Neuroinform
DOI 10.1007/s12021-012-9150-5

ORIGINAL ARTICLE

Morphological Homogeneity of Neurons: Searching for Outlier Neuronal Cells

Krissia Zawadzki · Christoph Feenders · Matheus P. Viana · Marcus Kaiser · Luciano da F. Costa

Abstract We report a morphology-based approach for the automatic identification of outlier neurons, as well as its application to the *NeuroMorpho.org* database, with more than 5,000 neurons. Each neuron in a given analysis is represented by a feature vector composed of 20 measurements, which are then projected into a two-dimensional space by applying principal component analysis. Bivariate kernel density estimation is then used to obtain the probability distribution for the group of cells, so that the cells with highest probabilities are understood as archetypes while those with the smallest probabilities are classified as outliers. The potential of the methodology is illustrated in several cases involving uniform cell types as well as cell types for specific animal species. The results provide insights regarding the distribution of cells, yielding single and multi-variate clusters, and they suggest that outlier cells tend to be more planar and tortuous. The proposed methodology can be used in several situations involving one or more categories of cells, as well as for detection of new categories and possible artifacts.

Keywords neuromorphometry · Archetypes · Outliers · *NeuroMorpho.org* · Neuroscience

K. Zawadzki · M. P. Viana · L. da F. Costa (✉)
Institute of Physics at São Carlos, University of São Paulo, PO Box 369, São Carlos, SP, CEP 13.560-970, Brazil
e-mail: luciano@ifsc.usp.br

C. Feenders · M. Kaiser (✉)
School of Computing Science, Newcastle University, Claremont Tower, Newcastle-upon-Tyne, NE1 7RU, UK
e-mail: m.kaiser@ncl.ac.uk

C. Feenders
Institute for Chemistry and Biology of the Marine Environment, Carl von Ossietzky University, Oldenburg, Germany

M. Kaiser
Institute of Neuroscience, Newcastle University, Newcastle-upon-Tyne, NE2 4HH, UK

M. Kaiser
Department of Brain and Cognitive Sciences, Seoul National University, Seoul, Republic of Korea

Introduction

Developments in neuroscience during the last decades have mainly focused on electrophysiological and functional recordings such as EEG and fMRI and relatively little attention has been given to the role of neuronal morphology for behavioral characteristics of an organism. For a long time, when eletrophysiology was the main tool of neuronal data acquisition, the neuroscientific interest was focused on describing a mature system based on its functional properties, which are mostly accounted for by the respective synaptic connections and respective strengths (Stepanyants and Chklovskii 2005; Kaiser and Hilgetag 2006). However, such a general connectivity follows the geometry of neurons as well as their sizes and positions (Stepanyants et al. 2002; Sporns et al. 2004). This fact implies that the full understanding of the nervous system's organization and behavior depends critically on these three geometrical elements (size, position, and shape), and several related works have been reported recently (Costa et al. 2002; Stepanyants and Chklovskii 2005; Kaiser et al. 2009). The comprehensive investigation of the relationship between neuronal geometry, con-

Published online: 22 May 2012

nectivity, and function was also constrained by the lack of experimental data and powerful concepts and methods for the respective characterization. Ramon y Cajal (Cajal 1989) was one of the first to systematically investigate neuronal shape and subsequent works studied the electrophysiology of nerve cells and chemical transmission of nerve impulses (Loewi 1921, 1955). Regarding the neuroanatomical approach, there was a trend of focusing on long-distance connectivity by techniques such as those of Nissl and Marchi, precursors of Cajal. More recently, we have witnessed a surge of interest in neuronal morphology, including Automated Sholl Analysis (Sholl 1953), fractal dimension characterization (Montague and Friedlander 1991; Binzegger et al. 2005), influence area analysis (Toris et al. 1995), and dendrogram representation (Poznanski 1992). The growing importance of neuromorphology is further substantiated by the fact that it has become a very important starting point for reference programs like Neuron (www.neuron.yale.edu/neuron) and Neuroconstruct (www.neuroconstruct.org) (Gleeson et al. 2007).

In a series of seminal works, McGhee (2006) introduced the concept of biological shape space, where different measures are applied to neurons in order to represent the cells' features quantitatively (*feature vector* of a cell). Each feature vector corresponds to a point in the *morphological space*, but not every point in this space corresponds to an existing shape—i.e. a biologically viable one. The morphological space is thus split into viable and impracticable geometrical shapes. Further restrictions of this space have been identified by McGhee based on functionality, development, and phylogenetics. The remaining subset corresponds to the biological shapes found in nature. Costa et al. (2010) used a generic model of neuronal shape to estimate the subspace of geometrically viable neurons. By using the data available in the *NeuroMorpho.org* database, the subspace of real neurons could be approximated and was found to occupy a relatively small portion of the geometrically viable subspace. This suggests that in addition to geometrical restrictions, several biological constraints determine the shape of real neurons. To better understand what these biological constraints are, we identify both prototypical and outlier neurons in the morphological space. This is done by noticing that neurons with similar geometrical properties lie close to each other in morphological space. Many cells with similar properties—usually well-defined neuronal cell categories—give rise to dense *clusters* (Costa et al. 2010). The most central points of a cluster have features most similar to other members of the cluster and can thus be seen as *archetypes*; those at the border of the respective cluster will be termed *outliers*. Differences between typical archetypes and deviating outlier cells can provide indications of biological limitations on geometry.

Improvements on techniques of visualization, acquisition, and sharing of data allowed the development of automatic methods to classify neurons according to their morphological features (Costa and Velte 1999; Cook 1998; Costa et al. 2002; Schierwagen 2008; Wen and Chklovskii 2008). We previously developed and applied a methodology in order to automatically identify outlier nodes in complex networks (Echtermeyer et al. 2011; Costa et al. 2009; Echtermeyer et al. 2011), which we adapted to analyze neuronal features.[1] The method is based on multivariate statistics and concepts of pattern recognition, which operate on feature vectors of each network node. When analyzing complex networks, feature vectors contain measurements of a node's local topology, such as degree and clustering coefficient. Instead of these network features, here we use morphological features of neurons. The corresponding vectors are then projected into a lower dimensional space (using Principal Component Analysis) (Härdle and Simar 2007) to create a density map (using a bivariate kernel density estimator) (Botev et al. 2010) that reflects the point distribution in the PCA-plane. Regions of high and low density then indicate archetypes and outliers in morphospace, respectively.

