8-12-2016

# Spatial Analysis of Retinal Pigment Epithelium Morphology

Haitao Huang

Recommended Citation
Huang, Haitao, "Spatial Analysis of Retinal Pigment Epithelium Morphology." Thesis, Georgia State University, 2016.
doi: https://doi.org/10.57709/8684852

# SPATIAL ANALYSIS OF RETINAL PIGMENT EPITHELIUM MORPHOMETRY

by

HAITAO HUANG

Under the Direction of Yi Jiang, PhD

## ABSTRACT

In patients with age-related macular degeneration, a monolayer of cells in the eyes called retinal pigment epithelium differ from healthy ones in morphology. It is therefore important to quantify the morphological changes, which will help us better understand the physiology, disease progression and classification. Classification of the RPE morphometry has been accomplished with whole tissue data. In this work, we focused on the spatial aspect of RPE morphometric analysis. We used the second-order spatial analysis to reveal the distinct patterns of cell clustering between normal and diseased eyes for both simulated and experimental human RPE data. We classified the mouse genotype and age by the k-Nearest Neighbors algorithm. Radially aligned regions showed different classification power for several cell shape variables. Our proposed methods provide a useful addition to classification and prognosis of eye disease noninvasively.

INDEX WORDS: Retinal pigment epithelium, Age-related macular degeneration, Cell morphometric data, Spatial analysis, k-Nearest Neighbors algorithm, Classification

# SPATIAL ANALYSIS OF RETINAL PIGMENT EPITHELIUM MORPHOMETRY

by

HAITAO HUANG

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2016

# SPATIAL ANALYSIS OF RETINAL PIGMENT EPITHELIUM MORPHOMETRY

by

HAITAO HUANG

Committee Chair: Yi Jiang

Committee: Xin Qi
Yichuan Zhao

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
May 2016

# DEDICATION

For my parents.

And for those who accompanied me along the journey.

# ACKNOWLEDGEMENTS

I would like to first thank my thesis advisor, Dr. Yi Jiang, without whom this thesis could not have been completed. She is knowledgeable, joyous and kind-hearted, always encouraging and supporting me to explore new realms. Most importantly, she led me into the world of statistical research and showed me how to conduct rigorous research in her own exemplary manner, which I think shall bear influence on me throughout my academic career. It has been my honor to work with her.

Thank Dr. John Nickerson and various others at Emory Eye Center and Dr. Karina Mazzitello for providing data and for their helpful discussions. This work has been built on their efforts. Thank Dr. Yichuan Zhao and Dr. Xin Qi for their advising during the completion of this thesis. I benifited a lot from their courses as well.

I am grateful for Dr. Liang Peng and Dr. Gengsheng Qin for all their help in many different forms during my graduate study in this department.

My friends both at Emory and GSU, Juncheng, Nancy, Hao, Xiuxiu, Xing, Yi, etc., have made my life in US more colorful.

Finally, thank my parents for always being there guiding me. They are my motivation and hope.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

- RPE - retinal pigment epithelium
- AMD - age-related macular degeneration
- SPP - spatial point patterns
- CSR - complete spatial randomness
- PCF - pair correlation function
- PCA - principal component analysis
- kNN - k-Nearest Neighbors

# Chapter 1

# INTRODUCTION

## 1.1 Retinal Pigment Epithelium (RPE)

Retinal pigment epithelium (RPE) is a multifunctional monolayer of cells attached to the overlying retinal photoreceptor cells and the underlying choroid [1]. In healthy eyes, the RPE is a single sheet of densely packed hexagonal cells that maintains and nourishes the photoreceptors. Some functions of RPE include absorbing light, supplying nutrients and oxygen to photoreceptors, removing waste byproducts from the retina and communicating between inner space of the eye and blood side of the epithelium, such as cells of the immune system [2]. Disruptions in RPE are prevalent in aged eyes and in the pathogenesis of a number of ocular disorders, including age-related macular degeneration (AMD) [3].

