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.beckman.fr/resources/reading-material/application-notes/cytobank-cytoflex-20-color-panel
- 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.

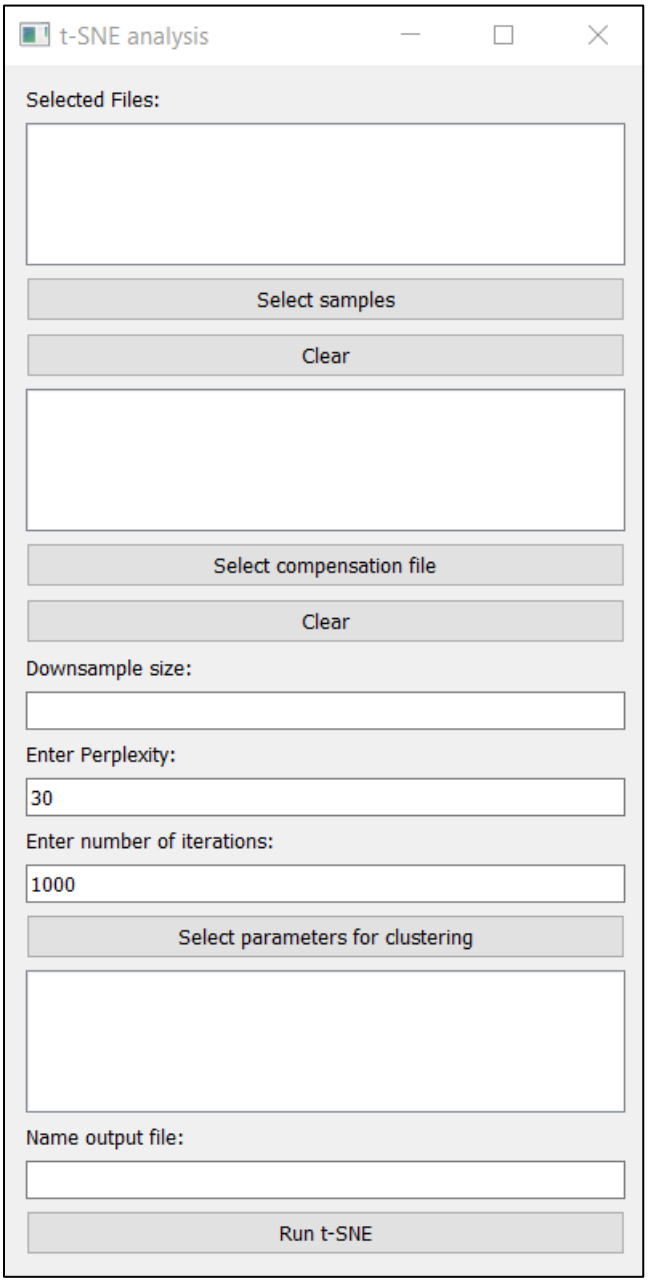
Save .fcs files

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 "FSC-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

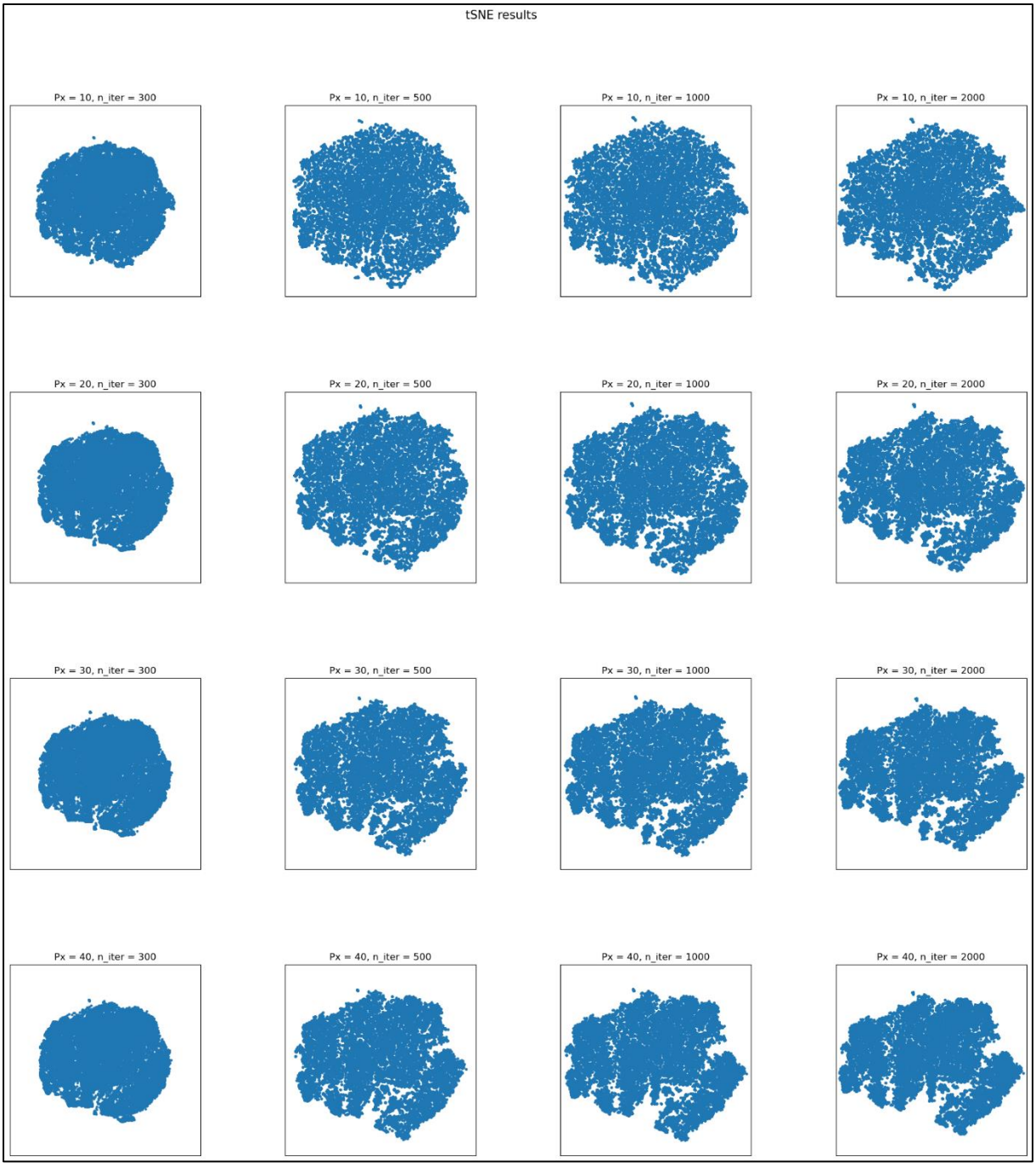
Create t-SNE map of your samples

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

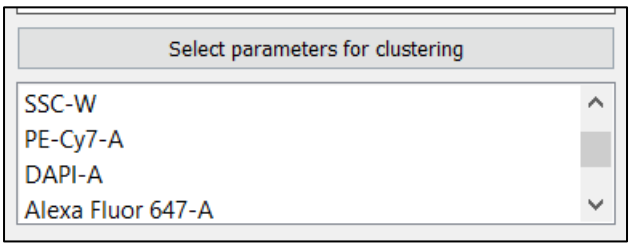


3. Select your samples (.fcs) using “Select samples” button. You can select multiple files or one by one. You can delete all selected samples by clicking “Clear”
4. Select ONE compensation file (point 4 of previous paragraph)
5. Select downsample size. It is the number of events that will be randomly 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.