Archetype cells are interesting because they can serve as a "blueprint" for a particular cell-type. Identifying outliers could not only be useful for artifact detection, in the sense that a cell that has been improperly sampled appears as an outlier, but outlier neurons further provide insights about the most extreme viable neuronal shapes, since they mark corresponding boundaries. The proposed algorithm could be applied for diagnosis. As an example, we could apply the method for measurements of cells extracted from a tissue, classifying them as healthy or not. Additionally, the method fosters understanding of which features cause a neuron to be an outlier. In the present work, we therefore identify both archetypes and outliers for neuronal cells in the *NeuroMorpho.org* database. Two types of experiments are reported. First, we consider four cell types irrespective of species, laboratory or brain region. Then, we investigate six cell types respective to a specific species.

The following section starts with the presentation of the *NeuroMorpho.org* database, followed by a descrip-

[1]The software DONE (Detection of Outlier NEurons) is an implementation for MATLAB that is available on our web-site (http://www.biological-networks.org/p/outliers/).

tion of the available morphological measurements and the explanation of the analysis methods used. Results obtained using data from all neuronal cells as well as insights from the group-wise analysis follow. Finally, we discuss and summarize our findings.

Materials and Methods

The *NeuroMorpho.org* Database

The *NeuroMorpho.org* database is a repository of digitally reconstructed neurons that have been made available by neuroscience laboratories around the world. Publicly available online, this database is intended for research in neuroscience, such as visualization, analysis, and modeling of neuronal cells. When the *NeuroMorpho.org* (Alpha) was released on 1st August 2006, data of 932 neurons were provided; at the time of access (9th March 2011), the database already contained measurements of 5,673 cells according the pie-charts shown in Fig. 1. It should be noted that some bias in such a database may arise as neuroscientists may choose neurons that can be more readily identified. Even if we were able to find out if each cell is healthy or not from the respective literature, there would be no means to identify abnormal cells coming from the normal tissues. The fact that our analysis does not distinguish between these categories allow its potential application for identifying potentially abnormal cells among the outliers. In other words, the universe considered for the neurons are all possible cells found in living animals.

Data-sets contain original and standardized files of the morphological reconstruction, general information of each neuron, such as cell type, region, species, classes, laboratory, researcher, age scale, staining and reconstruction methods, magnification, visualizations in 2D and 3D, as well as related papers, references, and numerical data related to the geometrical parameters of neuronal shape (Scorcioni et al. 2008). The database is maintained by the Computational Neuroanatomy Group (Krasnow Institute for Advanced Study, George Mason University) (Ascoli 2002; Ascoli et al. 2007). The *NeuroMorpho.org* project is linked to the Neuroscience Information Framework (NIF) (Halavi et al. 2008).

Digital reconstructions, from which the measurements are acquired, represent the morphology of neuronal cells as a series of points along the neurites with their positions, radius, connectivity, and process type (e.g. soma, dendrite, and axon). The reconstruction process involves a microscope connected to a computer, which combines captured images into virtual representation of axonal and dendritic processes, which are then traced (Donohue and Ascoli 2010). Indeed, this technique is time consuming and implies distortions (e.g. loss of focus, mechanical shear and crushing of imaged tissue, specular reflections, among others). Therefore, many studies of automatic data collection have been developed to overcome bias and errors (Srinivasan et al. 2007; Lu et al. 2009; Donohue and Ascoli 2010).

Measurements

The *NeuroMorpho.org* repository provides a large set of measurements (or variables) already extracted from each neuronal cell. Measurements include the soma surface area (S), number of stems (N_S), number of

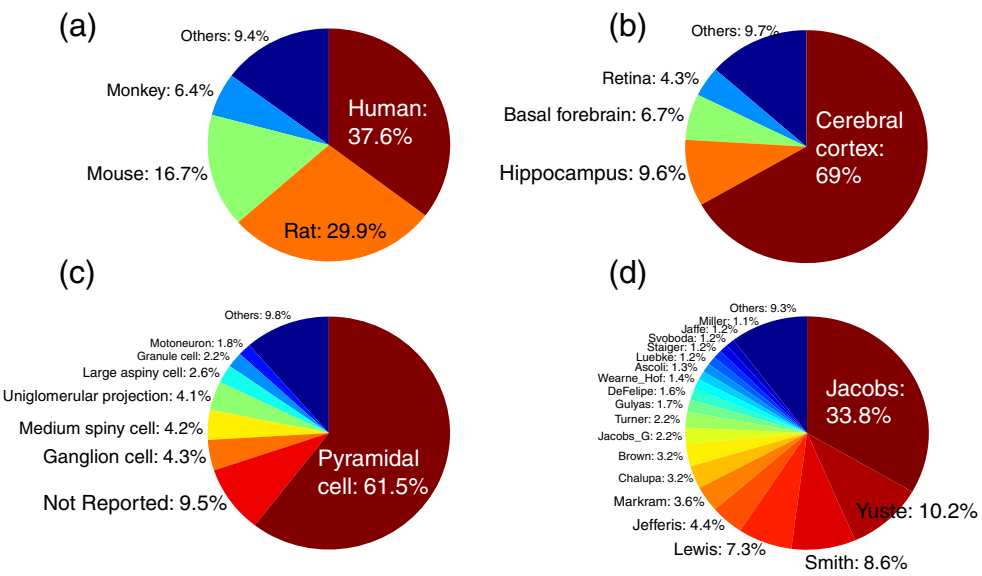

Fig. 1 Organization of cells in the *NeuroMorpho.org* database. *Pie-charts* show the percentage of cells classified according to **a** animal species, **b** brain region, **c** cell type, and **d** research group. Data from 9th March 2011

bifurcations (N_b), number of branches (N_{bc}), overall width (W), overall height (H), overall depth (D), average diameter (d), total length (L), total surface area (A), total volume (V), the maximum Euclidean distance between the soma and each of the nodes (E_d), the maximum arc length distance between the soma and each of the nodes (P_d), maximum branch order (O_b), average contraction—i.e. the maximum of the ratios between the Euclidean distance and the respective arc length distance (C), the total number of compartments in the neuron (F), and the average Rall's ratio—i.e. the ratio between the sum of the diameters of the daughter compartment and the diameter of the respective parent compartment (R). Another considered feature is the partition asymmetry (P_s). Given a node, the number of tips n_1 and n_2 of each of its two subtrees are determined and the partition asymmetry is then calculated as $P_s = |n_1 - n_2|/(n_1 + n_2 - 2)$. The last two measurements adopted in this work correspond to the average local (α_l) and remote (α_r) bifurcation angles. The former corresponds to the angle between two adjacent compartments. The latter is the angle between a bifurcation and the tips of the two respective daughter branches. More information about these measurements can be found in Ascoli (2002) and Costa et al. (2010). The respective statistics of these measurements can be found in Table 1, where variables related to size, such as volume and surface, are those with the largest coefficient of variation. To avoid size effects in our results, the variables S, W, H, D, d, L, A, V, E_d and P_d were normalized by $[1/3(W + H + D)]^{\nu}$, where ν is the exponent that makes the variable dimensionless.

