Distribution-free deconvolution of spatial transcriptomic data using heterogeneous single-cell datasets

J. Sodicoff[1], A. Kriebel[1], J. Welch[1,2]
[1]Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI
[2]Department of Computer Science and Engineering, University of Michigan College of Engineering, Ann Arbor, MI

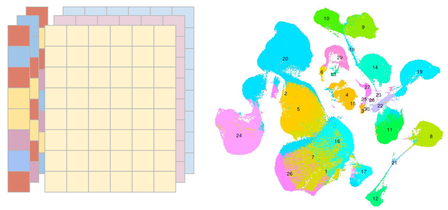
MEDICAL SCHOOL
UNIVERSITY OF MICHIGAN

Introduction

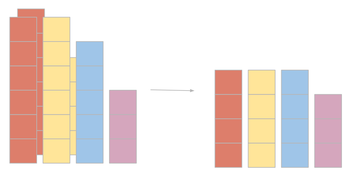
- Cell type deconvolution encompasses the set of computational methods attempting to determine the proportions and distributions of cell types within a tissue.
- The complex and heterogeneous structure of the brain is an ideal challenge for a cell type deconvolution method, and generated findings could provide insight into the relationship between its architecture and function in cognition, movement, and response to disease.
- *In situ* hybridization (ISH) allows for the imaging of gene expression by complementation of cellular RNA to a highly specific nucleic acid probe.
- The Allen Mouse Brain Atlas is a publicly available repository of ISH images for over 20,000 genes across the whole mouse brain.
- Quantification of staining into “expression energies” is available at the 200 μm^3 resolution, providing a more comprehensive, lower sparsity spatial expression reference than any other to date.
- Utilization of expression from the Allen Mouse Brain Atlas allows for association of derived data in the Allen Common Coordinate Framework (CCF), a 3D reference atlas of anatomical regions that many physiological modalities are also registered within.
- Here, we describe a cell type deconvolution method, Deconvolution using Matrix Factorization (DUMFound).
- DUMFound utilizes multiple scRNA-seq reference datasets to find maximally separating features between cell types, learn type-specific gene signatures, and deconvolve spatial expression data to determine absolute and relative expression of those types.
- Results generated with DUMFound from BICCN data processed into region-specific cell type atlases with LIRIpipe (see poster #426, Session 3) recapitulate known neural structure and show promise in answering new questions on the association between molecular and anatomical features of the brain.
- While DUMFound was developed for the unique characteristics of quantified ISH data, it provides robust results across spatial transcriptomic modalities.

Methodology

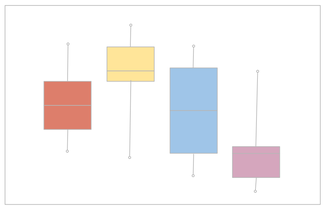
1. Single cell region atlas generation with LIRIpipe , yielding high resolution cell type annotations



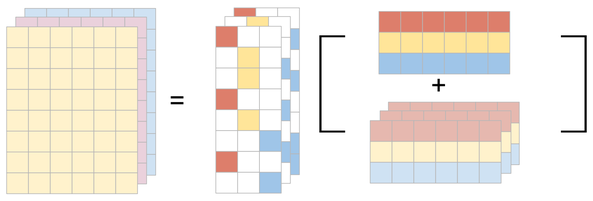
2. Sample RNA-seq cells datasets to reduce bias in gene signature calculation



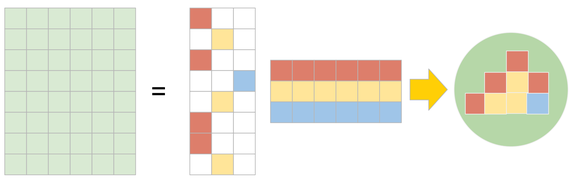
3. Kruskal Wallis one-way ANOVA to select genes that best separate cell types from all datasets



4. Alternating nonnegative least squares algorithm to determine shared gene signature that best reconstructs expression when multiplied by one hot matrix of cell types



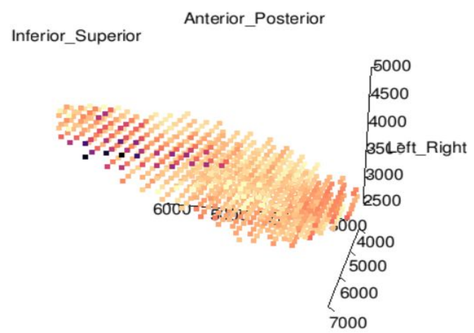
5. Nonnegative least squares algorithm to deconvolve spatial data



