Probabilistic inference of anatomical region underlying whole brain tiling

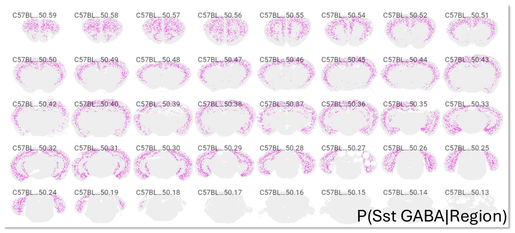
Inkar Kapen, Matthew Schmitz, Aaron Garcia, Rebecca D. Hodge, Trygve E. Bakken, Nelson J. Johansen, Allen Institute for Brain Science

Motivation

Building molecular atlases of the human brain requires dissecting complex anatomical regions in tiles which can contain multiple diverse functional. While this strategy enables comprehensive coverage of large brain regions, it can obscure the regional identities of the dissected and sequenced nuclei. The targeted dissection of brain regions can inadvertently include cells from neighboring anatomical regions which can complicate building regional cell type taxonomies.

- Region dissections are an assumed ground truth for an entire sequencing library.
- Cells from outside of dissected “ROI” can initially appear to be novel/distinct cell types when building regional taxonomies.

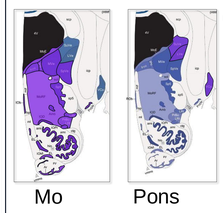
Research Outline

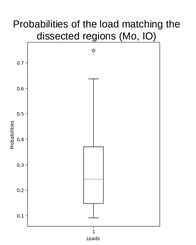


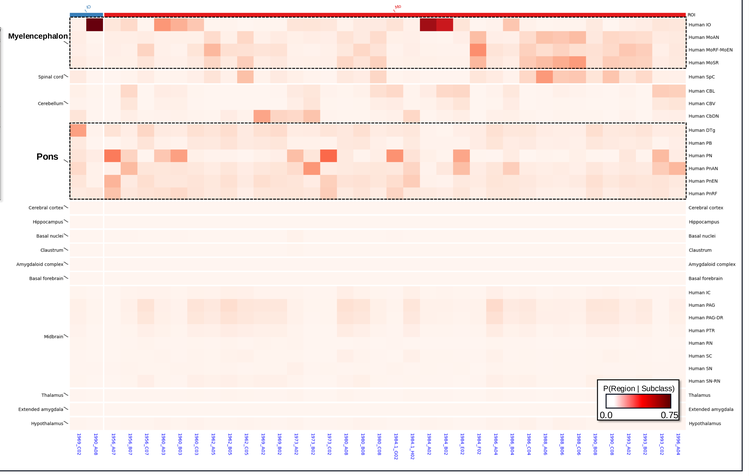
We introduce a probabilistic framework for inferring the anatomical region from which a cell was isolated. Using Bayesian statistics paired with a mouse whole brain reference atlas of spatially localized cell types, we can infer the likely brain region of origin for cells sampled from other mammalian brains, including human and non-human primates. Such a strategy enables the categorization of cells into the most likely captured anatomic regions.

Regional prediction for Midbrain dissections

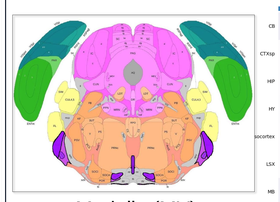
Using Siletti Human only to calculate P(Subclasses | Region) for Myelencephalon (Mo) and Inferior olive (IO).