For each neuron, the 20 variables mentioned above yield its feature vector $\vec{w}$ of size $M = 20$ that works as a morphological descriptor for the neuron. Each neuron is thus represented by a vector in M-dimensional morphospace. The morphospace has too many dimensions for direct visual inspection, which is why we apply principal component analysis (PCA) to these data (Härdle and Simar 2007), which yields a low-dimensional (in this work, two-dimensions) representation approximating the morphospace. In the PCA method, the number of components required to represent the whole dataset with minimal loss of information depends on the explained variance v, which is given by

$$v(d) = 100 \times \frac{\sum_{i=1}^{d} \lambda_i}{\sum_{i=1}^{M} \lambda_i}, \tag{1}$$

where d is the reduced dimension, and λ_i is the i-th eigenvalue of the covariance matrix. We are assuming that the eigenvalues are sorted in deceasing order, i.e. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$. Although the reported method can be straightforwardly extended to any value of d, all the results presented here correspond to $d = 2$. It is important to note that the higher the value of d, the higher should be the number of used samples in order to obtain a good estimate of the probability density described in Section "Outlier and Archetype Detection". In our experiments, the first two components

Table 1 Average, μ, and standard deviation, σ, of the 20 measurements of the overall data of the *NeuroMorpho.org* repository

Measurement	μ	σ	σ/μ
Soma surface—S (μm^2)	1,227.40	1,982.65	1.62
Number of stems—N_s	5.43	3.17	0.58
Number of bifurcations—N_b	46.52	96.87	2.08
Number of branches—N_{bc}	103.12	194.50	1.89
Width—W (μm)	365.00	388.30	1.06
Height—H (μm)	381.19	327.76	0.86
Depth—D (μm)	119.87	208.11	1.74
Diameter—d (μm)	1.13	0.80	0.71
Total length—L (μm)	6,110.45	11,216.35	1.84
Area—A (μm^2)	14,577.62	34,019.79	2.33
Volume—V (μm^3)	7,972.27	72,513.86	9.10
Euclidean distance—E_d (μm)	386.71	364.70	0.94
Path distance—P_d (μm)	589.79	1,029.01	1.74
Branch order—O_b	8.20	8.12	0.99
Contraction—C	0.87	0.07	0.08
Number of compartments—F	2,060.29	5,177.73	2.51
Partition asymmetry—P_s	0.47	0.12	0.25
Rall's ratio—R	1.50	0.46	0.31
Local bifurcation angle—α_l (°)	64.88	16.50	0.25
Remote bifurcation angle—α_r (°)	56.52	16.10	0.28

The last column corresponds to the coefficient of variation

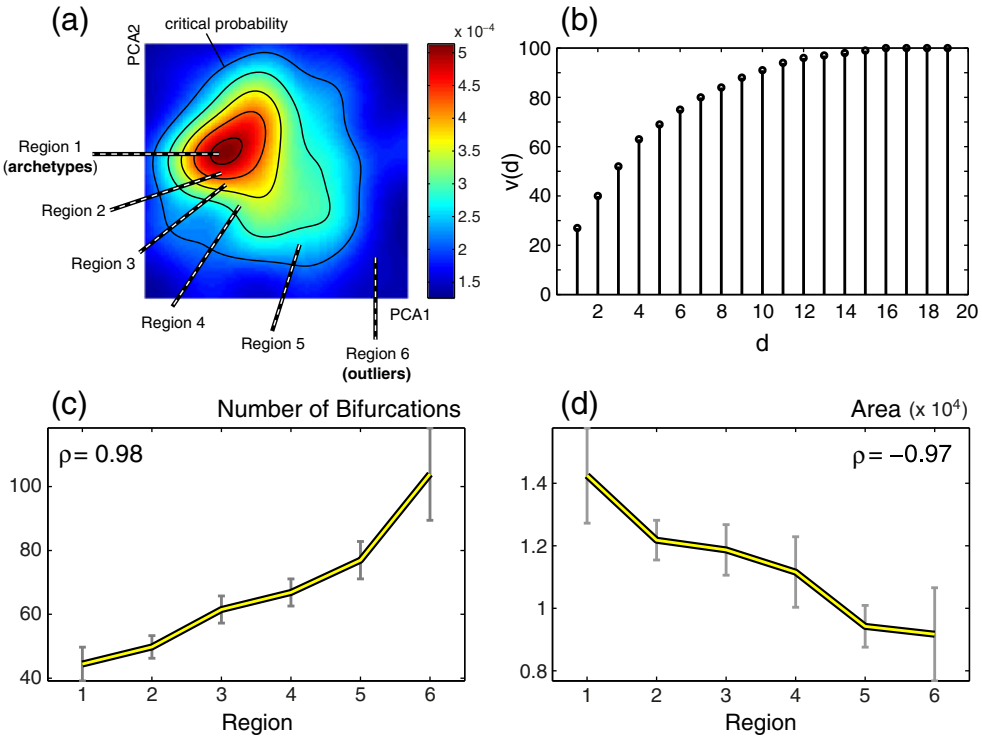

Fig. 2 Probability distribution estimated for ganglion cells of mouse. **a** The critical probability defines the contour outside which all cells are classified as outliers. Five equaly spaced regions divide the remaining 2D space, and the archetype cells are located at the smallest region, where the highest probabilities are found. **b** Total variance that can be explained given by Eq. 1 as a function of the number of dimensions d. **c**–**d** Average values of N_b and A as a function of the region *id*

are sufficient to represent on average 43% of the total variance. Despite some information is lost by reducing 20 to 2 dimensions, the most relevant outliers would still have high chances of being identified in the projected space.

Outlier and Archetype Detection

The neuronal projections in 2D space generated by PCA is used to estimate a probability distribution of finding a neuronal cell in a particular point of this space. This is done by using a bivariate kernel density estimator (Botev et al. 2010). The critical probability below which all cells are classified as outliers was estimated by using the procedure described by Echtermeyer et al. (2011). To illustrate how our method works, we show the results obtained for ganglion cells of a mouse in Fig. 2. Figure 2a shows the 2D probability distribution estimated over the space defined by the first two principal components. Figure 2b shows the total variance depending on the number of dimensions d where the first two axes explain about 41% of the total variance.

