Poster presentation

NeuroBlast: a 3D spatial homology search tool for gene expression

Lydia Ng*, Chris Lau, Rob Young, Sayan Pathak, Leonard Kuan, Andrew Sodt, Madhavi Sutram, Chang-Kyu Lee, Chinh Dang and Michael Hawrylycz

Address: Allen Institute for Brain Science, Seattle, WA 98103, USA

Email: Lydia Ng* - lydian@alleninstitute.org

* Corresponding author

Background
Finding genes with a similar spatial expression pattern as a known gene could potentially reveal novel or unknown genes involved in similar processes or pathways. The Allen Brain Atlas (ABA) [1] is an effort to produce a genome-wide mapping of the gene expression in the adult C57BL/6J mouse brain. To date, more than 21,000 genes have been assayed using a high-throughput in situ hybridization (ISH) platform and the resulting image data is publicly available at http://brain-map.org. A major goal of the ABA project is to employ image analysis techniques to search the ABA data for particular expression patterns such as spatial gene expression homologues.

Methods
Each ISH image series is processed through an automated anatomic mapping pipeline with the goal of determining expression sites and the spatial localization of these sites with respect to a 3D reference brain. Expression statistics is then aggregated with respect to individual 200 μm^3 cubes in reference space thereby reducing data complexity from $> 2 \times 10^8$ pixels per series to $\sim 2.5 \times 10^4$ cubes. Since every image series is spatially mapped with accurate registration to the same 3D reference space, we can compare expression statistics on a global scale in approximately the same 200 μm^3 spatial extent for all series. The most straightforward approach to finding spatial homologues
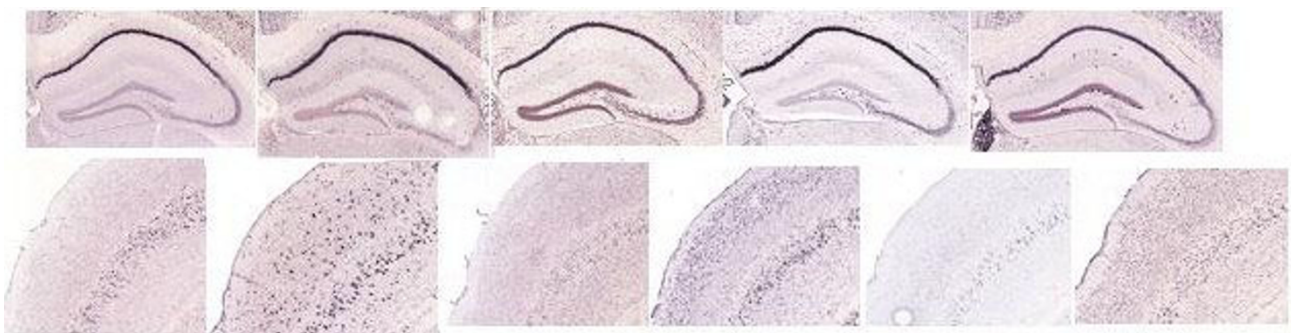
Figure 1
Search result examples.

is to compute the similarity between 200 μm^3 grid statistics of two genes. We conducted a pilot study computing the Pearson similarity score between every pair of over 4200 coronal series images. Two example searches are shown in Figure 1 with the initial seed gene shown in the left most panel: *Nov* (row 1) showing differential expression in CA1 of the hippocampus and *Etv1* (row 2) with enriched expression in layer 5.

Conclusion

We presented the functionality of NeuroBlast, a spatial search tool for finding genes with similar expression patterns within the ABA dataset. Preliminary testing of our pilot study has demonstrated efficacy of searches over different expression patterns and domain of interest. A comprehensive version of NeuroBlast spanning the entire ABA data will be publicly available in 2007.

References

1. Lein E, Hawrylycz M, *et al.*: **Genome-wide Atlas of Gene Expression in the Adult Mouse Brain.** *Nature* 2007, **445:**168-176.
2. Ng L, Pathak S, Kuan L, Lau C, Dong HW, Sodt A, Dang C, Avants B, Yushkevich P, Gee J, Haynor D, Lein E, Jones A, Hawrylycz M: **Genomic Scale Neuroinformatics for 3-D Gene Expression Mapping.** *IEEE/ACM Transaction on Computational Biology and Bioinformatics* 2007 in press.