Reconstruction of the mouse brain regions from the spatial gene expression data.

Tsvetkov A.I. (1) & Burtsev M.S. (2,3)

(1) Department of Cybernetics, National Nuclear Research University “MEPhI”, Moscow, Russia; (2) Non-linear dynamics department, Keldysh Institute of Applied Mathematics RAS, Moscow, Russia; (3) Anokhin Institute of Normal Physiology RAMS, Moscow, Russia

Abstract

Understanding the structural basis of nervous system still remains one of the primary challenges in neuroscience. New perspectives for the studies of brain anatomy are emerging with the availability of large-scale spatial gene expression data. Exploration of these data promises to deliver new insights into the understanding of relations between genes and brain structure.

We performed an analysis of the gene expression for more than 4, 000 genes in the mouse brain with the spatial resolution of 200 µm. These data are freely available from the Allen Brain Atlas [1] project (AGEA, http://brain-map.org/AGEA). After data filtering and scaling, distance matrices between all brain voxels (volumetric pixels) were calculated for a number of common distance measures. We also proposed and utilized a novel distance measure based on the locality of gene expression. Then the brain volume was hierarchically clustered with a variety of techniques [2]. In order to verify the resulting brain parcellations we compared them by means of a quantitative measure to the histologically-defined anatomical reference atlas and found significant overlap between automatically derived spatial clusters and anatomical structures.

We also compared brain regions by the number of genes expressed. Here the number of simultaneously activated genes was considered to be proportional to the amount of accumulated evolutionary changes. This allows us to identify dentate gyrus and adjoining hippocampal regions as the structures with the richest expression and, hence, less conservative in the evolutionary sense. Other regions of the cerebral cortex show intermediate levels of expression variability while other brain structures have accumulated fewer changes in the process of evolution, with olfactory bulbs and cerebellum being most conservative.

This work was supported by Russian FTP "Scientific and scientific-pedagogical personnel".

1. Ng L., Bernard A., Lau C. et. al., “An anatomic gene expression atlas of the adult mouse brain, ” Nature Neuroscience, vol. 12, no. 3, pp. 356-362, 2009.
2. Xu R. and Wunsch D., “Survey of clustering algorithms, ” IEEE Transactions on Neural Networks, vol. 16, no. 3, pp. 645-678, 2005.

Correlation maps

The spatial locality of gene expression is defined for each gene as follows:

$$\overline{E}_g = \frac{1}{N}\sum_{i=0}^{N} E_{ig} \qquad s_g = \sqrt{\frac{1}{N}\sum_{i=0}^{N}(E_{ij}-\overline{E}_g)^2} \qquad sp_g = \frac{1}{\overline{E}_g s_g}$$

Several common distance formulas were modified in order to take into account the locality of gene expression:

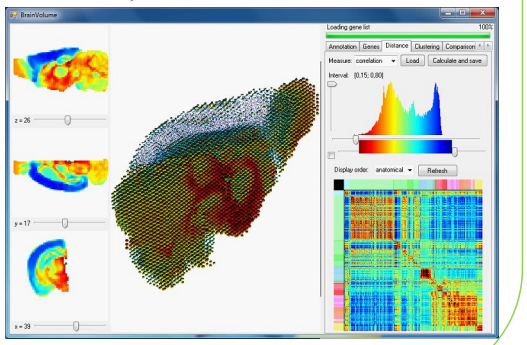
$$D_{ij} = \sqrt{\sum_{g=1}^{N_g}((E_{ig}-E_{jg})^2 sp_g)} \qquad D_{ij} = \sum_{g=1}^{N_g}\left|E_{ig}-E_{jg}\right|\cdot sp_g$$

We also developed several distance measures based on quantization of expression values:

$$Sp_{ijg} = \begin{cases} sp_g & if\ E_{ig}, E_{jg} \in [E_{\min}; E_{\max}], \\ 0 & else, \end{cases}$$

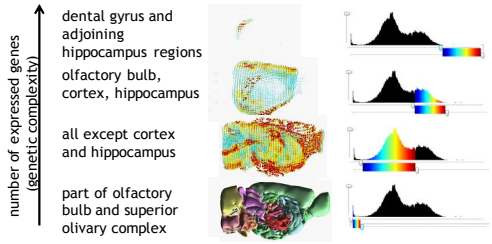
$$D_{ij} = 1 - \frac{1}{N_{ij}^{Expr}}\sum_{g=1}^{N_g} Sp_{ijg},$$

$$D_{ij} = \frac{1}{N_g}\sum_{g=1}^{N_g} S_{ijg} sp_g$$

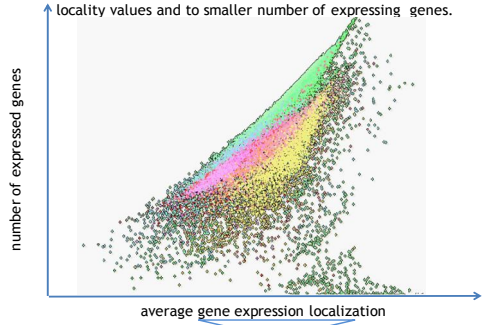


Results

Estimating the genetic complexity of brain regions by gene expression data analysis.



