

Deep Learning for Single Cell Analysis

Single cell analysis dedicates to apply data driven methods to cluster and classify cells using unsupervised learning in order to help scientist to discover new cell types. The project suggests using autoencoder as deep learning approach for single cell analysis and builds a user interface for single cell analysis uses.

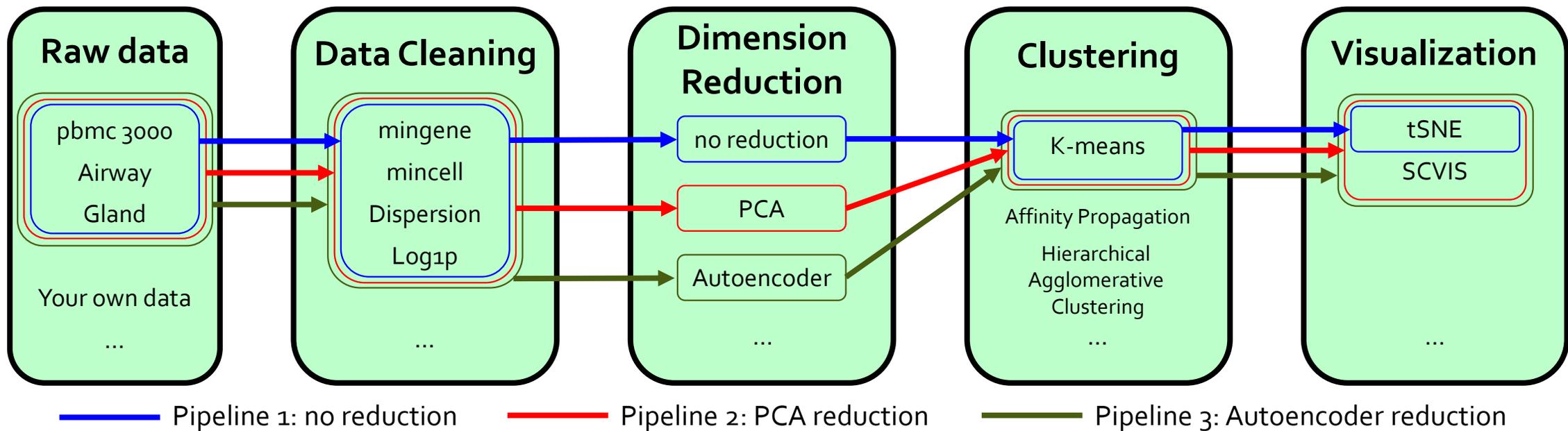
Group Members

Ziyang Ding , Undergraduate,
Ziyang.ding@duke.edu

Chaofan Tao, Undergraduate,
chaofan.tao@duke.edu

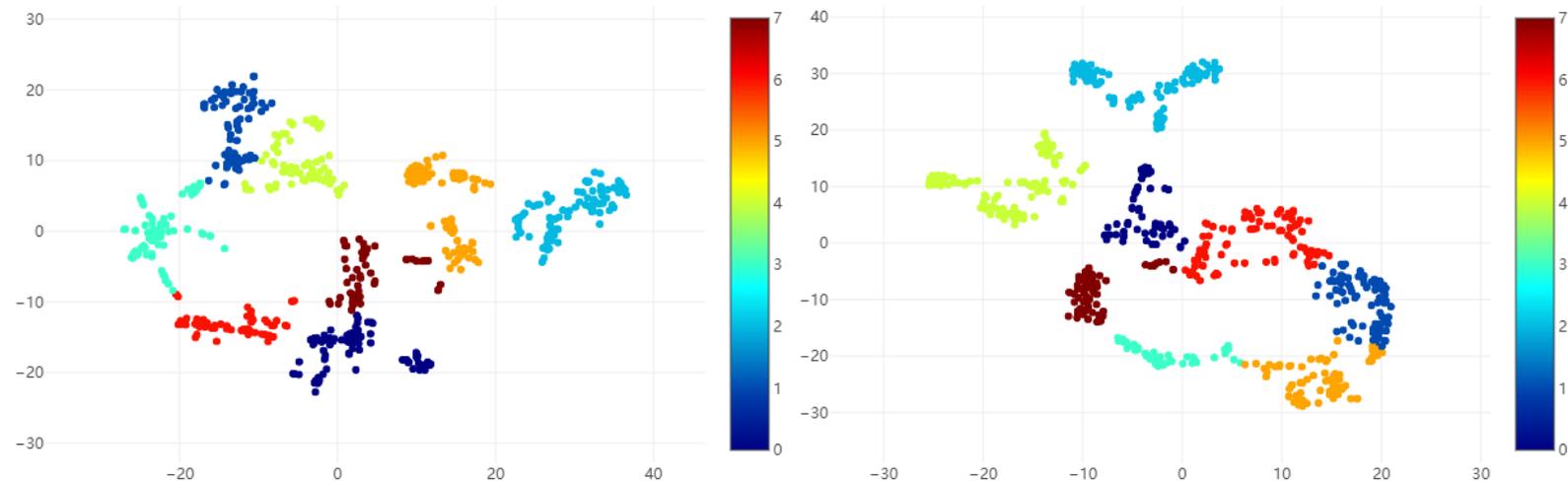
Kuei Yue Ko, project manager,
kuei.yueh.ko@duke.edu