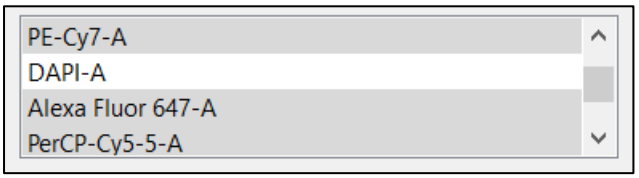
6. The perplexity and iteration values are set to be 30 and 1000 as default, but can be changed if needed. Higher perplexity leads to more separate clusters. Here is an example of their impact on tSNE plot:



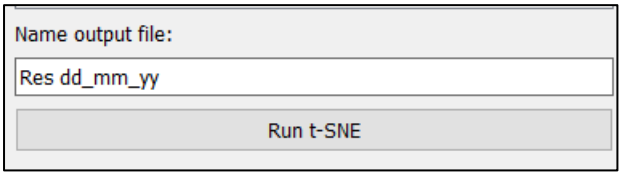
7. Click “Select parameters for clustering”



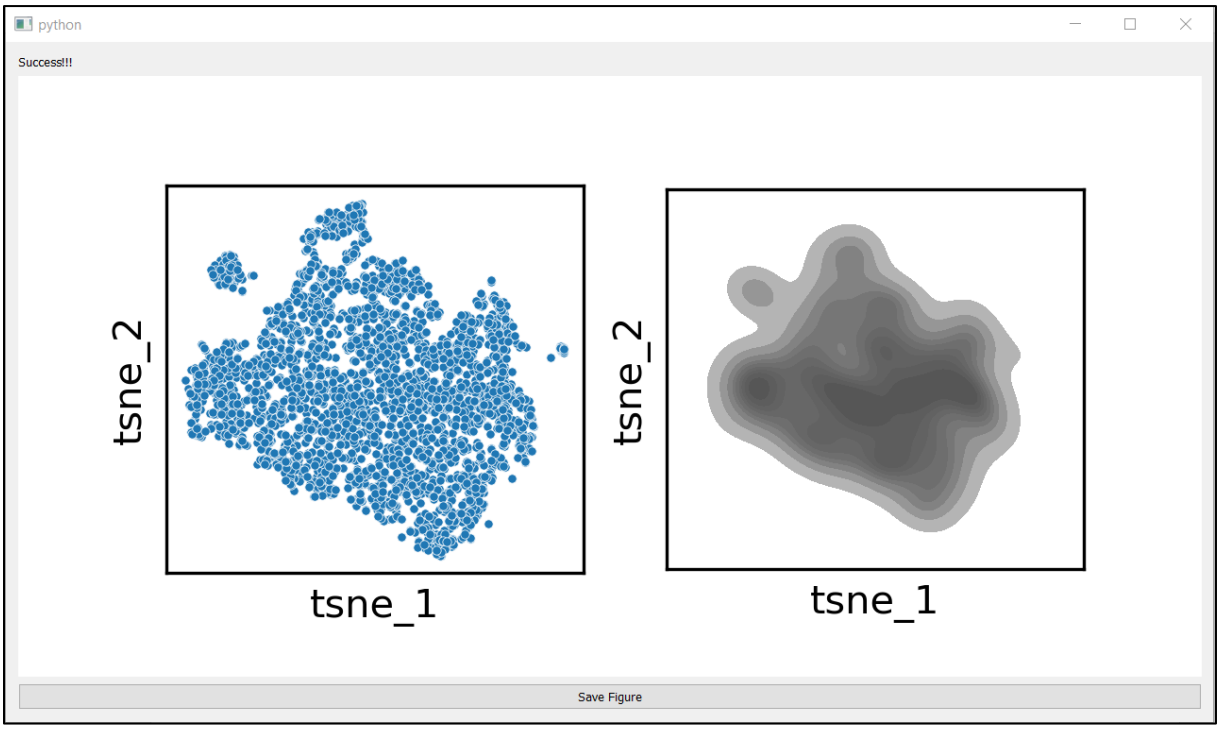
Select parameters that will be used for the tSNE (can use Ctrl or Shift). Usually if it is known that all events have similar values in one channel, this parameter is not used (for example FSC-A for one cell type)



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



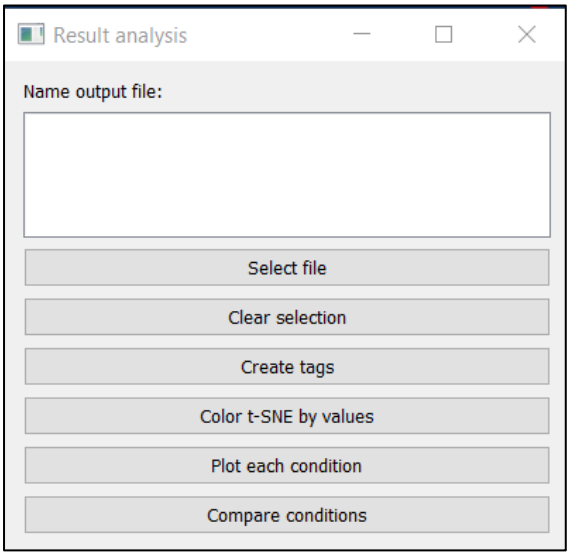
9. After some time you will get the resulting plot



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

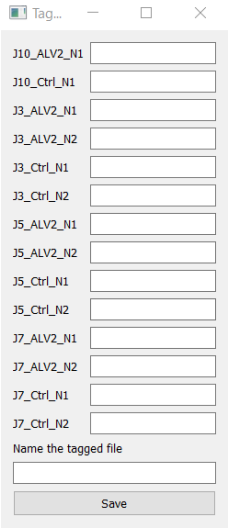
1. Open “Analysis.py”
2. It will open the following window



