Lab Internship
Summer 2019

Jaewoong Lee

July 25, 2019

Contents

List of Tables

List of Figures

List of Listings

1 Introduction

1.1 Single Cell Analysis

1.2 Principal Component Analysis

1.3 T-Distributed Stochastic Neighbor Embedding

1.4 Clustering

1.4.1 K-means Algorithm

1.5 Single Cell Trajectory

2 Method

2.1 Cell Ranger

Cell Ranger is a set of analysis pipelines that process Chromium single-cell RNA-seq output to align reads, generate feature-barcode matrices and perform clustering and gene expression analysis. [Zheng et al., 2017]

2.2 Seurat

Seurat is an R package designed for QC, analysis, and exploration of single-cell RNA-seq data. Seurat aims to enable users to identify and interpret sources of heterogeneity from single-cell transcriptomic measurements, and to integrate diverse types of single-cell data. [Butler et al., 2018, Stuart et al., 2019]

2.3 Scikit-Learn

Scikit-Learn is a Python module for machine learning built on top of SciPy. [Pedregosa et al., 2011]

3 Result

3.1 TSNE Map

4 Discussion

5 Acknowledgment

References

[Butler et al., 2018] Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology*, 36(5):411.

[Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830.

[Stuart et al., 2019] Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck III, W. M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell*.

[Zheng et al., 2017] Zheng, G. X., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature communications*, 8:14049.