Cliburn Chan, project lead,
cliburn.chan@duke.edu



Pipelines

We suggest 3 pipelines for single cell analysis, among which differs in dimension reduction methods. The PCA and Autoencoder did well in preserving local structures.

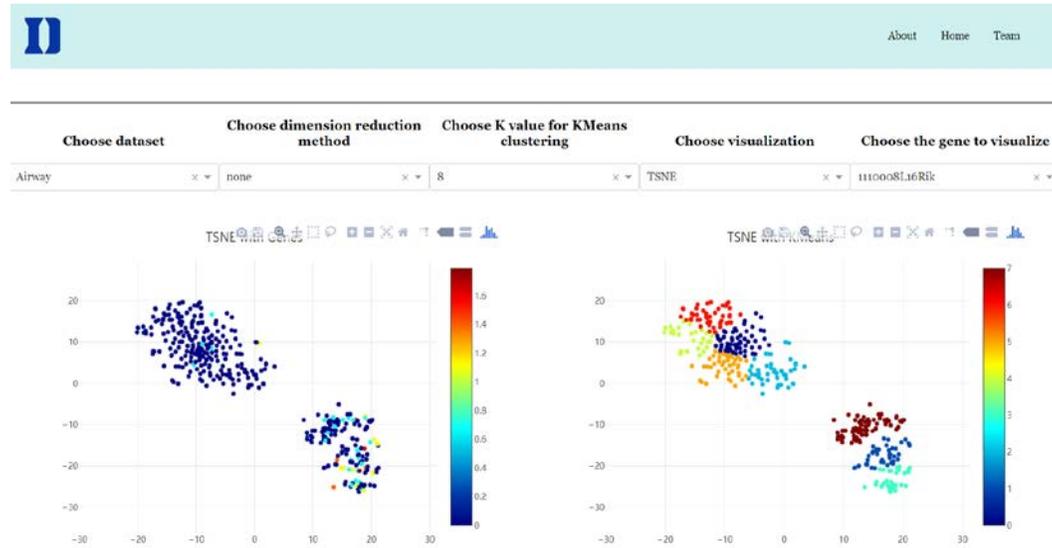


Autoencoder (left) and PCA (right) for Gland

User Interface

We built a user interface(web app) for biologist to visualize their own single-cell sequencing data. The user interface consists of three pages: Home, About and Home.

We also allow users to upload dataset as well as using preloaded dataset, change dimension reduction methods, change K-value for clustering, and change visualization method. The mean UMI value for each cluster is also demonstrated as a table, which can be automatically sorted by user's choice. This helps biologists to identify the principle component gene of each cluster, so that the new cell type can be more easily identified and understood.



The data pass through the entire 3 pipelines and will be visualized on the UI panel. The left panel shows UMI count for in each cells for a specified gene on the upper right; the right panel shows clustering color mask for the dataset with K specified above.

Genes	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7	Cluster8
1110008L16Rik	0.388	0.012	0.116	0.025	0.013	0.053	0.29	0
1110017D15Rik	0.073	0	0.013	0	0.013	0.143	0.02	0
1110038E12Rik	0.376	0.099	1.065	0.05	0.038	0.074	1.438	0.269
1600014C10Rik	0.277	0	0.418	0.05	0	0.015	0.218	0.036
1700003E16Rik	0.039	0	0	0.025	0	0.029	0.02	0
1700007K13Rik	0.075	0	0.039	0	0.02	0.088	0.059	0.018
1700016K19Rik	0.069	0	0.013	0	0	0.106	0.04	0
1700023F06Rik	0.013	0	0.039	0	0	0	0.091	0
1700088E04Rik	0.068	0	0.013	0.025	0	0.015	0.119	0
1810010H24Rik	0.404	0	0.097	0	0.05	0	0.031	0.036

Upload your own dataset. Please wait for about 30 minutes for program to finish computing.

Please upload a csv file

The rows should be observations/cells, and the columns be genes.

Drag and Drop or Select Files