Introduction

t-SNE (t-distributed Stochastic Neighbor Embedding) is a technique for reduction high-dimensional data in order to visualize it and have more clear view for analysis.

For further reading and applications in flow cytometry please refer to:

- https://www.datacamp.com/tutorial/introduction-t-sne
- https://www.ptglab.com/news/blog/introduction-to-t-sne-for-flow-cytometry/
- https://marissafahlberg.com/a-basic-overview-of-using-t-sne-to-analyze-flow-cytometry-data/
- https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2022.873315/full
- https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2019.01194/full

t-SNE analysis of raw cytometry data

According to the needs you can perform t-SNE analysis on different cytometry data, which are usually:

1. Different gated subpopulations within one experiment
2. Different conditions of one sample
3. Different time points of the same sample
   … or combination of the above

The code was adjusted to work with raw ".fcs" files from BD FACSDiva, however it should work with ".fcs" files of any origin.

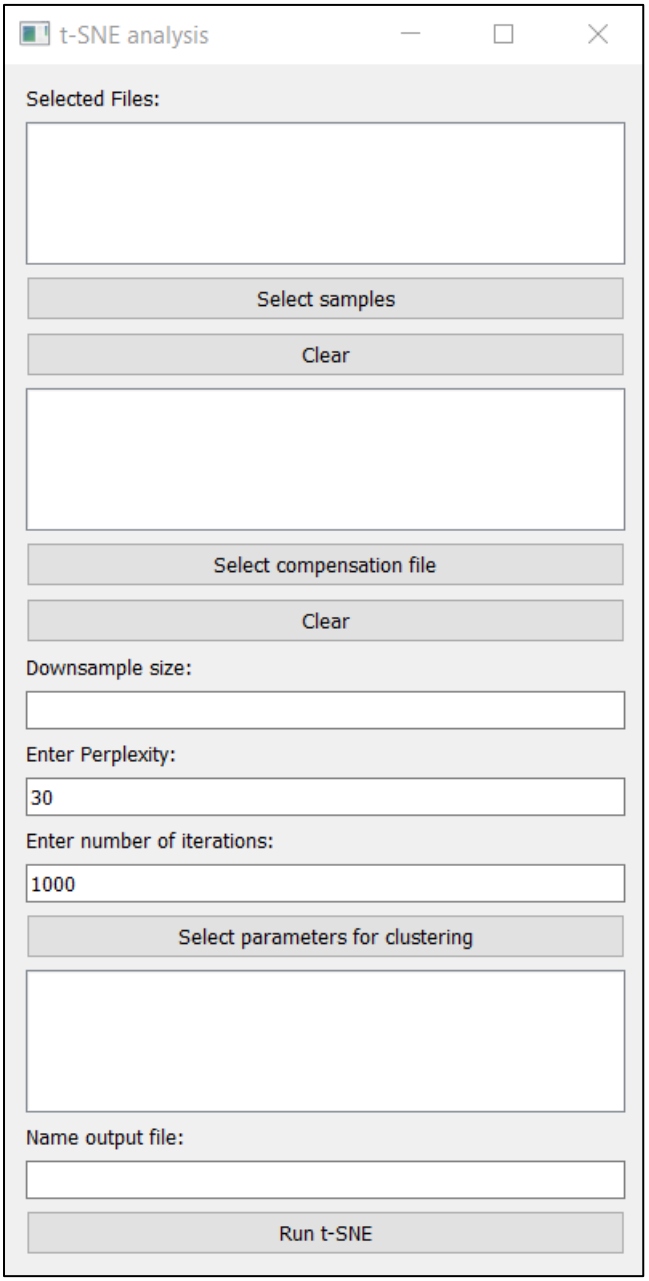
Save .fcs files

In order to have proper analysis the following conditions should be met:

