Pseudotime Based Analysis of Cancer Dynamics

Tasmia Aqila and Ananda Mohan Mondal†
School of Computing and Information Sciences
Florida International University
Miami, FL, USA
taqil001@fiu.edu, amondal@fiu.edu

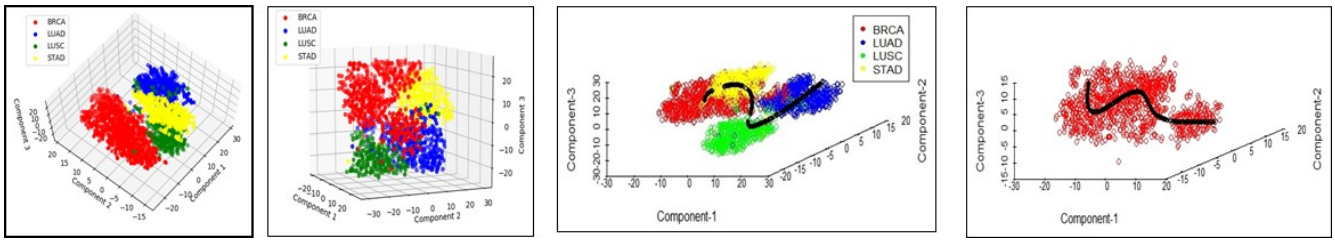

Figure 1. a) t-SNE plot for mRNA expression, b) t-SNE plot for miRNA expression, c) The trajectory of cancer development quantifying pseudotime in four cancers, and d) The trajectory of cancer development quantifying pseudotime in BRCA.

ABSTRACT

Longitudinal or time-series multiomics data for the same cohort of patients are necessary to understand the cancer dynamics or development. But no such data is available for cancer. Most of the time-series gene expression data for cancer genomes available in Gene Expression Omnibus (GEO) database contain gene expression values from different samples at different stages and each stage contains few samples, mostly 3 to 10. So, these datasets are not suitable for analyzing cancer dynamics. Recent studies show that single-cell gene expression with no temporal information can be analyzed to discover the mechanism of cell development by inferring pesudotime. Researchers have developed models for the dynamic behavior of cell development and differentiation using single cell expression profiles with both temporal [1] and no temporal [2] information. This motivates us that the static multiomics data of more than 12,000 cancer patients from TCGA could be analyzed by inferring pseudotime to decipher the cancer dynamics.

For experiment, we used four different types of data - mRNA expression, miRNA expression, lncRNA expression and DNA methylation for four different cancers - BRCA, LUAD, LUSC and STAD from LinkedOmics. t-Distributed Stochastic Neighbor Embedding (t-SNE), a well-known technique for nonlinear dimensionality reduction, particularly well-suited for the visualization of high-dimensional datasets has been used for dimensionality reduction as well as finding the similarities or clusters among samples over the dimension of genes. Finally, principal curve fitting is used to infer the pseudotime for each sample in a cluster. The inferred pseudotime is used to find the trajectory of cancer development in clustered samples.

These preliminary results show that it is possible to find the trajectory of cancer development using static multiomics data from TCGA. In future, the developed trajectory will be analyzed to find the possible tipping point for metastasis.

KEYWORDS

Cancer Dynamics, Multiomics, Pseudotime, Principal Curve, t-SNE.

ACKNOWLEDGMENTS

This research is funded by NSF CAREER award #1651917 (transferred to #1901628) to AMM.

REFERENCES

[1] G. Guo *et al.*, "Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire," *Cell Stem Cell*, vol. 13, no. 4, pp. 492–505, Oct. 2013.
[2] S. C. Bendall *et al.*, "Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development.," *Cell*, vol. 157, no. 3, pp. 714–25, Apr. 2014.

ACM-BCB '19, September 7–10, 2019, Niagara Falls, NY, USA

ACM ISBN 978-1-4503-6666-3/19/09.
https://doi.org/10.1145/3307339.3343253