Analysis of flow-cytometry microbial biofilm datasets with viSNE and Matlab

Anze Zupanic (✉ anze.zupanic@eawag.ch)
  Department of Environmental Toxicology, Eawag
Linn Sgier
  Department of Environmental Toxicology, Eawag

Abstract

In the present protocol, we explain how microbial biofilm stain-free flow-cytometry datasets can be analyzed using visual stochastic network embedding. This makes visual interpretation of the microbial community structure possible and allows detection of shifts in microbial biofilms after perturbations.

Equipment

PC with installed: Matlab \(R2012-R2014), with cyt \(download from: http://www.c2b2.columbia.edu/danapeerlab/html/cyt.html).

Procedure

Data Import and Transformation 1. Run cyt in Matlab \(select cyt directory as working directory and type run_cyt into Matlab prompt) and choose add data \(plus sign) in the GUI. Cyt enables import in the .fcs \(default, standard flow-cytometry format) as well as .csv format. To import from the .csv format, change file filter to *.csv. The .csv file needs to be formatted with the columns representing the measured parameters and each row representing a different measured particle/cell. 2. Transform the data using hyperbolic arcsin transformation, with cofactor of 150. 3. Merge or subsample the datasets. Running viSNE and post-processing in cyt 4. Select the samples and measurements that you want analyzed by choosing the appropriate Gates and Channels in cyt. Right-click on the selected ones and run viSNE \(bh-SNE). Once the algorithm is finished, new channels, named bh-SNE1 and bh-SNE2, appear. These are the SNE coordinates used for visualization of the data. 5. Visualize viSNE maps as scatter plots \(select bh-SNE1 and bh-SNE2 channels and click scatterplot), and play with the visualization options. 6. Categorize visual clusters into subpopulations, based on expert knowledge and optical and autofluorescent properties. For example, by comparing the optical scatter and fluorescence intensities at different wavelengths it is possible to assign a category \(e.g. cells with low forward scatter and high 660 nm intensity is likely to consist of cyanobacteria). 7. For quantification, the identified subpopulations need to be delineated, using the polygonal region selection tool in cyt. Counting the number of particles/cells in each delineated region is done automatically by cyt. Post-processing in Matlab 8. The analysis can be exported from cyt into Matlab by saving the session in cyt and then loading the session .mat file into Matlab prompt. 9. To calculate the similarity between different viSNE maps, you can use maximum mean discrepancy as described in the file mmd_visne.m \(https://github.com/anzezupanic/FC_analysis). The compared maps should be of equal number of cells, and obtained from the analysis of the same parameters obtained from the same instrument. 10. To see where particles/cells NOT used during viSNE would be projected onto a viSNE map, use the script in overlay_visne.m \(https://github.com/anzezupanic/FC_analysis).

Timing

\(Per step) 1. ~ few minutes 2. ~ few minutes 3. ~ 10 minutes \(depends on complexity of data) 4. ~ 60 minutes \(150000 cells, on a desktop PC) 5. ~ 60 minutes \(depends on experience) 6. ~ 60 minutes \(depends on experience) 7. ~ 5 minutes 8. ~ 1 minute 9. ~ 1 minute \(to calculate similarity of two viSNE maps) 10. ~ 15 minutes \(depends on number of cells you want to project)

Troubleshooting

Step 2: After transformation, there is very little variability left in my data. The cofactor should be chosen carefully, according to the properties of the datasets. We suggest that several different cofactors are tested, as well as different cofactors for different measured flow cytometry parameters, and that the analysis is repeated until an optimal cofactor is found. For stain-free flow cytometry data presented in our study, a cofactor of 150 was chosen. For mass cytometry datasets, a cofactor of 5 is usually used. Step 3: My viSNE maps look crowded. Crowding becomes an issues at ca. 150000 cells analyzed. Optimal visual results are usually achieved between 50000 and 150000 cells. Step 6: I don't see any clusters. Check your transformation and crowding. Step: 10. I have many cells, but see only a few on the viSNE map. It is possible that the cells whose properties you are projecting are very different from the cells in the viSNE map. In that case, many different cells would be projected onto the same parts of the viSNE and create the illusion that only a few have been projected. If this is the case, the projection should not be trusted.

References

Amir E-aD, et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat Biotech 31, 545-552 \(2013).
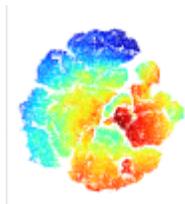
Figures

Figure 1

viSNE map viSNE map