single molecule localisation microscopy data set, in the format of **x, y, localisation precision as a .csv file for input.**

**Codes**

Data analysis is performed via a combination of Bayesian-based cluster analysis (as described in detailed in Nat Protocols, 11, 2499-2514, 2016) and self-written Matlab routines. To run the Matlab routines please use *Main Routine.m* file. Several functions are required to run the main routine, these functions are provided in the *Functions* folder. Note, that any change of user input required to run the functions can be made within the *Main Routine.m* code.

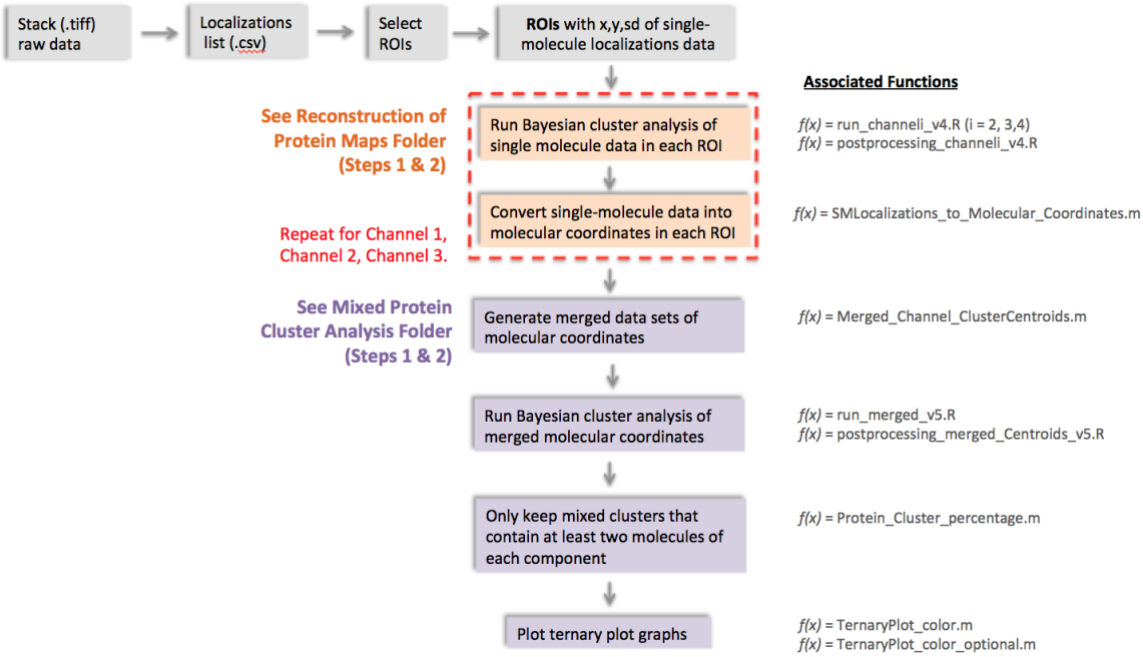


**Figure 1:** Analysis workflow depicting the main functions and their relation to the data processing stage.

## Implementation of Reconstruction of Protein Maps

1. Select ROIs (regions of interest) and stored each ROIs within ROI subfolders name as 1, 2, 3.. etc. Run Bayesian cluster analysis of single molecule data for each ROI following. To perform both of these steps refer to the guidelines described in Nature Protocols, 11, 2499-2514, 2016.
2. Start Matlab and open *Main Routine.m*. The **current folder** should contain the analysed ROIs (see Figure 2) and the *SMLocalizations_to_Molecular_Coordinates.m* function.

   Insert in **data_file_name_in** the name of .txt file that contains the x,y,sd localisation data (i.e: `data_channel2.txt`); in **data_file_name_out** the name of the file where the molecular coordinates will be saved (we recommend using the file name `ClusterCentroids_Ch2_v1.txt`, where `Ch2` can be replaced for `Ch3`, `Ch4`, etc if users are interested in running the posterior steps of the analysis, such as mixed cluster composition); insert in **summary_Bayesian_analysis** the name of the file where the summary of the Bayesian cluster analysis is saved (this would always be: `summary_channel2.txt, or summary_channel3.txt or summary_channel4.txt`). Set the number of localization per frame value in nCalibration and the cut-off value to define if a cluster of single molecule localizations should be attributed to a single binding site or not. Click Run and Advance in the Editor tab in Matlab.

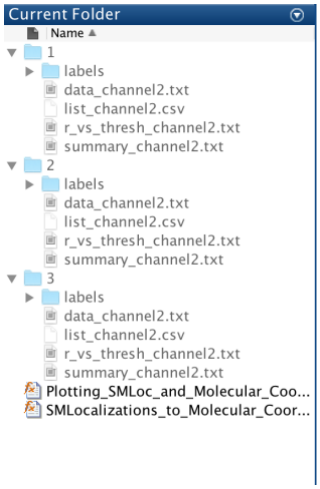


**Figure 2**. Setting files to convert in current folder.

3. (Optional) Use the "Plot SM localizations and molecular coordinates in the same graph" to visualize your single-molecule localization data together with the calculated molecular coordinates positions. Select which ROIs to plot by changing the number of **n_ROI**.