In Fig. 2a we also show the contour defined by the critical probability. The remaining space is divided into five equally spaced regions (regarding the probability values), where the most inner one, containing higher values of probability, is related to the archetype cells. Each region receives an *id* from 1 to 6. By studying how the variables change from the region with *id* 1 to the region with *id* 6, we can infer the main differences between archetype and outlier cells. These differences can be identified by considering the Pearson correlation coefficient (ρ) between average value of a given variable for all cells inside the same region, and the region *id*. For instance, variables with correlation near zero, do not show any significant change from archetype to outlier region. On the other hand, variables that present correlation near to ± 1, such as those shown in Fig. 2c and d, are good indicators for main differences between these cells. The statistical significance of such differences can be estimated by using the Kolmogorov-Smirnov test (Eadie et al. 1971) to compare the distributions for archetypes and outliers. The null hypothesis is that a given measurement comes from the same continuous distributions for both archetypes and outliers. This hypothesis was rejected at the 5% significance level for all results discussed in this paper.

Results and Discussion

In this section we present the main results obtained from ten different experiments in which we consider different sets of cells as input for the outlier detection. These experiments are divided into two main groups:

Table 2 Organization of all ten experiments performed in this work

Experiment	Description of the data	Number of cells
Cell types		
1	Unigromerular prejection neurons	233
2	Medium spiny cells	239
3	Ganglion cells	245
4	Pyramidal cells	3,490
Specific types		
Pyramidal cells		
5	Human	2,132
6	Monkey	344
7	Mouse	949
8	Rat	1,694
Ganglion cells		
9	Mouse	181
10	Salamander	64

cell types and *specific type*. In the former, we consider all cells of the same type in each experiment. In the second case we consider a specific type of cell of a given animal in each experiment. Details are summarized in Table 2.

Cell Types

In this analysis, we fixed the type of the considered cells and allowed the other parameters—like research group, animal species, and brain region—to change. Such an approach can be useful to investigate how the neuronal morphology changes along different species or along different brain regions. In addition, it is useful to see how the adopted procedure by each research-group affects the final measurements. We chose the four largest classes of cell types available in the *NeuroMorpho.org* database: *uniglomerular projection*, *medium spiny*, *ganglion*, and *pyramidal cells*. Figure 3

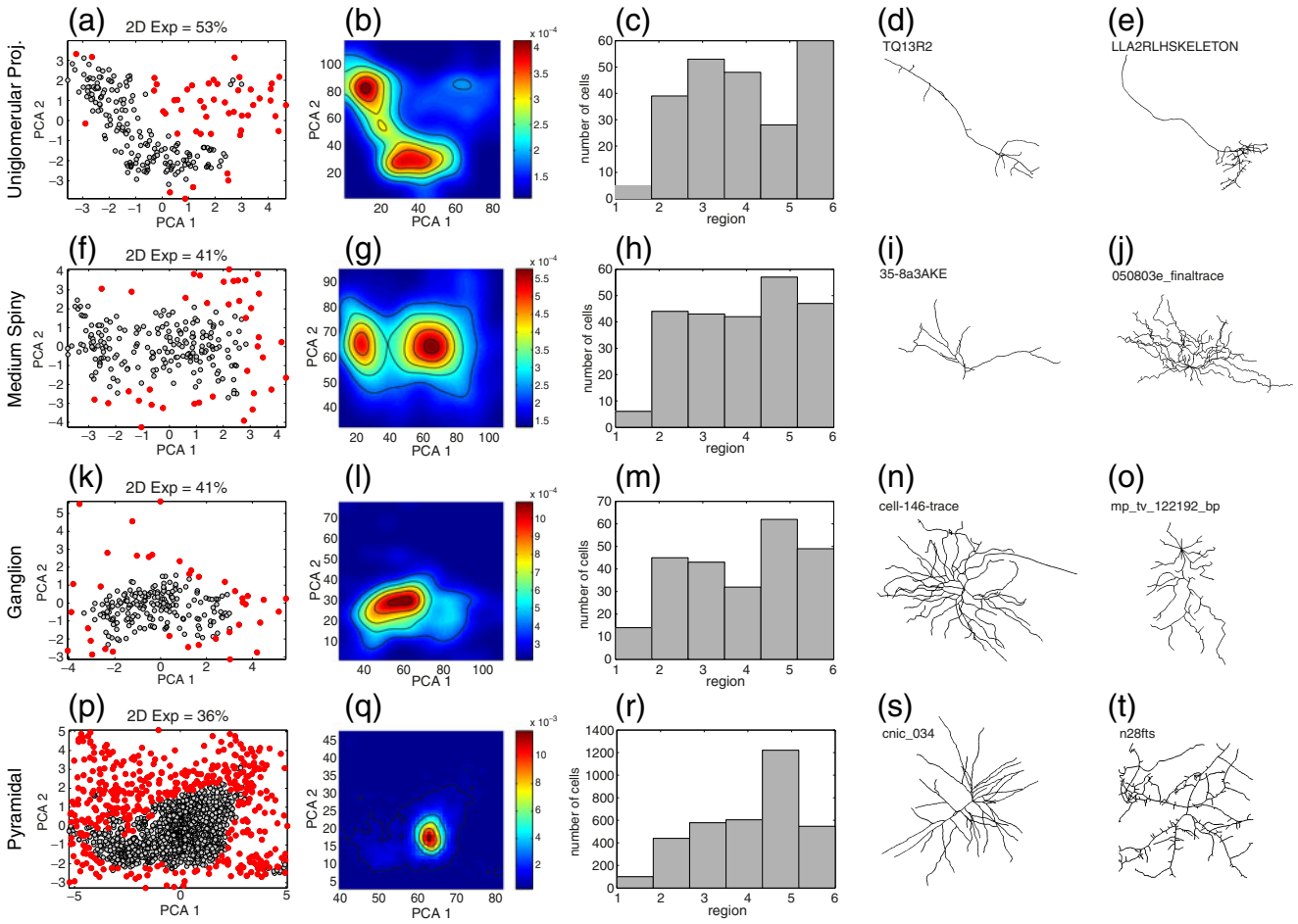

Fig. 3 Outliers detection for different cell types. The *first column* corresponds to the 2D PCA space. The *second column* is the estimated probability density and the regions in which the 2D space was divided. The number of cells inside each region is shown in the *third column*. Examples of archetype (the most probable cell) and outlier (the least probable cell) are shown in forth and fifth columns, respectively

summarizes the results obtained in this experiment. The first column corresponds to the PCA projection of all cells (detected outliers marked red). The percentage of the total variability explained by the two PCA axes is also provided in the first column. The probability distribution of the 2D PCA space as well as the five contours in which this space was divided are shown in the second column. Third column of Fig. 3 gives the number of cells inside each region, while the fourth and fifth column show a typical archetype and outlier cell, respectively.