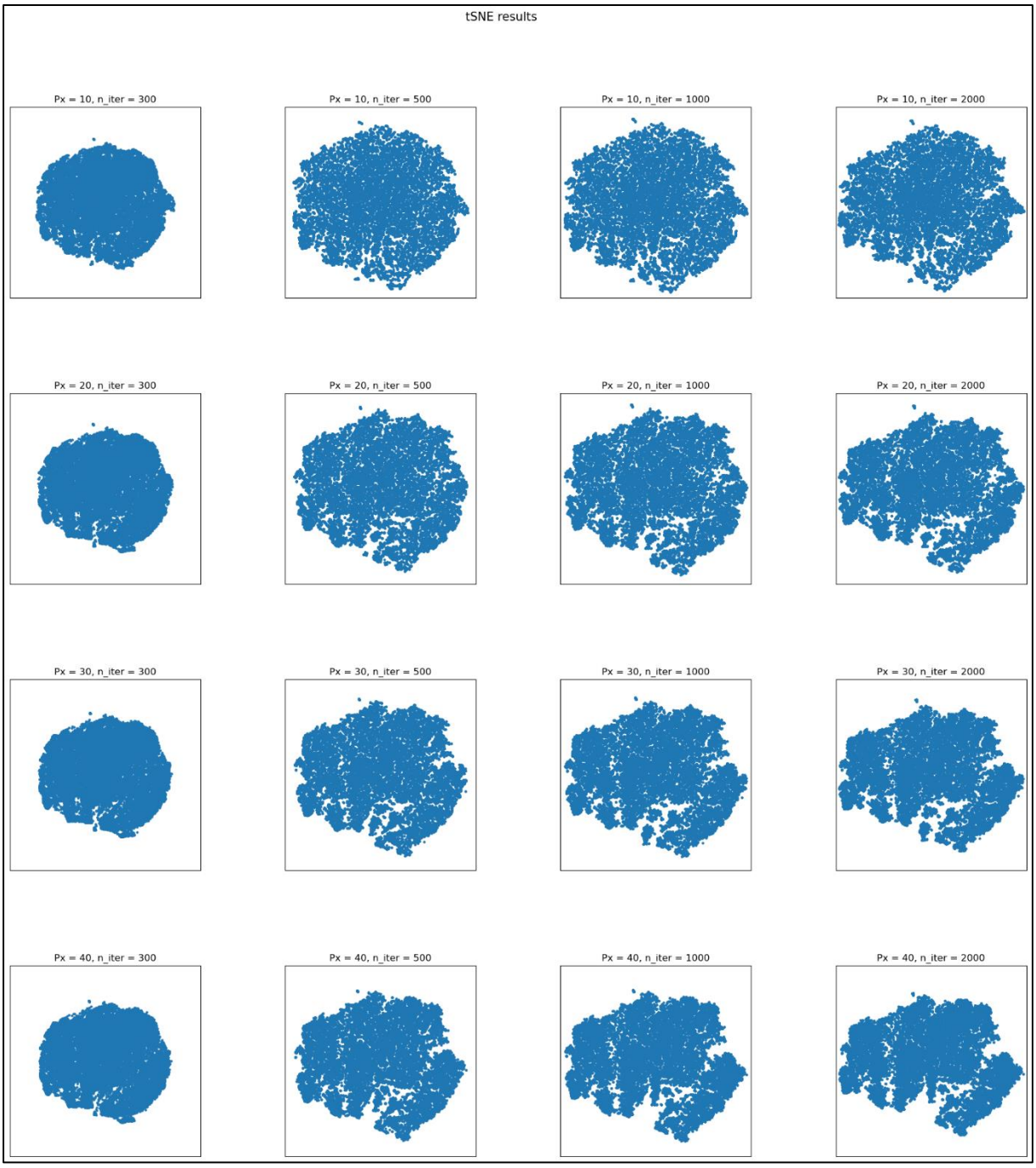
1. Export FCS3.0
2. The channels between all exports should match (example: if you don't export "FCS-W" for one sample it should not be exported in any other)
3. The compensation for all samples should be equal
4. Export one file without gating (All Events), it will be used as compensation file (does not matter which sample). This is due to BD FACSDiva problems with exporting compensation matrix in gated samples
5. Have an idea of the number of events in each sample, in particular the lowest number
6. Name the files as you want them to appear in the graphs. Better precise, but not too long

Create t-SNE map of your samples

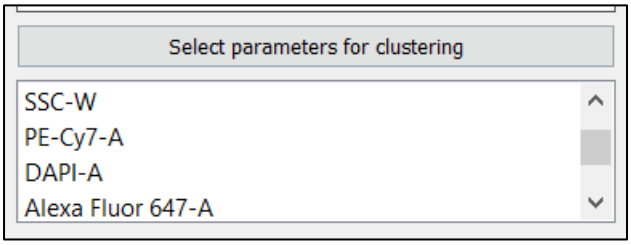
1. Open “Perform tSNE.py”
2. You will get into the following window



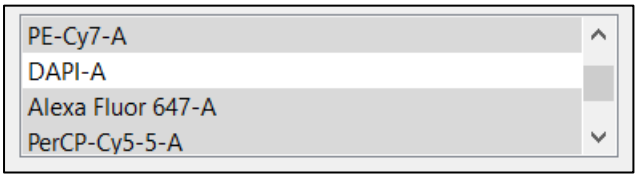
3. Select your samples (.fcs) using “Select samples” button. You can select multiple files or one by one. You can delete all selected samples by clicking “Clear”
4. Select ONE compensation file (point 4 of previous paragraph)
5. Select downsample size. It is the number of events that will be randomly taken from each sample. Usually it is around 5000 events. Caution: the downsample size should be lower than the number of events in each sample.
6. The perplexity and iteration values are set to be 30 and 1000 as default, but can be changed if needed. Higher perplexity leads to more separate clusters. Here is an example of their impact on tSNE plot:



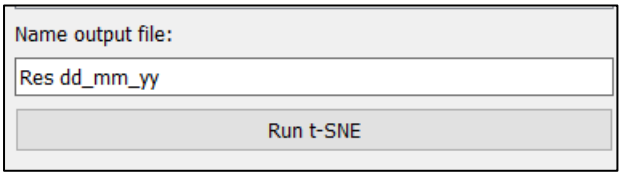
7. Click “Select parameters for clustering”



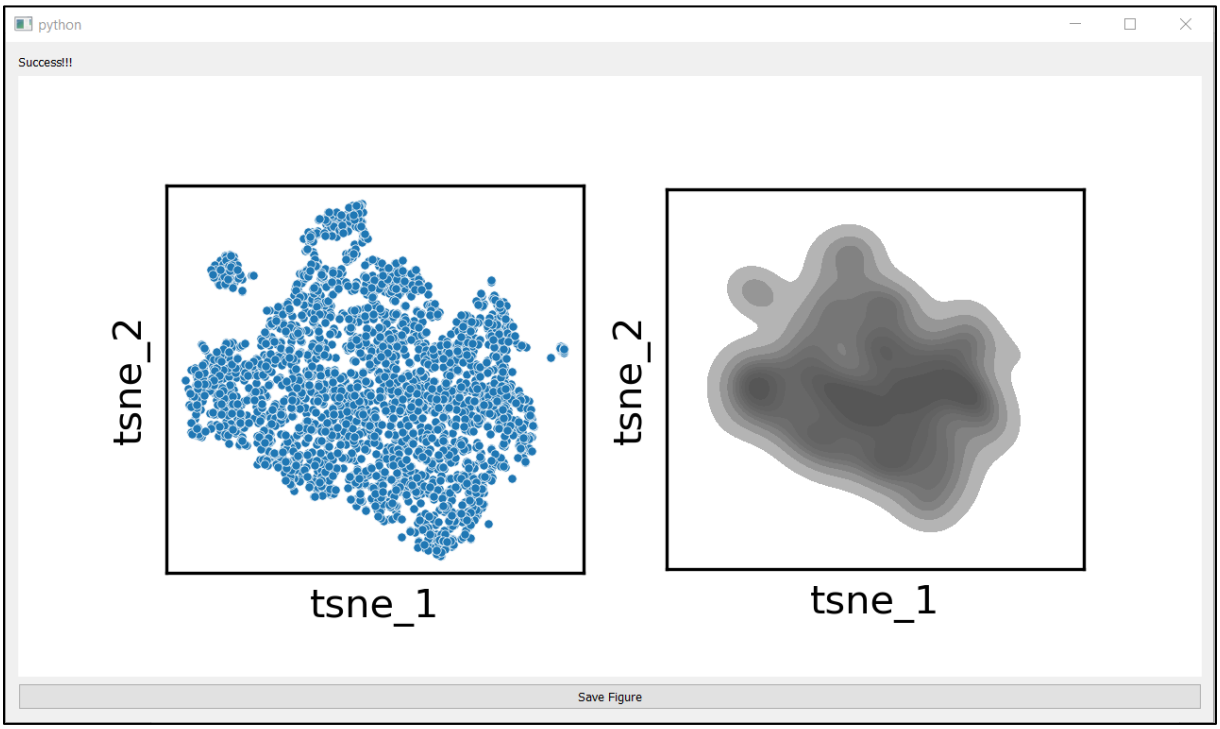
Select parameters that will be used for the tSNE (can use Ctrl or Shift). Usually if it is known that all events have similar values in one channel, this parameter is not used (for example FSC-A for one cell type)