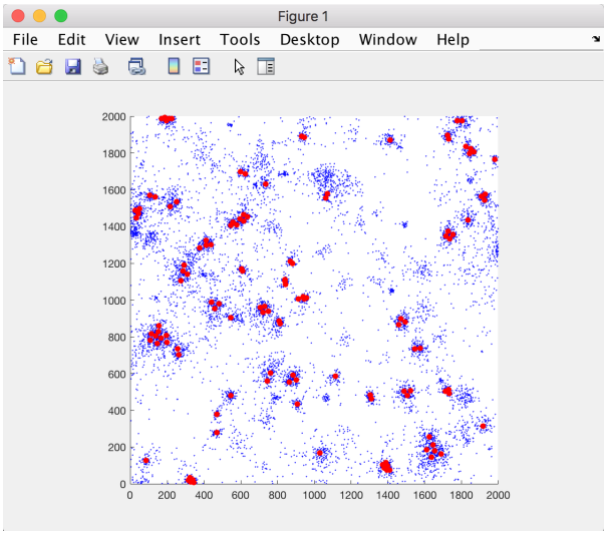


**Figure 3**. Visualize single-molecule localization together with the converted molecular coordinates.

4. Repeat this steps to convert single-molecule single-molecule localizations to molecular coordinates for each of your pseudo-colour data (i.e. `data_channel2, data_channel3, data_channel4`).

## Implementation of Mixed Protein Cluster Analysis

1. To assess the existence of mixed protein clusters, generate a single x,y coordinate list keeping the identity of the pseudo-colour for each point (merged dataset) using the *Merged_Channel_ClusterCentroids.m* function

which can be executed in the *Main Routine.m* file (lines 36-53). The **current folder** should contain the ClusterCentroids data sets for the different pseude-colours and the *Merged_Channel_ClusterCentroids.m* function (see Figure 4).

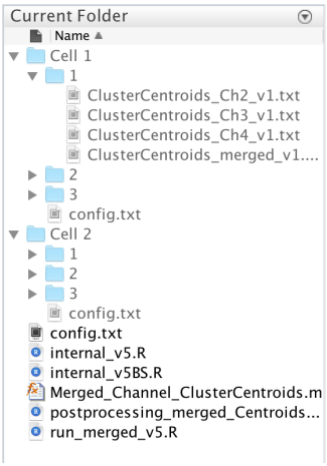


**Figure 4**. Preparing Cluster Centroids merged data sets to run Bayesian-based cluster analysis.

2. Run Bayesian cluster analysis of merged molecular coordinate data for each ROI following. Use the R code provided in: *Bayesian cluster analysis of merged molecular coordinates for 3 color data* folder (internal_v5.R; internal_v5B5.R; run_merged_v5.R and postprocessing_merged_Centroids_v5.R).

3. Return to the *Main Routine.m* file in Matlab and set the folder with the Bayesian-based cluster analysis results as the Current Folder (see Figure 5). Include in the Cell 1, Cell 2, etc folders the *Protein_Cluster_percentage.m* function. Run the function executing code provided between lines 54-69 of the Main Routine file, for each set of ROIs in Cell 1, Cell 2. If needed you can then combine data from the ROIs of different Cells using the code provided in lines 70-80.

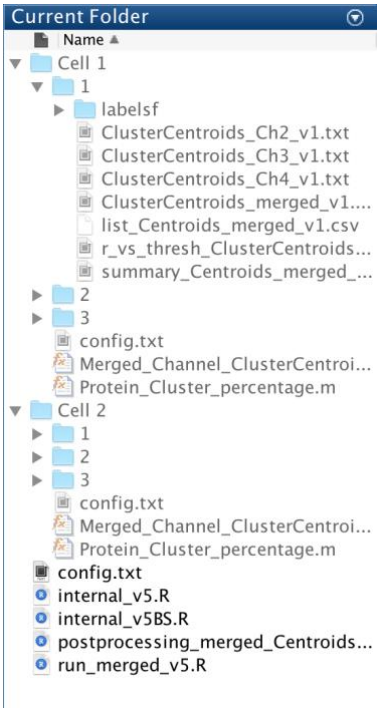


**Figure 5**. Setting files in current folder to generate Ternary Plot. Note: The only files required to run the next steps are the **list_Centroids_merged_v1.csv.**

4. Plot ternary plots by using the function *TernaryPlot_color.m*. To run this functions execute code provided in lines 81-99 of the *Main Routine.m* file. Use TernaryPlot_color_optional.m function to include the most likely cluster composition in the ternary plots.

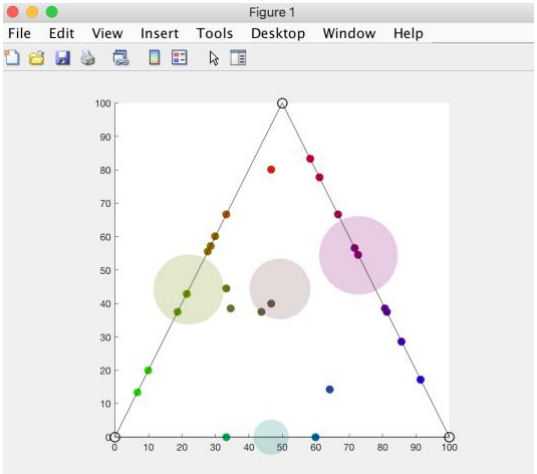


**Figure 6.** Example of output from Ternary Plot.