The first cell type considered was *uniglomerular projection neurons*. In the *NeuroMorpho.org* database, there are 233 cells of this type and we identified 26% of them as outliers. The emergence of a bimodal probability distribution, as shown in Fig. 3b, could not be explained by any simple characteristic, since all cells were obtained by the same research group from the same brain region of *Drosophila*. There are cells from two different ages, embryonic and larval, but this difference is not enough to explain the bimodal distribution, since we found embryonic and larval cells in both top and bottom peaks. We identified that the variables related to size and number of bifurcations were those that had the largest changes from archetypes to outliers cells. For instance, the Pearson correlation between the total length L and the region *id*, and between the number of bifurcations N_b and the region *id* are $\rho = 0.93$ and $\rho = 0.91$, respectively.

The second class concerns the *medium spiny cells*, which were obtained by the group *Smith* from adult animals of the *Long Evans Rat* species. As can be seen in Fig. 3g, the probability distribution is also bimodal and we found that the only difference between neurons located at the left and right peaks seems to be the date when the neurons were deposited. While the neurons at left were traced on 1st December 2007, the neurons at right were traced on 22nd February 2008. These differences are interesting because they suggest other possible sources of significant morphological variation, such as distinct individuals, or could also be an indication of cell type subgroups. For this type of cells, we identified 47 neurons as outliers, which correspond to 20% of the total of 239 medium spiny cells. When we looked at the variables changing from the archetypes to outliers, we also found the most relevant variables from archetypes to outliers to be the size and number of bifurcations. However, contrary to what happened in the previous case, for these cells, we observed a densification process, i.e. the linear variables related to size, such as height H and depth D are negatively correlated with the region *id* ($\rho = -0.85$ and $\rho = -0.8$, respectively). On the other hand, there is a strong positive correlation of 0.93 between total length L and the region *id*. The outliers thus tend to be longer than the archetypes in a smaller region.

We also analyzed the 245 *ganglion cells* of the database and identified 20% of them as outliers. In addition, the probability distribution is described by a single unimodal distribution (see Fig. 3l) even though we have cells from two different species obtained by two different research groups. These results indicate a consistency of the technique adopted by these two groups, which makes the data reliable for further investigations. Our results show that the main differences between archetype and outlier ganglion cells concern the soma surface S and the average contraction C, whose correlations with the region *id* are $\rho = -0.85$ and $\rho = -0.9$, respectively. The last result indicates that the outliers' branches tend to be more tortuous than those observed in the archetype cells.

Finally, the largest class of cell type is that of *pyramidal cells* containing 3,490 neurons from which 547 (16%) were identified as outliers. These cells were obtained by several different groups, from three different regions (amygdala, cerebral cortex, and hippocampus) and from five different animal species (cat, human, monkey, mouse, and rat). Also in this case, a unimodal distribution was found (Fig. 3q) and the measurements related to size, such as volume V and soma surface S were the most relevant to discriminate between archetype and outlier cells.

Specific Type

In this section, we focus on the cells belonging to the classes pyramidal and ganglion cells. For the first class, we considered four different animal species, *human*, *monkey*, *mouse*, and *rat*, while for ganglions we analyzed cells from *mouse* and *salamander*. The main results are shown in Figs. 4 and 5.

Pyramidal Cells

When we analyzed the 2,132 pyramidal cells of human obtained by three different groups, we found that 24% of them were classified as outliers. Moreover, a single unimodal distribution was able to describe the diversity of topologies observed in the database. Among the variables that change significantly from archetypes to outliers, we identified that the most relevant were the number of bifurcations N_b and the partition asymmetry P_s. The first variable has a negative correlation ($\rho = -0.93$) with the region *id*, indicating that the outliers are less bifurcated than the archetypes. On the other hand, the partition asymmetry has a positive correlation