Age is the largest risk factor for AMD, which plaques around 3% of American adults and its prevalence increases dramatically with age [3-4]. It is a progressive retinal degenerative disease that may result in blurred or loss of vision in the central visual field. Although the presence of a few small hard drusen is a normal, non-vision-impairing part of aging, the deposition of large diffuse (or soft) drusen in the macula impairs vision and is indicative of early AMD [5]. While the pathogenesis of AMD is not well understood, it typically involves imbalance between production of damaged cellular components and degradation, which leads to the accumulation of metabolic debris and death of photoreceptors. As AMD progresses to advanced-stages, it is categorized as either dry AMD ot wet AMD. Dry AMD begins with drusen in the macula between RPE and the choroid, and progresses to extensive atrophy of RPE and central geographic atrophy with photoreceptor loss. Wet AMD is exudative as proliferation of abnormal blood vessels in the retina (choroidal neovascularization) will lead to leak of blood and fluid, causing irreversible damage to photoreceptors and severe loss of central vision eventually [5-7]. Most AMD starts as the dry form and in some individuals, it progresses to the wet form with the passage of time. The dry form of macular degeneration is much more common than the wet form and makes up about 90% of all cases [8].

To maintain the RPE tissue structures, forces including adhesion, tension and contraction hold adjacent RPE cells together [9]. An actin-myosin cytoskeleton exerts contractile forces and leads to regular polygonal, mostly hexagonal shapes of RPE cells. The hexagonal network of cells is believed to be the most stable configuration of cells with the least amount of surface tension [10]. As AMD progresses, mechanical stress induced by drusens and RPE cell death will lead to shape deformation and re-organization in RPE cells. Disruptions in cell shape and cell packing would lead to measurable alteration of the RPE sheet patterns. Therefore, quantitative analysis of the morphology of RPE could be clinically useful to diagnose the disease stage and predict the progression for retinal degeneration and AMD. Moreover, due to the development of sophisticated non-invasive imaging technologies [11], analyses of RPE morphology may be available in the clinical setting for diagnosis and prognosis.

Extensive efforts have already been made in analyzing the morphological changes in RPE sheets with retinal degeneration. For example, RPE cell morphometrics in rd 10 mouse, such as number of neighbors, eccentricity and form factors, are analyzed by simply computing the mean and conducting nonparametric hyphothesis tests to compare the mean data in [12] and [13]. Such quantitative analyses are useful in characterizing the morphological changes of RPE cells with age. Jiang et al. [14, 15] proposed to discriminate mouse RPE genotypes and age through RPE cell morphology by functional principal component analysis and other classification methods. It was shown that RPE cells' area and aspect ratio are quantitative indicators of RPE age and genotype. These studies all used morphometric data for each flatmount as a whole, without discriminating the retinal regions. However, it is of pratical importance to describe the changes in RPE cell morphology with respect to location in the retina as RPE cell characteristics differ by location within each eye. In this thesis, we will further study RPE morphometry with regard to age, genotype and location of RPE with mouse models.

Another direction to pursue is the characterization of the spatial inhomogeneity within a single RPE flatmount as evident by visual inspection of degenerated eyes. In the RPE of mutant mice that mimic the development of AMD, the irregularity can occur as early as postnatal 30 days. While relatively large number of mouse eyes can be collected, it is prohibitively difficult for obtaining a large sample of human donor eyes. On the other hand, simple comparison of averaged morphometric variables may not be sensitive enough to detect the irregularity at the early stage of the disease. From these perspectives, new quantitative tools are urgently needed to be able to reflect the subtle disruption deviating from the normal cell pattern. Previous studies of human RPE morphology have found spatial differences in cell density: a high density of cells in the macula, decreasing peripherally. Recently, spatial point analysis has emerged as an appropriate technique in revealing the spatial irregularity of cells, for example, see [16]. This approach is further facillitated by the popular 'spatstat' R package. We propose to study RPE cell morphology in human eyes through spatial analysis of both simulated and experimental human RPE sheets.

