Cartography of Genomic Interactions Enables Deep Analysis of Single-Cell Expression Data

Md Tauhidul Islam and Lei Xing
Department of Radiation Oncology, Stanford University, Stanford, California-94305, USA

HAI FIVE
Celebrating 5 Years of Impact

Stanford MEDICINE

INTRODUCTION

Analyzing tabulated datasets such as genomics data is critically important in the diagnosis, prognosis, and treatment of different diseases. This work proposes a novel approach to transforming the tabulated genomics data into a 2D image format utilizing gene-gene interactions. 2D convolutional neural networks (CNNs) are then employed for deep data analysis. The developed approach enables high-performance analysis of the data for improved clinical decision-making.

MATERIALS AND METHODS

To reconfigure the tabulated data of m cells and n genes, we compute the pairwise interaction matrix among the genes of size nxn. We then compute the pairwise Euclidean distance matrix among the n locations in a 2D grid of size nxn. We optimize the Gromov-Wasserstein discrepancy between these two matrices to obtain a transformation matrix T. Multiplying the tabular expression data with the transformation matrix allows us to obtain m number of image representations of the samples. We then apply a 10-layer deep neural network with 1 convolutional layer (kernel size [3,3]), 3 dense layers, 2 relu layers, and 1 dropout layer for different downstream analyses.

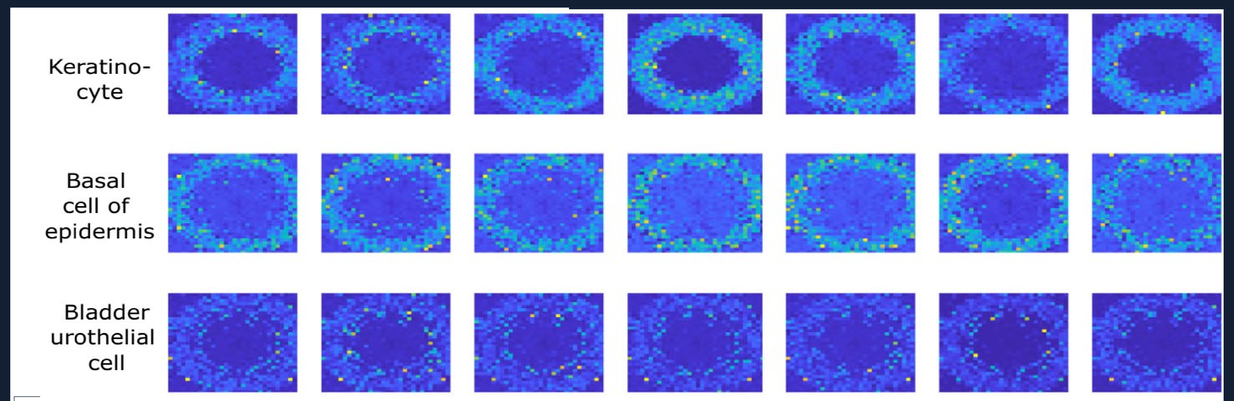
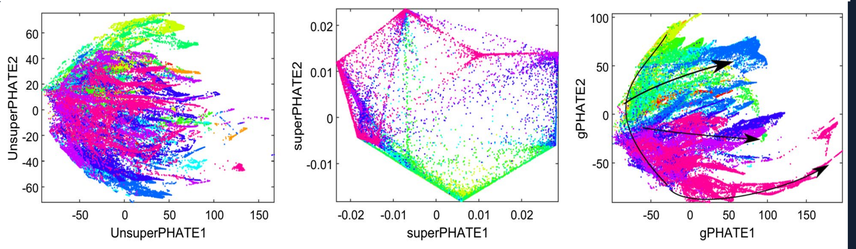


Fig 2. Genomaps created from Tabula Muris dataset.

