Bioinformatics, 2015, 1–8
doi: 10.1093/bioinformatics/btv097
Advance Access Publication Date: 16 February 2015
Original Paper

OXFORD

Bioimage informatics

Characterizing spatial distributions of astrocytes in the mammalian retina

Aruna Jammalamadaka[1,*], Panuakdet Suwannatat[2], Steven K. Fisher[3,4], B. S. Manjunath[1,*], Tobias Höllerer[2] and Gabriel Luna[3]

[1]Department of Electrical and Computer Engineering, [2]Department of Computer Science, [3]Neuroscience Research Institute and [4]Department of Molecular, Cellular, and Developmental Biology, University of California, Santa Barbara, Santa Barbara, CA, USA

*To whom correspondence should be addressed.

Associate Editor: Robert F. Murphy

Received on July 12, 2014; revised on January 5, 2015; accepted on January 31, 2015

Abstract

Motivation: In addition to being involved in retinal vascular growth, astrocytes play an important role in diseases and injuries, such as glaucomatous neuro-degeneration and retinal detachment. Studying astrocytes, their morphological cell characteristics and their spatial relationships to the surrounding vasculature in the retina may elucidate their role in these conditions.

Results: Our results show that in normal healthy retinas, the distribution of observed astrocyte cells does not follow a uniform distribution. The cells are significantly more densely packed around the blood vessels than a uniform distribution would predict. We also show that compared with the distribution of all cells, large cells are more dense in the vicinity of veins and toward the optic nerve head whereas smaller cells are often more dense in the vicinity of arteries. We hypothesize that since veinal astrocytes are known to transport toxic metabolic waste away from neurons they may be more critical than arterial astrocytes and therefore require larger cell bodies to process waste more efficiently.

Availability and implementation: A 1/8th size down-sampled version of the seven retinal image mosaics described in this article can be found on BISQUE (Kvilekval *et al.*, 2010) at http://bisque.ece.ucsb.edu/client_service/view?resource=http://bisque.ece.ucsb.edu/data_service/dataset/6566968.

Contact: arunaj@ece.ucsb.edu or manj@ece.ucsb.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

The large amounts of image data that can be obtained with recent advances in microscopy require mining in an objective and unbiased fashion, using automated and reproducible methods. In this way, it is possible to discover trends and hypotheses for systems that domain-experts may have very little prior knowledge about. These improvements have especially had a profound impact on examination of the spatial distribution and correlation of biological entities, as the locations of large amounts of cells and other structures can be imaged *in situ*. Studying the spatial arrangement and relationships in full tissue samples can improve our understanding of the various development or pathological processes that underlie proper organ or organism function (Whitney *et al.*, 2008). In particular, it has been found that neuronal or vascular structures are pervasive in many tissues, and oftentimes are spatially correlated with other cells (Armstrong, 2003; Suematsu *et al.*, 1994).

In this study, we are particularly interested in astrocytes, one of two types of glial cells found in the nerve fiber layer (NFL) of the mammalian retina. Astrocytes can be found between vasculature and retinal neurons and although many functions of astrocytes in the retina are poorly understood, it is widely accepted that they play an essential role in the development and function of the retinal

vasculature, blood flow and blood–retinal barrier (BRB) (Kur *et al.*, 2012). Studying astrocytes and their spatial distributions within the healthy retina may give us some insight into their function in disease cases such as glaucomatous neuro-degeneration and retinal detachment.

With regard to cell characteristics, one work determines that there are two types of astrocyte cells in developing rat white matter, and that they can be differentiated by their morphology, but only analyzes a total of <1000 cells and measurements to quantify the cells are made manually (Raff *et al.*, 1983). Another, more recent work reports that astrocyte cells which lie on or near blood vessels exhibit different morphological characteristics (specifically angles and lengths of primary processes) than those which do not (Zahs and Wu, 2001). This study involves a small dataset on the order of tens of astrocyte cells, hand-picked to have distinguishable processes that can be manually counted. Our current dataset of seven full retinal mosaics, each containing 3614–5499 cells, provides a much richer platform for testing hypotheses such as whether there exist distinct morphological classes of astrocytes and if so, how many. Our semi-automated processing pipeline attempts to minimize bias from manual measurements, while also testing the uncertainty of cell characteristic clustering results, which rely on cell segmentation.

Astrocyte processes physically contact the vascular structure (Paula *et al.*, 2010), and they also play a key functional role in the development of the retinal vasculature (Metea and Newman, 2006). An apparent spatial correlation between astrocytes and the blood vessels has been noted (Stone and Dreher, 1987), but only observational evidence of such a relationship has been provided. The vasculature is a large heterogeneous structure with specific arterial and venous delineations (Dorrell and Friedlander, 2006; Gariano and Gardner, 2004). In order to quantify spatial distributions of astrocyte cells with respect to it, we need to image the entire retina. Strong evidence of astrocyte spatial correlation with various vascular properties would lend further support to additional hypotheses of astrocyte function, such as the suspected role of astrocytes in vasodilation and constriction (Kimelberg, 2010). However, quantifying spatial properties of astrocytes in retinal tissue can be a challenging task. One must be careful in selecting spatial mining methods which are invariant to scale, rotation, and which use distance metrics that operate in geodesic spaces, such as along the linear network of the blood vessels. This rules out traditional spatial quantification methods such as colocation (Shekhar and Huang, 2001), or nearest neighbor methods (Cressie, 1992) which operate in Euclidean space.

