654.02

A deep-learning pipeline for studying olfactory bulb molecular anatomy at genomic scale

Alex J. Andonian[1], Daniel A. Paseltiner[2], and Jason B. Castro[2*]

1) Massachusetts Institute of Technology, Computer Science & Artificial Intelligence Laboratory, Cambridge MA

2) Bates College, Program in Neuroscience, Lewiston ME

1 Introduction

The olfactory bulb is organized into discrete columnar modules that are stereotyped across individuals, and defined by the convergence of idiotypic olfactory sensory neurons onto common glomeruli (1). Additionally, the bulb exhibits a coarser zonal organization, with functionally distinct subsystems defined by inputs from distinct receptor families (2). ***While the bulb's spatial heterogeneity at the level of inputs has long been appreciated, we understand comparatively less about whether this heterogeneity is preserved, discarded, or altered in the intrinsic circuits of the bulb.*** A major challenge in studying these and other possibilities of bulbar organization is the experimental throughput required to systematically investigate the many candidate molecules (ion channels, signaling molecules, etc) that may differ in expression across the bulb. To deal with this challenge, we have appealed to a genome-scale, neuroinformatic strategy employing the Allen Brain Atlas (ABA, (3)). Here, we describe our data-driven appraoch for registering thousands of in-situ hybridization images from the ABA, for the purpose of mapping and clustering gene expression in the bulb with laminar specificity.

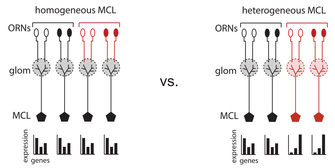

Figure 1. Possible models of bulbar organization. Mitral cells of the bulb may comprise a molecularly and physiolgically uniform population (left), such that readout of heterogeneous inputs is invariant. Alternatively, mitral cells may show molecular and functional specializations that allow for input-specific readout (right).

2 Concept and approach

Identifying candidate subregions of the mitral cell layer

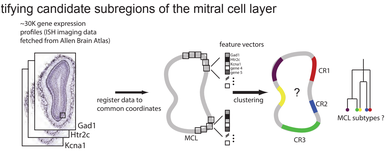

Image registration to a common coordinate frame

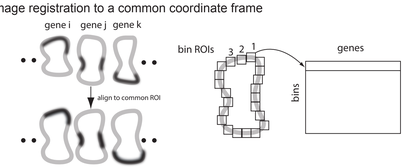

Figure 2. Identifying gene expression motifs in the mitral cell layer. Top: ISH data (all coronal sections comprising the bulb) were fetched from the Allen Brain Atlas using custom written scripts in Python, modified from (4). ISH data were registered to a set of common templates (using methods of (5), such that lamina are superimposable, and each point along the MCL can be conceived as a high-dimensional feature vector of gene expression. These feature vectors can then be clustered to reveal candidate subregions of MCL. Bottom: illustration of basic methods. ISH data for each gene is registered to a common coordinate frame, to populate a matrix of bins x genes for further analysis.

3 ISH Data from the ABA

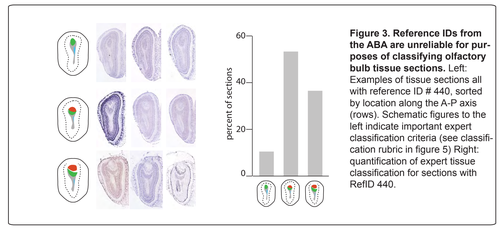

Figure 3. Reference IDs from the ABA are unreliable for purposes of classifying olfactory bulb tissue sections. Left: Examples of tissue sections all with reference ID # 440, sorted by location along the A-P axis (rows). Schematic figures to the left indicate important expert classification criteria (see classification rubric in figure 5) Right: quantification of expert tissue classification for sections with RefID 440.

4 Preprocessing

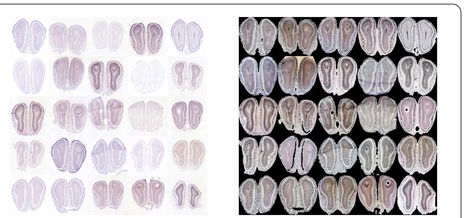

Figure 4. Examples of image preprocessing. Coronal sections fetched from the ABA were cropped, scaled, and downsampled before subsequent image classification and registration. Left: before pre-processing. Right: after pre-processing.

5 CNN Training

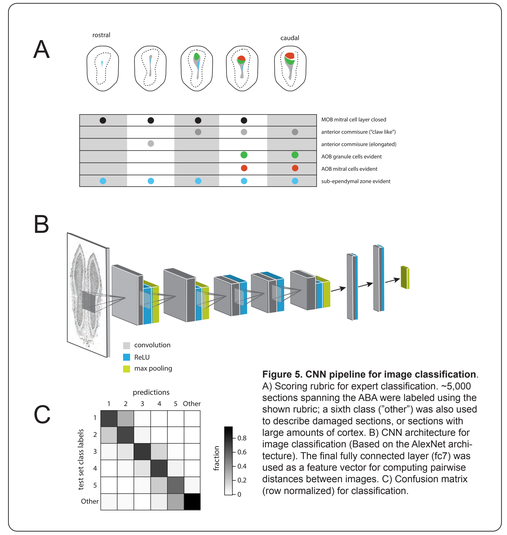

Figure 5. CNN pipeline for image classification. A) Scoring rubric for expert classification. ~5,000 sections spanning the ABA were labeled using the shown rubric; a sixth class ("other") was also used to describe damaged sections, or sections with large amounts of cortex. B) CNN architecture for image classification (Based on the AlexNet architecture). The final fully connected layer (*fc7*) was used as a feature vector for computing pairwise distances between images. C) Confusion matrix (row normalized) for classification.