Fig. 3. Trajectory mapping results



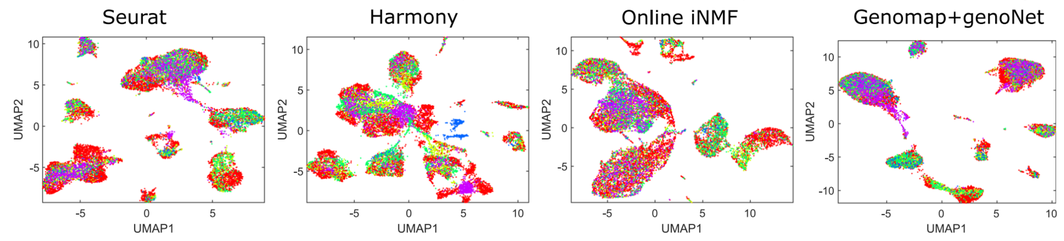


Fig. 4. Multi-omic data integration results

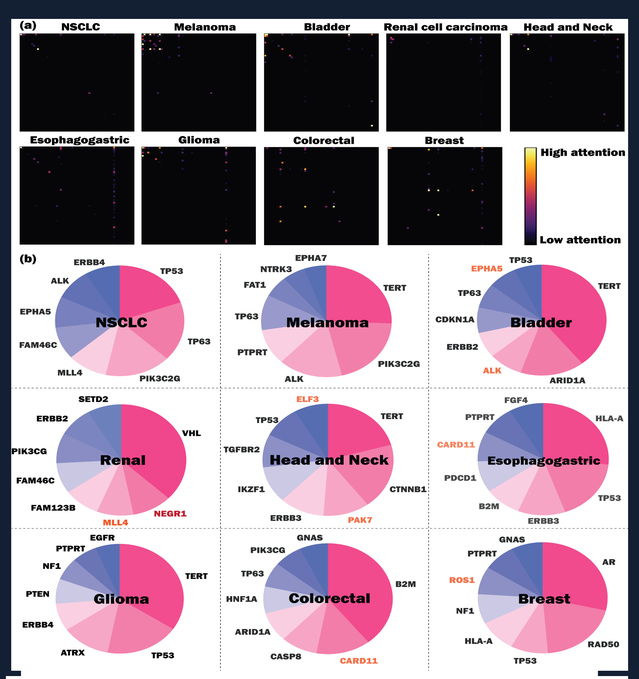


Fig. 6. Gene marker discovery using Shapley scores on the genomaps created from Memorial Sloan Kettering dataset.