8. Name your file (do not use /) and Run tSNE



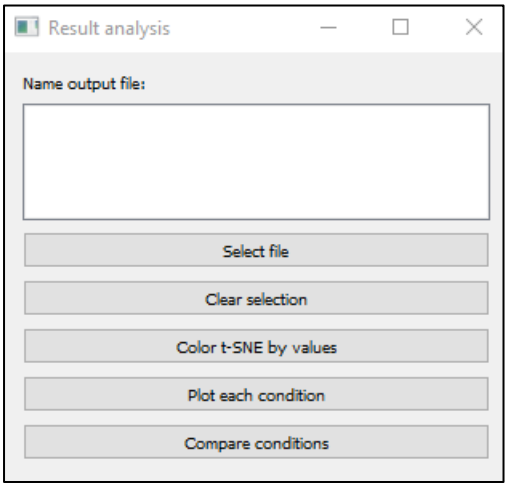
9. After some time you will get the resulting plot



10. You can save the figure
11. The resulting data with your chosen name will be stored in the folder “Results”
12. This procedure can be done once for the data of interest and analyzed after without performing t-SNE each time

Analysis of the t-SNE plot

1. Open “Analysis.py”
2. It will open the following window

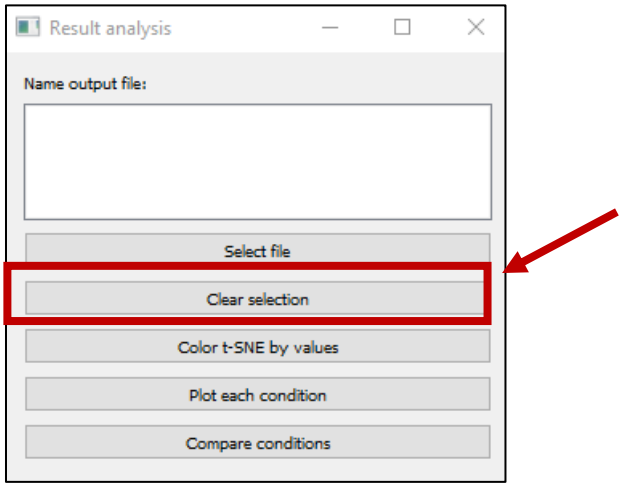


3. Using “Select file” open t-SNE results (“.pkl”), obtained with “Perform t-SNE.py”. You can open only one file. Here the results of “Xn.pkl” are shown. This file can be found in the “test/for analysis” folder.