To our knowledge, there has been only one previous work focused specifically on the analysis of spatial distributions of astrocytes in the retina (Ruttenberg *et al.*, 2013). Using manually marked cell centers and automatically traced blood vessels, their study attempts to determine existence of spatial relationships between the two in both detached and healthy retinas. Astrocytes and 2 μm blood vessel segments are represented by 4D features that quantify their respective locations. Astrocytes are mapped to their nearest blood vessel segment via Euclidean projection and 2D histograms of the features are computed with respect to the width of the blood vessel segment and the geodesic vascular distance to the optic nerve head (ONH). The article postulates that a correlation of astrocyte cell locations with the structure of the vasculature can be determined by comparing these histograms. The study concluded that it is unlikely that the astrocytes are randomly distributed along the structure of the vasculature, regardless of the retina being normal or detached. The article also reported that arterial astrocytes are spatially distributed as random samples from the arterial structure, whereas venous astrocytes spatially deviate from the venous spatial structure. While this study presents a novel method for spatial analysis of astrocyte and blood vessel distributions, there has been much previous work in point process modeling and spatial statistics dating back to 1986 (Berman, 1986; Foxall and Baddeley, 2002) regarding quantifying point distributions with respect to surrounding line segments or linear networks. We choose to make use of such spatial statistics tools, making our approach a more in-depth and intuitive way to represent and analyze this type of data. The conclusions arrived at here offer more detail but do not contradict those of Ruttenberg *et al.* (2013) and we attribute that to the use of more principled methodology, as mentioned above, as well as a larger and more accurate dataset due to staining of astrocyte nuclei.

The goal of our analysis is to determine if the spatial distribution of astrocytes in the retina correlates with the vascular network, or with morphological cell characteristics such as area and perimeter. To accomplish this, we extract astrocyte and blood vessel data for large image mosaics of the retinal NFL. We then segment the astrocytes using two different segmentation methods in order to assess the dependence of our resulting morphological characteristics on the cell segmentation step of the analysis pipeline. Following this, we perform an exploratory correlation analysis of cell characteristics to decide which to incorporate into our overall point process model and which are redundant. We also offer some possible spatial covariates for the cell distribution, whose importance will be tested on all cells and on each type of cell resulting from the study of cell characteristics. From these results we will arrive at an overall model for the density of astrocyte cells in the retina.

The methodology presented here can be used for spatial studies between other vasculature structures and cells, which occur in many places in the mammalian body. Using point process models we are able to systematically quantify the spatial distribution of astrocyte cells within the mammalian retina, and provide a foundation for future research aimed at studying the spatial distribution of various biological components in large tissue images. Our contributions lie in our choice of statistical models to describe the astrocyte distributions as well as the biological conclusions derived from these models. We will show evidence that in normal healthy retinas, the distribution of observed astrocyte cells is more densely packed around the blood vessels than a uniform distribution would predict. We also show that compared with the distribution of all cells, large cells are more densely packed in the vicinity of veins and toward the ONH whereas smaller cells are often more densely packed in the vicinity of arteries.

2 Materials and methods

We now present each step of our analysis as described above in detail.

2.1 Image acquisition

Our method is designed to quantify the spatial relationship between large biological structures and point data such as cellular locations. Hence, we acquired images of the entire retinal NFL, which allow us to capture the complete retinal vasculature and all astrocytes present in the retina. Images of mouse retinal NFL were viewed and collected on a laser scanning confocal microscope using an automated stage to capture optical sections at 0.5 μm intervals in the z-axis and pixel resolution of 1024×1024 in the x–y direction, with 20% overlap in the x–y plane. The objective used was an UPlanFLN $40 \times$ oil immersion lens, N.A. 1.30. Approximately 350–400 3D

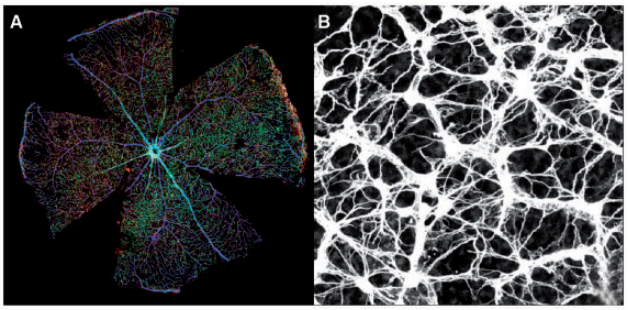

Fig. 1. (**A**) An example retinal mosaic used in the study. See online version for color figure, where astrocytes are in green, vasculature is in blue and nuclei are in red. (**B**) Magnified view of the Anti-GFAP astrocyte channel. (Color version of this figure is available at *Bioinformatics* online.)

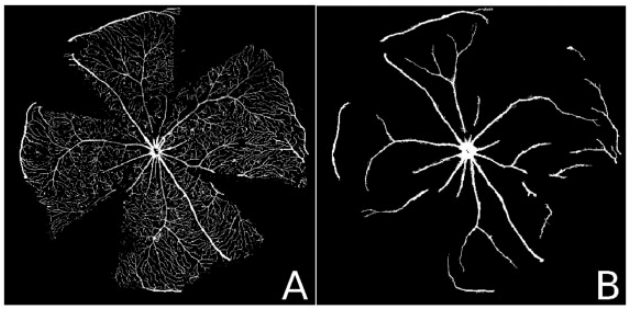

Fig. 2. Example manual binarization of major blood vessels for GFP11. Gray-scale and binary images have been inverted for better visibility. (**A**) Original blood vessel channel. (**B**) Binarized major blood vessels

images were acquired per retina, which were then used to create maximum-intensity projections using Olympus FluoView viewer version 1.7a. The resulting projections were then stitched together using the bio-imaging software Imago to create a seamless single mosaic on the order of $\sim 17\,500 \times 17\,500$ pixels ($\sim$54 002 µm).

A total of seven mosaics were created for this study, denoted as GFP1, GFP2, GFP3, GFP8, GFP11, GFP12 and GFP13. All retinas were stained with anti-glial fibrillary acidic protein (GFAP) and anti-collagen IV. Astrocytes express GFAP (Bignami and Dahl, 1974), outlining the cytoskeleton of each astrocyte in the retina. Because antibodies to GFAP label only the cytoskeleton of the cell (Bushong *et al.*, 2002), we also collected images of astrocytes in which the entire cell is filled with the intracellular dye, Lucifer yellow. We were able to verify using this data that the binarized area of cells arrived at from the Lucifer Yellow labeling plotted against the binarized area of cells arrived at from the GFAP labeling generally followed a positive linear trend. This indicated that if the GFAP staining shows a cell is large, the actual cytoskeleton will also be large.