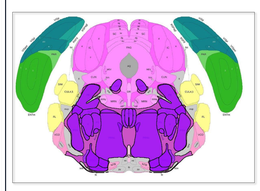


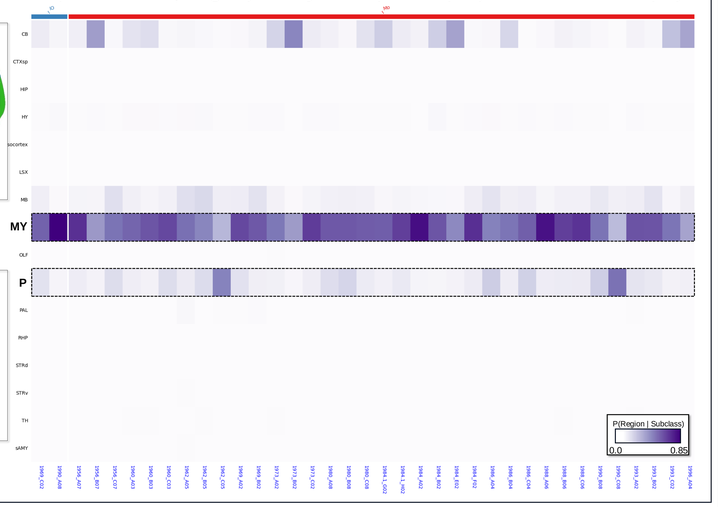




Using Mouse WB to calculate P(Subclasses | Region) for Myelencephalon (Mo) and Inferior olive (IO).

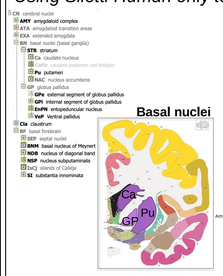


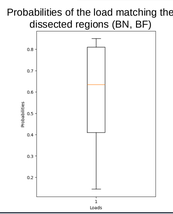


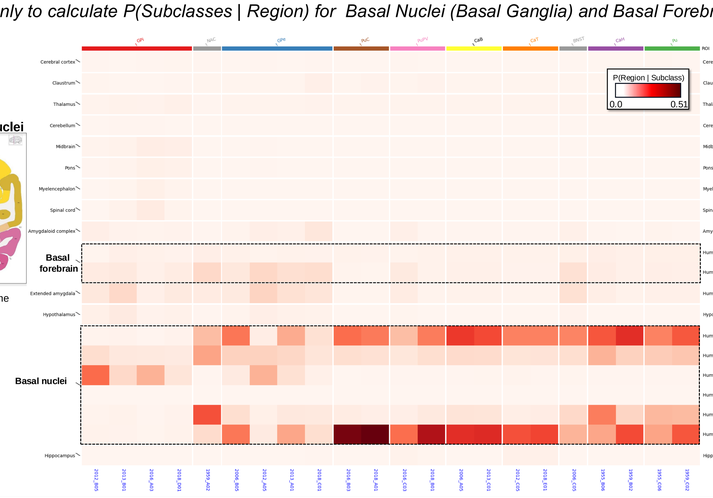


Regional prediction for Basal Ganglia+ dissections

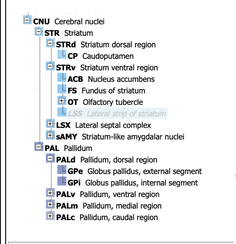
Using Siletti Human only to calculate P(Subclasses | Region) for Basal Nuclei (Basal Ganglia) and Basal Forebrain.