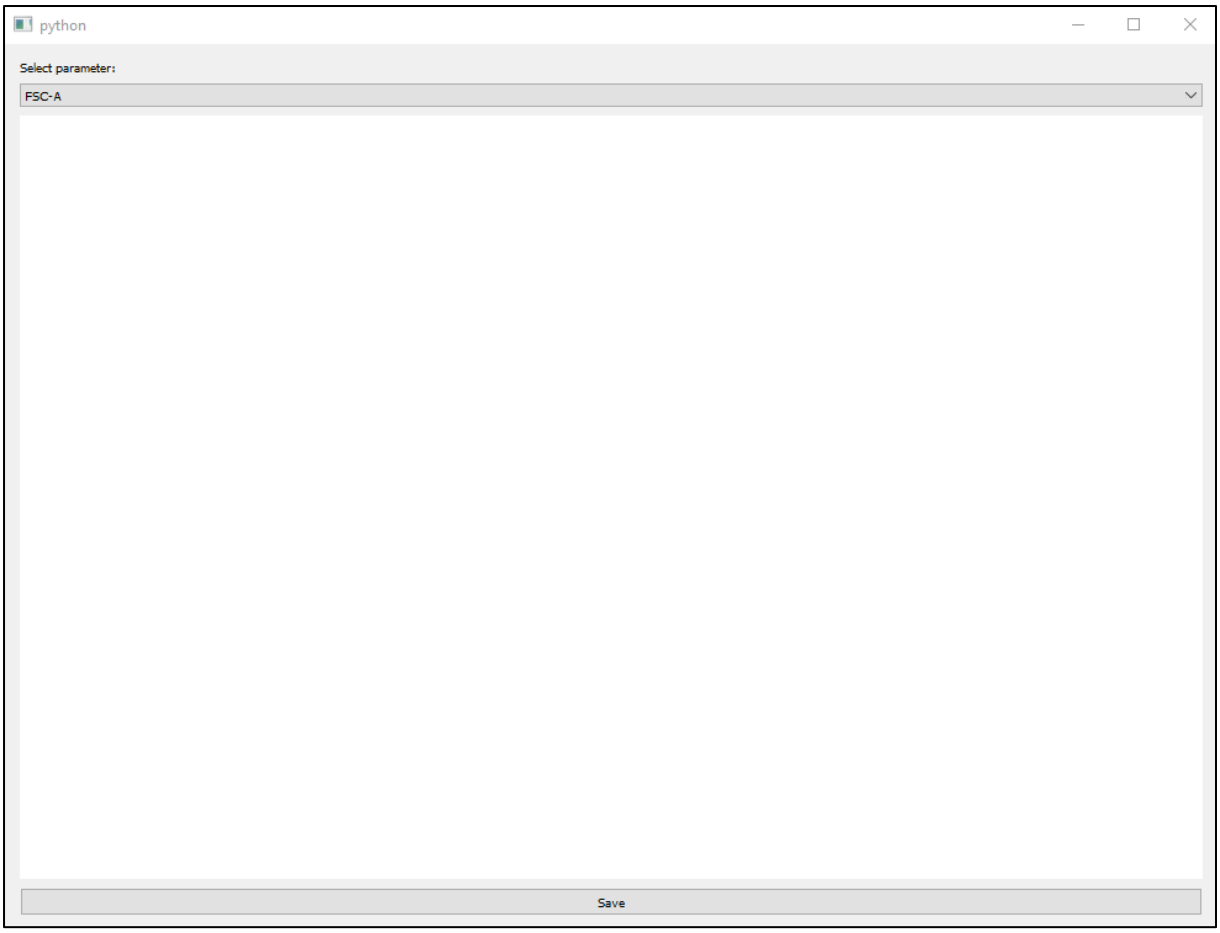
Visualize your t-SNE plot

There are three main functionalities for analysis of your t-SNE plot. The first one is to color your plot according to values, that can be numerical (MFI) or categorical (name of the sample).

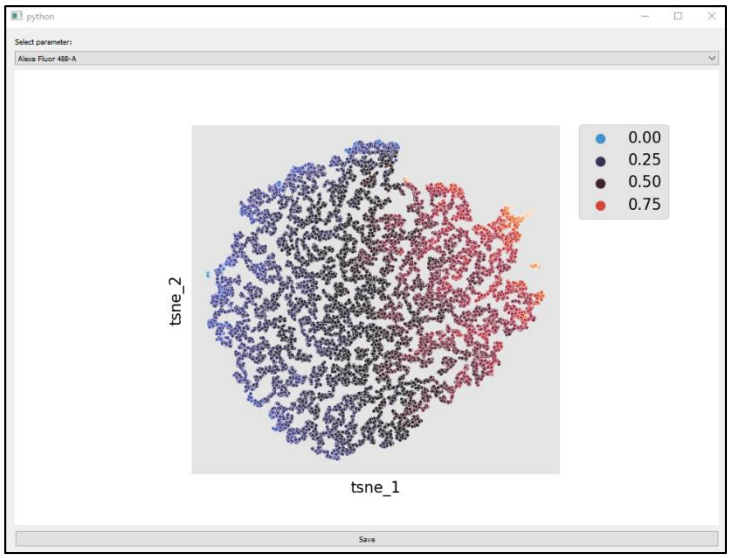
1. Click “Color t-SNE by values”.

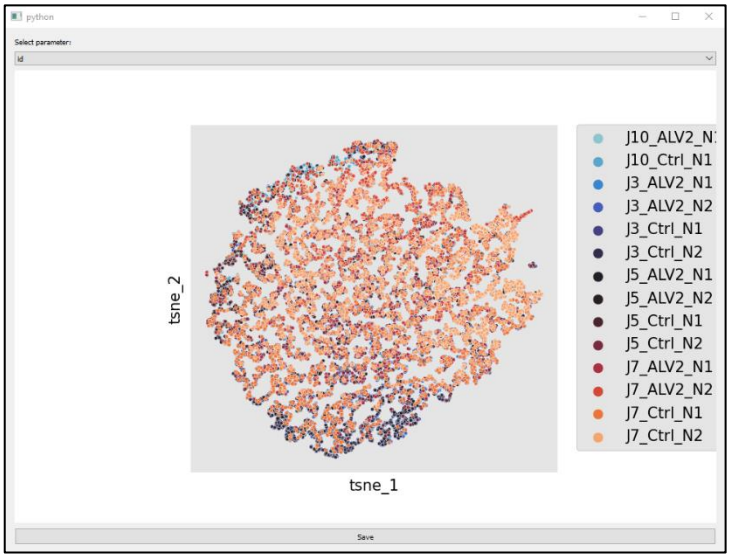


2. You will be transferred to a new empty window



3. From the box on the top select the parameter of interest and click on it



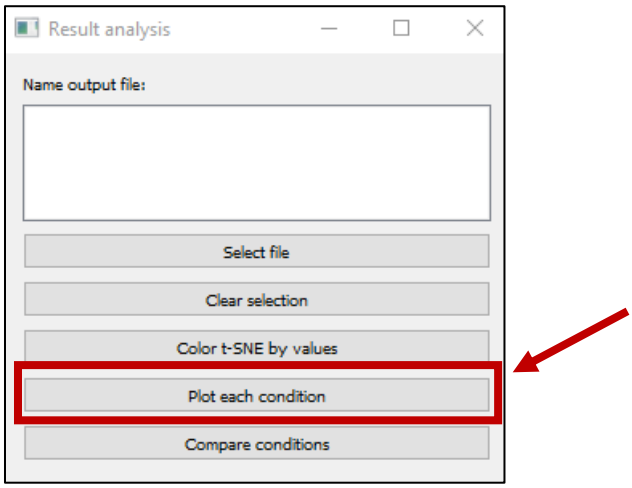


   1) All the numerical values (that were selected for clustering (point 7 ”Create t-SNE map of your samples”) were scaled to be in range between 0 and 1
   2) The categorical values and parameters that were not used in clustering can overlap. The upper layer corresponds to the last value (see second image)
4. You can save the plot using “Save” button on the bottom

