A data-driven study of comparative molecular neuroanatomy

Emma M. Myers[1], Jason W. Bohland[2]

[1] Graduate Program for Neuroscience, Boston University, Boston, MA 02215 emyers@bu.edu
[2] Department of Health Sciences, Boston University, Boston, MA 02215 jbohland@bu.edu

Spatially resolved gene expression profiles carry information about how genes are co-expressed, in what brain areas genes or gene sets are expressed, and about the similarity of the molecular architecture of different brain areas. For comparative studies, these data may also provide insights into relationships between the genetic networks and brain systems of different species. We take a data-driven approach, using these patterns to explore and characterize relationships between mouse and human brain, two species for which large systematic datasets are now available. The human and mouse expression data used in these analyses are made available by the Allen Institute for Brain Science (AIBS). The AIBS is collecting genome-wide microarray profiles from a large set of region-specific tissue samples from postmortem human brains donated by neurologically healthy adults. At the time of writing, the collection of about 900 samples from each of two donor brains is complete, and a smaller set of samples is available from four additional donors. Mouse expression data from the AIBS are derived from *in situ* hybridization (ISH) studies in adult C57BL/6J mice. Data used here describe the expression of 3670 genes across 25,155 voxels. From these datasets, we can make direct cross-species comparisons, for example to determine if specific brain regions carry a common genetic "signature" (i.e., a vector describing expression levels of a set of genes) in mouse and human. An initial exploratory analysis was conducted by calculating these signatures for a set of human brain regions (each defined as the average across samples from that region of one human donor brain) and quantifying its correlation with the expression signatures at each voxel throughout the mouse brain. We made this comparison for each of twelve gross brain regions including cerebral cortex, cerebellum, thalamus, and hippocampus. In most cases, the human brain region's signature tended to have distinctly higher correlations with mouse voxels in the presumed homologous region than voxels elsewhere in the mouse brain (see Figure 1). While this tendency points to relative similarity in the genetic signatures of homologous regions, the somewhat low correlation values suggest potentially important differences as well. These results raise several questions: Which genes drive these correlations? Do certain genes drag them down because they are differentially expressed in mouse and human? Are certain classes of co-expressed genes of particular importance? To explore these questions, we used weighted gene coexpression network analysis (WGCNA)[1], in which genes are represented as nodes in a network. Connection weights between node pairs are determined by the correlations between the corresponding genes' expression patterns, such that strongly connected genes have similar spatial expression patterns. By choosing subsets of genes that are strongly connected, we can re-compute correlations between human and mouse datasets to determine how different gene sets influence regional identity across species. The overall approach may provide a starting point for a molecular-level understanding of homologous brain regions, and may also highlight sets of genes and sets of brain areas that are particularly distinct between mouse and human. As more neuroanatomically-specific gene expression datasets become available, this approach may be extended to make quantitative comparisons between other species as well.

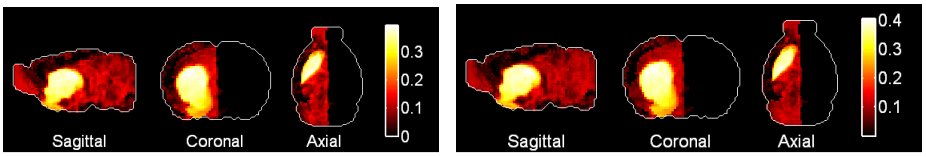

Figure 1. Maximum intensity projection images of the Pearson correlation between each left hemisphere mouse voxel and the signature of the human striatum, for two human donor brains. Voxels with the highest correlations are focused in the mouse striatum.

1. Zhang, B & Horvath, S. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology* 4, 17 (2005).

Support: NSF SMA-0835976 CELEST, PI Barbara Shinn-Cunningham.