Parcellation of the mouse brain using molecular imaging

Alex Lee

1 Specific Aims

Parallel technological advances in molecular biology and microscopy have dramatically increased the ease with which experimental biologists can make spatially resolved measurement of large numbers of proteomic or transcriptomic features. These high-fidelity molecular mapping techniques are promising to catalog the enormous diversity of cell types in the complex organs like the brain, which has been recognized since the early work of Ramon y Cajal[1]. These new capabilities are the focus of large scale collaborations facilitated by the NIH BRAIN initiative[2], which aim to precisely elucidate the structural oragnization of the brain by generating multimodal, cross-species datasets for comparison. The datasets that will result from these projects will be unprecedented in size and complexity and will require new and innovative computational techniques to organize and analyze.

Two principal focus areas in the neuroscience community's investigation of these structure-function relationships are variation in gene expression[3] and connectivity[4,5], and how they interact in the circuitry of the brain. A variety of unsupervised and supervised machine learning techniques have been developed for extraction of subregions from spatial imaging data[6], however these methods typically target 2D tissue sections, making them unsuited for analysis of 3D volumetric data. In addition, there are few techniques suited for multimodal data integration in the specific case of connectivity and transcriptomic data, with most methods focus on spatial transcriptomics exclusively.

I propose to address this unmet need by developing computational methods for machine learning-based spatial subregion detection, with application to existing and in-development datasets from the Allen Institute for Brain Science. I will use these techniques to map the organization of functional subregions in spatial transcriptomics datasets in the mouse brain, and use interpretable statistical techniques to understand which cell type contribute most to these subregions. I will then make use of these techniques to integrate this spatial transcriptomics dataset with the existing Allen Mouse Connectivity Atlas[4], identifying genetic correlates of differential connectivity in transcriptomically similar cells.

1.1 Aim 1: establish a method for the parcellation and analysis of transcriptomic imaging data

I hypothesize that the mouse brain's subregions vary in terms of only a small number of the different cell types. To investigate this I will build on previous models for interpretable, stability-driven automatic subregion detection[7] and extend them to the setting of spatial transcriptomics datasets, which are both high dimensional and very sparse. I will then characterize these subregions by analyzing the relative contribution of different cell types to these different identified regions.

1.2 Aim 2: establish an analysis pipeline for the integration of transcriptomic and connectomic imaging data

I will study the relationship between gene expression and connectivity by developing statistical methods to test whether particular genes are indicative of differential connectivity. I will benchmark methods on their ability to capture previously identified connectivity-genetic relationships and test whether single cell features or regional features are more predictive of differential connectivity.

1.3 Significance:

some other stuff

2 Background

From the first neuroanatomical studies[1], it has been recognized that although cells can be broadly grouped into high-level categories (like neurons, or glia), even visual inspection of silver-stained cells is sufficient to identify a diverse array of cellular subtypes that differ along morphologic and connectomic characteristics. Understanding the role of this diversity in facilitating the complex computations the brain performs remains one of the fundamental projects of modern neuroscience research. However, the scope of this project has fundamentally changed in the modern 'omics age due to the accessibility of high dimensional molecular measurement tools. In particular, the ubiquity of high-fidelity molecular imaging data provides a unique opportunity to study the correspondence between our current and previous understanding of the organizing principles of brain structure and the patterns that can be identified in new datasets using modern statistical machine learning techniques.

The Allen Brain Atlas[8] (ABA) and Allen Mouse Connectivity Atlas[4] (ACA) are ideal datasets to computationally study this correspondence. The integration of automated high-throughput data collection methods, comprehensive informatics, and fully open data sharing have facilitated an enormous number of recent studies (as measured by the 3474 citations of the original Allen Brain Atlas Paper [as of 2023-02-18]), and in themselves were useful to characterize cellular heterogeneity not observed in classical histological neuroanatomical studies. These datasets were integral to the development of the Allen Common Coordinate Framework v3@wang_allen_2020, a systematic effort to integrate gene expression, connectivity, and transgenic expression in the form of a generalized reference coordinate framework with neuroanatomical consensus structural labeling.

These datasets will be augmented by new datasets using spatially resolved transcriptomic techniques, such as from Yao et al.[9], using MERFISH[10], and Langlieb et al.[11], using Slide-Seq[12].

but existing techniques for spatial subregion detection are poorly suited for analysis of large-scale, 3D imaging data. For example, one common class of methods utilizes spatial Gaussian processes[13] with $O(n^3)$ scaling, which is undesirable for large spatial (3D) imaging datasets. Other classes of promising methods would require assumptions to be made about the size of the spatial neighborhood for cellular interaction, which is difficult to assess with serial sections[14].

Single-cell sequencing studies[3,15,16] in the mouse have made significant progress towards charting cell type gradients as defined by transcriptomic content; likewise, computational efforts have been developed for very fine-grained connectivity mapping and clustering[17]. Molecular atlases such as the Allen Institute's Common Coordinate Framework[18] (CCF) atlas are critical to help organize inquiries into these phenomena by facilitating comparisons across studies and spatial scales. One strength of these atlases is their extensive manual curation, starting from the level of image registration and quality-checking towards consensus decisions by expert anatomists to delineate subregions and their boundaries.

3 Aim 1: establish a method for the parcellation and analysis of transcriptomic imaging data

Aim 2: establish an analysis pipeline for the integration of transcriptomic and connectomic imaging data

1. Cajal, S. R. y. *Texture of the nervous system of man and the vertebrates*. (Springer, 1999). doi:10.1007/978-3-7091-6435-8.

2. NIH BRAIN initiative launches projects to develop cell atlases and molecular tools for cell access. National institute of mental health (NIMH). https://www.nimh.nih.gov/news/science-news/2022/nih-brain-initiative-launches-projects-to-develop-cell-atlases-and-molecular-tools-for-cell-access (2022).

3. Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).

4. Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).

5. Sun, Y.-C. *et al.* Integrating barcoded neuroanatomy with spatial transcriptional profiling enables identification of gene correlates of projections. *Nat Neurosci* **24**, 873–885 (2021).

6. Walker, B. L., Cang, Z., Ren, H., Bourgain-Chang, E. & Nie, Q. Deciphering tissue structure and function using spatial transcriptomics. *Commun Biol* **5**, 1–10 (2022).

7. Wu, S. *et al.* Stability-driven nonnegative matrix factorization to interpret spatial gene expression and build local gene networks. *Proceedings of the National Academy of Sciences* **113**, 4290–4295 (2016).

8. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).

9. Yao, Z. *et al.* A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. (2023) doi:10.1101/2023.03.06.531121.

10. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).

11. Langlieb, J. *et al.* The cell type composition of the adult mouse brain revealed by single cell and spatial genomics. (2023) doi:10.1101/2023.03.06.531307.

12. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).

13. Townes, F. W. & Engelhardt, B. E. Nonnegative spatial factorization applied to spatial genomics. *Nat Methods* **20**, 229–238 (2023).

14. Brbić, M. *et al.* Annotation of spatially resolved single-cell data with STELLAR. *Nat Methods* **19**, 1411–1418 (2022).

15. Zeisel, A. *et al.* Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014.e22 (2018).

16. Bandler, R. C. *et al.* Single-cell delineation of lineage and genetic identity in the mouse brain. *Nature* **601**, 404–409 (2022).

17. Knox, J. E. *et al.* High-resolution data-driven model of the mouse connectome. *Netw Neurosci* **3**, 217–236 (2018).

18. Wang, Q. *et al.* The allen mouse brain common coordinate framework: A 3D reference atlas. *Cell* **181**, 936–953.e20 (2020).