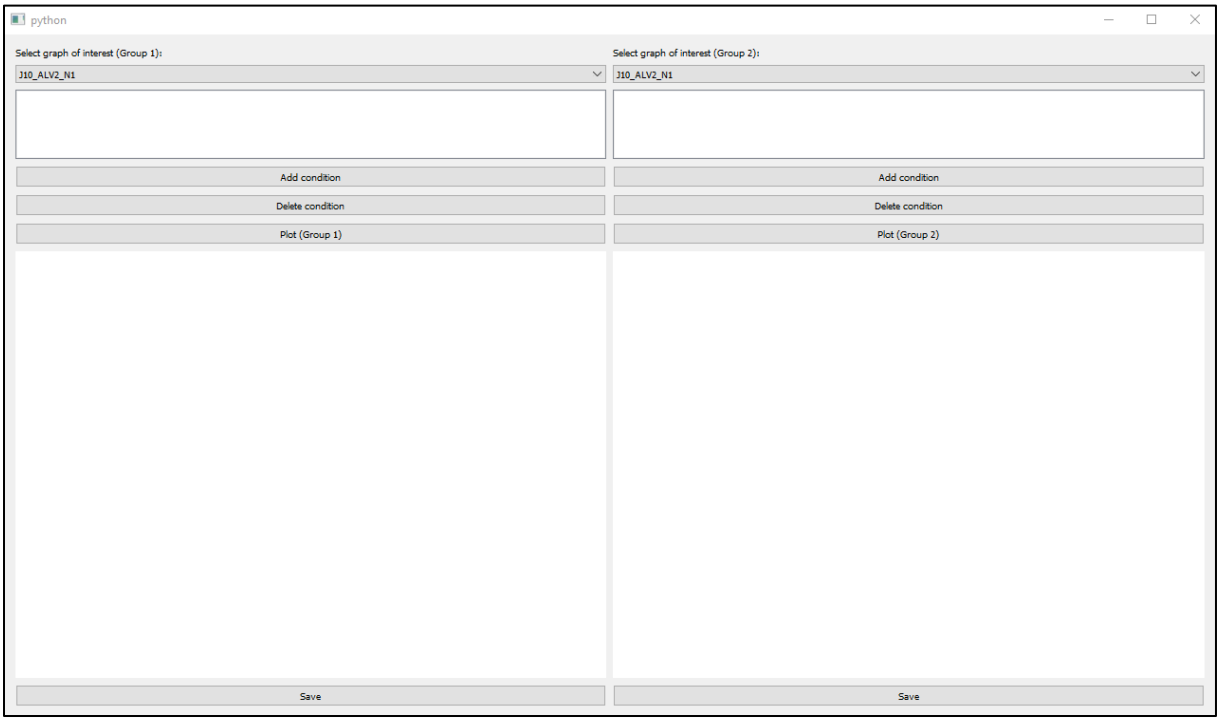
Plot different conditions

You can compare different conditions side by side as well as different timepoints

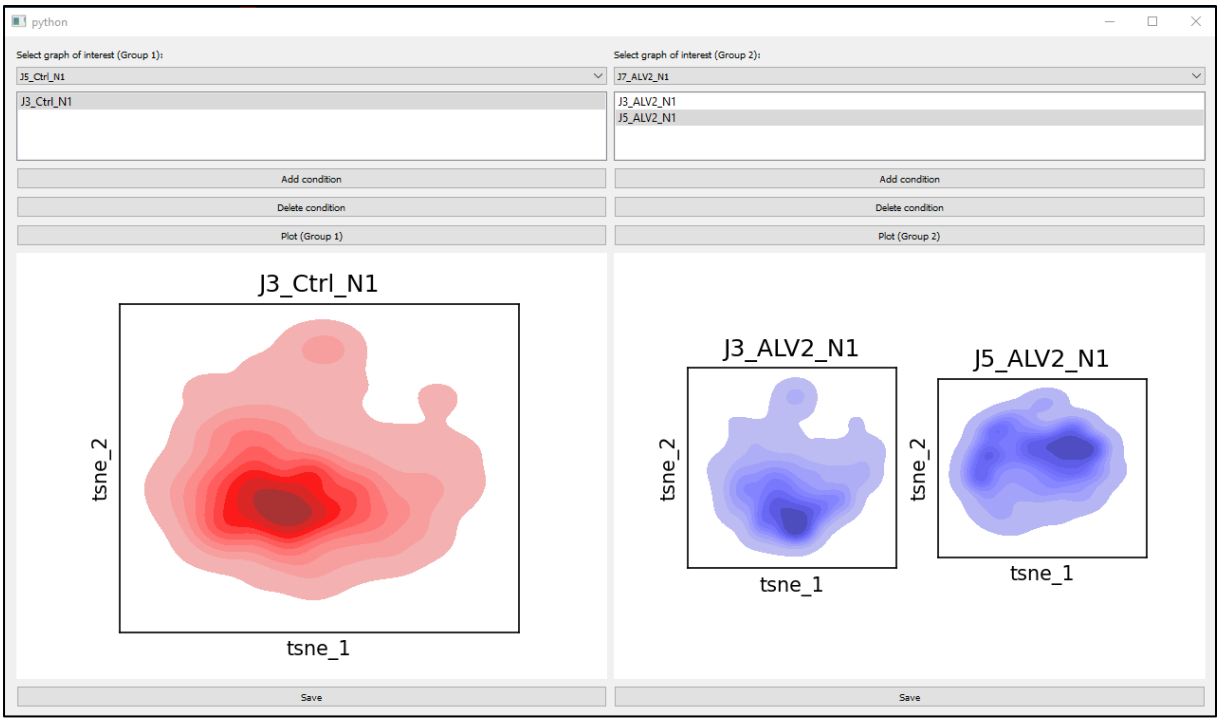
1. Click “Plot each condition”.