GFP-transgenic mice were injected at their embryonic stage such that their astrocyte nuclei were stained. However, due to false negative centers of the astrocytes are still manually marked. The retinal vasculature was captured by examining the anti-collagen IV labeling with anti-GFP. An example mosaic is shown in Fig. 1. Down-sampled versions of all seven retinas can be found on BISQUE (see Availability and implementation). Incisions were made in the retina to flatten it for imaging, hence the cross-shape. The retinal vasculature is clearly visible as the tree-like structures within the images. Individual astrocytes are visible as small star-shaped cells and can be seen in detail in Fig. 1B.

Several manual pre-processing steps were taken prior to the automated analysis which follows. These include manual marking of astrocyte cell centers, and tracing of the major blood vessels (see Fig. 2) using NeuronStudio (Wearne *et al.*, 2005). We estimated the average cell radius by measuring the distance from the cell center to the farthest point of the cell from the center. We also labeled major blood vessels as either arteries or veins, which are easily identified by their 'conveying; type branching (Ganesan *et al.*, 2010). Although an entirely automated method for analysis would be preferred we performed these steps manually in order to ensure the biological accuracy of the resulting models and spatial analysis.

2.2 Cell segmentation

We present a new segmentation algorithm for astrocytes and compare it with the results of random walk segmentation (Ljosa and Singh, 2006) in order to verify that our biological findings are robust to varying segmentation results.

2.2.1. Random walk segmentation

The random walk-based segmentation is an algorithm for probabilistic cell segmentation that requires taking a large number of jumps to various pixels within the image. The algorithm starts at the center of the cell of interest, and each at each step jumps to one of the 8 pixels neighboring the current location in the image, chosen so that the step is more likely to be in the direction of a bright neighbor pixel. Otherwise the algorithm will jump back to the cell center with 'restart' probability that attempts to prevent the algorithm from traveling to other nearby astrocytes, since a single cell segmentation is what we desire. A separate matrix of the same dimensions as the image is updated at each step to keep count of how many times each pixel has been visited and conclusions about which pixels belong to the cell with what probability will be based on this 'visit record' matrix. The number of visits to each pixel is normalized to provide an output segmentation probability map for the cell. The restart parameter for the random walk restart parameter was optimized for similar data in Ruttenberg *et al.* (2013) and found to be 5×10^{-5}.

2.2.2. Adaptive threshold segmentation

An alternative segmentation is arrived at using local adaptive thresholding (http://homepages.inf.ed.ac.uk/rbf/HIPR2/adpthrsh.htm), an algorithm which separates the cell from the background, allowing for non-uniform illumination. As opposed to applying a global threshold to the astrocyte (GFAP) + nuclei (Lucifer Yellow) channels of the entire 501×501 region of interest (ROI), we first normalize the ROI to the scale [0,1] and move through the image considering sliding windows of size $ws \times ws$, at each step thresholding the foreground at $mean - C$. For best results we used $ws = 501$, $C = 0.5$. We then mask the original ROI with the foreground and apply a Gaussian decay to the resulting image, which is centered at the cell center and has $\sigma = 50\,\mu m$, which is a slightly inflated manually calculated average cell radius. The masking step is analogous to the restart parameter of the random walk in that it is necessary to ensure that the adaptive thresholding does not leak outside the actual cell extent. We normalize the resulting image again to the scale [0,1] and treat this as a segmentation probability map. Note that this segmentation method is much lower in computational complexity than the random walk, and as we will see in later sections, produces statistically similar results.

Binarization was then performed using the same simple three step process for both probabilistic segmentation results. First we create the binary Mask 1, which is the local adaptive thresholding

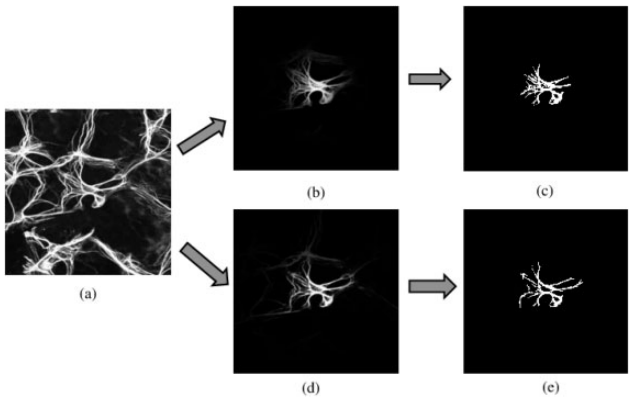

Fig. 3. Example segmentation pipeline. (**a**) Input ROI, (**b**) adaptive threshold segmentation, (**c**) binarized adaptive threshold segmentation, (**d**) Random walk segmentation, (**e**) binarized random walk segmentation

($ws = 501$, $C = 0.5$) of the output segmentation probability map. Then we create the binary Mask 2, which is a low global thresholding (at mean $+ 3\sigma$) of the probability map. The output binarization is the largest connected component of Mask 1 'AND' Mask 2. Parameters are adjusted such that the resulting binarized cells visually agreed with the input probability maps, as shown in Fig. 3.

2.3 Cell characteristics

Since we know very little about which morphological characteristics of astrocytes are important in differentiating classes of cells, we performed an exploratory search for possible relevant features which begins with a correlation analysis.

2.3.1 Correlation analysis

We started with seven attributes for each binarized cell which include Area, Perimeter, Eccentricity, Equivalent Diameter, Euler Number, Fraction of Convex Hull (also called 'Solidity'). For details on the definition of these attributes we refer the reader to the Matlab help page (http://www.mathworks.com/help/images/ref/regionprops.html). We normalized each of the characteristics with respect to its mean and standard deviation per retina to minimize minor imaging inconsistencies. Upon calculating correlation coefficients for each of these and dropping all correlations above 0.70 we arrive at only four characteristics, Area, Eccentricity, Fraction of the Convex Hull and Perimeter whose correlations are each no higher than ± 0.56. Equivalent Diameter was dropped due to its 0.99 correlation with Area, and Euler Number is dropped due to its 0.79 correlation with Area.

