Classification of Cell Type through Topological Data Analysis of Gene Co-expression Distance
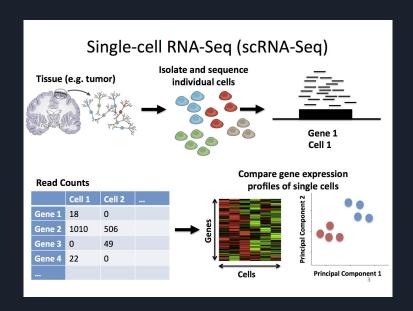
Cecilia Liu, Jason Chen, Zixi Yuan
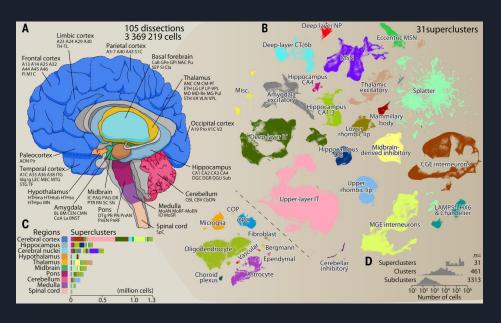
HDSI
UC SAN DIEGO

Background

- Complex structure
- Diverse Cell Types
- Critical Functions
- Health and Research

Single Cell RNA Sequencing

- Gene expression in single cell resolution
- High-throughput
- Comprehensive gene expression profiles
- Wide application in computational biology

The Human Brain Cell Atlas

- > 3,000,000 cells
- Comphrehensive transcriptomic landscape of full human brain
- Great source to investigate cell types
- Great source to investigate the molecular diversity of the human brain
- Potential to understand the epigenetic fundations of diseases in human brain

Methods

- Data Collection
    - Open source data

- Data Pre-processing
    - Subset
    - Scale
    - Log-transformation

Methods

- Gene Co-expression Distance Construction:
    - Pearson Correlation
        - Convert to proper distance metrics
    - Euclidean Distance
        - PCA
        - UMAP

Methods

- Distance to Persistent Diagrams
    - Create point cloud data with genes as vertices
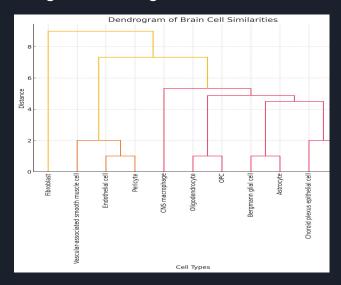   - Performe Rips filtration
    - Calculate persistent diagrams

- Comparing Cell Types via Persistent Diagrams
    - Bottleneck Distance
    - Persistence Images and p-norms

Methods

- Hierarchical Clustering
    - Single Linkage
    - Average Linkage
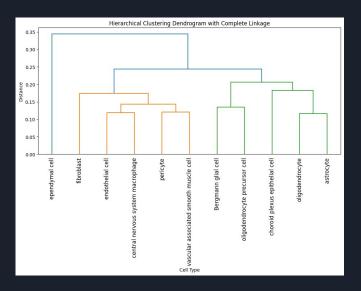   - Complete Linkage

- Aggregate Hierarchical Clustering
    - Sum up multiple normalized distance matrices

Results & Conclusion

Hierarchical Clustering based on previous biological knowledge:

Aggregate Bottleneck Distance Hierarchical Distance

Results & Conclusion

- Bottleneck Distance vs Persistence image

- Hierarchical clustering result

- How to use our findings

Discussion

- Limitations of dataset

- Limitations of Hierarchical clustering

- Future direction

Reference

[1] Siletti, K. et al. (2023) 'Transcriptomic diversity of cell types across the adult human brain', Science, 382(6667). doi:10.1126/science.add7046.

[2] Wolf, F.A., Angerer, P. and Theis, F.J. (2018) 'Scanpy: Large-scale single-cell gene expression data analysis', Genome Biology, 19(1). doi:10.1186/s13059-017-1382-0

[3] Our code is available at https://github.com/cecilialmw/DSC-214-Final-Project