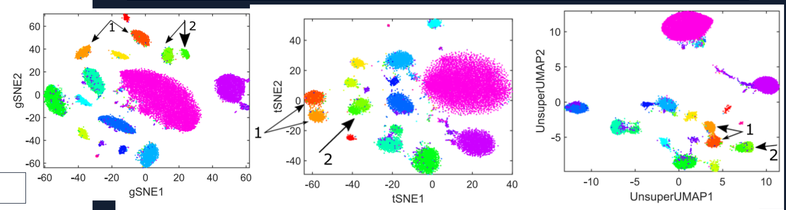


Fig. 7. Dimensionality reduction results

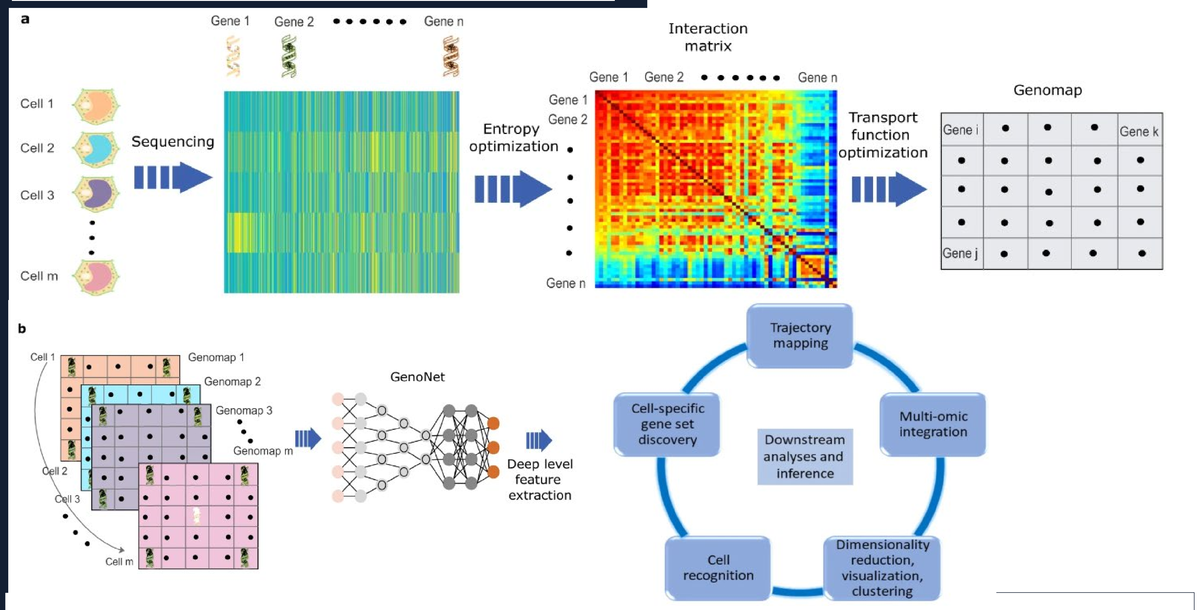
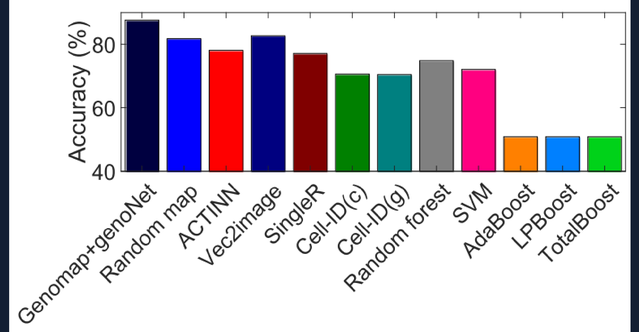


Fig 1. Deep analysis of gene expression data by using genomap and genoNet. (a) Workflow of genomap generation from scRNA-seq data. Note that the genomap is dataset dependent and the gene distribution in the genomaps vary with dataset. b) GenoNet is applied on the genomaps to extract deep level features for downstream analyses.

RESULTS

Genomaps created from Tabula Muris dataset are shown Fig. 2. The results of genomap-based approach for cellular trajectory mapping, multi-omic integration, cell classification, biomarker discovery and dimensionality reduction are shown in Figs. 3-7.

Fig. 5. Cell classification accuracy of the proposed approach and eleven existing techniques for T-cell landscape dataset



CONCLUSIONS

This work presents a novel method for converting genomics data in tabular format into 2D images, guided by the gene-gene interactions. These structured images are then analyzed using 2D convolutional neural networks (CNNs) for learning the underlying patterns. This innovative approach facilitates high-quality data analysis, resulting in improved downstream analysis and clinical decision-making.

REFERENCES

1. Islam, M.T. and Xing, L., 2023. Cartography of Genomic Interactions Enables Deep Analysis of Single-Cell Expression Data. Nature Communications, 14(1), p.679.
2. Islam, M.T., Wang, J.Y., Ren, H., Li, X., Khuzani, M.B., Sang, S., Yu, L., Shen, L., Zhao, W. and Xing, L., 2022. Leveraging data-driven self-consistency for high-fidelity gene expression recovery. Nature Communications, 13(1), p.7142.
3. Islam, M.T. and Xing, L., 2023. Leveraging cell-cell similarity for high-performance spatial and temporal cellular mappings from gene expression data. In Press, Cell Patterns.
4. https://github.com/xinglab-ai/genomap