6 ISH images in CNN feature space

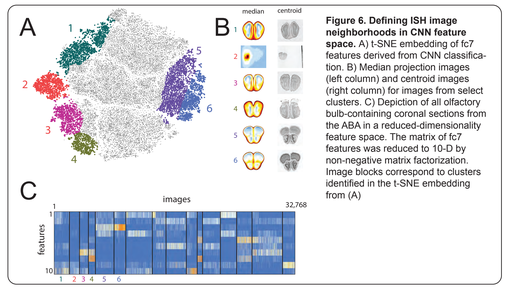

Figure 6. Defining ISH image neighborhoods in CNN feature space. A) t-SNE embedding of fc7 features derived from CNN classification. B) Median projection images (left column) and centroid images (right column) for images from select clusters. C) Depiction of all olfactory bulb-containing coronal sections from the ABA in a reduced-dimensionality feature space. The matrix of fc7 features was reduced to 10-D by non-negative matrix factorization. Image blocks correspond to clusters identified in the t-SNE embedding from (A)

7 Registration template design

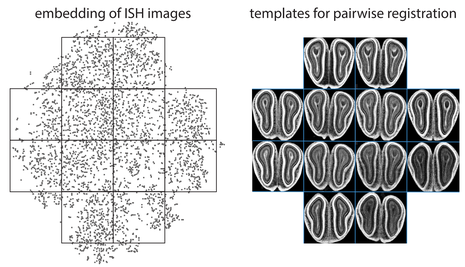

Figure 7. **Registration template design. Left:** t-SNE embedding of fc7 features for all ISH images with 'closed' mitral cell layers. Feature space was gridded into 12 equal-sized areas, and registration templates were constructed from 10 hand-chosen ISH images within each area. Right: templates constructed from group-wise registrations, and used for subsequent pairwise registrations of all images.

8 Image Registration

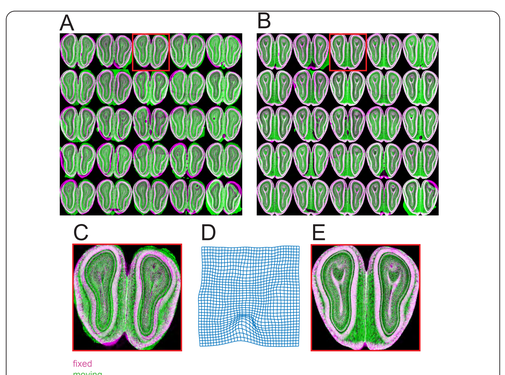

Figure 8. **Representative results of pairwise registration:** A) Montage of coronal sections before registration. Pink sections are fixed (i.e. template image), and green sections are those to be registered. B) Same montage after registration. C) Sample image pair (see red box in (A) before registration. D) Mesh showing transformation applied to the moving image in (C). E) Sample image pair after registration.

9 Preliminary clustering results

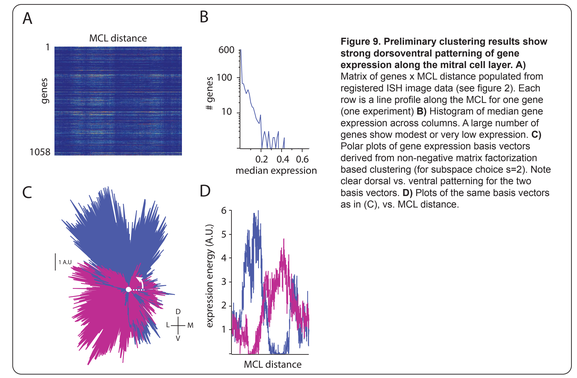

Figure 9. Preliminary clustering results show strong dorsoventral patterning of gene expression along the mitral cell layer. A) Matrix of genes x MCL distance populated from registered ISH image data (see figure 2). Each row is a line profile along the MCL for one gene (one experiment) **B)** Histogram of median gene expression across columns. A large number of genes show modest or very low expression. **C)** Polar plots of gene expression basis vectors derived from non-negative matrix factorization based clustering (for subspace choice s=2). Note clear dorsal vs. ventral patterning for the two basis vectors. **D)** Plots of the same basis vectors as in (C), vs. MCL distance.

10 Conclusions

>> We are studying spatial heterogeneity of gene expression, at genomic scale, along the mitral cell layer using the ISH data of the Allen Brain Atlas.

>> This endeavor requires image registration, which in turn necessitates robust classification of images. We have used a convolutional neural network (CNN) to derive a distance metric for olfactory bulb coronal sections.

>> We achieve ~85% success in pairwise image registration using 12 templates that span olfactory bulb feature space.

>> Our preliminary clustering results indicate strong dorsoventral patterning of genes along the MCL.

11 References

1) Buck, L., and Axel, R. (1991) A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell 65(1):175-87.

2) Munger, SD., Leinders-Zufall, T., and Zufall F. (2009). Subsystem organization of the mammalian sense of smell. Ann. Rev. Physiol. 71:115-40.

3) Lein et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature. 445(7124):168-76.

4) Ramsden, HL., Surmeli, G., McDonagh, SG., and Nolan, MF. (2015). Laminar and dorsoventral molecular organization of the medial entorhinal cortex revealed by large-scale anatomical analysis of gene expression. PLoS Comput. Biol. 11(1): e1004032.

5) Myronenko, A., and Song, X. (2010). Intensity-based Image Registration by minimizing residual complexity. IEEE Trans. on Medical Imaging. 29(11): 1882-1891.

Funding

JBC is supported by an NSF-CAREER award (1553279).