plusTipGroupAnalysis

TABLE OF CONTENTS

OVERVIEW

plusTipGroupAnalysis is a graphical interface allowing the user to create and analyze MT subpopulations grouped by project, sub-cellular location or dynamics.

This document explains how to use the plusTipGroupAnalysis interface. The main workflow can be decomposed into three major steps:

1) Load project(s) and setup groups for analysis
2) Divide projects into sub-regions of interests
3) Compare MT dynamics between groups using statistical tools
4) Create quadrant scatter plots splitting the microtubules into four categories.

h

ANALYSIS TOOLS

The plusTipGroupAnalysis currently implements three analysis tools:

- Sub-ROI tool (see Applegate *et al.* 2011, Fig. 7A&B):
  Divide the cell area into sub-regions of interest and extract sub-tracks in the correponding regions.

- Group analysis (see Applegate *et al.* 2011, Fig. 7C):
  Perform two statistical tests on different groups of cells or sub-regions of interest.

- Quadrant scatter plot (see Applegate et al. 2011, Fig. 8):
  Classify MT subpopulations by their dynamics (*e.g.* by growth speed and growth lifetime) and generate quadrants and overlays for each subpopulation.

GROUP SETUP

SELECT PROJECTS

This step allows you to choose one or more projects (i.e. roi_1 folders) for analysis. You may analyze many movies as a batch, provided they should use the same control parameter set.

- If "**Load projList**" is checked, you will be asked to select one or more projList.mat files containing the directory paths to various projects you have previously created. This is a shortcut, as generating the project list can be time-consuming with large directory trees.
  The projList.mat file is generated by the getProj function, which is called during project setup from plusTipAnalysis. You can also run getProj from the command line to generate more specific project lists. (see function header for details.) If "Load projList" is unchecked, you will select a parent directory containing previously-created projects.
- If "**Narrow down list**" is checked, a window will pop up asking for one or more search strings. These are strings of characters that can be used to narrow down the number of projects you have to scroll through when selecting from a long list. For example, if "ctrl" appears anywhere in the file path to your control movies, you may enter "ctrl" into the search string list. Only those projects matching all the query strings will appear. If "Narrow down list" is unchecked, this step is bypassed and all the projects will appear in the list.

From the resultant list of projects, select one or more and use the arrow to move them from the left to the right. The selected projects will be loaded into the Matlab workspace as a cell array in case you want to reference them.

- If "**Create groups from projects**" is checked, groups of selected projects will be created using the selected projects.
- If your data is arranged in a data hierarchy such that projects from different groups are stored at the same level, you may generate groups automatically by checking "**Auto group from hierarchy**." You will be prompted to select which levels of the directory tree should be used to create unique group names. If this option is unchecked, you will be prompted to choose groups of projects and name them. Avoid using spaces and hyphens in the group names.

The output can be saved in a file called "groupList.mat"

LOAD EXISTING GROUP(S)

If groups have been previously created, they can be loaded directly into the workspace by selecting the MAT file saved at the end of the group creation operation. This option allows the selection of multiple group lists which are then combined together.

SELECT OUTPUT DIRECTORY

This button allows the user to select a directory where to save the output of the analysis tools (see below).

SUB-ROIS

If **Use project mask to create subROIs** is checked, the analysis will automatically load the roiYX.mat file created when setting up the project to divide the cell into sub regions of interest. Else the user will be prompted to draw the mask or manually select a file containing a masl.

There are currently two modes to divide the cell into sub-regions:

- If **Manual selection** is selected, the user may select a variable number of ROIs. If the regions overlap when selected, they will be automatically adjusted so no overlap occurs during track extraction.
- If **Auto divide center/periphery** is selected , the cell is split into a central and a peripheral sub-ROI. The peripheral region can be further sub-divided by checking the "**Also divide periphery into quadrants**" option. The thickness of the peripheral band is chosen by the user in microns or as a fraction of the largest distance from the cell edge to the center of mass. Thus, automatic sub-ROI selection creates 2, 4, or 5 sub_x folders.