2.3.2 Clustering

We used an unsupervised Gaussian Mixture model clustering on each characteristic separately using the BIC criterion (Fraley *et al.*, 2012) and allowing for anywhere from 1 to 15 clusters in order to find the naturally occurring classes of cells. We arrived at three clusters for Area, five classes for Eccentricity, four classes for FracHull and four classes for Perimeter. We also verified that the distributions of clustered classes remain the same regardless of which segmentation method was used.

2.4 Multitype point processes

An unmarked homogeneous Poisson process with intensity λ is characterized by a few important properties. First, $N(A)$, the number of points in region A, is Poisson distributed with mean $\lambda|A|$ for all A, and has the probability function $P(N(A) = k) = e^{-}(\lambda|A|)(\lambda|A|)^k/k!$. Conditional on $N(A) = n$, the n points are independent and uniformly distributed in A.

In the case of marked point processes, a mark $m_i \in M$ is associated with each point x_i, in our case this mark corresponds to the cluster numbers for each of the various cell characteristics. Since our mark takes on discrete values our process is termed a 'multitype point process'. We leave the higher complexity full evaluation of continuous marks for another analysis, as the trends of interest are still accurately displayed through discretization. The intensity can then be evaluated per mark, and their inter-dependencies studied. A homogeneous multi-type Poisson process is one where each component process X_i has a constant intensity $\lambda_i > 0$ for all $m_i \in M$. The unmarked process X^* has constant intensity $\lambda^* = \sum_{i=1}^{M} \lambda_i$. The marks are independently and identically distributed (iid) with probability $p_i = \lambda_i/\lambda^*$.

The homogeneous Poisson process may be generalized to an inhomogeneous process by making λ depend on the point u in space i.e. $\lambda = \lambda(u)$. We can then model the inhomogeneous Poisson process as being dependent on various spatial covariates, such as a distance function $d(u)$ i.e. $\lambda = \lambda(d(u))$. It is possible to model certain marks as homogenous and others as spatially varying, in various combinations, to arrive at an appropriate model. For a more detailed introduction on multi-type point processes we refer the reader to Moller and Waagepetersen (2003).

In the case of the Queensland dataset (Berman, 1986) where ore deposits surrounding geological lineaments are modeled, the covariates considered include $d(u, L)$, the distance from any location u to the nearest lineament L and $\theta(u, L)$ the orientation of that lineament. In our case we would like to use the following spatial covariates, which for the case of GFP11 are visualized as heat-maps in Section 1 of Supplementary Data.

- **d0(u,ONH)**, the radial distance from any location to the ONH
- **d1(u,V)**, the distance from any location to the nearest point on the vasculature, V
- **d2(u,V,ONH)** the geodesic distance from the projected point on the vasculature to the optic disk along the linear network of the vasculature

For simplicity of notation we will refer to them as d0, d1 and d2, dropping the explicit dependencies on ONH and V. Note that these quantities are dependent on the retinal window, and the structure of the vasculature, and therefore modeling of these quantities within the retina must be done on a retina-by-retina basis. There are a variety of methods that can be used to evaluate the dependence of the point process intensity on these covariates, including visual tests such as Q–Q plots and kernel-smoothing estimates, and formal tests of constant intensity such as Kolmogorov-Smirnov test. We choose to use the Q–Q plot for visualization, and for d2 we use the linear network version of the Q–Q-plot as described in Baddeley *et al.* (2014) and Jammalamadaka *et al.* (2013).

3 Results

3.1 Cell clustering

3.1.1 Segmentation similarity

We ran a 2-sample Kolmogorov-Smirnov test between the cell characteristics derived from the two different segmentation methods to evaluate the dependence of our analysis on the segmentation step. For this test we pooled values for each attribute (Area, Eccentricity, Fraction of the Convex Hull and Perimeter) over all retinas and found there was no statistically significant difference for any of

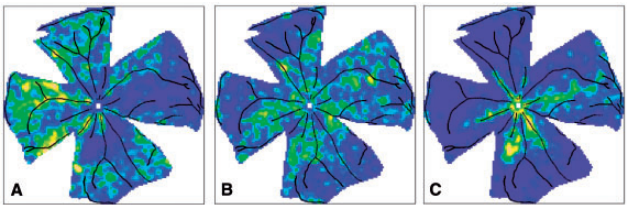

Fig. 4. Estimation of astrocyte intensity for Mark 1 (**A**), Mark 2 (**B**) and Mark 3 (**C**) in GFP11. The heat-map scale bar is in points per μm^2. Blood vessels are drawn in black. Large cell intensity is greater near veins, and toward the ONH, while small cell intensity is greater near arteries (Color version of this figure is available at *Bioinformatics* online.)

them ($P > 0.9993$). The exceptional similarity between resulting attributes verifies the robustness of our conclusions to variance in the segmentation step and allows us to proceed using either one of the methods for further analysis steps. We chose to use the adaptive segmentation method as opposed to the random walk due to its simplicity and computational efficiency.

For brevity, and due to the spatial similarities found with regard to clusters of each attribute described below, we chose to continue our analysis using the Area attribute only. We find that in general the cells which are lower in eccentricity, higher in fraction of convex hull, and lower in perimeter follow a similar pattern to the large area cells, i.e. close to the ONH and oftentimes more dense around veins. This can be seen in Section 2 of Supplementary Data, where we have included color-graded plots of the various classes for the three remaining characteristics along with their normalized means in the legend. The analysis that follows is repeated for each of these three remaining characteristics with no modification however the biological insights gained were determined to be minimal because of the aforementioned similarity in patterns. Possible future work includes exploring other cell characteristics such as the angle, length and number of primary processes, although these can be hard to define.

3.1.2 Kernel smoothed intensity estimation

To get an overall view of how the classes of cells described above vary spatially within the retina, we first computed a kernel smoothed intensity function for each class separately using a Gaussian kernel with standard deviation 150 μm. We chose this value because it is three times the average nearest neighbor distance between cell centers, an approximation to the 3σ value normally used.

In Fig. 4 we show the intensity estimation results for the characteristic 'Area' and make note of some observed spatial trends for both small cells (Mark 1) and large cells (Mark 3). Specifically, we can see that large area cells seem to hug the veinal structure and seem to have greater intensity closer to the ONH, whereas small area cells seem to hug the arterial structure and do not necessarily have greater intensity closer to the ONH. Note that these patterns recur in all six remaining retinas shown in Section 3 of Supplementary Data.