2. You will be transferred to new window



3. From the list above you can select a condition of interest for each group and then click “Add condition”.
4. If you made an error, you can select one condition and delete it from the list using “Delete condition”.
5. The number of conditions between the groups can be different, starting from 1 condition per group
6. Plot each condition using “Plot (Group 1 or 2)” button
7. You will get the resulting plots

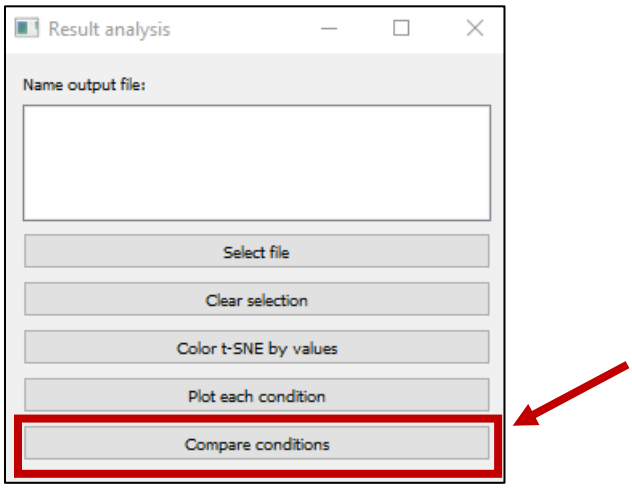


8. They can be saved using “Save” button. It is recommended to save the plots of interest as they will serve you as a guide for the next part.

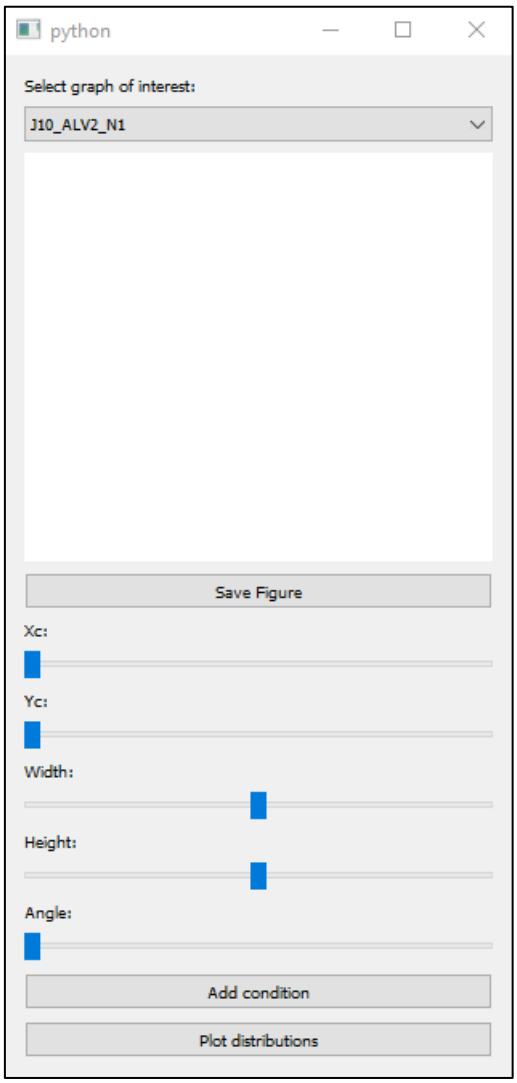
Compare conditions

You can compare different parts of one t-SNE plot or compare the areas between different samples

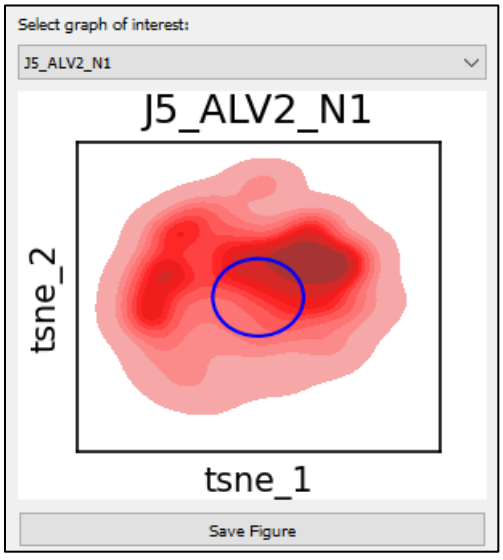
1. Click “Compare conditions”.



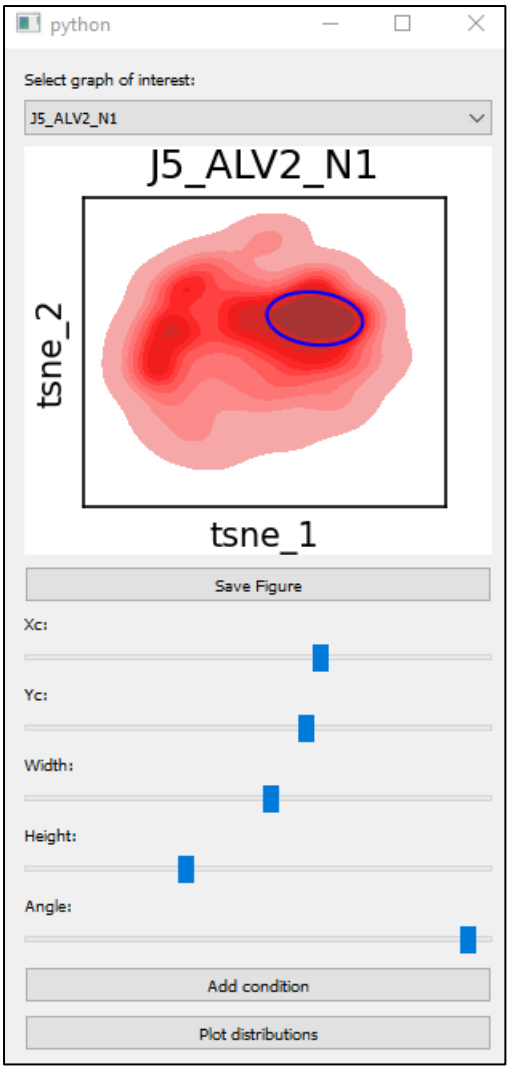
2. You will be transferred to new window