If it is desirable to exclude tracks from some region of the cell (e.g. from a previously-selected sub-ROI), check the "**Choose exclude regions**" option and either load a mask or draw the region(s) for exclusion when prompted.

Next, define how long a track must exist in the ROI to be included. This duration is given either as a fraction of the track's lifetime or as some number of seconds (use the dropdown menu to select the lifetime unit)

.

To begin, select one or more projects and press "**Select Sub-ROIs**."A 'subROIs' folder will be created under the roi_x directory and will contain info for all sub-ROIs.

Previously-created sub-ROI projects (sub_x) may be included during project selection; for these, new sub-regions cannot be selected, but tracks will be re-extracted according to the lifetime fraction/seconds.

Sub-ROI 'meta' folders will contain data for GROWTH PHASES ONLY pulled from the original ROI's data. Sub-ROI projects can then be selected for maps and movie generation, group analysis...

GROUP ANALYSIS

If **Remove tracks at the beginning/end of movie**"is checked, any track not entirely contained within the frame range will be excluded. If this option is unchecked, any track which ends before the frame range begins or begins after the frame range ends will be excluded.

Currently we have two different ways to compare between groups.

1. **Pool data from all cells in group**
   Pool per track parameters (gs,gl,gd, fs,fl,fd, bs,bl,bd) for all cells treated under a given condition. N ~1000 (on the order of the number of sub-tracks you pool)

Pros to perTrack Perspective: Have lots of stats: lots of information. One can also perform a mean subtracted KS to test for changes in a specific population of microtubules.

Cons to perTrack Perspective: This form of analysis is only useful for a parameter that can be calculated per track. Here one also loses information regarding cell to cell heterogeneity as information from all cells in a given group are combined. Also the fact that the underlying distributions come from a potentially heterogeneous cell populations makes this approach potentially sensitive to false positives. (See below)

2. **Perform per cell analysis**
   Test for differences among two cellular populations

Pros to perCellAnalysis: Get information regarding cell-to-cell heterogeneity. Can look for cellular 'outliers' and partial knockdown effects.

Cons to perCellAnalysis: No information about subtrack distributions besides mean/median. Therefore, in some respects less information, which makes it harder to detect more nuanced changes in a specific population of microtubules.

After selecting a group comparison method, the user can choose two statistical tests between a series of predefined test using the drop-down menus. See below for a discussion of which test to apply to each group comparison method.

First statistical test (below diagonal)
he result of the test will be save in the output directory as “hitTest1.mat”. All hits for this first statistical test will be exported under graphic format.

Second statistical test (above diagonal)
The result of the test will be save in the output directory as “hitTest2.mat”

Test stringency (alpha value)
This allows the user to choose the value of the stringency of the statistical tests. A post-processing statistics is considered as statistically significant is at least one of the p-values is less than this stringency (alpha value). Histograms of the corresponding hits will be generated and saved in the output directory.

POOL DATA FROM ALL CELLS IN GROUP

The default test for this analysis is the permutation test of the means (as these distributions are almost certainly NOT normally distributed) and the calibrated mean subtracted KS test which allows one to test for differences in total microtubule distribution among the different groups. It is recommended to use the calibrated KS test as the large sample size often makes the statistics hypersensitive. These default tests reflect the tests of the discrimination matrix of Figure 7C in Applegate et al. Results of these stats are similar to those documented in Thoma et al. Figure 1H left-hand plot. (It is of note that in the Thoma study there a consistency test was applied before pooling data. See Materials and Methods of the Thoma paper- Merging of data from multiple videos per condition. No such conistency test is implemented here.