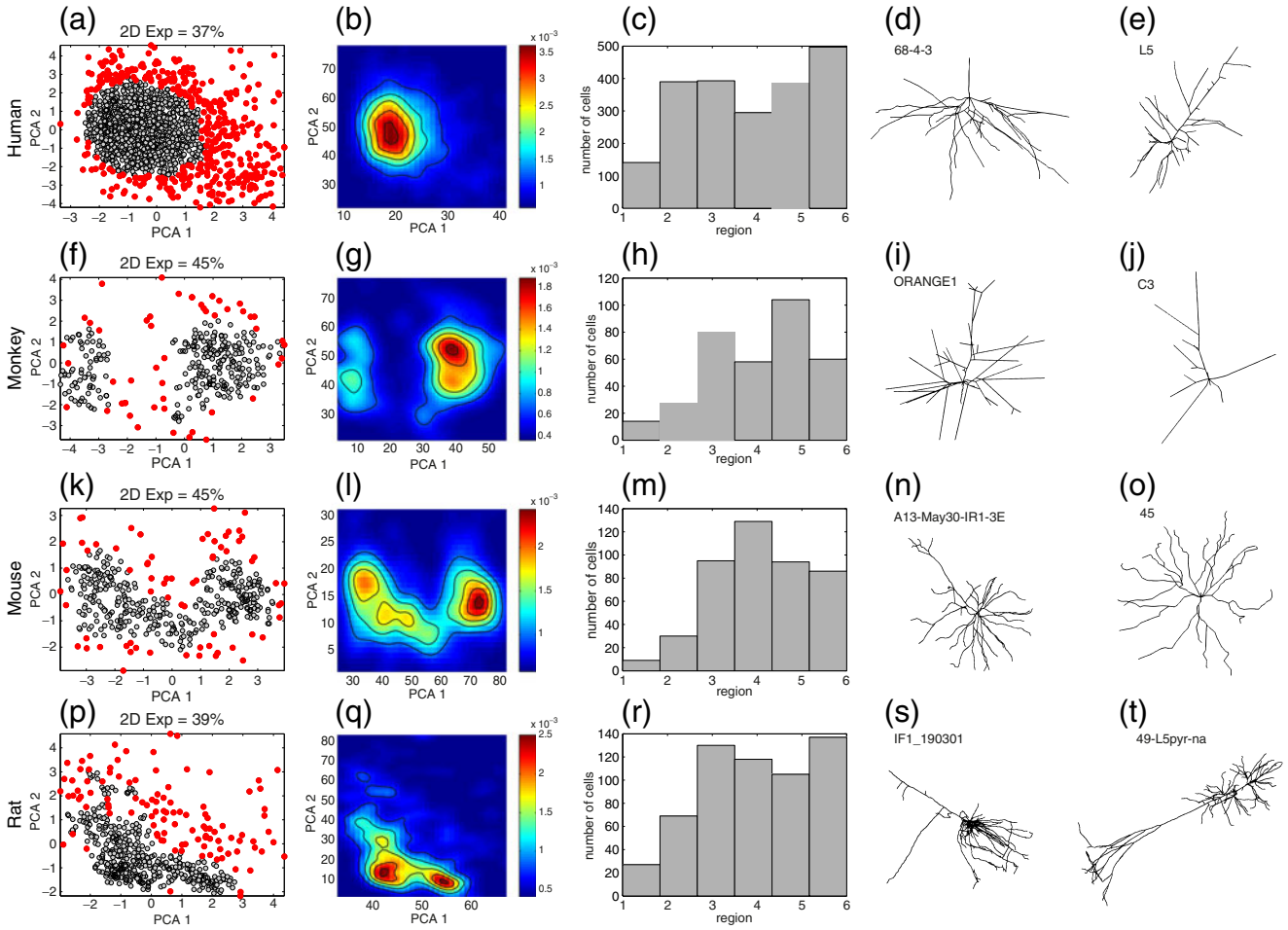

Fig. 4 Outliers detection for pyramidal cells of different species. For description of the sub-plots please refer to Fig. 3

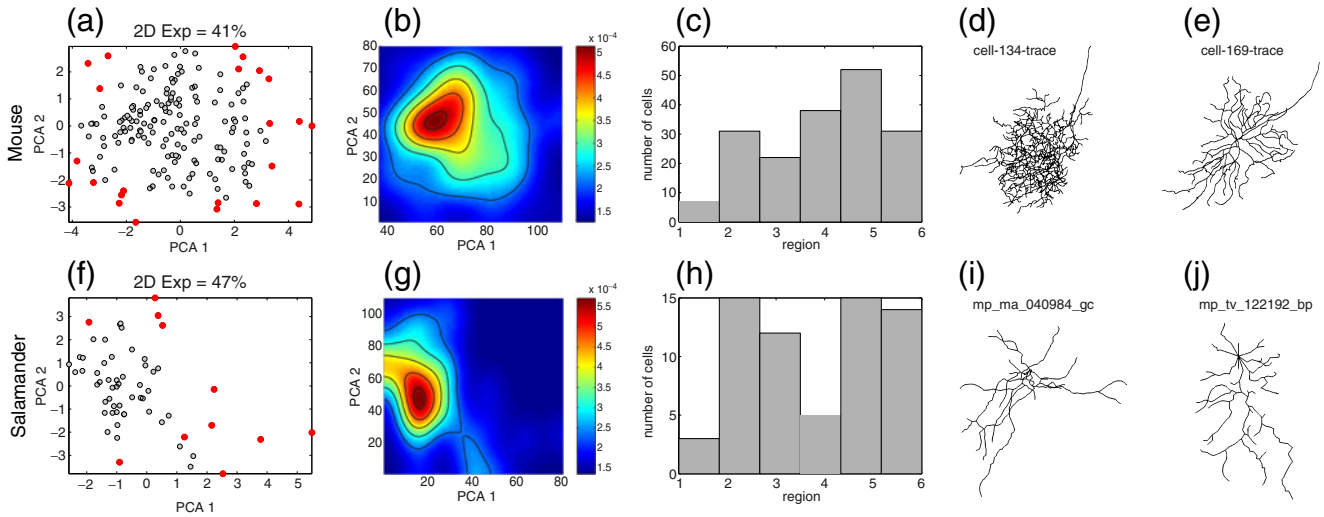

Fig. 5 Outliers detection for ganglion cells of different species. For description of the sub-plots please refer to Fig. 3

($\rho = 0.94$), showing that archetypes are more symmetric than outlier cells.

We also found a bimodal distribution for pyramidal cells of monkeys. On the left peak, we only observed cells obtained by the research group *Lewis* from the specie *Rhesus Monkey*, which were *Lucifer yellow*-stained. However, on the right peak, we found cells from both groups *Lewis* and *WearneHof* from *Rhesus Monkey*. In addition, the cells at this peak were obtained by two different staining methods, *Golgi* and *Lucifer yellow*. We also found that these two peaks are distinct by the thickness of slice in which the cells were obtained. While the left peak contains only cells with thickness of slice of $90\mu m$, 80% of cells in the right peak have thickness of 300 or $400\mu m$. The main differences observed between archetypes and outliers were related to the number of stems and contraction, whose correlations are $\rho = -0.91$ and $\rho = -0.88$, respectively. On the other hand, we found that the distances E_d and P_d are positively correlated with the region *id* ($\rho = 0.88$ and $\rho = 0.89$ respectively).

Considering mouse cells, for which we have 949 pyramidal samples, about twenty percent of them were identified as outliers and the 2D PCA space also showed a bimodal distribution. On the left side, we have a more elongated peak, in which we found cells from the group *Yuste*. In the right peak, cells from both *DeFelipe* and *Luebke* groups were found. Differently from the previous case, no significant changes were observed in the variables from archetypes to outliers mouse pyramidal cells.

Rat was the last species considered. From the total of 586 pyramidal cells, we found that 24% of them were classified as outliers. Also here, a bimodal distribution was found. There is a weak trend of cells located at the right peak coming from the same research group (*Smith*) and from the sub species *Long Evans Rat*. The remaining 2D space is equally filled by cells from 16 different groups and four different sub species. In this case, we found the most relevant differences between archetypes and outliers are their volume V (Pearson correlation coefficient $\rho = 0.97$) and number of bifurcations N_b ($\rho = 0.95$), indicating that outlier cells are longer and more bifurcated than archetypes.

It is interesting to observe the opposite behavior of N_b when human or rat cells are considered. While for the former, the number of bifurcations decreases from archetypes to outliers, it increases for the latter.