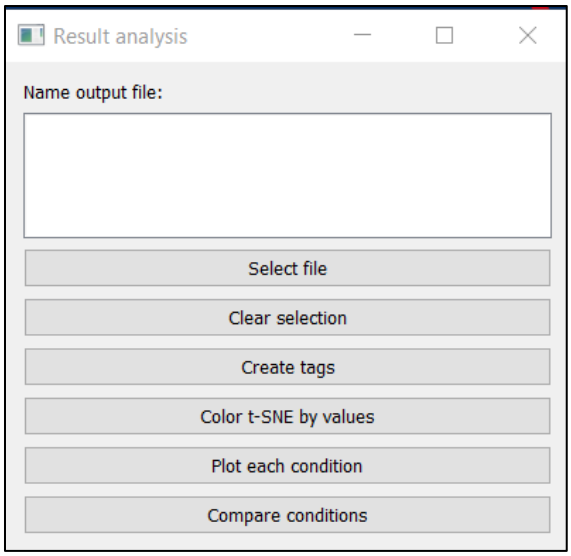
3. Using “Select file” open t-SNE results (“.pkl”), obtained with “Perform t-SNE.py”. You can open only one file. Here the results of “Xn.pkl” are shown. This file can be found in the “test/for analysis” folder.

Create tags

1. If you want to create a tag select the result of interest and push “Create tags” button
2. The following window will open



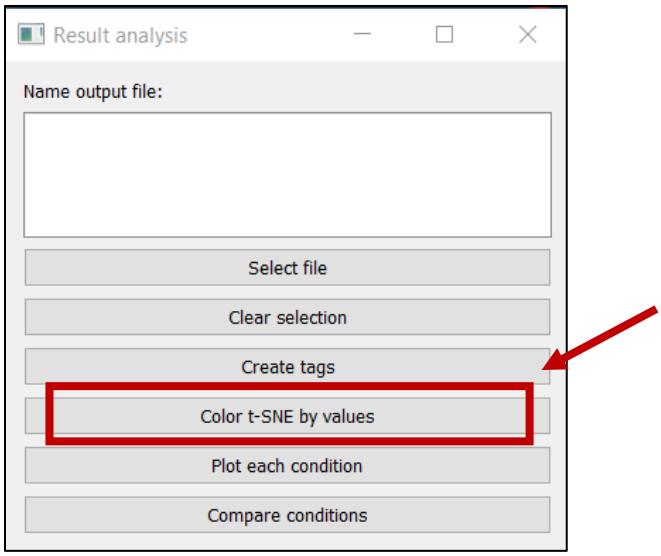
3. Create a tag for each sample, name the new results file and save it. After saving you can close this window
4. Now you can select the tagged file from the previous menu. The file will be in the “Results” folder



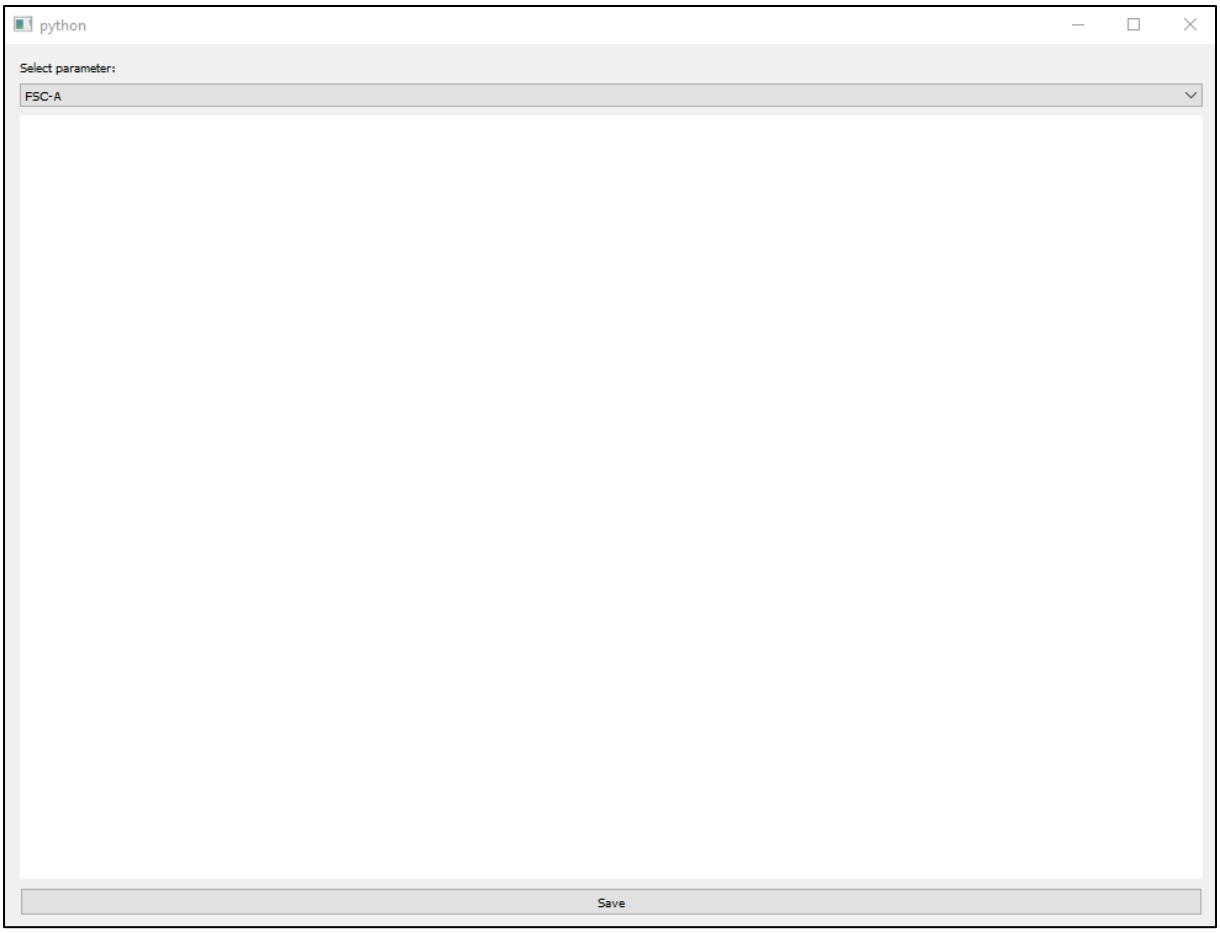
Visualize your t-SNE plot

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).