### 1.2 Description of Data

For the spatial analysis, we used both simulated and real human RPE data. Simulated RPE images were generated by Dr. Karina Mazzitello, where a normal RPE sheet undergoes varied degrees of clustered cell death, as shown in Figure 1.1. The black hollow area is where clustered cell death occurs. From these simulated images we used the ImageJ software to export cell centroid coordinates. We obtained real human RPE data by first segmenting the RPE images and then extracting the cell centroid data in ImageJ (NIH).

For the classification study, all RPE images and cell morphometric data came from our collaborators at the Emory Eye Center, especially from Dr. John Nickerson's Lab. A total of 107 mouse RPE flatmounts were collected. Two strains of mice were considered, namely the C57BL/6J genotype as the wild type and rd10 as the mutant, at ages postnatal 30 (labelled as P30), 45, 60, 100, 180 and 330 days. Each flatmount was dissected manually into 4 flaps (Inferior, Superior, Nasal, Temporal) and further overlaid with concentric rings in

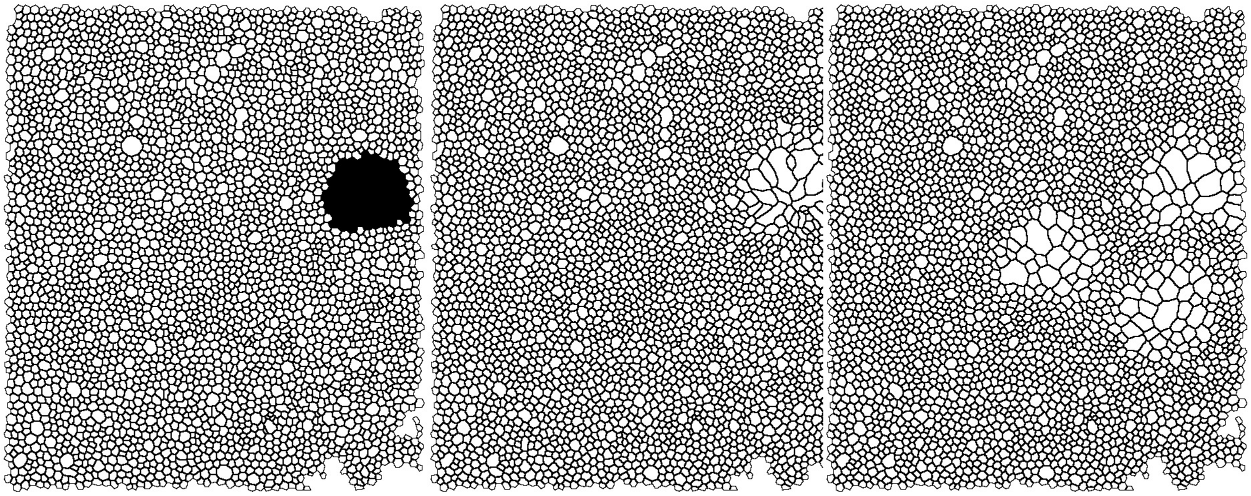


**Figure 1.1** Simulated sequence of RPE sheets showing cell patterns resulting from clustered cell death (black hole in the left image, around 100 cells are killed). Courtesy of Dr. Karina Mazzitello.

image processing softwares. Cutboxes with dimensions of 265×180 pixels were automatically selected within different geographical regions and cell morphometric data were measured for all cells from each cutbox. Figure 1.2 shows an RPE flatmount image of a C57BL/6J mouse eye at age P60 overlaid with cutboxes and concentric rings. All eyes were divided into three age groups: young age group (group 1) for eyes less than or equal to P45, medium age group (group 2) for eyes greater than P45 but less than P120, and old age group for eyes greater than P120. Thus, each cell was labelled with four categorical variables: age, genotype, flap and zone. Around 1000-10,000 cells were measured in each RPE flatmount. Table 1.1 shows the number of mice in different groups of ages and genotypes. Roughly, the number of RPE sheets in different ages and genotypes are balanced to avoid systematic biases.