Note very often the p-values for the **permutation t-test** of the means are quite low due to the large sample size and cell-to-cell heterogeneity that **may not** be dependent upon your condition of interest. (Note that the p-value in Thoma *et al*. Figure 1H first plot is for the calibrated KS test NOT the perm t-test). Running statistics on two same day control populations is a good initial way to determine how much variability one observes in the subtrack population in the absence of the perturbation of interest.

Do within group comparison: if this option is checked, histograms and boxplots are plotted comparing elements from each group.

PERFORM PER CELL ANALYSIS_

This option allows you to test for differences among two cellular populations. For example, if we have a perTrack Parameter, such as growth velocity, we average over all subtracks for a given cell. It is this avg value per cell is the number we feed into the statistical test. Therefore, N becomes on the order of cell sample size (~8) We expect these cellular distributions to be closer to a normal distribution, though as we are not sure of this, we often run both the t-test and the perm-t- test to test for differences in the means of these cellular populations. (These two tests should yield similar results as the cellular distribution should be approximately normal). These statistics correspond to the analysis performed in Thoma et al Figure 1H (right side). The groupAnalysis GUI automatically screens all per cell statistics as saved in projData.stats for differences that are statistically different between the two groups. Hits are plotted in two forms 1) as per project (per cell) plots, which allow one to examine the cellular distribution of the hits. And 2) as bar graphs that show the mean of the cellular population for the various groups tested. In the later case the error bars represent the std of the cellular distribution. Hits are always defined using the first test.

Note it is easy to add your own field of interest to plusTipDynamParam. If you save it as projData.stats.yourField this new field will automatically be considered in the group analysis stats.

We typically do not run a KS test here as this will be metric of the mean subtracted cellular distribution which in many cases one would expect to stay fairly constant (though it can potentially be used as a metric to evaluate the degree of knockdown under a given treatment).

Note that default for both the t-test and the perm t-test are two-tailed (note in the original version the default for the t-test was set to two-tailed, while the default for the perm-test was one-tailed).

QUADRANT SCATTER PLOTS

Use the Quadrant Scatter Plots panel to color-code tracks falling within specified ranges of various parameters.

Select parameters to be plotted from the x- and y-axis drop-down menus, e.g growth speed and growth lifetime. Adjust the data values or percentiles for each parameter independently and provide min/max limits (if desired) for each. Data outside this range will be excluded from the analysis.
Because the values on the x- and y- axes must be paired, only certain combinations of parameters work. The track type (e.g. "fgap") must bethe same for x- and y- axes.

If **Remove tracks at start/end**"is checked, any track not entirely contained within the frame range will be excluded. (Lifetime measurements can be biased especially in short movies where most long tracks will exist at the beginning or end, thereby getting discarded.) If this option is unchecked, any track which ends before the frame range begins or begins after the frame range ends will be excluded.

If projects from different groups should be compared, use the "**Batch process on groups**" option and select the appropriate groupList (see Step 1 above).

Seven figures for each project will appear:

5) a scatter plot
6) five images with tracks overlaid (four colors separately and together)
7) a percentage bar plot

For the track overlays, the colors of the tracks correspond to the color map of the scatter plot. For example, if we take the 50th percentile each for growth speed and growth lifetime we will see four populations in four colors: fast and short-lived, slow and short-lived, fast and long-lived, and slow and long-lived. The four populations will appear separately in four images and merged together in a fifth image. The percentage bar plot will show the relative proportion of the four populations.
If running in batch mode, summary percentage bars and the raw data of the four colors will be saved for each group. The percentage bars will be stacked in the order of the group names (grp1, grp2 etc.), and the data will be stored in "btwGrpQuadStats" file. To speed up processing during batch mode, choose the "Make summary plots only" option to bypass making track overlays.

NOTE

It is also possible to divide the population of tracks based on one parameter into three groups. For example, if we choose growth speed for both the x- and y- axes, and select the 25th and 50th percentiles, respectively, we will see three populations in three colors: tracks in Q1, tracks in Q4, and tracks in both Q2 and Q3. In this case one figure will simply show the raw image.