Readme for main R codes in folder “Methods’ codes in R”

The scripts in this folder are somewhat important and standard to what we already generated now, such as tables, lists, and pictures. For most of the part, I have made these main codes encompassing, expanding, or reducing in a more convenient way to use. You can also excerpt part of them to make your own use.

* “Kmeans Pipe”:

My stardard script to complete the post procedures of k means clustering, including heat map images, gene sets FDR calculation, and cluster inheritance tree. This is the second code that I processed afterward part of K means analysis.

Please pay attention to the directory setting to change either mean or zero format of the input folders.

* “NegSilhoutte Score Pipe Case Template”:

My standard script to complete the silhouette calculation and plotting, including silhouette value for each selected GO terms from the viewpoint of patient identity, their corresponding silhouette and tsne plots for candidate GO gene sets, and the significance of the NSV score that helps us discern what GO terms makes cells more separable or intermingle.

* RejSamping and Heterogeneity Scoring:

My standard script to complete the heterogeneity scoring calculation using rejection sampling method. Some cases may encounter “the standard deviation is zero”. Please ignore, it won’t affect the overall evaluation OR it will return NA that indicate some cluster are unable to calculate.

* “ServerFilesPipe\_R\_sh\_bat”:

This script is for the big data case to run ISS cluster. Since the computation of silhouette p value is time-consuming (for the case having more than 1000 single cells), I split the GO gene sets which waited to be evaluated into different R files. After I chopped it into several similar copied R files, I was able to upload these files to run ISS cluster. The instruction of this script is clear inside of it. The first part is to modify the waiting GO terms length and write out as many as you chop into the length you think it is proper. The second part is to correspondingly generate sh files for each R file you wrote out previously. The third part is to create a bat file mimic executing order for linux server to run multiple jobs with only just one submission. Detailed steps are indicated inside of this script.

* “SinCell9cases\_tsne\_Pipe”:

This script aims for producing overall pictures of how the single cells distribution and spreading looks like by showing in tsne plotting in different labeling, such as patient identity and cell types. It already includes normal cells only plotting, cancer cells plotting, and mixture of normal and cancer cells plotting.