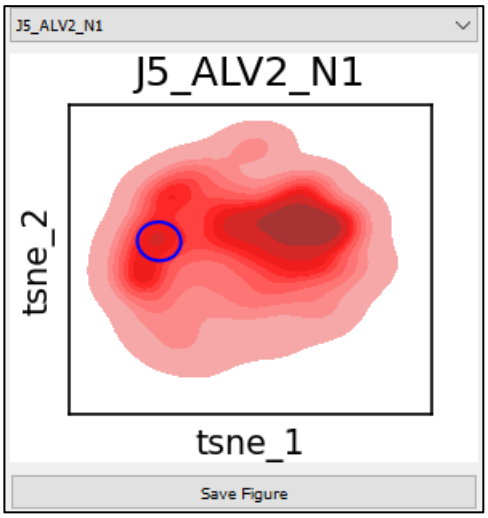
3. Select condition from the box above. You will get something like this

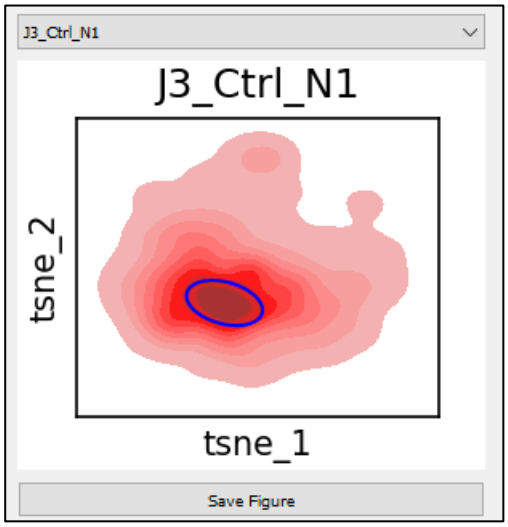


4. Using sliders move and transform the ellipse, so the area of interest appears inside it.

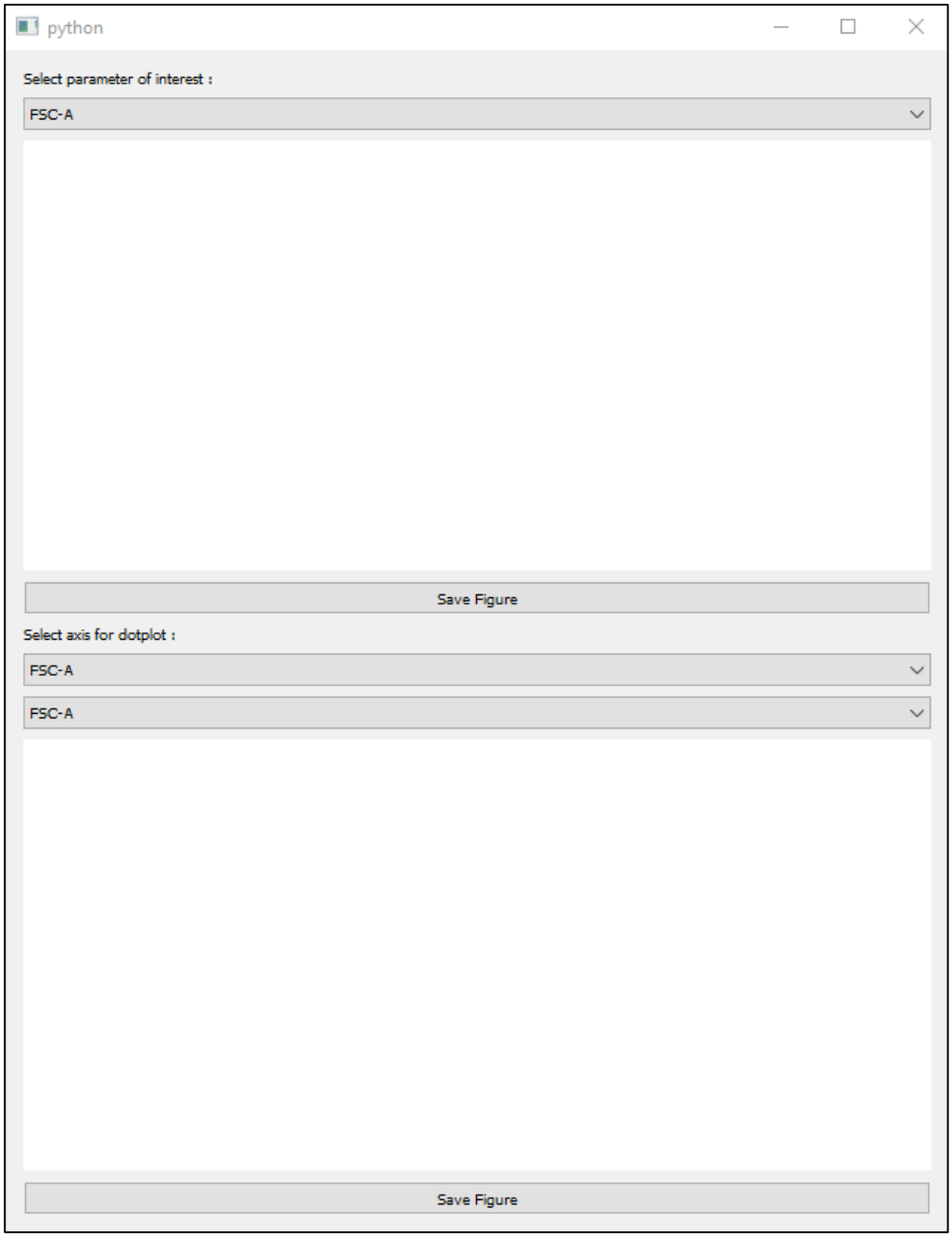


5. You can save the figure with resulting selection using “Save Figure” button.
6. Using “Add condition” add this selection to comparison. You will get small “Success!!!” window that you can close
7. Select another sample from the box above or move the ellipse to select another area on the same plot