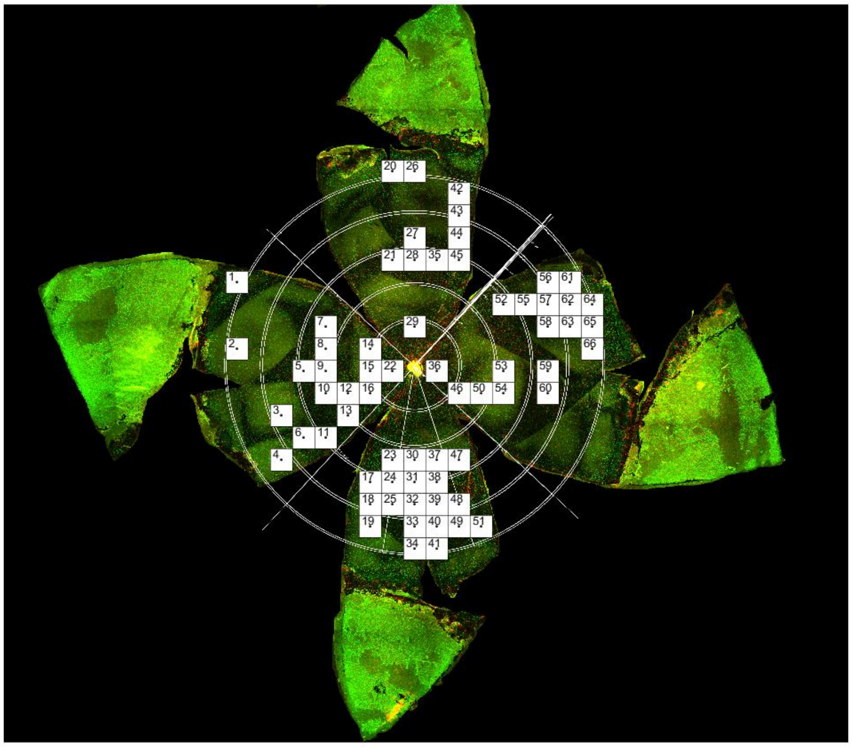


**Figure 1.2** An RPE flatmount image for a C57BL/6J mouse at age P60. The top, bottom, left, right flaps are called Superior, Inferior, Temporal and Nasal flaps, respectively. The concentric rings label zones from 1 to 4, from the center radially outward. The white cubes are cutboxes tracked by arabic number labels. Courtesy of Kevin Donaldson.

**Table 1.1** Number of mouse eyes in different genotype and age groups

| Genotype | Age | | | | | | Total |
|---|---|---|---|---|---|---|---|
| | Group 1 | | Group 2 | | Group 3 | | |
| | P30 | P45 | P60 | P100 | P180 | P330 | |
| C57BL/6J | 15 | 2 | 7 | 7 | 8 | 4 | 43 |
| rd10 | 12 | 13 | 11 | 13 | 13 | 2 | 64 |
| Total | 32 | | 38 | | 27 | | 107 |

Table 1.2 displays the cell morphometric variables we considered in this study as well as

their definitions. Area and perimeter are size parameters for cells whilst solidity, eccentricity, extent, form factor and major/minor ratio are measures of cell shape. Shape measures have been shown to be good indicators for the changes in rd10 morphology. Eccentricity shows how elongated a cell is. A perfect circle has an eccentricity value of 0, while an ellypse has a larger eccentricity value approaching 1. The definition of major/minor ratio is consistent with that of eccentricity. Form factor measures how far a cell deviates from the shape of a perfect hexagon. A cell with perfect equilateral hexagon shape has a form factor of about 0.64, while less regularly shaped hexagons have a smaller value. Solidity and extent are both differentiators of cells with protrusions or irregular shape compared to generally round cells.