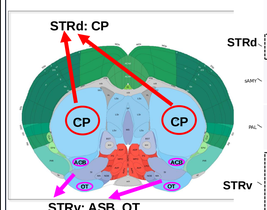


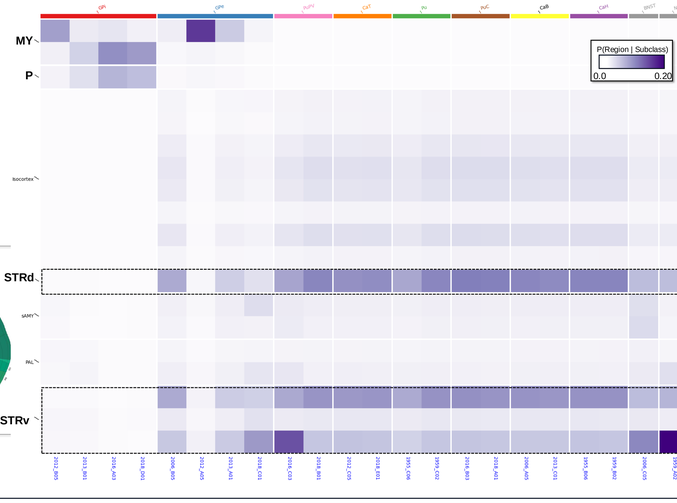




Using Mouse WB to calculate P(Subclasses | Region) for Basal Nuclei (Basal Ganglia) and Basal Forebrain.







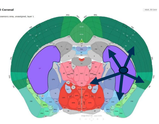
Research Outline

To computationally infer the anatomical region for individual cells from RNA-seq (10X Multiome), we are calculating the probability of the regions given the cells’ subclasses $P(Region \mid Subclasses)$. Each dataset has multiple loads for each dissected region.

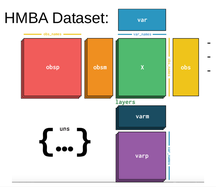
$P(Subclasses \mid Region)$: Calculated using MERFISH Mouse Whole Brain data.

$P(Subclasses)$: Model’s confidence score in predicted subclass labels per cell from the **MapMyCells** hierarchical algorithm.

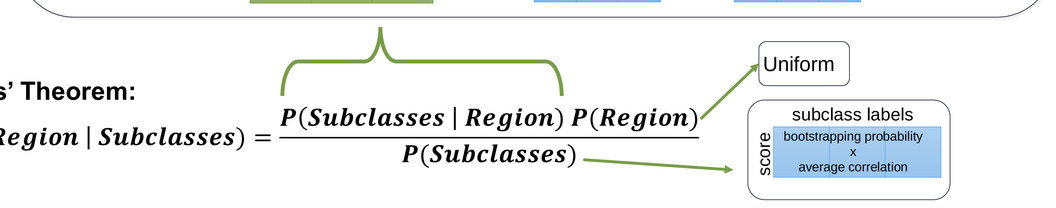
$P(Region)$: Uniform or based on neighboring regions.



HMBA Dataset:



- Cell_ids
- Load names
- Predicted subclass labels
- …



Bayes’ Theorem:

$$P(Region \mid Subclasses) = \frac{P(Subclasses \mid Region)\,P(Region)}{P(Subclasses)}$$

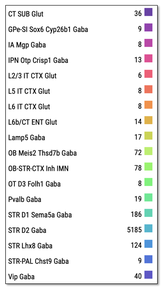
Uniform

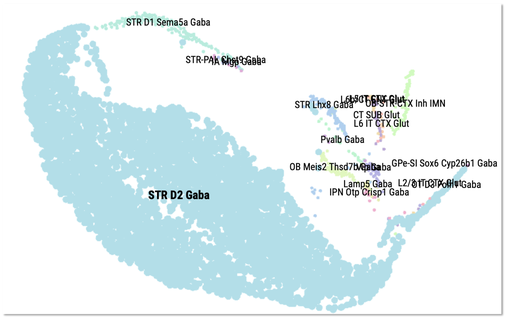
subclass labels

score: bootstrapping probability x average correlation

UMAP

Dataset: HMBA Basal Ganglia
Load name: 2018_A01





Siletti, Kimberly, et al. “Transcriptomic diversity of cell types across the adult human brain.” *Science* 382.6667 (2023): eadd7046.

Yao, Zizhen, et al. “A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain.” *Nature* 624.7991 (2023): 317-332.

Zhang, Meng, et al. “Molecularly defined and spatially resolved cell atlas of the whole mouse brain.” *Nature* 624.7991 (2023): 343-354.

brain-map.org

BICAN GRANT: 5U01MH114812-04