

Running spatial_transcriptomics.ipynb for StLearn normalization analysis and comparison against clinically derived ground truth data

The program spatial_transcriptomics.ipynb is to be run within a Jupyter notebook python environment. To begin using the program, the working directory should be set to the desired location for the output plot comparing normalized and non-normalized data. You should also change the folder path in module labeled "Specify folder that all sample sub folders are listed under" to the folder containing your sample folders. By default, the program is set to analyze all 12 tissue samples from the human DLPFC data set, if you comment out the code segment labeled "Run sample analysis on all data sets:" and uncomment the module labeled "Test running on sample from stLearn paper alone:", the data analysis can be ran on the example file for 151673 alone. Once these changes have been made, all code segments are to be run for proper functioning of the program.

The segments running the stlearn clustering with and without normalization will take approximately 6 minutes per sample each. The rest of the program takes approximately 5 minutes.

There should be an output for each sample of the ground truth images, the normalized clustering results, and the non-normalized clustering results. Additionally, a plot comparing normalized and non-normalized clustering statistics will be produced. It should be noted that the plot produces colored in box and whisker plots, that will be an uncolored line if only one sample is input. Additional note should be made that the program will place a folder labeled "tmp" in the computer base directory containing image tiles for the spatial analysis.