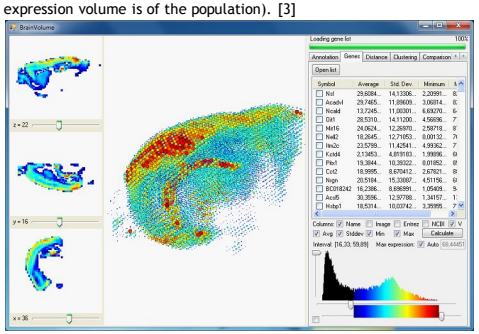
Voxels are plotted in coordinates of average gene expression locality by the number of expressed genes. Brain regions are color coded by the Allen Reference Atlas colormap. We suppose that earlier evolutionary origin of anatomical structures corresponds to lower spatial locality values and to smaller number of expressing genes.



Experimental data

While substantial quality control measures were employed in the creation of the ABA, it remains possible that data for individual genes contain artifacts or other problems that result in poor reproducibility. For example, tissue folding, dust, or air bubbles on slides could potentially give rise to incorrect determination of expressing cells, or particular genes or probes may simply produce highly variable results. For this study, we sought to use a suitably large set of genes while maintaining at least some control over the quality of the individual data volumes (e.g. how representative a particular expression volume is of the population). [3]

The list of genes considered in further analysis can be filtered by the values of statistical characteristics, such as statistical moments, correlation coefficient between datasets for the same gene, total volume of regions with identified expression.

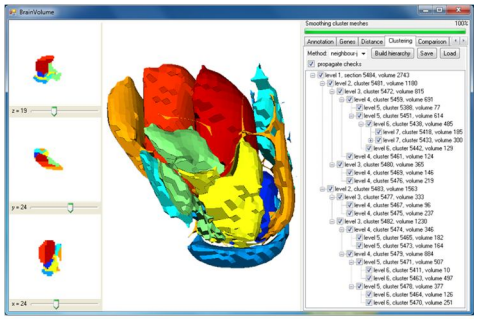


3. J.W. Bohland et al., Methods (2009), doi:10.1016/j.ymeth.2009.09.001

Hierarchical clustering

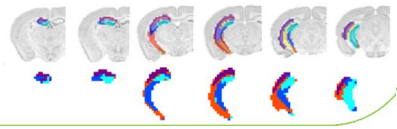
distance as average locality of coexpressed genes
+
agglomerative clustering with neighbor-joining measure

$$I(A,B) = \frac{V(A \cap B)}{V(A \cup B)}$$



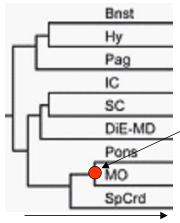
Striatum dorsal region	82%
Caudoputamen	82%
Striatum ventral region	93%
Nucleus accumbens	88%
Fundus of striatum	73%
Olfactory tubercle	72%
Lateral septal complex	75%
Lateral septal nucleus	80%
Septofimbrial nucleus	46%
Striatum-like amygdalar nuclei	54%
Central amygdalar nucleus	50%
Anterior amygdalar area	62%
Medial amygdalar nucleus	63%
Pallidum	52%
Pallidum dorsal region	84%
Pallidum ventral region	74%
Magnocellular nucleus	76%
Substantia innominata	79%
Pallidum caudal region	62%
Bed nucleus of the commissure	100%
Bed nuclei of the stria terminalis	61%
Pallidum medial region	68%
Average	**72%**
Standart deviation	**14%**

“Genomic Anatomy of the Hippocampus”
C.Thompson,S.Pathak,A.Jeromin et al., Neuron, 2008
HIP ARA region clustering results



Conclusion and perspectives

- The software tool for 3D visualization, analysis and clustering of the mouse brain model using the gene expression data was developed and tested with experimental data.
- Clustering results in the form of spatial region masks closely match the reference anatomical atlas.
- Gene expression data analysis performed in the project highlighted the difference in gene expression complexity between brain regions.
- The resulting brain region tree can also be analyzed using gene ontology and gene duplication data.
- By associating gene duplication times with brain region evolutionary origin times we may refine the evolutionary history of brain functions.



Gene Sym	Correlation	Gene duplication time
Etv1	1.0	5,4
Etv1	0.877498	4,3
9130213B05Rik	0.720919	3,8
Chst8	0.703844	3,1
Serpine2	0.699663	2,6
6430573F11Rik	0.688416	1,6
9130213B05Rik	0.664911	3,7