NEVADA NASA SPACE GRANT CONSORTIUM

SINGLE-CELL ANALYSIS USING HIERARCHICAL AUTOENCODER

UNIVERSITY OF NEVADA, RENO

Duc Tran, Hung Nguyen, Bang Tran, and Tin Nguyen*

Department of Computer Science and Engineering, University of Nevada, Reno

Contact: tinn@unr.edu, Website: https://bioinformatics.cse.unr.edu/

05/08/2020

BACKGROUND

Defining cell types through unsupervised learning is considered one of the mainstay for single-cell RNA sequencing (scRNA-seq) analysis. However, the ever-increasing number of cells, technical noise, and high dropout rate [1] pose significant computational challenges in scRNA-seq analysis [2].

OBJECTIVES

Developing an accurate and scalable clustering method, single-cell Decomposition using Hierarchical Autoencoder (scDHA), for single-cell sequencing data.

METHODS

Normalization

Data is log transformed and rescaled to 0 and 1.

Gene Filtering

scDHA use a non-negative kernel autoencoder to filter out genes or components that have insignificant contribution to the data.

Dimension Reduction

scDHA uses Stacked Bayesian Self-learning Network that is built upon the Variational Autoencoder (VAE) to project the data onto a low dimensional space.

Clustering

scDHA uses k-nearest neighbor spectral clustering to separate the cells into different groups.

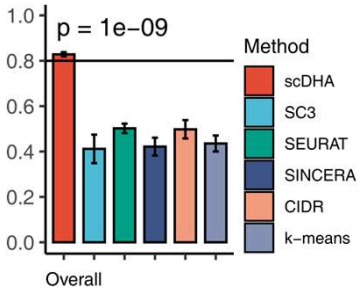
Visualization: Using the compressed data, scDHA project the data to 2-dimensional space for visualization.

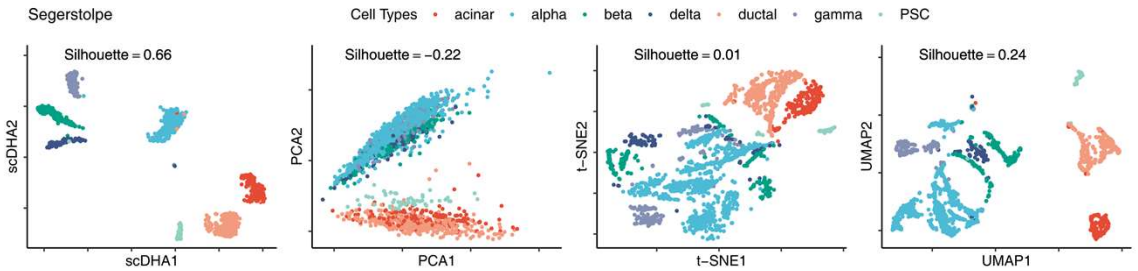
RESULTS

Data: The method is validated using 24 real single-cell datasets.

Clustering: scDHA outperforms current state-of-the-art (SC3 [3], Seurat [4], SINCERA [5], CIDR [6]) in clustering accuracy. Accuracy is measured by adjusted Rand Index (ARI).

Visualization: Using compressed data, scDHA is more efficient than both t-SNE and UMAP, as well as the classical principal component analysis (PCA) in visualizing single-cell data.





CONCLUSION

scDHA is an accurate method for scRNA-seq data clustering. Using two autoencoders in a hierarchical setting, scDHA can effectively transform high dimensional scRNA-seq data to a low dimensional space. This compressed data can be used in downstream applications.

FUTURE WORK

- Expanding scDHA to work with other data type such as Hi-C [7], network [8], or multi-omics data [9, 10].
- Applying hierarchical autoencoder concept to other areas such as predicting cellular spatial position [11], or pathway analysis [12].

FUNDING SOURCE

This material is based upon work supported by the National Aeronautics and Space Administration under Grant No. 80NSSC19M0170.

References

1. Tran et al. (2019). "RIA: a novel Regression-based Imputation Approach for single-cell RNA sequencing." *2019 11th International Conference on Knowledge and Systems Engineering (KSE)*.
2. Tran et al. (2019). Fast and precise single-cell data analysis using hierarchical autoencoder. *bioRxiv*, 799817.
3. Kiselev et al. (2017). "SC3: consensus clustering of single-cell RNA-seq data." *Nature methods* 14(5), 483-486.
4. Satija et al. (2015). "Spatial reconstruction of single-cell gene expression data." *Nature biotechnology* 33(5), 495-502.
5. Guo et al. (2015). "SINCERA: a pipeline for single-cell RNA-seq profiling analysis." *PLoS computational biology* 11(11).
6. Lin et al. (2017). "CIDR: Ultrafast and accurate clustering through imputation for single-cell RNA-seq data." *Genome biology* 18(1), 59.
7. . Stansfield et al. (2019). "R Tutorial: Detection of Differentially Interacting Chromatin Regions From Multiple Hi-C Datasets." *Current protocols in bioinformatics* 66(1), e76.
8. Nguyen et al. (2019), "A comprehensive survey of tools and software for active subnetwork identification." *Frontiers in genetics* 10, 155.
9. Nguyen et al. (2017). A novel approach for data integration and disease subtyping. *Genome research*, *27*(12), 2025-2039.
10. Nguyen et al. (2019), "PINSPlus: a tool for tumor subtype discovery in integrated genomic data." *Bioinformatics* 35(16), 2843-2846.
11. Tanevski et al. (2019). "Predicting cellular position in the Drosophila embryo from Single-Cell Transcriptomics data." *bioRxiv,* 796029.
12. Nguyen et al. (2020). NBIA: a network-based integrative analysis framework–applied to pathway analysis. *Scientific Reports*, *10*(1), 1-11.