Analyzing Cell Type Diversity in the Human Middle Temporal Gyrus

Skaggs Graduate School at Scripps Research

Arthur J. Olson, PhD
Shirley King

Authors - Michelle Fox, Jacob Bruggemann
Scripps Research and Su Lab, San Diego, California

LSSI LIFE SCIENCES SUMMER INSTITUTE

Introduction

The realm of bioinformatics is an increasingly essential element of the life sciences. The increasing ease of use and decreasing cost of data analysis programs and gene sequencing has vastly facilitated the process of examining and integrating large datasets within the field of computational biology.

Over the course of this program, I developed programming skills in the Python and R programming languages and learned how to interact with a Linux terminal in bash. Using both languages, I ran several essential data science programs for gene expression analysis using Garibaldi, the Scripps Research Institute's supercomputer. The data set used was a matrix from the Allen Brain Atlas Data Portal that contained over 50,000 gene-level exonic read count values for RNA sequences within the middle temporal gyrus (MTG) of the human brain.

Fig 1. The MTG, encompassed by Brodmann's area 21 in the temporal lobe of the human brain, is involved in a number of functions including sensory integration, semantic memory, visual perception, and language. 15,928 nuclei from 8 human tissue donors (ages 24-66) were analyzed for RNA sequences.

Materials and Methods

- RNA-seq data were derived from the Allen Brain Atlas Data Portal.
- After converting the data into a matrix, Principal Component Analysis (PCA) was run to identify a smaller-dimensitonality subset, both for de-noising and computational expediency.
- The reduced data were embedded to two dimensions using t-Distributed Stochastic Neighbor Embedding (t-SNE) and subpopulations were identified with both k-means and Louvain clustering. (Fig. 2 & 3)

Results

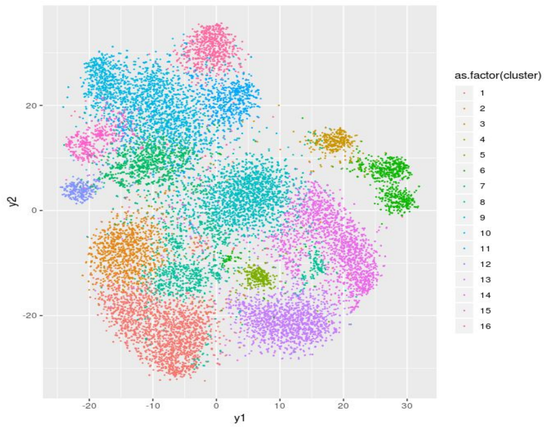


Fig 2. t-SNE plot colored by Louvain clustering showing 16 different clusters, based on shared nearest neighbors between samples.

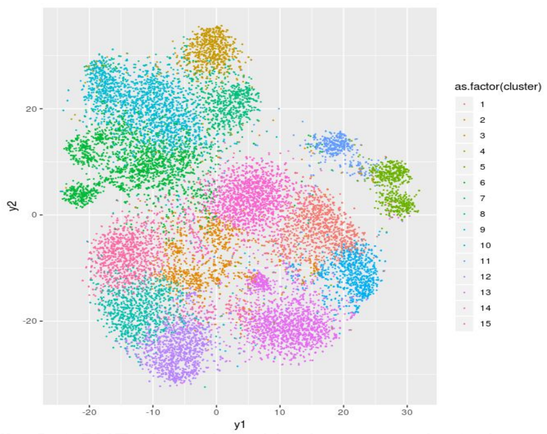
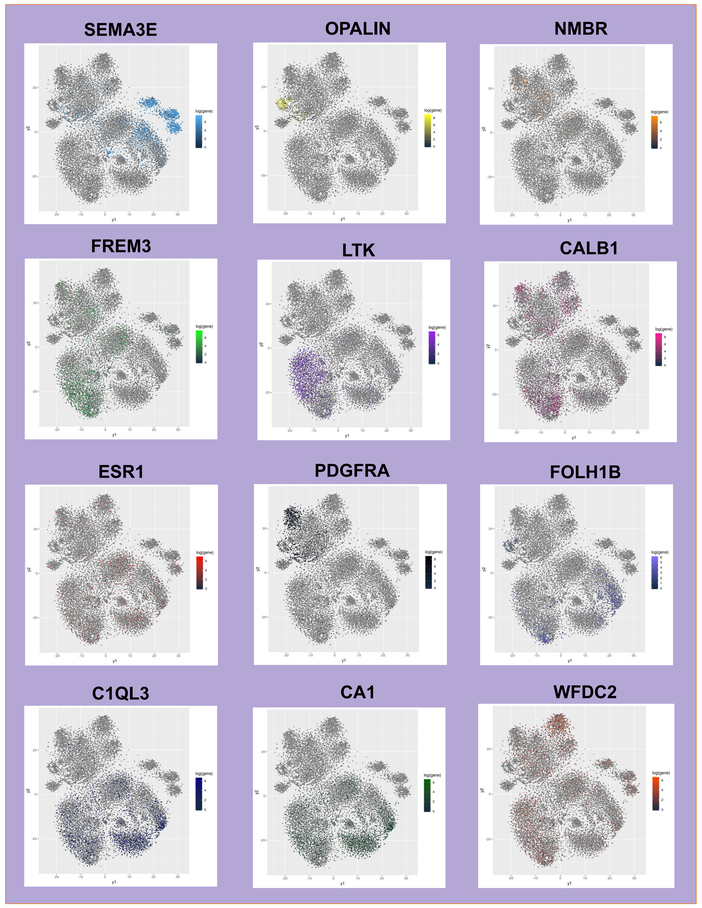


Fig 3. t-SNE plot colored by k-means clustering algorithm for k=15. Similar results were obtained through both Louvain and k-means clustering.



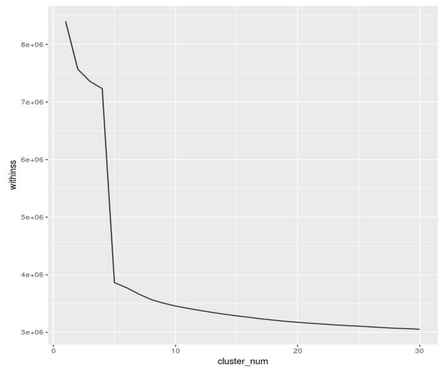


Fig 4. (above) Scree plot of k-means for Allen Brain Atlas Data. K was set to equal 15.

Fig 5. (left) Each graph shows the counts of a gene within each sample. 75 genes were analyzed based on Allen Brain Atlas transcriptional profiles. Shown are the 12 most common of those genes from all samples.

Conclusions

- K-means and Louvain clustering correlate with previously identified neuronal marker genes.
- Initial clustering revealed 15-16 clearly delineated clusters of human neurons.
- Identification of marker genes between different samples is required to elucidate the identity of the uncovered clusters.
- Marker genes can signify different functional types of neurons and their importance in brain structure.
- Iterative clustering and application of metadata is a possible next step for revealing further subpopulations of neurons.

Acknowledgments

Scripps Research
Life Sciences Summer Institute
San Diego Workforce Partnership
Arthur J. Olson, PhD
Shirley King
Andrew I. Su, PhD
Jacob Bruggemann
Fatima Perez

Citations

"Cell Diversity in the Human Cortex" Allen Brain Atlas, August 1, 2018

L.J.P. van der Maaten and G.E. Hinton. **Visualizing High-Dimensional Data Using t-SNE**. *Journal of Machine Learning Research* 9(Nov):2579-2605, 2008.