Journal of Neuroscience Methods 203 (2012) 212–219

Contents lists available at SciVerse ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Study of cerebral gene expression densities using Voronoi analysis

Mauro Miazaki[a], Luciano da F. Costa[a,b,*]

[a] Institute of Physics at São Carlos, University of São Paulo, São Carlos, SP, Brazil
[b] National Institute of Science and Technology of Complex Systems, Niterói, RJ, Brazil

ARTICLE INFO

Article history:
Received 11 May 2011
Received in revised form 5 September 2011
Accepted 14 September 2011

Keywords:
Voronoi analysis
Spatial density correlation
Gene expression
Mouse brain

ABSTRACT

As the available public cerebral gene expression image data increasingly grows, the demand for automated methods to analyze such large amount of data also increases. An important study that can be carried out on these data is related to the spatial relationship between gene expressions. Similar spatial density distribution of expression between genes may indicate they are functionally correlated, thus the identification of these similarities is useful in suggesting directions of investigation to discover gene interactions and their correlated functions. In this paper, we describe the use of a high-throughput methodology based on Voronoi diagrams to automatically analyze and search for possible local spatial density relationships between gene expression images. We tested this method using mouse brain section images from the Allen Mouse Brain Atlas public database. This methodology provided measurements able to characterize the similarity of the density distribution between gene expressions and allowed the visualization of the results through networks and Principal Component Analysis (PCA). These visualizations are useful to analyze the similarity level between gene expression patterns, as well as to compare connection patterns between region networks. Some genes were found to have the same type of function and to be near each other in the PCA visualizations. These results suggest cerebral density correlations between gene expressions that could be further explored.

1. Introduction

Since the pioneering work of Cajal (1989), the "unaccustomed" variety of neuronal morphology, i.e. the diversity of neuronal cell shapes, has attracted increasing attention. Indeed, the multiplicity of shapes range from relatively simple forms, such as bipolar cells, to complex structures, like Purkinje cells (Masland, 2004; Bota and Swanson, 2007). This variety in morphology has influence in the emerging dynamics of a neuronal system, whose functions are constrained and even defined by the individual neuronal shapes (Agmon-Snir et al., 1998; Segev, 1998; Costa et al., 2002; Jan and Jan, 2003; Wen and Chklovskii, 2008; Pérez-Reche et al., 2010). The reason of such diversity is that although all cells in an organism share the same DNA, each cell type expresses only a specific subset of genes, resulting in cellular differentiation. This subset of expressed genes in a cell (the expression pattern) is influenced by the environment, internal signals and signals from other cells (Alberts et al., 2002). Therefore, the spatial distribution of gene expression densities can yield important insights about cell morphology and physiology. In addition, genes with similar expression patterns can present similar and connected function (Eisen et al., 1998; Ben-Dor et al., 1999; Hastie et al., 2000; Ross et al., 2000; Costa et al., 2007a; Namand and Kim, 2008).

The important role of the expression pattern in the organism formation can be observed in a well studied case: the development of the head to tail (anterior–posterior) axis in the *Drosophila melanogaster* embryo. A few hours after the fertilization, it is possible to observe the pattern distribution of the Bicoid, Even-skipped and Caudal proteins. These proteins are synthesized by three different genes and their presence indicates that their respective genes are activated. The *bicoid* gene expression is more concentrated in the anterior region, while the *caudal* gene expression is more active on the opposite side (posterior region). The *even-skipped* gene expression form a striped pattern of seven bands perpendicular to the anterior–posterior axis. Later, these stripes will regulate several other genes that will define the embryo division in segments (Gilbert, 2003). In the brain, the gene families involved in the complex and dynamic processes of development and regionalization, starting with precursor cells, which divide and originate neurons that migrate, differentiate and create synaptic connections, have been gradually identified by the research community (Kammermeier and Reichert, 2001; O'Leary et al., 2007; Krubitzer, 2009). For instance, the pioneering studies in the *Drosophila melanogaster* allowed the identification of the *ems* and *otd* homeobox gene families as having important roles in brain

* Corresponding author at: Institute of Physics at São Carlos, University of São Paulo, PO Box 369, CEP 13.560-970, São Carlos, SP, Brazil. Tel.: +55 16 3373 9858; fax: +55 16 3373 9879.
E-mail addresses: mauro@ursa.ifsc.usp.br (M. Miazaki), luciano@ifsc.usp.br (L.d.F. Costa).