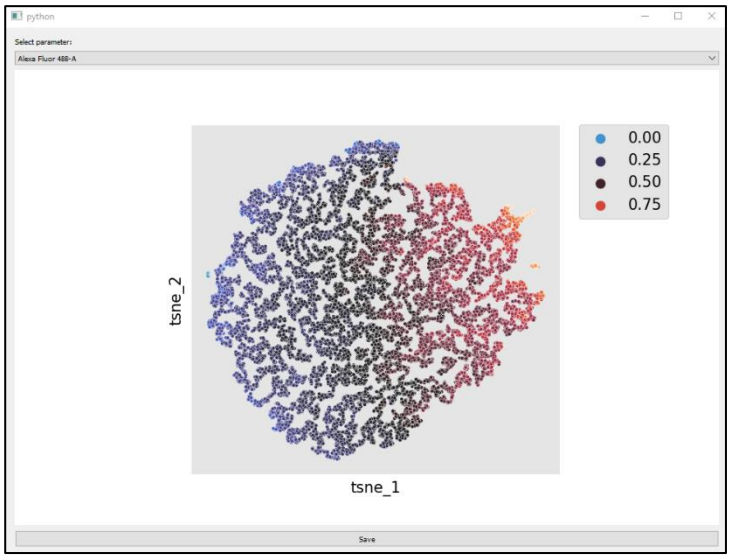
1. Click “Color t-SNE by values”.

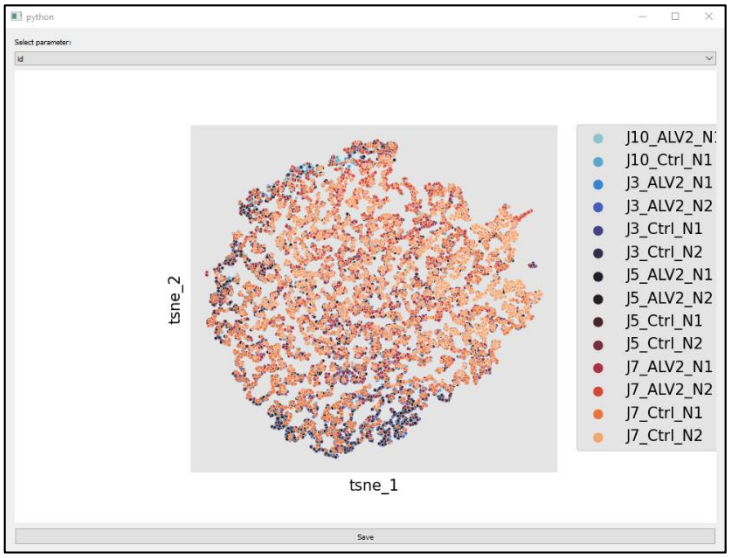


2. You will be transferred to a new empty window



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





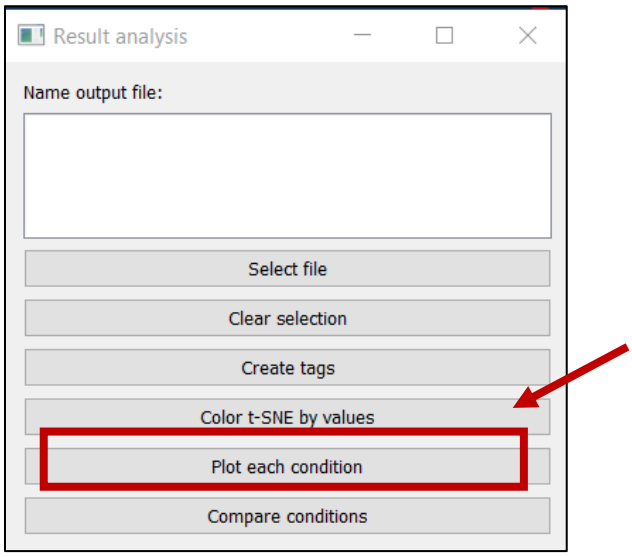
1) All the numerical values that do not have linear distribution (log) were scaled to be in range between 0 and 1
2) The categorical values and parameters that were not used in clustering can overlap. The upper layer corresponds to the last value (see second image)

