SINGLE-CELL RNA-SEQUENCING TO IDENTIFY AND EXAMINE CELL-TYPES WITHIN THE MOTOR CORTEX

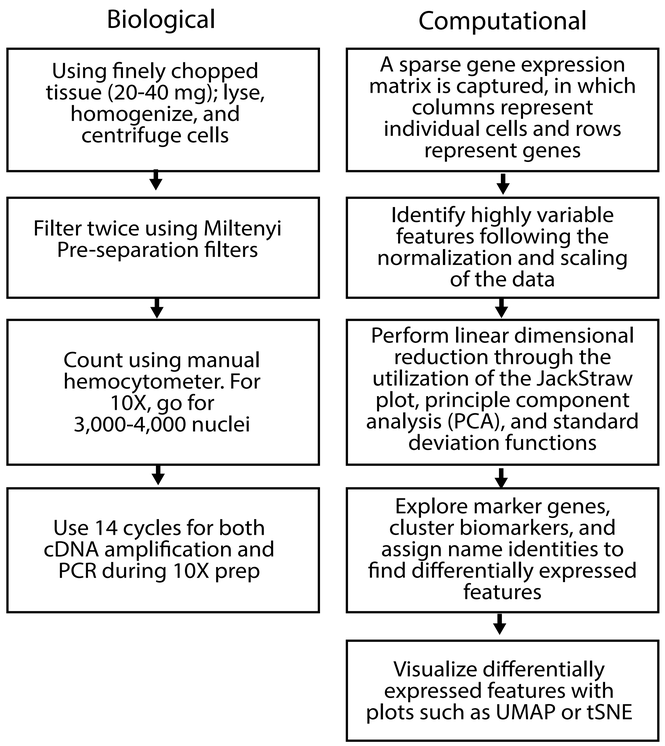
J. Hemela[1], Z. Wang[1,2,5], J. Lowell[3,5], J. Bixby[3,4,5], V. Lemmon[1,3,4,5]

[1]Computing for Structure REU, [2]Department of Computer Science, [3]Miami Project to Cure Paralysis, [4]Center for Computational Science, University of Miami

Introduction

Different cells express different proteins, depending on the cells' roles. Consequently, cells express mRNAs that code for these different proteins, and expression levels of these mRNAs can vary dramatically among cell types. Single-cell RNA sequencing (scRNA-seq) provides a robust way to interrogate the mRNAs expressed in 1000s of individual cells. Using a clustering algorithm for scRNA-seq, we can identify specific cell types present during cell differentiation and development. Use of this platform to construct developmental trajectories of cellular fates and identify shared populations across data sets has contributed novel insights, particularly in the fields of neuroscience, cancer, and immunology. We analyzed two experiments, each of which compared hippocampal neurons treated either with DMSO (control) or a kinase inhibitor, RO48, that promotes axon growth. For my single-cell RNA computation, I utilized the 'Seurat' package for quality control, analysis, and data exploration. Seurat is based on 'R', a language and environment for statistical computing and graphics. Seurat provides users with informative visualizations of dimensionally-reduced single-cell transcriptional expression data. This cutting edge tool aids in the recognition and examination of the origins of cellular heterogeneity.

Workflow



Results

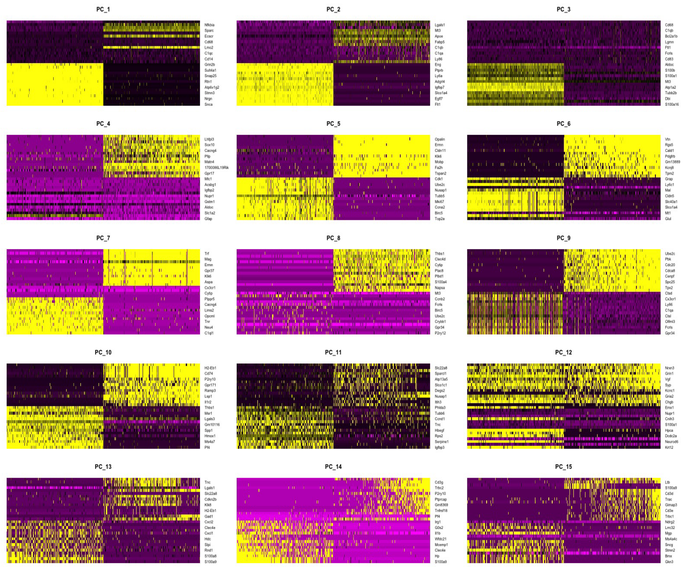
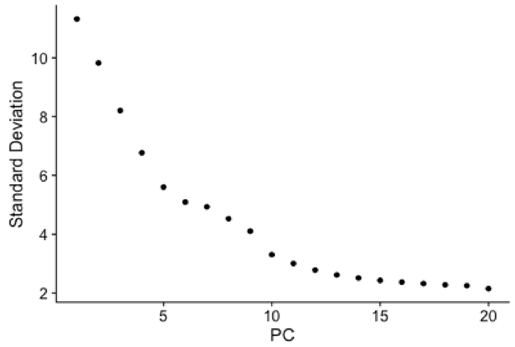