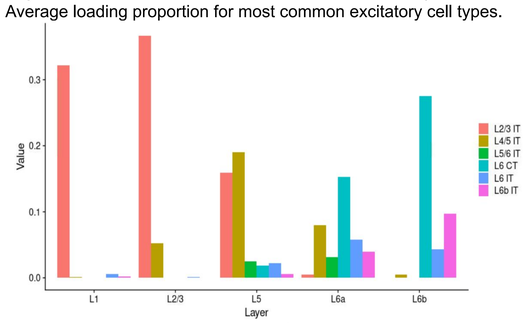
Results

All data in this section comes from an analysis of the primary motor cortex.

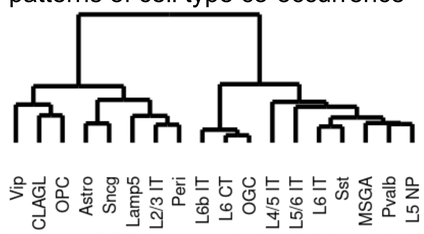
DUMFound predicts the distribution of cell types within the Allen Common Coordinate Framework.
Representation of Vip expression in one frame of summary 3D animation.



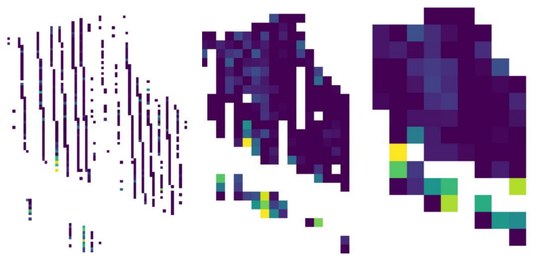
The method accurately recapitulates the layer-specific distribution of excitatory neurons.
Average loading proportion for most common excitatory cell types.



Hierarchical clustering of resulting distributions yields patterns of cell type co-occurrence



Voxelization of Slide-seq data yielded similar results to analysis of ISH data at multiple resolutions.
Predicted expression of L2/3 IT cells along an anterior-posterior slice at bead, 100 μm^3, and 200 μm^3 resolution



Discussion

- DUMFound builds on the LIRIpipe tool for cell type atlas generation to deconvolve expression data across the whole murine brain by region
- Results demonstrate biological accuracy of predictions and potential for novel insights
- The algorithm has several advantages over other existing deconvolution methods
  - Robustness to sparsity in spatial data
  - Ability to incorporate multiple reference datasets, matching the scale and complexity of current single cell analyses
  - Interpretability of gene signatures on account of non-negativity constraint
- While designed for deconvolution of expression as determined by quantified ISH, it shows promise in application to a range of spatial transcriptomic modalities
- Code for the pipeline can be found at https://github.com/Akriebs/BICCN/blob/main/R/deconv.R in advance of release as a package

Future Directions

- Adaptation of algorithm for direct, accurate prediction of bead and spot based modalities
- Implementation of advanced plotting utilities and a GUI for simplified interpretation
- Application to analysis of ISH data corresponding to 19 anatomical regions across the cortex, midbrain, and hindbrain
- Linkage of derived cell type distributions with anatomical modalities

Acknowledgement

Special thanks to:
- Chao Gao for technical support
- The Macosko Laboratory at the Broad Institute for providing Slide-seq data and technical support

References

Cable, D. M., Murray, E., Zou, L. S., Goeva, A., Macosko, E. Z., Chen, F., & Irizarry, R. A. (2022). Robust decomposition of cell type mixtures in spatial transcriptomics. *Nature Biotechnology*, *40*(4), 517-526.
Gao, C., Liu, J., Kriebel, A.R. *et al.* Iterative single-cell multi-omic integration using online learning. *Nat Biotechnol* 39, 1000–1007 (2021). https://doi.org/10.1038/s41587-021-00867-x
Mezias, C., Torok, J., Maia, P. D., Markley, E., & Raj, A. (2022). Matrix Inversion and Subset Selection (MISS): A pipeline for mapping of diverse cell types across the murine brain. *Proceedings of the National Academy of Sciences*, *119*(14), e2111786119.
Wang, Q., Ding, S. L., Li, Y., Royall, J., Feng, D., Lesnar, P., ... & Ng, L. (2020). The Allen mouse brain common coordinate framework: a 3D reference atlas. *Cell*, *181*(4), 936-953.
Yao, Z., Liu, H., Xie, F. *et al.* A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* 598, 103–110 (2021). https://doi.org/10.1038/s41586-021-03500-8