3.2 Evaluation of spatial covariates

Our next goal is to evaluate the previously proposed spatial covariates using the tools described above.

3.2.1 Test of homogeneity

We first ran a test of homogeneity for each mark separately to determine whether the usage of spatial covariates was necessary or if a model with a constant lambda will suffice. We performed a test of Complete Spatial Randomness (CSR) for the observed point pattern corresponding to each mark, based on quadrature counts (Baddeley and Turner, 2005). The retinal window was divided into tiles and the number of data points in each tile was counted. The expected number of points in each quadrature according to CSR was calculated and a χ^2 goodness-of-fit test was performed. The resulting P-values were highly significant ($P < 2.2e^{-16}$) for each cell size and each retina, therefore rejecting the null hypothesis of homogeneity.

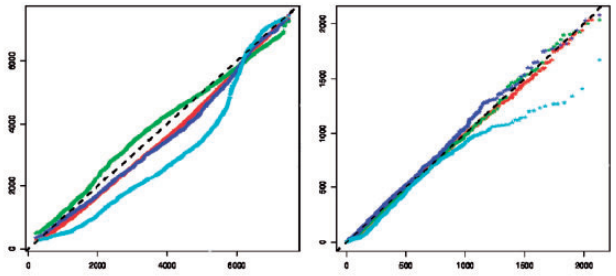

Fig. 5. Superimposed d0 (**A**) and d1 (**B**) Q–Q plots for each mark, the unmarked point process and the uniform distribution for GFP11. The quantiles of the covariate with respect to the observed astrocytes are on the *y*-axis, and with respect to a uniform simulation on the *x*-axis. Large cells are in light blue, medium size cells in dark blue, small cells in green, all cells in red, and the uniform $y = x$ line is dashed and black. (Color version of this figure is available at *Bioinformatics* online.)

d0, radial distance from ONH

The Q–Q plots in Fig. 5 show that the various mark distributions are dependent on the quantity d0, much more so than the unmarked process (in red). The large area cells (Mark 3) seem to be more clustered around the ONH than would be expected from a uniform distribution of cells, and the small cells are repulsed by the ONH. The medium size cells (Mark 2) seem to be closer to uniform and follow the pattern of the unmarked process fairly closely. The d0 Q–Q plots for the remaining retinas can be found in Section 4 of Supplementary Data. Figure 5A also shows an example of the large and small cells reversing their roles at the highest distances from the ONH, and upon inspection of the original image data it is clear that this is due to the presence of major veins running along the outer border of the retina. For this reason, and because we find that d2 captures this trend in a more biologically accurate way, we decided to drop d0 from our analysis.

d1, distance to the nearest blood vessel

We first calculated the d1 variable over all cells and the results are shown in Fig. 5B. The plot does not show much difference between marks, although the difference seems apparent in the kernel density smoothing shown in Fig. 4. Since we noticed that the arteries and veins seem to have different kernel density intensities, we decide to separate results from arteries and veins and these more informative results are shown in Section 5 of Supplementary Data. An example artery and vein Q–Q plot can be found in Fig. 6.

Now that the blood vessels are separated into veins and arteries, we see here that there is actually a difference in spatial distribution of cell sizes between the two. The graphs in Fig. 6 show that small size cells are found closer to the arteries and large size cells are found near the veins. This is verified by the kernel smoothing images in Fig. 4 and by the remaining retinas in Section 5 of Supplementary Data.

d2, geodesic distance from the projected point along the blood vessel to the ONH

For d2 blood vessels must necessarily be measured separately due to the definition of the distance. Note that there are gaps in the

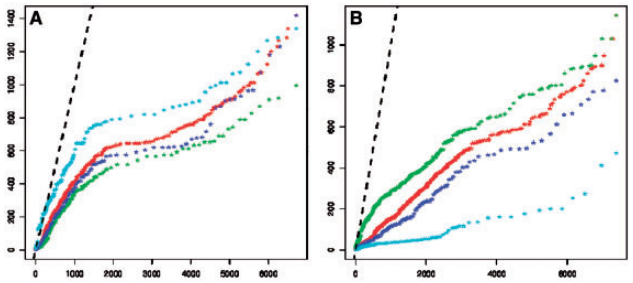

Fig. 6. Superimposed d1 Q–Q plots for each mark, the unmarked point process and the uniform distribution for two major blood vessels, an artery (**A**) and a vein (**B**), of GFP11. See Fig. 5 caption for legend. The plots for all 12 veins and arteries can be found in Section 5 of Supplementary Data. (Color version of this figure is available at *Bioinformatics* online.)

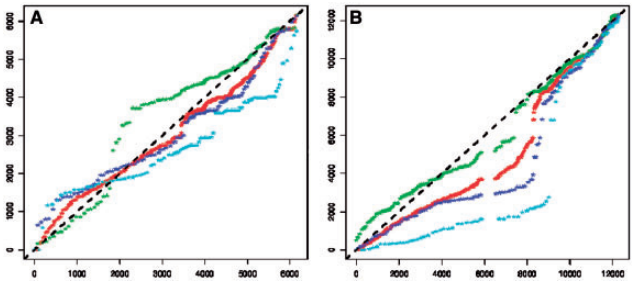

Fig. 7. Superimposed d2 Q–Q plots for each mark, the unmarked point process and the uniform distribution for the same artery (**A**), and vein (**B**), of GFP11 that are shown in Fig. 6. See Fig. 5 caption for legend. The plots for all 12 veins and arteries can be found in Section 6 of Supplementary Data. (Color version of this figure is available at Bioinformatics online.)