4. You can save the plot using “Save” button on the bottom

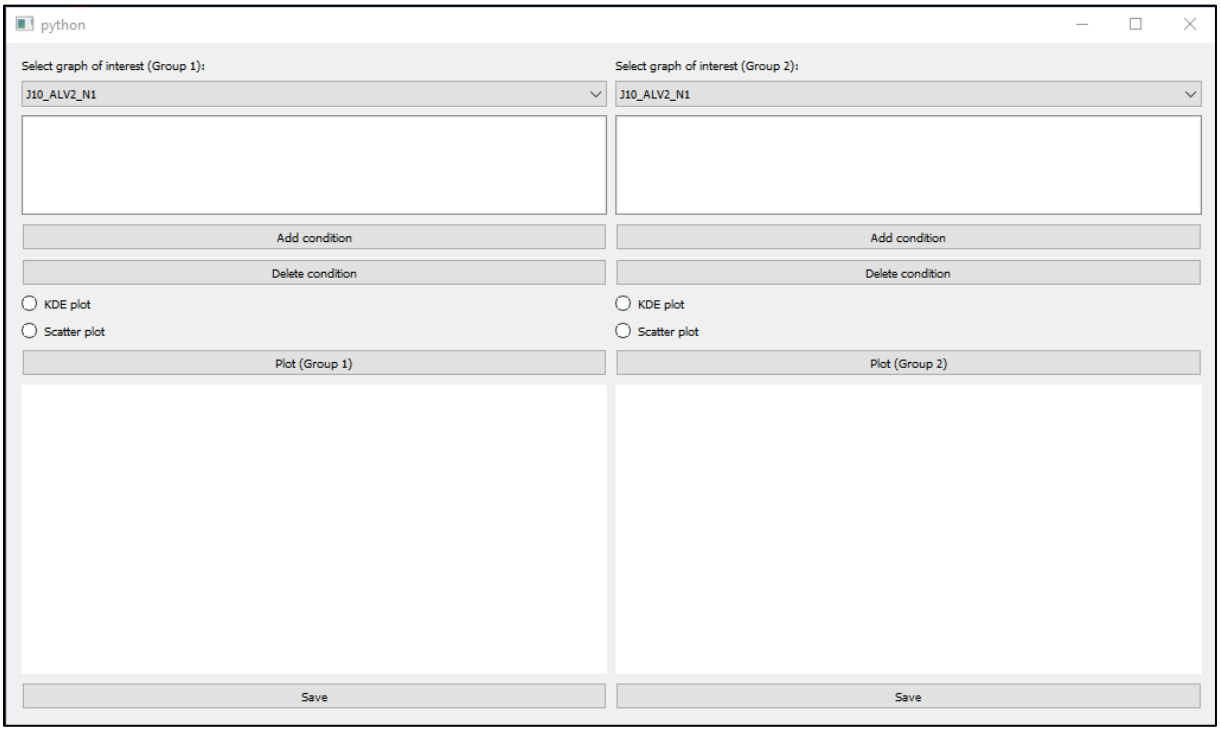
Plot different conditions

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

1. Click “Plot each condition”.

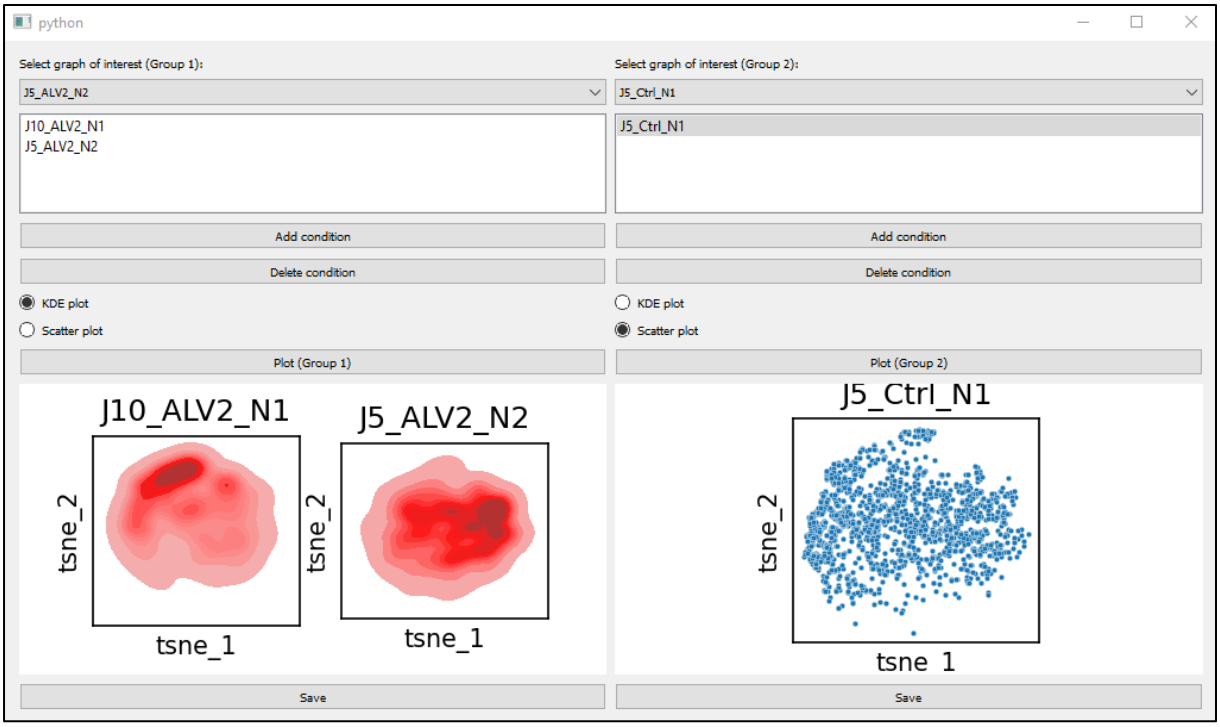


2. You will be transferred to new window



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. Select whether you want scatter plot or density plot
7. Plot each condition using “Plot (Group 1 or 2)” button
8. You will get the resulting plots

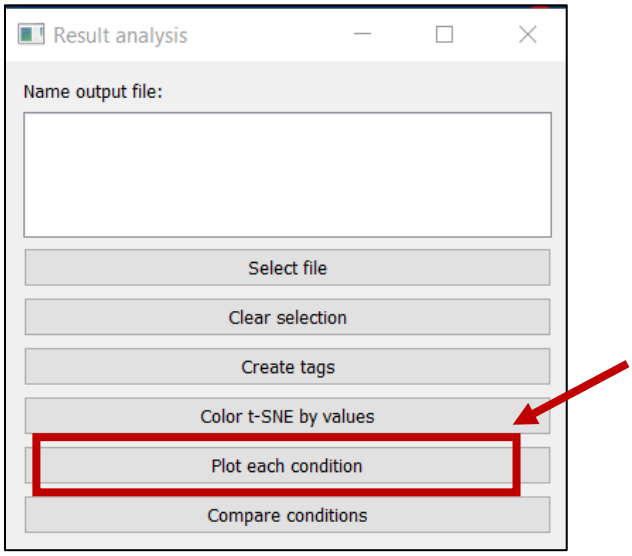


9. They can be saved using “Save” button. It is recommended to save the plots of interest, as they will serve you as a guide for the next part.

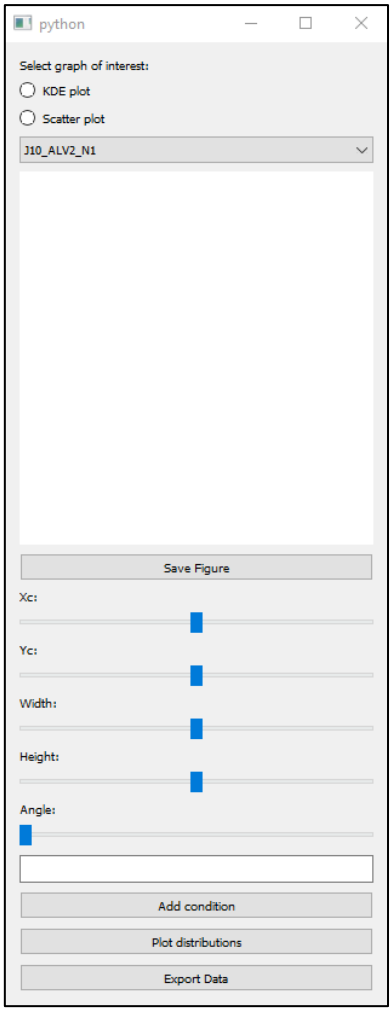
Compare conditions

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

1. Click “Compare conditions”.

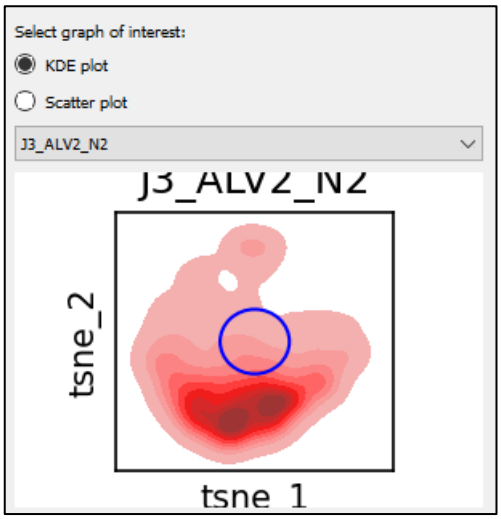


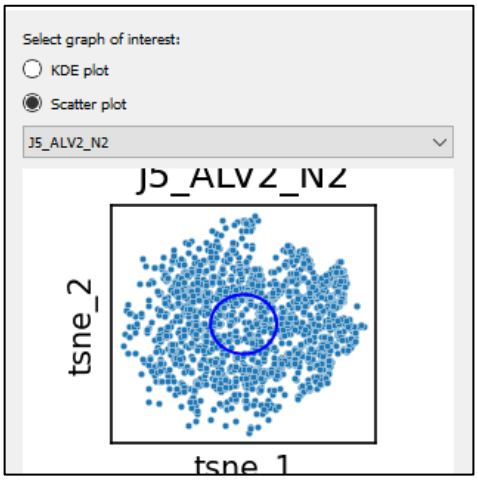
2. You will be transferred to new window