Ganglion Cells

The database contains 181 ganglion cells of mouse that were obtained by the group *Chalupa*. Seventeen percent of these cells were classified as outliers. Our results also indicate that several measurements change significantly from archetypes to outliers. Area (Pearson correlation coefficient $\rho = -0.97$), diameter ($\rho = -0.94$) and contraction ($\rho = -0.98$) tend to decrease in outlier cells, while number of bifurcation ($\rho = 0.98$) and bifurcation angle remote ($\rho = 0.95$) tend to increase.

In the last experiment, a total of 64 ganglion cells of salamander traced by the group *Miller* were considered. Our results indicate that about 22% were classified as outliers. We noted that in this case, the main differences between archetypes and outliers concern the number of bifurcations and diameter, which tend to increase in outlier cells. The correlation coefficients for these measurements were 0.89 and 0.93, respectively.

Although the behavior of variable N_b seems to be the same for outlier cells of both mouse and salamander, this is not true for the average diameter. As we can see, outliers of ganglion cells of mouse tend to be thinner than archetypes, while the diameter of outlier cells is greater than that observed in archetypes for salamander cells.

Concluding Remarks

The morphology of neurons provides an important aspect for their classification, function, and patterns of connectivity. Due to the large variety of neuronal shapes it is not straightforward to identify archetype and exceptional cells. In this work we analyzed morphological measurements provided by the *NeuroMorpho.org* database and identified archetype as well as outlier cells.

The technique applied in this paper estimates a probability density function over combinations of morphological features, which is then used to identify cells with typical and unusual properties. Apart from cells with high and low probability, we divided the morphometric space into regions of similar densities. This allowed us to characterize cells in the transition area between archetypes and outliers. To our best knowledge, this is the first work systematically addressing the automatic identification and characterization of outlier neuronal cells according to their morphological shapes. From the ten different experiments performed, we found that often the probability distribution on the 2D PCA space is given by a bimodal distribution, in which the peaks can be related to different animal species, different research groups, different individuals or even the presence of so far unidentified sub-groups of certain cell-types.

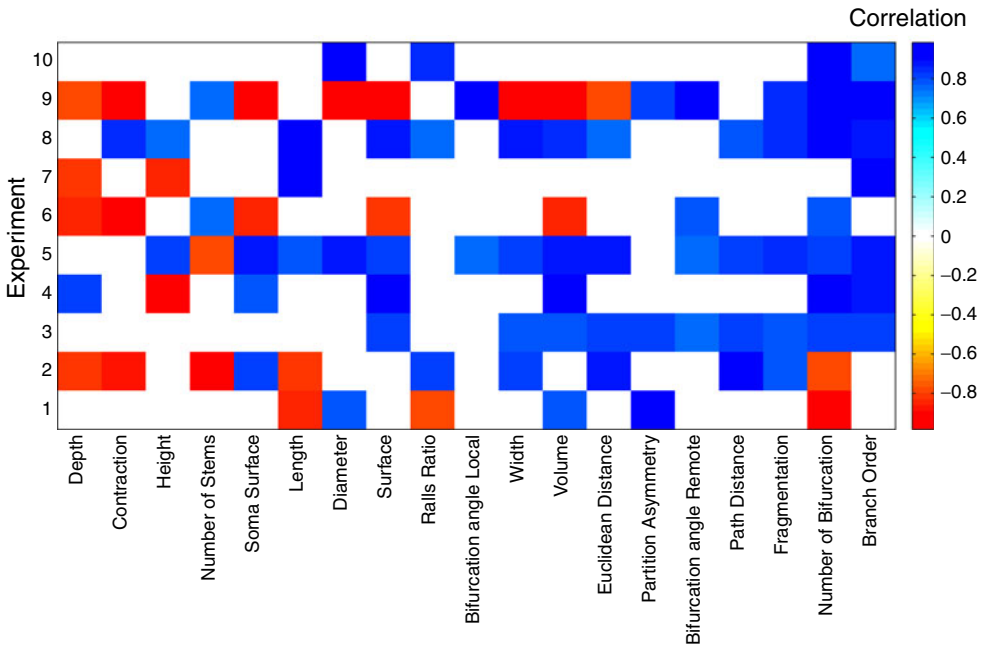

Fig. 6 Relevant differences between archetypes and outliers. *Each column* corresponds to a given measurement, while the *rows* indicate different experiments for which the Pearson correlation coefficients between measurements and the region *id* were calculated. In addition, all cases in which the absolute correlation is smaller than 0.75 are shown in *white*. For instance, from the ten different experiments, a positive correlation was observed between the measurement *number of bifurcations* and the region *id* in seven cases. Two experiments yielded a negative correlation, while no significant correlation was observed in just one of the experiments. The measurements along the x-axis were ordered according to increasing values of total correlations (i.e. the sum of the correlations along each column)

To investigate the differences between archetypes and outliers, we observed how the variables change from the region with *id* 1 to the region with *id* 6. The Pearson correlation coefficient between the average of each variable and the region *id* was obtained in ten different experiments, as shown in Fig. 6. It is clear that the variables *number of bifurcations* and *branch order* O_b are those which presented positive correlation in most analyses, indicating that, in general, the outliers have more branches than archetypes. On the other hand, *depth* D and *contraction* C are those which presented negative correlation in most analyses. Compared to archetypes, outliers tend to be planar with tortuous branches. Moreover, all measurements presented highly positive correlation in at least one experiment, while some of them never presented highly negative correlation.

Building on our results, future works could contribute to the knowledge regarding the relationship between the function and shape of neuronal cells. Ultimately, a joint investigation of geometrical, functional, and phylogenetic factors involved in the differentiation process could help in understanding important issues, such as evolution of the nervous system or diseases related to morphological anomalies. Also, some of the outliers could correspond to normal neurons affected by microscope type and calibration, laboratory of origin, scaling errors, among other artifacts. In fact, the methodology proposed in the current work can also help experts in specific areas to check for such effects.

Information Sharing Statement

The software DONE (Detection of Outlier NEurons) is an implementation for MATLAB that is available on our web-site http://www.biological-org/p/outliers/. Redistribution and modification of the DONE package is permitted under the terms of the GNU General Public License http://www.gnu.org/licenses/gpl.html.

Acknowledgements Luciano da F. Costa is grateful to FAPESP (05/00587-5) and CNPq (301303/06-1 and 573583/2008-0) for sponsorship. Krissia Zawadzki is grateful to FAPESP sponsorship (2010/01994-1). Matheus P. Viana is grateful to FAPESP sponsorship (2010/16310-0). Marcus Kaiser and Christoph Feenders acknowledge support by EPSRC (EP/G03950X/1) and the CARMEN e-science Neuroinformatics project (http://www.carmen.org.uk) funded by EPSRC (EP/E002331/1). Marcus Kaiser is also funded through the WCU program of the National Research Foundation of Korea

funded by the Ministry of Education, Science and Technology (R32-10142).

