Recognizing Antigen Specificity and Functional Status of T cells by a Graph Deep Learning Model

Phi Le[1,2], Hai Yang[2],Tao He[3], Aixin Tan[4], Bridget P. Keenan[1,2], Li Zhang[1,2,5]

[1]*Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco;* [2]*Department of Medicine, University of California San Francisco, San Francisco;*

[3]*Department of Mathematics, San Francisco State University, San Francisco;* [4]*Department of Statistics, The University of Iowa,* [5]*Department of Epidemiology & Biostatistics, University of California San Francisco, San Francisco;*

BACKGROUND

- Immunotherapy has emerged as a promising approach in the treatment of cancer due to its ability to use the immune system to target cancer cells, and to combine with other treatment modalities, such as chemotherapy or targeted therapies, to enhance treatment effectiveness
- T cells are crucial components of the adaptive immune system, mediating anti-tumoral immunity and immune response to infections.
- T cell receptors (TCR) target specific antigens based on protein structure of nucleotide sequence. TCRs are highly varied and adaptable to antigens.

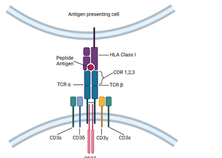


Image from [7]

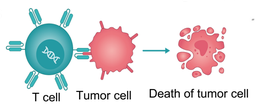


Image from [6]

- Investigating which TCRs can effectively bind to antigens or cancer cells has become a significant area of interest in immunotherapy research.
- Gene expression is the process by which the information encoded in genes is converted into functional products, scRNA-seq gene expression presents the number of RNA sequence reads corresponding to each cell
- The integration of genomics, proteomics, and other omics technologies has provided invaluable insights into the underlying molecular mechanisms of the disease, paving the way for personalized and targeted approaches to patient care.

OBJECTIVE

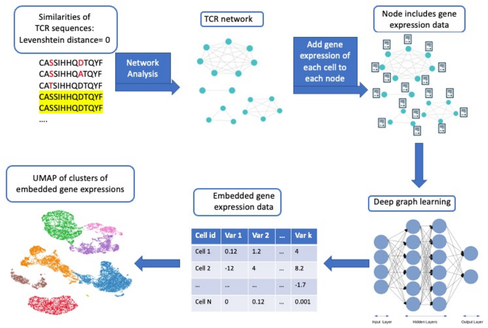
- To build a model that seamlessly links gene expressions and TCR sequencing data to refine clusters of functional T cells and their associated gene list

Illustration Datasets

- **CD40 agonist clinical trial**[9,10]:
  Esophageal/Gastroesophageal junction cancer patients received Sotigalimab (CD40 Agonist) treatments
- **10X Genomic datasets**[1]**:**
  10X data set is a collection of CD8+ T cells from four healthy donors. For each cell, it has TCR sequence, cell surface protein expression, and antigen binding information.

METHODS

Pipeline of Integrating TCRs and Gene Expression



- Construct a TCR network based on TCR similarities.
- Import gene expressions to graph as node features
- Build an unsupervised graph neural network on the built graph
- Extract embedded gene expression values
- Find clusters on embedded gene expression by Leiden cluster method

Deep Graph Learning Model

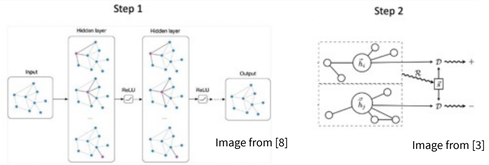


Image from [8]

Image from [3]

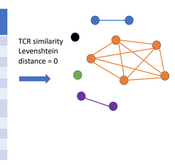
- Step 1: To capture the nonlinear relationships and differences in cell functions within the similarity network[2].
- Step 2: To learn the variations and matches in cell functions across different parts of the network through gene expressions[3].

Levenshtein Distance

Distance = 0: Cat = Cat
Distance = 1: Hat → Cat
at → Cat
Cat → at
Distance =2: Hate → Cat

TCR Network Analysis[4]

	Cell	TCR sequence	TCR Cluster id
1	CASSIHHQDTQYF	1
2	CASSIHHQDTQYF	1
3	CSAEGGGTEVFF	2
4	CSAEGGGTEVFF	2
5	CSAEGGDTVSFT	2
6	CASSPGTDSGNTIYF	3
7	CASSHPGGSDVQDTQYF	4
8	CASRRGAETLYF	5
9	CASRRGAETLYF	5
10	CSAEGGGTEVFF	2
11	CSAEGGGTEVFF	2



Leiden Cluster Algorithm[5]

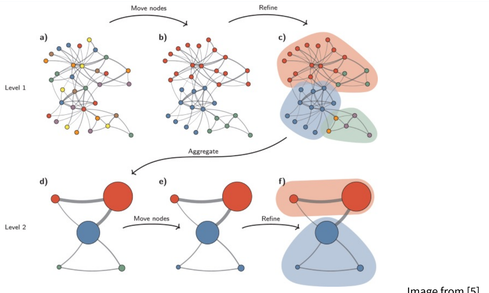
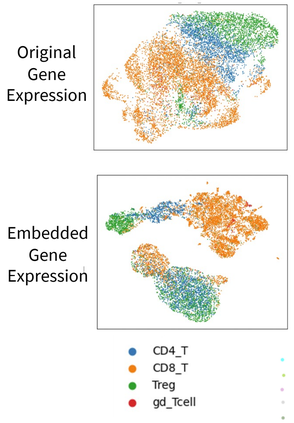


Image from [5]