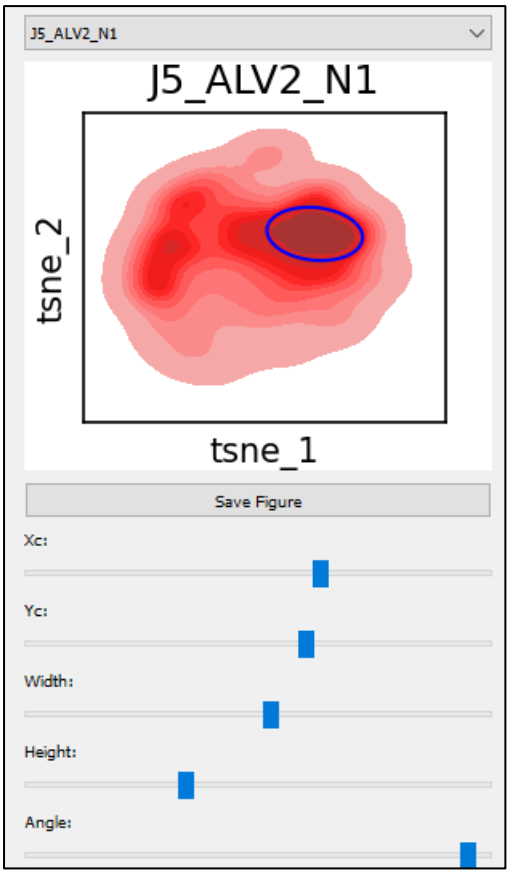
3. Select type of the plot

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



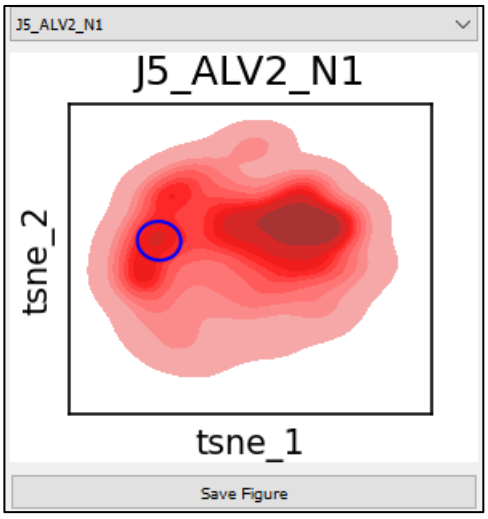


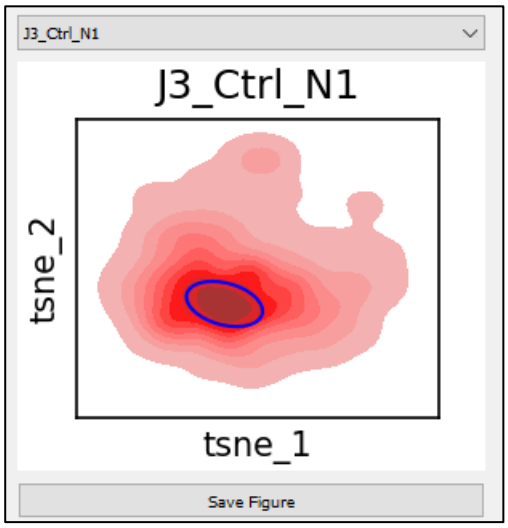
5. Using sliders move and transform the ellipse, so the area of interest appears inside it.



6. You can save the figure with resulting selection using “Save Figure” button.
7. Choose name for your selection and using “Add condition” add this selection to comparison. You will get small “Success!!!” window that you can close

