# Lab3

Please download the three scRNA-seq gene expression matrix and corresponding cell cluster labels from iSpace.
For the given dataset, please finish the following task and submit your report in pdf.

1.  Data preprocessing, including filtering and normalization
2.  Imputation( refers to r package scImpute https://github.com/Vivianstats/scImpute or python package MAGIC https://github.com/KrishnaswamyLab/MAGIC)
3.  t-SNE plot to show the cells in 2D space (before and after imputation)
4.  Dimension reduction (PCA or UMAP)
5.  After dimension reduction, apply Kmeans and hierarchical clustering to do cell clustering (use Euclidean distance)
6.  Compare your clustering result with the given cluster labels by calculate the ARI of clustering.
7.  Draw barplot to compare the ARI of different clustering method across different datasets.

8.  Do you think which distance measurement or clustering method can better clustering the cells? Please use experiment to show your idea. [Optional]