

Implementing Single Cell Mapping Algorithm

Authors - Fatima Perez, Jacob Bruggemann
Scripps Research and Su Lab, San Diego, California

Introduction

The Su Lab is a Computational Biology lab dedicated to the acceleration of biomedical research through the use of computational systems. Before interning in the lab, I had no experience with programming or data science. During my time here, I learned the principles of programming and the basic skills of manipulating biological data. Through first-hand exposure to methods used by bioinformaticians I gained important insights to the significance of bioinformatics in modern biology.

After becoming acquainted with basic programming skills, focus was turned to an RNA sequencing data set from Allen Brain Atlas. This data set is derived from 15,928 nuclei of human brain cells. Using subsets of this dataset, I tested scmap, an algorithm for single-cell sequencing data for comparing datasets from different sources.

Methods

- Retrieved RNA sequence data set derived from nuclei of human brains from Allen Brain Atlas
- Used scmap algorithm to compare slices of the Allen Brain Atlas data set and produce graphs.
- Scmap takes cells from the RNA sequence and projects them onto reference data to identify which cells or clusters are most similar.

Results

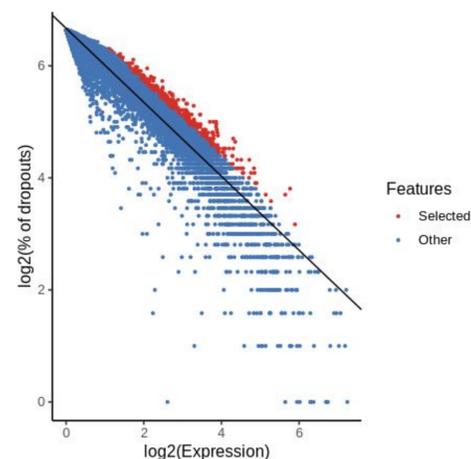


Fig 1 Feature selection for scmap was performed to select a subset of informative genes. Genes were evaluated based off expected dropout vs median expression.

Unselected	Selected
49781	500

Fig 2 Table of features the algorithm selected.

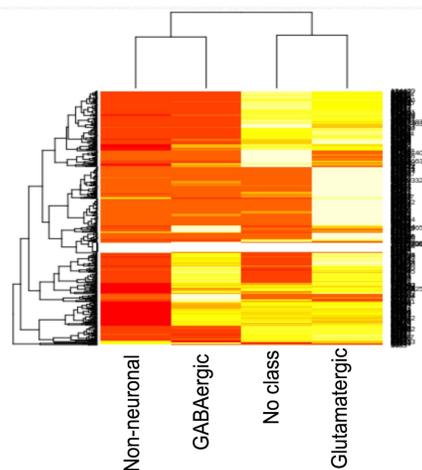


Fig 3 Heatmap of selected features vs cell class. The selected features distinguish the pre-defined cell classes.



Fig 4 Sankey plot of self-projected data. GABAergic cells were unable to be reassigned to their own identity.

Conclusions

- The set of genes identified by scmap well-characterized cell class in a single-cell dataset
- Inconsistencies arised when self-projecting data, particularly with regards to smaller classes
- Further work is required to determine the reproducibility and efficacy of scmap for assigning cell identities across datasets.

Acknowledgments

Scripps Research
Life Sciences Summer Institute
San Diego Workforce Partnership
Dr. Arthur Olson
Ms. Shirley King
Dr. Andrew Su
Jacob Bruggemann
Su Lab
Michelle Fox

Citations

“Cell Diversity in the Human Cortex” Allen Brain Atlas, August 1, 2018.

Kiselev, Vladimir Yu and Martin Hemberg. “scmap-A tool for unsupervised projection of single cell RNA-seq data.” *Nature Methods* 15.5 (2018): 359.