8. Select another sample from the box above or move the ellipse to select another area on the same plot

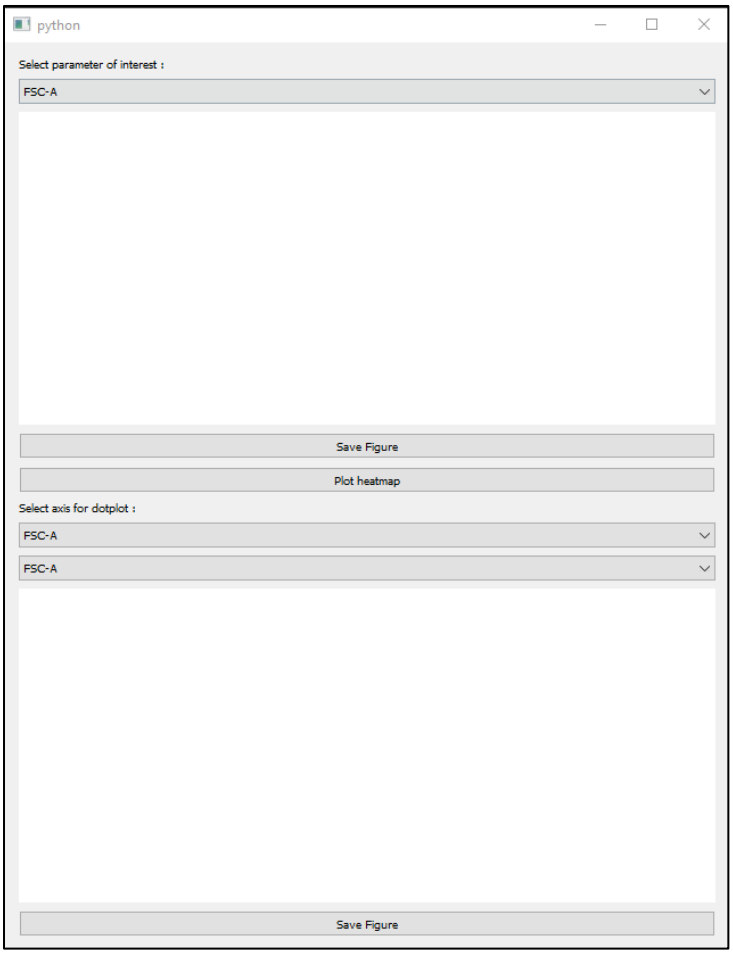




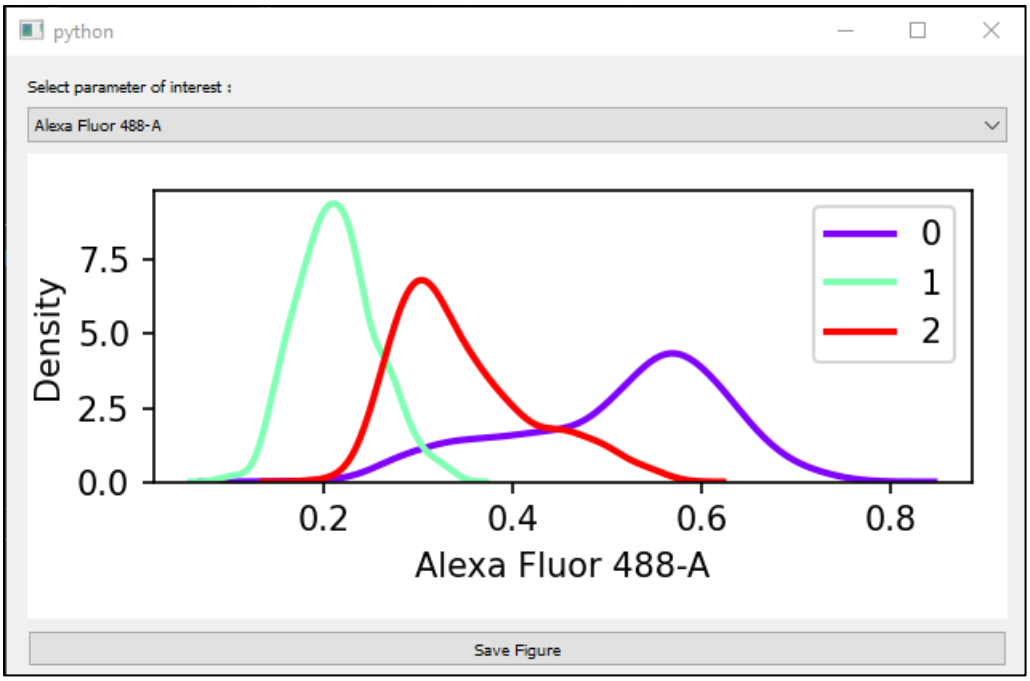
9. Add as many conditions as you want
10. You can use the bottom button to save data for all areas as an excel file. These data is visualized in “Plot distributions”
11. It will look something like this, where the sheet names are the names of the area

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		FSC-A	SSC-A	Alexa Fluo	PE-Cy7-A	Alexa Fluo	id	tsne_1	tsne_2	p							
2		95200,17	43003	0,289917	0,737286	0,253203	J3_Ctrl_N2	13,42197	-101,106	VRAI							
3		84176,11	35113	0,223936	0,706006	0,042277	J3_Ctrl_N2	-48,4308	-100,195	VRAI							
4		80656,08	40928	0,2279	0,663276	0,154012	J3_Ctrl_N2	-55,6691	-67,5163	VRAI							
5		100108	47030	0,186805	0,622323	0,057096	J3_Ctrl_N2	-89,4149	-57,1766	VRAI							
6		157441	66880	0,291893	0,726002	0,1718	J3_Ctrl_N2	-12,0423	-98,4133	VRAI							
7		75208,79	37247	0,141141	0,690079	0,153145	J3_Ctrl_N2	-63,0169	-90,2185	VRAI							
8		107036,8	39979	0,291893	0,667455	0,100231	J3_Ctrl_N2	-34,8902	-74,732	VRAI							
9		100349,5	50847	0,226326	0,695093	0,216292	J3_Ctrl_N2	-21,4162	-108,224	VRAI							
10		118170,4	41800	0,231771	0,665031	0,077835	J3_Ctrl_N2	-77,832	-60,8123	VRAI							
11		70631,34	31126	0,194409	0,672685	0,148784	J3_Ctrl_N2	-58,9342	-80,3938	VRAI							
12		104123,5	77232	0,276783	0,69784	0,145269	J3_Ctrl_N2	-17,7532	-93,0744	VRAI							
13		78890,67	35189	0,209739	0,694807	0,119275	J3_Ctrl_N2	-49,9289	-92,2519	VRAI							
14		76116,8	37674	0,234049	0,626666	0,187856	J3_Ctrl_N2	-53,5454	-51,4092	VRAI							
15		96748,12	40915	0,326379	0,711571	0,180747	J3_Ctrl_N2	10,51305	-91,5251	VRAI							
16		169738,3	82623	0,512447	0,767869	0,261024	J3_Ctrl_N2	33,3094	-82,0654	VRAI							
17		118232,7	34272	0,261499	0,622585	0,155742	J3_Ctrl_N2	-62,8065	-49,7936	VRAI							
18		95816,03	42765	0,332027	0,670927	0,122896	J3_Ctrl_N2	-5,63397	-58,9091	VRAI							
19		106880,8	41145	0,268186	0,657995	0,043081	J3_Ctrl_N2	-38,3338	-79,2022	VRAI							
20		63671,79	33409	0,199036	0,618562	0,127412	J3_Ctrl_N2	-89,3715	-42,895	VRAI							
21		140215,2	56950	0,24774	0,636437	0,148784	J3_Ctrl_N2	-62,078	-55,0277	VRAI							
22		110170	36206	0,263356	0,672714	0,24829	J3_Ctrl_N2	-12,8223	-73,5334	VRAI							
23		98813,98	54059	0,314539	0,6843	0,142618	J3_Ctrl_N2	1,559662	-86,7786	VRAI							
24		98576,61	46778	0,237036	0,678227	0,165126	J3_Ctrl_N2	-27,7332	-83,318	VRAI							
25		113998,8	40949	0,199036	0,709705	0,16848	J3_Ctrl_N2	-27,4826	-107,139	VRAI							
26		136665,3	53802	0,223132	0,714941	0,076953	J3_Ctrl_N2	-43,7198	-97,9641	VRAI							
27		93991,69	30311	0,437534	0,627213	0,093909	J3_Ctrl_N2	7,023098	-49,0545	VRAI							