0165-0270/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jneumeth.2011.09.009

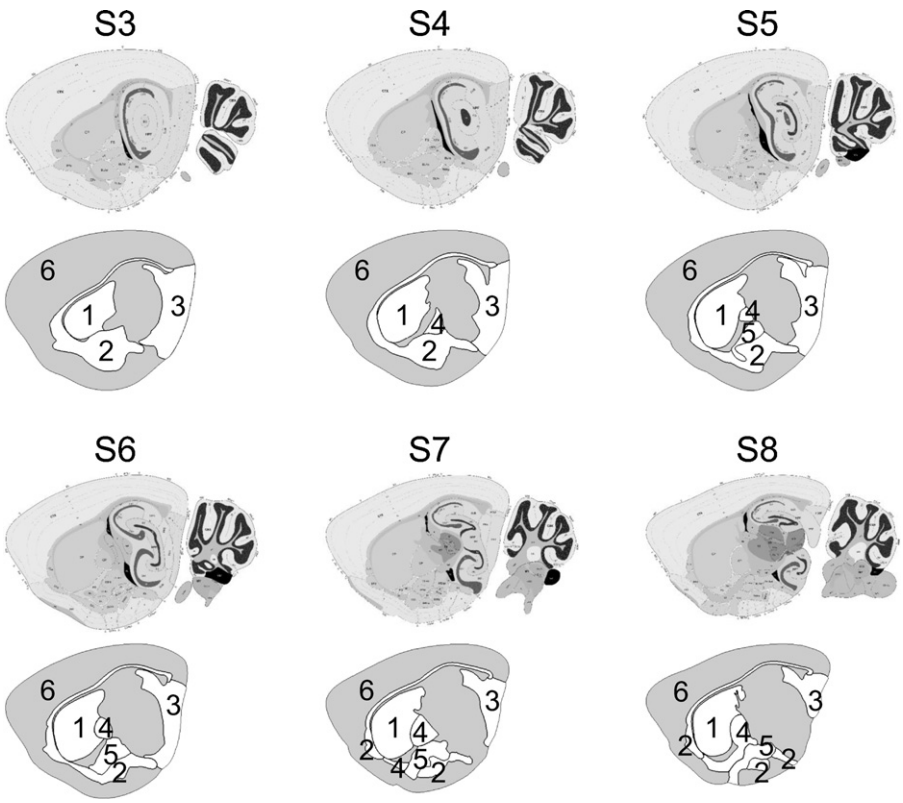

Fig. 1. Reference atlas of the selected sagittal sections from S3 to S8 (Section 3 to Section 8) and the chosen regions: (1) CP, (2) CTXsp, (3) PAL, (4) RHP, (5) sAMY and (6) brain (all previous regions and the gray area).

development. Later, the mouse homologues were also identified: *Otx1*, *Otx2*, *Emx1* and *Emx2* (Boncinelli et al., 1993; Cecchi et al., 2000).

Therefore, in order to contribute with an automated procedure of expression pattern investigation, we propose in this work a high-throughput methodology to compare the spatial density distribution of gene expressions based on the Voronoi Local Density Analysis (VLDA) method and procedures to analyze the generated information using networks and Principal Component Analysis (PCA) (Duda et al., 2001). The VLDA was first proposed by Costa et al. (2007b). This methodology uses Voronoi diagrams to quantify local density and uncover possible spatial relationships. This method was compared (Costa et al., 2007b) with the Ripley *K* function in a simulated mosaic image and in the analysis of the spatial interrelation between the normally placed INL (inner nuclear layer) and displaced GCL (ganglion cell layer) amacrime cells in the thraira retina. The Voronoi-based method was able to detect the interrelationship in both cases. The compared retinas were two-dimensional images. In this work, we use the Voronoi-based methodology to analyze the spatial density distribution and compare gene expressions considering three-dimensional volumes, composed of several image sections. In addition, we also compare the expression of eight different genes in six cerebral regions. The analyzed data was obtained in the Allen Mouse Brain Atlas, which comprises high-resolution expression images for about 20,000 genes (Lein et al., 2007). Large public online databases like this one are becoming common on the web with the recent advances in technologies for data storage, management and transmission. Such databases allow worldwide researchers from any field to easily access these materials and carry out their studies. This also permits the replication of experiments and results, as anyone can access the same data. Moreover, it brings a challenging task in the exploration and analysis of such large databases, indicating that the only suitable possibility is the use of high-throughput automated processing, as proposed in this work.

In the following section, we present the database and the selected data, as well as the methodology for the analysis (Section 2). Afterwards, we present and discuss the results (Section 3) and concluding remarks (Section 4).

2. Materials and methods

The data used in this work are images from the Allen Mouse Brain Atlas database (see Section 2.1). Before the analysis, we performed the image registration as described in Section 2.2. In the following Section 2.3, we introduce the concept of a Voronoi diagram, which is used on the VLDA methodology (Section 2.4). The results are generated as a network that can also be visualized in PCA (Section 2.5).

2.1. Allen Mouse Brain Atlas database

The images used in this study are from the public database Allen Mouse Brain Atlas (Lein et al., 2007), which is maintained by the Allen Institute for Brain Science (Jones et al., 2009). This database comprises high resolution images of brain slices that show the distribution of the gene expression intensity of about 20,000 genes, as well as documentations, papers and tools for data exploration and visualization. The tissues were collected from 56-day old adult male C57BL6/6J mice purchased from the Jackson Laboratory. The brains were dissected, frozen, sectioned, stained using *in situ* hybridization (ISH) technique, and photographed. ISH uses nucleic acid probes that are complementary to and hybridize with a specific DNA or mRNA sequence, allowing the location of a particular gene expression (Hicks et al., 2004). More details about the