Date: 20 February 2020
SSF Project Report

Performance and Benchmarking

Comparison of bioinformatic tools for pseudo time analysis

Vanja Börjesson[1]

[1]Bioinformatics Core Facility, Sahlgrenska Academy, University of Gothenburg, Box 115, 405 30, Sweden.

Summary

This project compares four of the most commonly used tools; Monocle, SLICER, destiny and Scanpy, to define the most efficient tool based on resulting trajectory, running time, required computational resources, and ability to manipulate plots and pre-work such as filtering, normalization etc.

Contact: vanja.borjesson@gu.se

1 Introduction

Cells specialize to do different and essential things in our body. Some cells pro- duce proteins important for building our body, such as tissues and organs, and others might specialize to perform a function such as communicating through chemicals. In research, studying the function of cells are very important in order to understand, for an example, the cause of disease or other phenotypes. Single cell RNA sequencing (scRNA-seq) allows researchers to measure the expression levels of RNA in individual cells. Studying how cells change in their specialization, i.e. the change in gene expression, over time is also very important to understand the whole cell process. Several bioinformatics tools exist for pseudotime analysis, all based on different algorithms for dimensional reduction, clustering and trajectory creation. This project aims to compare four of the most commonly used tools; Monocle (Trapnell *et al.*, 2014; Qiu X, Hill A *et al.*, 2017; Qiu X, Mao Q *et al.*, 2017), SLICER (Welch 2017), Destiny (Angerer *et al.*, 2015) and Scanpy (Wolf *et al.*, 2018), to define the most efficient tool based on resulting trajectory, running time, required computational resources, and ability to manipulate plots and pre-work such as filtering, normalization etc.

2 Methods

The data used for this analysis was two different scRNA-seq datasets generated by 10x genomics. The two different datasets represent two different places in the intestine. We analyzed the datasets separately.

We did the filtering, normalization and scaling in Seurat (Stuart *et al.,* 2019) for both datasets. Then we converted the datasets into a cell data set (CDS) object. These objects were then loaded as input into Monocle2, Monocle3, SLICER, destiny and Scanpy. We performed the dimensional reduction and clustering following the tool specific tutorials.

3 Result

All tools deliver a plot showing the same pseudotime, indicating a two-branch differentiation. Monocle2, Monocle3 and Scanpy were the only tools that distinguished the proliferated cells as a separate branch. Monocle was also the only tool that had most variability when it comes to plotting genes and expression onto the trajectory.

SLICER and Destiny (Figure 1) were easily installed and executed in R. SLICER was the tool that took the longest time to run; almost 24 hours using 4 cores. The other tools took about 5 to 30 minutes to run. Neither SLICER nor Destiny provided functions to calculate differentially expressed genes, to plot genes nor groups of interest.

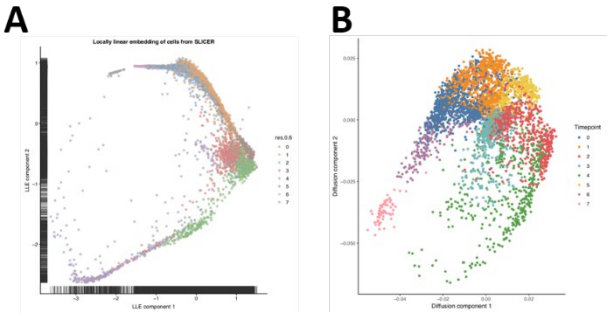

Figure 1. A) SLICER. **B)** Destiny

Monocle3 alpha version was more difficult to install but plots a smoother trajectory, better looking compared to Monocle2, and also introduce UMAP for dimensional reduction (Figure 2). However, in our test, dimensional reduction performed by tSNE returns the best-looking trajectory.

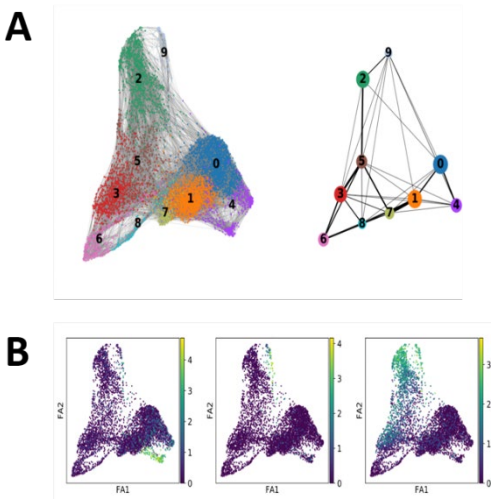

Figure 2. **A)** Monocle 2. **B)** Monocle3

Scanpy (Figure 3) returns similar results as Monocle and is very easy to install and use. It provides all functions needed for pre-processing data and visualizing the results. It is written in python

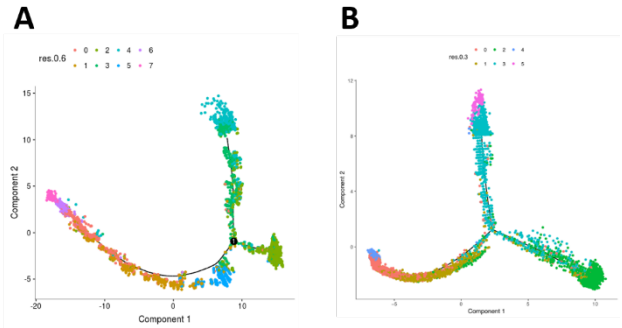

Figure 2. Scanpy. A) Clustering and lineage tracing. B) Gene expression

4 Conclusion

Although all five tools were easy to install and use, and predicted the same trajectory, their performance and the options provided vary. Monocle and Scanpy are the easiest and most cost-efficient tools to use for pseudotime analysis. SLICER was the most time-consuming tool. Monocle3 is still just a beta version and the final version 3 has not yet been released. The beta version has a lot of dependencies and are very difficult to install and run. At this moment Scanpy is the tool (out of the five tools that we tested) that is best in practice.

Availability of data and materials

No datasets generated during the current study. However, use case data may be available from the authors upon reasonable request and with permission of Prof. Malin Johansson, at the University of Gothenburg.

Acknowledgements

The authors thank the Bioinformatics Core Facility at the University of Gothenburg for providing computational resources for data analysis and storage. Special thanks to Prof. Malin Johansson (University of Gothenburg) for kindly providing datasets used as use cases.

Funding

This study was supported by the Swedish Foundation for Strategic Research (RIF14–0081). The funding body did not play any role in the design of the study, or collection, analysis, or interpretation of data, or in writing the manuscript.

References

Angerer P, Haghverdi L, Büttner M, Theis F, Marr C, Büttner F (2015). “destiny: diffusion maps for large-scale single-cell data in R.” Bioinformatics.

Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL (2014). “The dynamics and regulators of cell fate decisions are revealed by pseudo-temporal ordering of single cells.” Nature Biotechnology.

Qiu X, Hill A, Packer J, Lin D, Ma Y, Trapnell C (2017). “Single-cell mRNA quantification and differential analysis with Census.” Nature Methods.

Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner H, Trapnell C (2017). “Reverse graph embedding resolves complex single-cell developmental trajectories.” BioRxiv.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell.

Welch J (2017). “SLICER: Selective Locally Linear Inference of Cellular Expression Relationships”. R package version 0.2.0.

Wolf, F., Angerer, P. & Theis, F. (2018). “SCANPY: large-scale single-cell gene expression data analysis”. Genome Biology.