CSIC5011 Final Project

CHEN Yiyun (ychenab@connect.ust.hk) ZENG Wenshu (wzengaa@connect.ust.hk)
Division of Life Science, Hong Kong University of Science and Technology

1. Introduction

Glioblastoma (GBM), is the most aggressive tumor that begins in the brain. Radiation, chemotherapy, and targeted therapy for GBM have shown limited success. In recent years, increasing studies showed that GBM evolves temporally and spatially, resulting in drug resistance and poor patient survival. The technological advancements in genome sequencing, RNA expression profiling, and single-cell platforms give us power to study cancer evolution in individual patients and discover precise targets for cancer treatment.

2. Single cell data and method

Single-cell data was obtained from Raul Rabadan's paper published on Nature Genetic in 2017[1], which contained 305 single cells RNA sequencing data from 3 patients, namely GBM2, GBM9, GBM10. We performed a case study using 133 cells from patient GBM9, which was obtained from the initial tumor on the right hemisphere, the initial tumor on the left hemisphere, and the recurrent tumor on the left (which recurred at the same location as the initial tumor on the left, after the initial one was cut out and treated).

From raw sequencing file, we generated read count of every gene of each single cells. Read count is normalized by the total reads in the cell, thus generating the expression matrix with transcript per million (TPM) of each gene. The expression matrix was subjected to analysis using PCA, t-SNE, sc-TDA, and other methods. We demonstrated the power of all these methods in analyzing single-cell sequencing data from GBM9, and traced the cancer evolution in time and space.

3. Results

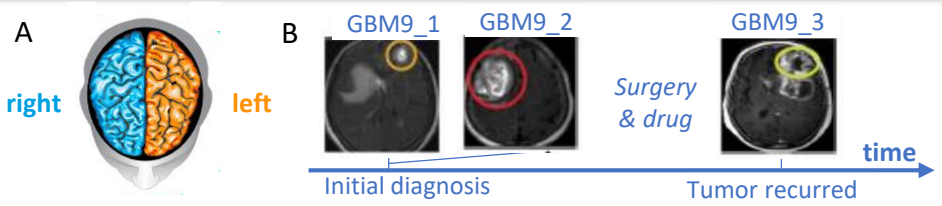

Figure 1. (A) Schematic diagram of left and right hemisphere of human brain. (B) Single cell is derived from left initial (GBM9_1), right initial (GBM9_2), and left recurrent (GBM9_3) tumor.

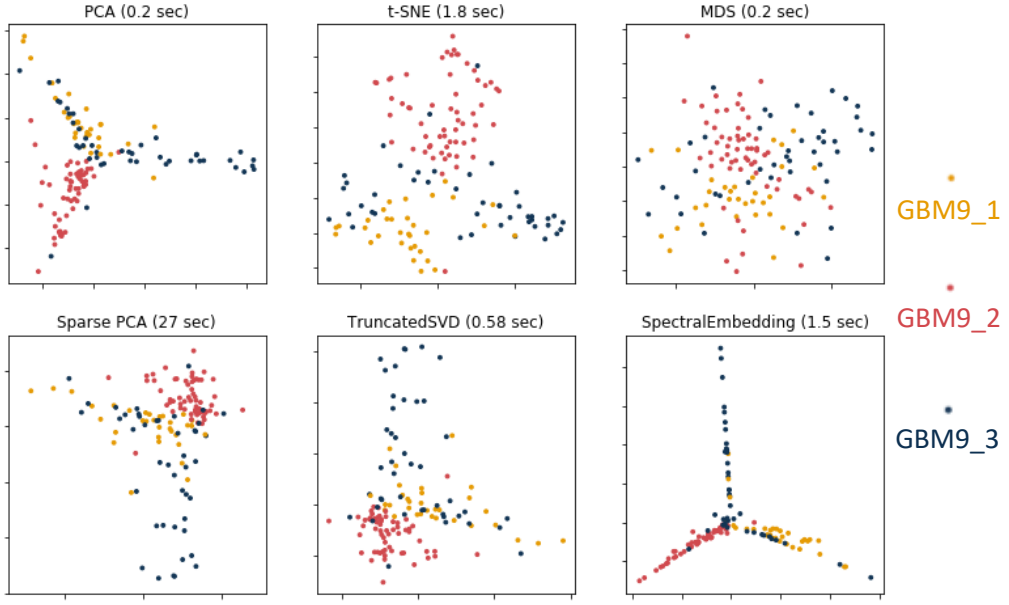

Figure 2. Comparison of the most common analysis methods applied in single-cells data.

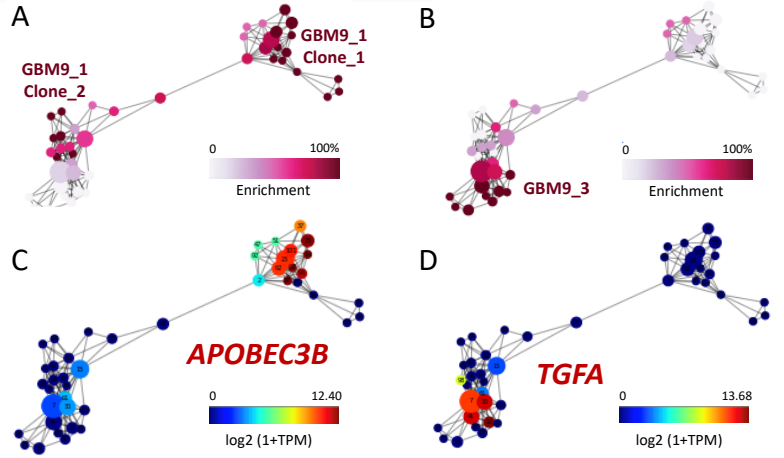

Figure 3. Single-cell TDA results. (A) Nodes enriched in cells from GBM9_1 are separated into 2 different groups, representing 2 clones with distinct expression features. (B) Nodes enriched in cells from GBM9_3. (C) The expression level of APOBEC3B gene is significantly high in GBM9_1 Clone_1. (D) TGFA gene expression level is significantly high in GBM9_3.

4. Conclusion

- Comparing commonly used methods in single-cell data (Figure 2), PCA and spectral embedding are more robust, and both showed a close connection between the initial and recurred tumor in the left brain, while the tumor in the right brain is very different. However, the evolution trajectory of tumor is not presented by PCA or t-SNE.
- Single-cell Topological Data Analysis (scTDA)[2] generated the major connected component, containing only cells from GBM9_1 and GBM9_3, which is consistent with the fact that GBM9_2 on the right brain is very distinct from tumors in the left brain.
- GBM9_1 contains 2 tumor clones with distinct expression profiles (Clone_1 and Clone_2). GBM9_3 is similar to Clone_2, indicating potential progression from Clone_2 to the recurred GBM9_3.
- APOBEC3B and TGFA are enriched in different clones in the tumor, suggesting potential combination therapy to eliminate both clones.

5. Reference

1. Lee JK, Wang J et al. Spatiotemporal genomic architecture informs precision oncology in glioblastoma. Nat Genet 2017; 49(4):594-9.
2. Rizvi AH, Camara PG et al. Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development. Nat Biotech 2017; 35(6):551-560.

6. Contribution

CHEN Yiyun:
Method comparison (PCA, MDS, t-SNE and scTDA), report writing

ZENG Wenshu:
Report writing, data preprocessing, gene expression analysis