12. After all selections are made click “Plot distributions”
13. You will be transferred to the following window

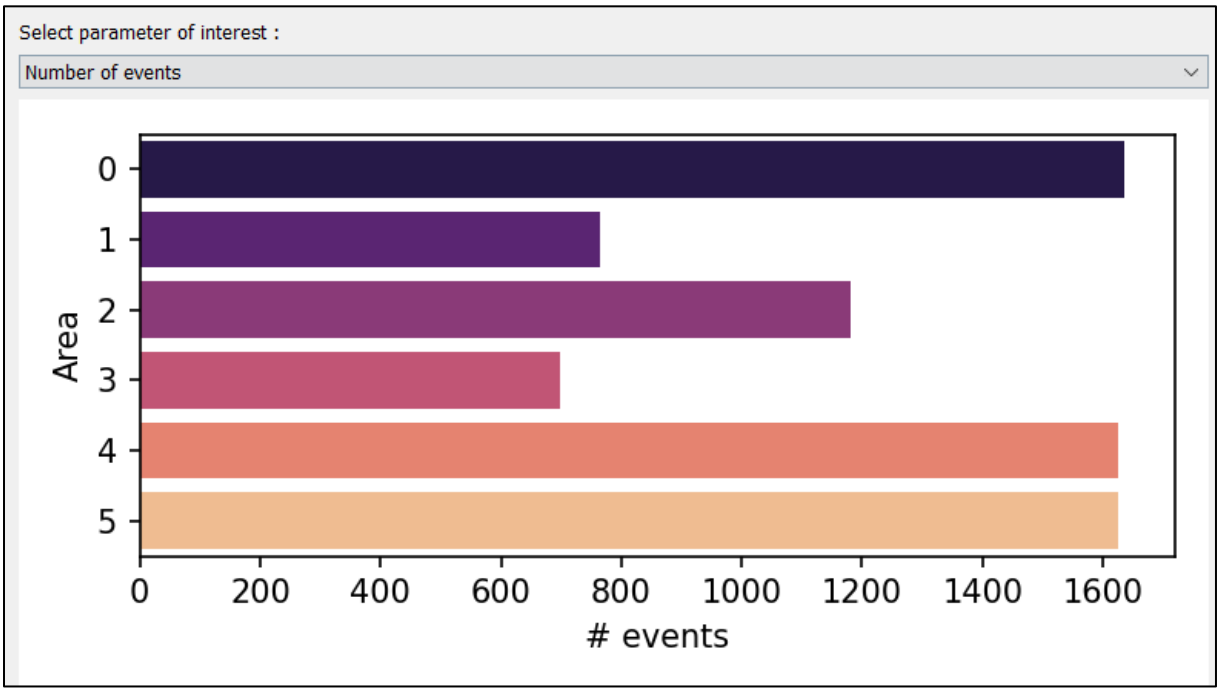


14. Using the top box you can select the parameter for plotting the distribution within your areas

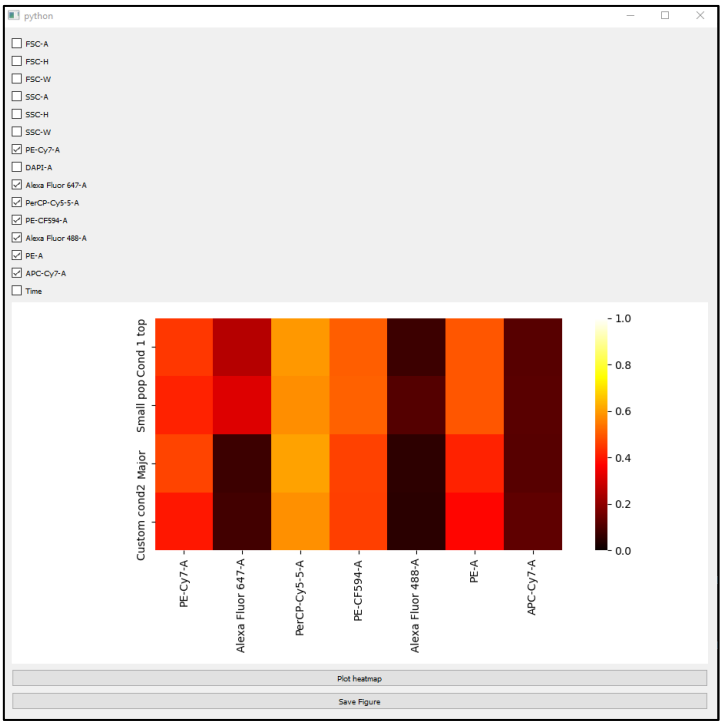


15. You can save the figure using “Save Figure” button

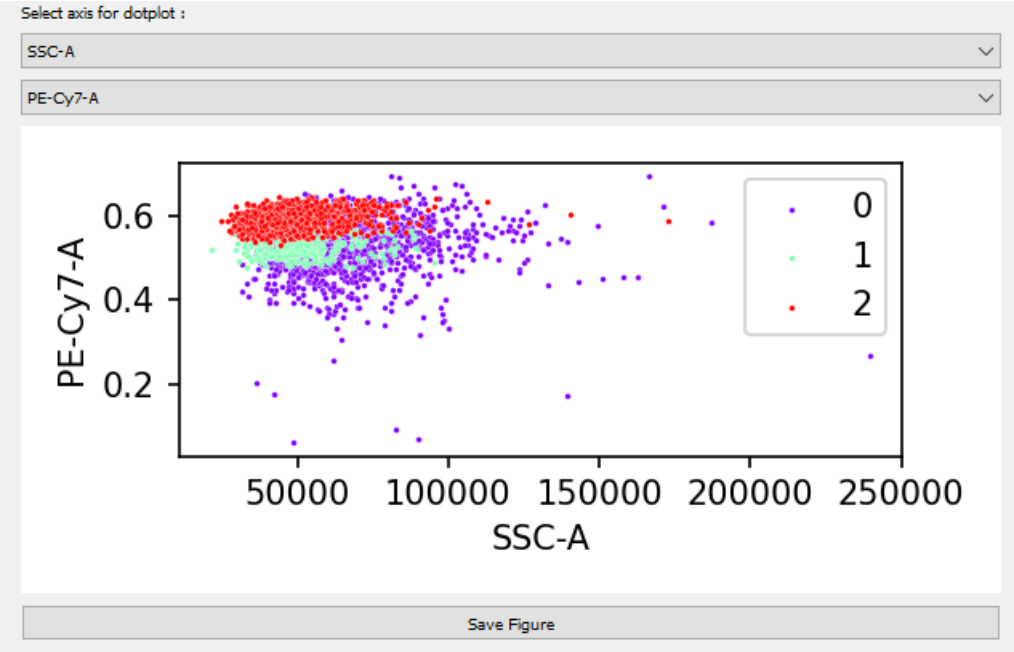
16. You can also plot the number of events in each area



17. Remember: The numerical values, that do not have linear distribution (log) were scaled to be in range between 0 and 1. The parameters that are linear were not scaled
18. For scaled data (usually all channels, excluding FSC, SSC and Time), a heatmap can be plotted with “Plot heatmap” button. This will transfer you to new window. This is done so you can scale it and save in the most appropriate way. The graph depicts median values across all areas and channels.



19. 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” is scaled, whereas SSC-A not



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