Figure 1: The heat map allows us to visualize similarities between replicates in large datasets. The individual tiles on a heat map are scaled with colors that correspond to gene expression values. A colored display represents the matrix of values as a grid; the number of rows correlating to the number of genes being analyzed, and the number of columns to the number of samples.

Figure 2: A heuristic method available with Seurat is the 'Elbow plot'. This plot generates a ranking of principle components based on the percentage of variance explained by each one. Seurat clusters cells based on their PCA scores, with each PC representing a 'metafeature' that combines information across a correlated feature set. In this example, we can observe an 'elbow' around PC9-10, suggesting that the majority of significance is captured in the first 10 PCs.



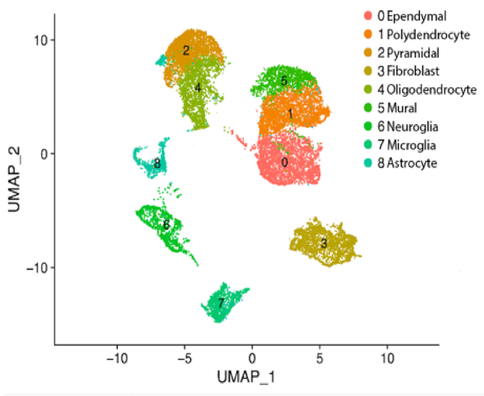


Figure 3: The UMAP plotting function in Seurat preserves global structure, relative distances, and creates cluster according to cell type. This tool uses genes that are highly expressed in each cluster, not genes solely expressed in a given cluster, to do the clustering. Cells in the pyramidal cell cluster (neurons found in the cerebral cortex and hippocampus) express relatively small amounts of genes expressed in cells in the Oligodendrocyte cluster. This UMAP plot illustrates that Oligodendrocytes and Pyramidal cells are proximate, as are Polydendrocytes and Microglia.

Conclusion

- Our goal was to analyze and identify cell types that are present in the datasets, obtain biologically significant cell type markers, and cluster cells based on these markers.
- Through computation, we have been able to identify cells expressing high levels of marker genes, cluster cells according to gene expression, and perform differential expression analysis to identify genes differentially regulated in neurons by the kinase inhibitor RO48.
- In order to determine the extent of the role of kinase inhibitors in the context of regeneration, future research in this area may examine other kinase inhibitors in addition to RO48, as well as the effect on other areas such as the spinal cord.

Acknowledgments

I would like to thank Burt Rosenberg for coordinating this Research Experience for Undergraduates program. Much appreciation to Vance Lemmon and all others from The LemBix Lab for fostering my learning this summer. Special thanks to Victor Jeffery Jann for oodles of moral support. Finally, I would like to thank the National Science Foundation (CNS-1659144) and NIH (NS100531-01A1) for funding this project.

References

[1] Butler, A., & Satija, R. (2017). Integrated analysis of single cell transcriptomic data across conditions, technologies, and species. Nature. doi:10.1101/164889

[2] Chen, H., Albergante, L., Hsu, J. Y., Lareau, C. A., Bosco, G. L., Guan, J., . . . Pinello, L. (2019). Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. Nature Communications,10(1). doi:10.1038/s41467-019-09670-4

[3] Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., . . . Satija, R. (2018). Comprehensive integration of single cell data. doi:10.1101/460147