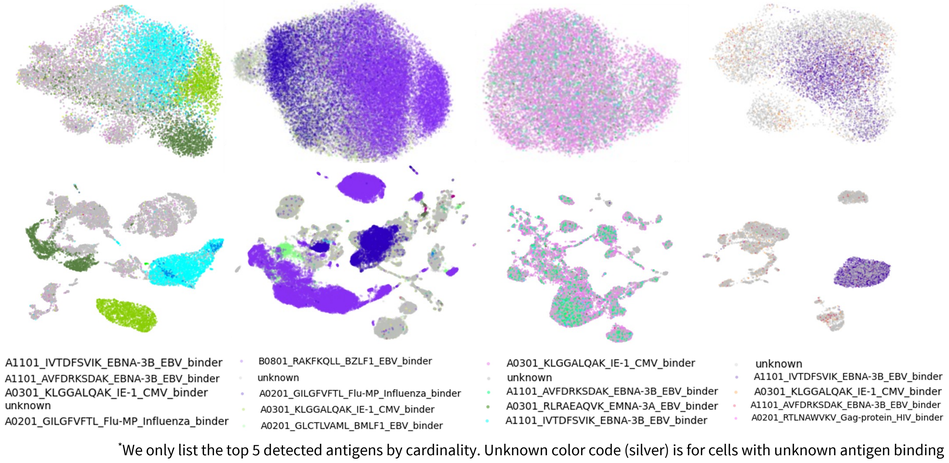
- Construct a graph where edge connection is the correlation between embedded gene expression profiles.
- Apply community detection to identify clusters of cells within the graph.
- Optimize cluster assignments by iteratively moving cells between clusters.
- Repeat optimization steps to find the optimal cluster assignments that maximize the within-cluster similarity and minimize the between-cluster similarity.

RESULTS

CD40 agonist clinical trial[9,10]



10x Genomics Data[1] with Antigen Binding Information*



*We only list the top 5 detected antigens by cardinality. Unknown color code (silver) is for cells with unknown antigen binding

Summary of Findings

- The embedded gene expression preserves the clusters of cell type calling for expanded TCRs similarity clusters
- The embedded gene expression can refine and separate different types of antigens clearly better then just using gene expression data
- The clusters obtained by the proposed pipeline outperforms than only using TCR clusters or gene expression clusters
- Performance check by Adjusted Rand Index** (ARI)

Index	Gene expression	TCR sequence	Embedded gene expression
Donor 1	0.562	0.508	0.76
Donor 2	0.159	0.163	0.242
Donor 3	-0.0003	-0.001	0.003
Donor 4	0.168	0.894	0.902

** ARI is a well-known index to measure cluster performance, its value is in [-1,1], larger is better

CONCLUSIONS

We developed a deep learning model that can seamlessly integrate single-cell gene expressions and T cell receptors. Our approach can

- Preserve the cell type calling for expanded TCR clusters
- Cluster cells into groups to predict antigen binding better than TCR network analysis

Limitation

- We can get up to 50% improvement of finding antigen binding clusters comparing to TCR network from 10X data
- In rare situation, we do not see significant differences between our method and TCR network
- Limited available data to test our model

REFERENCES

1. 10X genomics. *A New Way of Exploring Immunity. https://pages.10xgenomics.com/rs/446-PBO-704/images/10x_AN047_IP_A_New_Way_of_Exploring_Immunity_Digital.pdf.*
2. Hamilton WL, Ying R, Leskovec J. *Inductive Representation Learning on Large Graphs.* Published online September 10, 2018. Accessed June 3, 2023. http://arxiv.org/abs/1706.02216.
3. Veličković P, Fedus W, et al. *Deep Graph Infomax.* Published online 2018. doi:10.48550/ARXIV.1809.10341.
4. Yang H, Cham J, Fan Z, et al. NAIR: *Network Analysis of Immune Repertoire. Frontiers in Immunology* (Accepted). https://github.com/mlizhangx/Network-Analysis-for-Repertoire-Sequencing-.
5. Traag VA, Waltman L, Van Eck NJ. *From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep.* 2019;9(1):5233. doi:10.1038/s41598-019-41695-z
6. CAR T-cell therapy, *https://www.elicera.com/technology*
7. Sun Y, Li F, Sonnemann H, et al. *Evolution of CD8+ T Cell Receptor (TCR) Engineered Therapies for the Treatment of Cancer. Cells.* 2021;10(9):2379. doi:10.3390/cells10092379
8. https://theaisummer.com/Graph_Neural_Networks/
9. Ko A NM, Chao J, Sohal DPS, et al. *A Multicenter Phase 2 Study of Sotigalimab (CD40 Agonist) in Combination with Neoadjuvant Chemoradiation for Resectable Esophageal and Gastroesophageal Junction (GEJ) Cancers*. 2022 September 9-13, 2022. ; Paris, France. ESMO Annual Conference.
10. Maira Soto, Erin L Filbert, et al. *Use of high-dimensional and spatial immune profiling to explore sotigalimab (CD40 agonist) activation of antigen presenting cells and T cells in the tumor microenvironment in patients with esophageal/gastroesophageal junction cancer*. Journal of Clinical Oncology 2023 41:4_suppl, 450-450

Acknowledgement

- PL, YH, TH, BK and LZ are partially supported by NIH/NCI R21CA264381 and NIH/NLM R01LM013763-01A1.
- Bridget Keenan is supported by a UCSF Helen Diller Family Comprehensive Cancer Center Cancer Immunology and Immunotherapy grant and 1K12CA260225-01.

Corresponding author: philong.le@ucsf.edu