Q–Q-plots where the blood vessel trace falls outside of the retinal window, since we dealt with the empty sections of blood vessel traces caused by retinal incisions by simply also removing the simulated uniform points along the blood vessel which fall outside of the retinal window. This phenomena are especially apparent in the d2 Q–Q plots for GFP2, GFP8, GFP12 and GFP13 (see Section 6 of Supplementary Data). In the Q–Q plots of the two major blood vessels shown in Fig. 7 we can see that large type cells typically exist closer to the ONH than small cells, medium cells or all cells. This distinction is clearer for veins than for arteries. In addition, the distributions of all cell types seem closer to uniform when the nearest blood vessel is an artery instead of a vein. This is also generally true of the remaining retinas shown in Section 6 of Supplementary Data. Please note that in these figures the arteries and veins are differentiated in the title of each plot, where A stands for artery, and V stands for vein. Some of the shortest veins have a lower number of cells total and this could also attribute to the somewhat erratic nature of the corresponding graphs.

3.3 Inhomogeneous Poisson intensity model

Based on our preliminary analyses above, we fit a point process model to an observed point pattern using the covariates d1, d2, and whether the nearest blood vessel is an artery or vein. A quadrature scheme was constructed which specifies both the data point pattern and a dense grid of dummy points. The model was fit by maximizing the pseudolikelihood (Besag, 1975) using the Berman-Turner computational approximation (Baddeley and Turner, 2000; Berman and Turner, 1992). Maximum pseudolikelihood is equivalent to maximum likelihood if the model is a Poisson process, either homogeneous or inhomogeneous, as is the case in our implementation. We used standard border correction, in which the quadrature window (the domain of integration of the pseudolikelihood) is obtained by trimming off a margin of fixed width from the observation window of the data pattern.

The total intensity model for astrocyte cells in the retina is

$$\lambda_{\text{all}} = \lambda_1 + \lambda_2 + \lambda_3 \tag{1}$$

$$\lambda_i(u, \text{ONH}, V) = \beta_i e^{\alpha 1_i * d1(u,V) + \alpha 2_i * d2(u,V) + \alpha 3_i * AV} \tag{2}$$

where the parameters $\beta, \alpha 1, \alpha 2, \alpha 3$ were estimated separately for each cell size. The resulting estimated parameters are shown in Tables 1–3. From these tables we can see that using a Z-test all coefficients are significant ('***' implies $P < 0.001$, '**' implies $P < 0.01$, '*' implies $P < 0.05$) except for the artery-vein variable for medium size cells, which was previously noted. We can also see that

Table 1. Parameters for point process model of small cells

	Estimate	S.E.	Z-test
(Intercept)	−1.127e + 01	7.204e − 02	NA
d1	−1.759e − 04	8.130e − 05	*
d2	−1.682e − 05	7.719e − 06	*
av	3.673e − 01	5.848e − 02	***

Table 2. Parameters for point process model of medium cells

	Estimate	S.E.	Z-test
(Intercept)	−1.112e + 01	6.861e − 02	Na
d1	1.826e − 04	8.422e − 05	*
d2	−5.028e − 05	7.913e − 06	***
av	−1.100e − 01	5.884e − 02	

Table 3. Parameters for point process model of large cells

	Estimate	S.E.	Z-test
(Intercept)	−1.050e + 01	6.675e − 02	na
d1	−4.386e − 04	1.179e − 04	***
d2	−1.176e − 04	9.143e − 06	***
av	−9.135e − 01	6.562e − 02	***

the large cells have the most distinct pattern as influenced by the spatial covariates listed. As we noticed in the kernel smoothing estimate and in the Q–Q plot of d1, small cells have a positive coefficient for the artery (av = 1) or vein (av = 0) variable whereas large cells have a highly negative coefficient, and medium cells are closer to the neutral coefficient of 0. Small and large cells have a negative coefficient for d1, meaning that as one moves away from the blood vessels the intensity drops exponentially. Large cells drop off more quickly than small cells, and medium cells are more uniformly distributed in the retina with less regard to the vasculature. As for the covariate d2, it is negative for small and large cells, and more negative for medium cells. This implies that medium cells are more clustered around the ONH along the blood vessels and we postulate that this is just an artifact of the random distribution of medium-size cells in this particular retina, as we do not observe this in many of the six remaining retinas as shown in Section 7 of Supplementary Data. We refrain from remarking in detail on the parameters estimated for the rest of the retinas as we find it more useful and

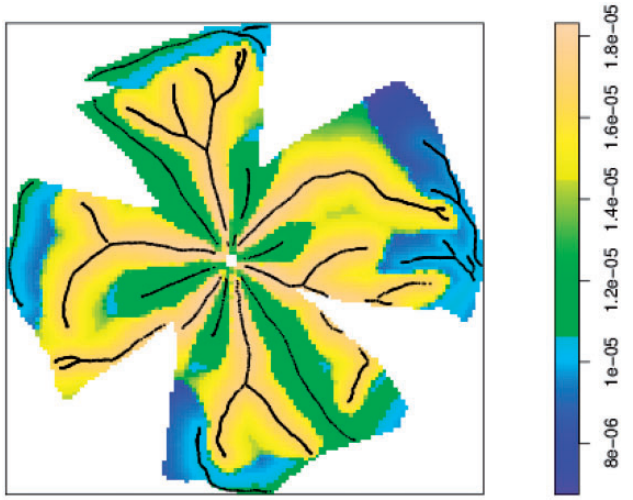

Fig. 8. Heatmap of fitted point process model conditional density for Mark 1 on GFP11. (Color version of this figure is available at *Bioinformatics* online.)

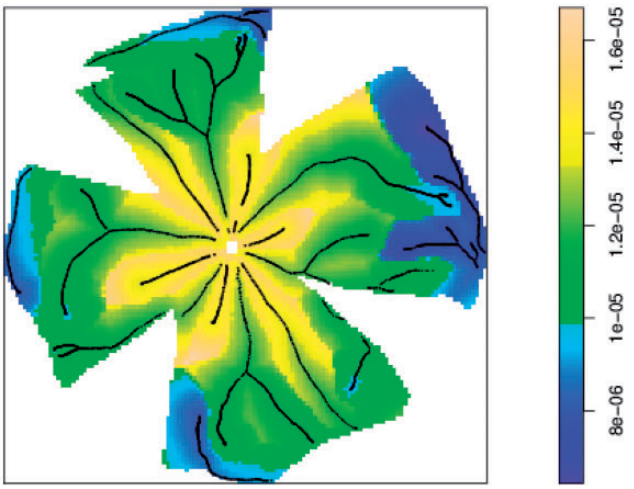

Fig. 9. Heatmap of fitted point process model conditional density for Mark 2 on GFP11 (Color version of this figure is available at *Bioinformatics* online.)