Conflict of interests The authors declare that the research was conducted commercial or financial relationships that could be construed as a potential conflict of interest.

References

Ascoli, G. A. (Ed.) (2002). *Computational neuroanatomy: Principles and methods*. Totawa, NJ: Humana Press.

Ascoli, G. A., Donohue, D. E., & Halavi, M. (2007). Neuromorpho.org: A central resource for neuronal morphologies. *Journal of Neuroscience, 27*(35), 9247.

Binzegger, T., Douglas, R. J., & Martin, K. A. (2005). Axons in cat visual cortex are topologically self-similar. *Cereb Cortex, 15*(2), 152–165.

Botev, Z. I., Grotowski, J. F., & Krose, D. P. (2010). Kernel density estimation via diffusion. *The Annals of Statistics, 38*(5), 2916.

Cajal, S. R. (1989). *Recollections of my life*. Massachussetts: MIT Press.

Cook, J. E. (1998). Getting to grips with neuronal diversity: What is a neuronal type? In L. Chalupa & B. Finlay (Eds.), *Development and organization of the retina* (pp. 91). New York: Plenum.

Costa, L. da F., Manoel, E. T. M., Faucereau, F., Chelly, J., van Pelt, J., & Ramakers, G. (2002). A shape analysis framework for neuromorphometry. *Network: Computation in Neural Systems, 13*(3), 283.

Costa, L. da F., Rodrigues, F. A., Hilgetag, C. C., & Kaiser, M. (2009). Beyond the average: Detecting global singular nodes from local features in complex networks. *Europhysics Letters, 87*, 18008.

Costa, L. da F. & Velte, T. J. (1999). Automatic characterization and classification of ganglion cells from salamander retina. *Journal of Comparative Neurology, 404*, 33.

Costa, L. da F., Zawadzki, K., Miazaki, M., Viana, M. P., & Taraskin, S. N. (2010). Unveiling the neuromorphological space. *Frontiers in Neuroscience, 4*, 1.

Donohue, D. E., & Ascoli, G. A. (2010). Automated reconstruction of neuronal morphology: An overview. *Brain Research Review, 67*,(1–2), 165.

Eadie, W. T., Drijard, D., James, F. E., Roos, M., & Sadoulet, B. (1971). *Statistical methods in experimental physics*. North-Holland, Amsterdam.

Echtermeyer, C., Costa, L. da F., Rodrigues, F. A., & Kaiser, M. (2011). Automatic network fingerprinting through single-node motifs. *PLoS ONE, 6*(1), 9.

Echtermeyer, C., Han, C. E., Rotarska-Jagiela, A., Mohr, H., Uhlhass, P., & Kaiser, M. (2011). Integrating temporal and spatial scales: Human structural network motifs across age and region of interest size. *Frontiers in Neuroinformatics, 5*, 14.

Gleeson, P., Steuber, V., & Silver, R. A. (2007). Neuroconstruct: A tool for modeling networks of neurons in 3D space. *Neuron, 54*(2), 219.

Halavi, M., Polavaram, S., Donohue, D. E., Hamilton, G., Hoyt, J., Smith, K. P., et al. (2008). Neuromorpho.org implementation of digital neuroscience: Dense coverage and integration with the NIF. *Journal of Neuroinformatics, 6*(3), 241–252.

Härdle, W. K., & Simar, L. (2007). *Applied multivariate statistical analysis* (2nd ed.). Springer.

Kaiser, M., & Hilgetag, C. C. (2006). Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS Computational Biology, 7*(95), 806.

Kaiser, M., Hilgetag, C. C., & van Ooyen, A. (2009). A simple rule for axon outgrowth and synaptic competition generates realistic connection lengths and filling fractions. *Cerebral Cortex, 19*(12), 3001.

Loewi, O. (1921). Über humorale Übertragbarkeit der Herznervenwirkung. *Pflügers Archiv, 189*, 239.

Loewi, O. (1955). Salute to Henry Hallet Dale. *The British Medical Journal, 1*(4926), 1356.

Lu, J., Fiala, F. C., & Lichtman, J. W. (2009). Semi-automated reconstruction of neural processes from large numbers of fluorescence images. *PLoS ONE, 4*(5), e5655.

McGhee, G. R. (2006). *The geometry of evolution: Adaptive landscapes and theoretical morphospaces*. Cambridge University Press.

Montague, P. R., & Friedlander, M. J. (1991). Morphogenesis and territorial coverage by isolated mammalian retinal ganglion cells. *Journal of Neuroscience, 11*, 1440.

Poznanski, R. R. (1992). Modelling the electronic structure of starburst amacrine cells in the rabbit retina: Functional interpretation of dendritic morphology. *Bulletin of Mathematical Biology, 54*, 905.

Schierwagen, A. (2008). Neuronal morphology: Shape characteristics and model. *Neurophysiology, 40*(4), 310.

Scorcioni, R., Polavaram, S., & Ascoli, G. A. (2008). L-measure: A web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nature Protocols, 3*, 866.

Sholl, D. A. (1953). Dendritic organization in the neurons of the visual and motor cortices of the cat. *Journal of Anatomy, 87*, 387.

Sporns, O., Chialvo, D. R., Kaiser, M., & Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends in Cognitive Sciences, 8*(9), 418.

Srinivasan, R., Zhou, X., Miller, E., Lu, J., Litchman, J., & Wong, S. T. C. (2007). Automated axon tracking of 3D confocal laser scanning microscopy images using guided probabilistic region merging. *Neuroinformatics, 5*(3), 189.

Stepanyants, A., & Chklovskii, D. B. (2005). Neurogeometry and potential synaptic connectivity. *Trends in Neuroscience, 28*(7), 387.

Stepanyants, A., Hof, P. R., & Chklovskii, D. B. (2002). Geometry and structural plasticity of synaptic connectivity. *Neuron, 34*(2), 275.

Toris, C. B., Eiesland, J. L., & Miller, R. F. (1995). Morphology of ganglion cells in the neotenous tiger salamander retina. *Journal of Comparative Neurology, 352*(4), 535.

Wen, Q., & Chklovskii, D. B. (2008). A cost-benefit analysis of neuronal morphology. *Journal of Neurophysiology, 99*, 497.