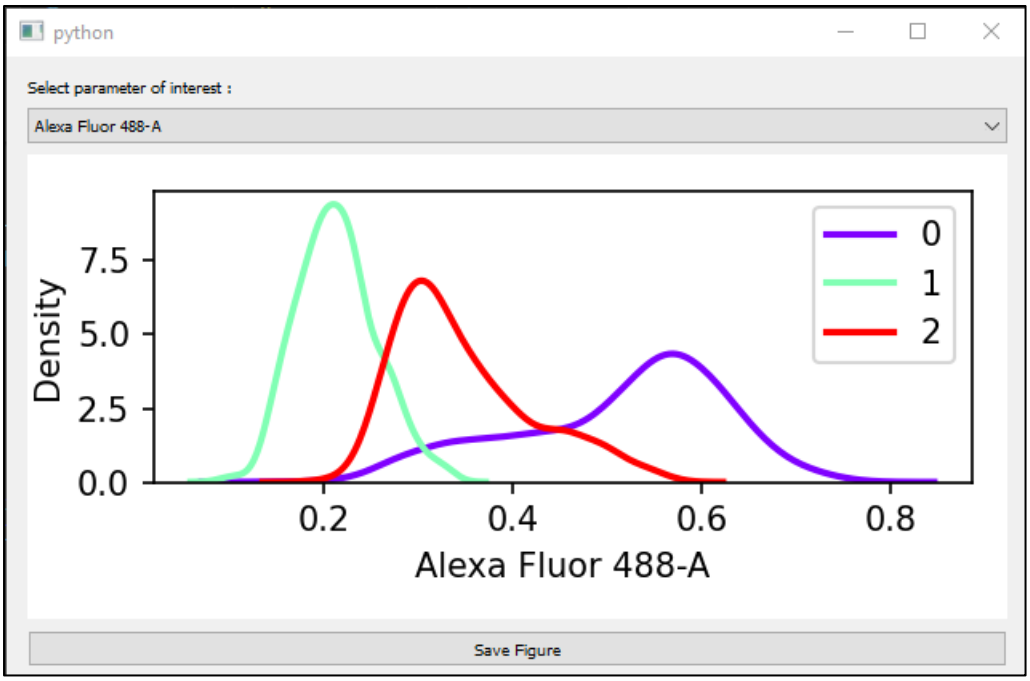




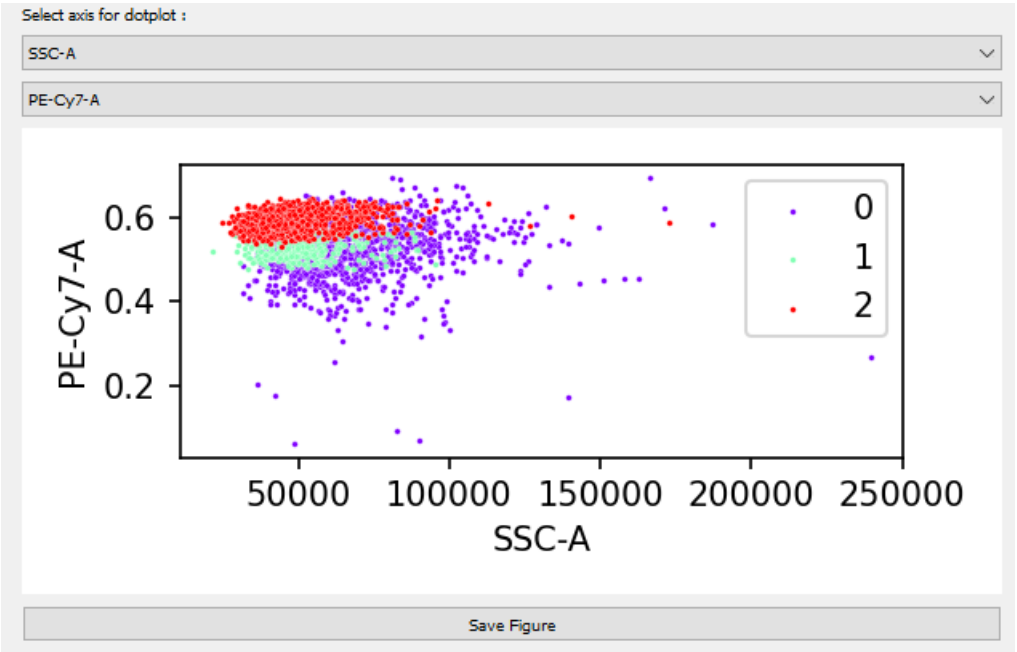
8. Add as many conditions as you want
9. It is recommended to save each selection as a figure, because in the comparison plot they will be referred as “1”, “2”, “3”, etc., according to the sequence of their selection
10. After all selections are made click “Plot distributions”
11. You will be transferred to the following window



12. Using the top box you can select the parameter for plotting the distribution within your areas



13. You can save the figure using “Save Figure” button
14. Remember: The numerical values, that were selected for clustering (point 7 ”Create t-SNE map of your samples”) were scaled to be in range between 0 and 1. The parameters that were not used in clustering were not scaled
15. For the bottom part you can select x-axis (top) and y-axis (bottom) parameters for the scatter plot. It should give an idea of distribution of the populations of interest on classical 2D cytometry dot plot. IN the example you can see that “PE-Cy7-A” was used to perform t-SNE and therefore is scaled, whereas SSC-A not



16. You can save the figure using “Save Figure” button