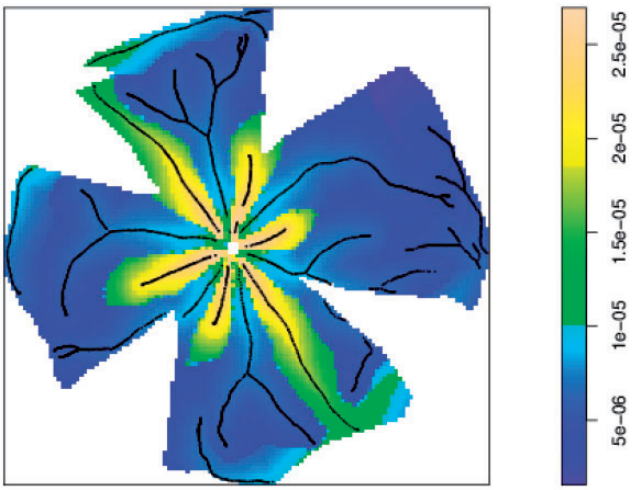

Fig. 10. Heatmap of fitted point process model conditional density for Mark 3 on GFP11 (Color version of this figure is available at *Bioinformatics* online.)

comprehensible to compare the conditional density maps, estimated as described below.

Given each point process model fitted to its corresponding point pattern, we compute the fitted conditional intensity (Baddeley *et al.*, 2005) of the model at the points of the quadrature scheme used to fit the model. From this we obtain $\lambda(u, \text{ONH}, BV)$ values for each of the quadrature points. We then perform spatial smoothing of lambda values observed at the set of quadrature locations using Gaussian kernel smoothing (Nadaraya, 1964; Watson, 1964). From this we obtain a heatmap of conditional density for each point process model, as shown in Fig. 8 (and Section 8 of Supplementary Data). From these heat-maps we can see that the large cells (Mark 3) seem to have a high point-process intensity in a flower-like shape around the ONH, with petals centered on the veins. The likelihood of finding small cells in this area is low, but is higher around the arteries. The veins and arteries are drawn in black in these figures for the reader's benefit.

4 Discussion

In this study, we presented what we believe to be the most appropriate choice of statistical model to represent astrocyte locations and sizes in healthy mammalian retinas and most importantly, the biological conclusions derived from those models. Due to very little prior knowledge regarding these distributions, we performed an exploratory analysis which necessitates steps such as hypothesis testing and model selection. These resulting trends regarding the size of astrocytes and their spatial locations within the retina, particularly with respect to the surrounding arteries and veins are new biological discoveries that have not yet been published to our knowledge. Although some of the trends were visible from the kernel smoothing estimate of the astrocyte intensity, the full quantification of the distribution of cells and the statements regarding the statistical significance of the dependence on the proposed spatial covariates could not have been made without usage of the appropriate point process models.

Our results show that in normal healthy retinas, the distribution of observed astrocyte cells is more densely packed around the blood vessels than a uniform distribution. We also show that compared with the distribution of all cells, large cells are more densely packed in the vicinity of veins and toward the ONH whereas smaller cells are often more densely packed in the vicinity of arteries. The conditional density maps shown in Figures. 8–10 and in Section 7 of Supplementary Data show that the density of large cells is clearly higher in a flower-shaped region with the petals centered on the retinal veins, whereas small cells have low density in these regions. Due to this clear differentiation between arteries and veins, the results of this study could potentially aid in automated differentiation between arteries and veins in the retina, which has proven to be a difficult task (Pilat *et al.*, 2014).

A possible explanation for these phenomena is related to the vascular function within the retina. The retinal vasculature enters the retina through the central retinal artery via the ONH, and after being distributed through the retinal tissue, it leaves the tissue through the retinal vein. Astrocytes sit between vasculature and retinal neurons and although many functions of astrocytes in healthy retinas are poorly understood, it is widely accepted that they play an essential role in the development and function of the retinal vasculature, blood flow and BRB (Kur *et al.*, 2012). In fact glial cell dysfunction in retinal pathologies is associated with retinal swelling and BRB breakdown (Bringmann *et al.*, 2006; Klaassen *et al.*, 2013;

Shen *et al.*, 2010). Astrocytes transport small molecules (glucose, glutamate, small proteins or polypeptides) from the blood stream to the neurons at the arteries and take metabolic byproducts away from the neurons back into the blood stream at the veins. We hypothesize that the latter process may be more critical, since for example it is well known that too many metabolic waste products can be toxic to neurons. We speculate that this criticality may be the reason for the astrocytes residing on the veins to be the largest, as larger cells can presumably process waste more efficiently.

In this article, we have argued that the explanatory power of each covariate considered follows a common trend across all seven retinas studied, but have not yet shown the performance of the generative spatial point process model derived here on 'hold-out' retinal data. Future work will include validating the generative model that we have created on a hold-out area from each retina, using the parameters generated from the rest of that retina, as well as holding out a few entire retinas and re-using parameters derived from other retinas. The latter requires normalization for the varying physical quantities between retinas such as shape, orientation, size, number of arteries, veins and overall intensity of cells. We believe that once the retinal normalization issue is resolved the resulting models will allow us to predict the locations and sizes of the cells in a new retina based on its blood vessel structure and directly compare parameters between retinas in other animals or stages of treatment, disease or detachment without needing these manual markings. This study can serve as a solid foundation for an end-to-end script which analyzes and simulates retinal cell distributions as well as spatial dependencies between other vasculature structures and cells, which occur many places in the body.

Funding

This work was supported by the National Science Foundation (NSF) award [grant number III-0808772] and a grant from the International Retina Research Foundation.

Conflict of Interest: none declared.

References

Armstrong,R.A. (2003) Measuring the degree of spatial correlation between histological features in thin sections of brain tissue. *Neuropathology,* **23**, 245–253.

Baddeley,A. and Turner,R. (2000) Practical maximum pseudolikelihood for spatial point patterns. *Aust. N. Z. J. Stat.,* **42**, 283–322.

Baddeley,A. and Turner,R. (2005) spatstat: an R package for analyzing spatial point patterns. *J. Stat. Softw.,* **12**, 1–42.

Baddeley,A., *et al.* (2005) Residual analysis for spatial point processes (with discussion). *J. R. Stat. Soc. B,* **67**, 617–666.

Baddeley,A., *et al.* (2014) Multitype point process analysis of spines on the dendrite network of a neuron. *J. R. Stat. Soc.*, **63**, 673–694.

Berman,M. (1986) Testing for spatial association between a point process and another stochastic process. *Appl. Stat.,* **35**, 54–62.

Berman,M. and Turner,T.R. (1992) Approximating point process likelihoods with glim. *Appl. Stat.,* **41**, 31–38.

Besag,J. (1975) Statistical analysis of non-lattice data. *The Statistician,* **24**, 179–195.

Bignami,A. and Dahl, D. (1974) Astrocyte-specific protein and radial glia in the cerebral cortex of newborn rat. *Nature,* **252**, 55–56.

Bringmann,A. *et al.* (2006) Müller cells in the healthy and diseased retina. *Prog. Retin. Eye Res.,* **25**, 397–424.

Bushong,E.A. *et al.* (2002) Protoplasmic astrocytes in CA1 stratum radiatum occupy separate anatomical domains. *J. Neurosci.,* **22**, 183–192.

Cressie,N. (1992) Statistics for spatial data. *Terra Nova,* **4**, 613–617.

Dorrell,M.I. and Friedlander,M. (2006) Mechanisms of endothelial cell guidance and vascular patterning in the developing mouse retina. *Prog. Retin. Eye Res.,* **25**, 277–295.

Foxall,R. and Baddeley,A. (2002) Nonparametric measures of association between a spatial point process and a random set, with geological applications. *J. R. Stat. Soc. C,* **51**, 165–182.

Fraley,C. *et al.* (2012) *mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation.* Technical Report no. 597, Department of Statistics, University of Washington.

Ganesan,P. *et al.* (2010) Development of an image-based network model of retinal vasculature. *Ann. Biomed. Eng.,* **38**, 1566–1585.

Gariano,R.F. and Gardner,T.W. (2004) Retinal angiogenesis in development and disease. *Nature,* **438**, 960–966.

Jammalamadaka,A., *et al.* (2013) Statistical analysis of dendritic spine distributions in rat hippocampal cultures. *BMC Bioinformatics,* **14**, 287.

Kimelberg,H.K. (2010) Functions of mature mammalian astrocytes: a current view. *The Neuroscientist,* **16**, 79–106.

Klaassen,I. *et al.* (2013) Molecular basis of the inner blood-retinal barrier and its breakdown in diabetic macular edema and other pathological conditions. *Prog. Retin. Eye Res.,* **34**, 19–48.

Kur,J. *et al.* (2012) Cellular and physiological mechanisms underlying blood flow regulation in the retina and choroid in health and disease. *Prog. Retin. Eye Res.,* **31**, 377–406.

Kvilekval,K. *et al.* (2010) Bisque: a platform for bioimage analysis and management. *Bioinformatics,* **26**, 544–552.

Ljosa,V. and Singh,A.K. (2006) Probabilistic segmentation and analysis of horizontal cells. In: *Data Mining, 2006.* ICDM'06. Sixth International Conference on, pp. 980–985. IEEE.

Metea,M.R. and Newman,E.A. (2006) Glial cells dilate and constrict blood vessels: a mechanism of neurovascular coupling. *J. Neurosci.,* **26**, 2862–2870.

Moller,J. and Waagepetersen,R.P. (2003) *Statistical Inference and Simulation for Spatial Point Processes*. CRC Press, Boca Raton.

Nadaraya,E. (1964) On estimating regression. *Theory Prob. Appl.,* **9**, 141–142.

Paula,K.Y. *et al.* (2010) The structural relationship between the microvasculature, neurons, and glia in the human retina. *Invest. Ophthalmol. Vis. Sci.,* **51**, 447–458.

Pilat,A.V. *et al.* (2014) Morphology of retinal vessels in patients with optic nerve head drusen and optic disc edema. *Invest. Ophthalmol. Vis. Sci.,* **55**, 3484–3490.

Raff,M. *et al.* (1983) Two types of astrocytes in cultures of developing rat white matter: differences in morphology, surface gangliosides, and growth characteristics. *J. Neurosci.,* **3**, 1289–1300.

Ruttenberg,B.E. *et al.* (2013) Quantifying spatial relationships from whole retinal images. *Bioinformatics,* **29**, 940–946.

Shekhar,S. and Huang,Y. (2001) Discovering spatial co-location patterns: a summary of results. In Jensen, C.S. (eds.) *Advances in Spatial and Temporal Databases,* pp. 236–256. Springer, Berlin.

Shen,W. *et al.* (2010) Retinal vascular changes after glial disruption in rats. *J. Neurosci. Res.,* **88**, 1485–1499.

Stone,J. and Dreher,Z. (1987) Relationship between astrocytes, ganglion cells and vasculature of the retina. *J. Compar. Neurol.,* **255**, 35–49.

Suematsu,M. *et al.* (1994) Spatial and temporal correlation between leukocyte behavior and cell injury in postischemic rat skeletal muscle microcirculation. *Lab. Invest.,* **70**, 684–695.

Watson,G.S. (1964) Smooth regression analysis. *Sankhya Indian J. Stat., A,* **26**, 359–372.

Wearne,S. *et al.* (2005) New techniques for imaging, digitization and analysis of three-dimensional neural morphology on multiple scales. *Neuroscience,* **136**, 661–680.

Whitney,I.E. *et al.* (2008) Spatial patterning of cholinergic amacrine cells in the mouse retina. *J. Compar. Neurol.,* **508**, 1–12.

Zahs,K.R. and Wu,T. (2001) Confocal microscopic study of glial-vascular relationships in the retinas of pigmented rats. *J. Compar. Neurol.,* **429**, 253–269.