**Table 1.2** Cell morphometric variables

| Morphometric Variable | Definition |
|---|---|
| Area | Area of the cell |
| Perimeter | Perimeter of the cell |
| Solidity | Area/Area of the convex hull |
| Eccentricity | Eccentricity of an ellipse fitted to the cell |
| Extent | Area/Area of the smallest enclosing rectangle |
| Form Factor | Area/Area of a circle with the same perimeter |
| Major/Minor Ratio | Major axis length over minor axis length |

## 1.3 Purpose of the Thesis

We examined the applications of several statistical methods in analyzing human RPE cell center data and mouse RPE cell morphometric data from the spatial aspect. Our purpose of this study is two-fold. First, we hypothesize that in AMD eyes, RPE cells exhibit more irregular shapes and we seek to quantify the globally present spatial disorder by spatial point analysis. Second, we hypothesize that the RPE morphology changes differently in different locations (macula, mid-periphery and far-periphery) of the retina and in different age/disease status groups. We aim to establish the relationships among RPE morphometric measures and age, RPE location and disease status of normal and AMD eyes that highlight morphological irregularities.

# Chapter 2

# METHODOLOGIES

## 2.1 RPE Flatmount and Data Preparation

Mice were euthanized with $CO_2$ in accordance with Emory IACUC guidelines and The Association for Research in Vision and Ophthalmology guidelines for treatment of animals. Eyes were marked on the superior side with a blue sharpie and then enucleated, fixed for 10 min in 10% neutral buffered formalin, and then washed 3 times with PBS. Extra tissue was removed from the outside of the globe. Flatmounting was done by making 4 radial cuts from the center of the cornea back towards the optic nerve. A drop of PBS was placed on the eye to keep it moist. The flaps were pealed away from the lens and the lens removed. The iris and retina were removed using forceps. Tension from the sclera was relieved by making cuts halfway through each flap at the ciliary body/cornea margin and small cuts through the ciliary body. The RPE sheet was imaged under fluorescence confocal microscopy after staining for ZO-1 to identify RPE cell boundaries. Substantial fractions of the RPE sheet (usually 20-50% of the sheet) could be analyzed. Details of the experimental procedure and the recent methodological advances for RPE imaging data preparation can be found in [11].

Human RPE images were obtained in similar procedures as above. For obtaining human RPE cell centroid data, accurate segmentation of RPE images is necessary. Our segmentation algorithm in ImageJ consists of the following steps:

1. Convert an RGB image to 8-bit grayscale.
2. Apply the bandpass filter to reduce noises (filter large structures down to 20 pixels).
3. Binarize the image using the mean-value threshold.
4. Segment the binary image by the watershed segmentation.

5. Obtain the final polygonal cell borders by the Voronoi diagram.
6. Extract the coordinates of center of mass (maxima) for further analysis.

## 2.2 Second-order Spatial Analysis

In this subsection, we review the fundamental concepts and theories underlying second-order spatial analysis. Spatial data abound in diverse fields, including forestry, agriculture, astronomy, epidemiology and cellular biology. Statistical models and related inference have been developed for random patterns of events. Diggle [17] gives the theoretical developments in spatial statistics. Illian et al. [18] and Waller and Gotway [19] include various applications of spatial analysis.

We define an event as an occurrence of interest (e.g., an incident case of a disease) and a point as any location in the study area where an event could occur. As the most basic model for random patterns, complete spatial randomness (CSR) defines a situation where events follow the uniform distribution within the study area, and are independent of one another. It serves as a boundary condition between spatial processes more clustered than random and processes more regular than random. Spatial scale plays a critical role in describing the clustering and regularity in observed patterns. It is possible that at one spatial scale the point patterns are clustered whereas at another spatial scale, they are regular. In our application, we define an event as the observed location of a cell centroid.

Spatial point processes (SPP) describe a family of stochastic process models where each random variable represents the location of an event in space. A realization of the process is a collection of locations generated under the spatial point process model. Typically, one models spatial clustering as a departure from spatial homogeneous Poisson process, an equivalent but more rigorous model for CSR, which is defined by the following criteria:

1. The numbers of events in nonoverlapping regions are statistically independent.

2. For any region $A \subseteq D$,

$$\lambda := \lim_{|A| \to 0} \frac{\Pr[\text{exactly one event in } A]}{|A|} > 0,$$

   where $|A|$ is the area of region A, D is the study area, and

3. $$\lim_{|A| \to 0} \frac{\Pr[\text{two or more events in } A]}{|A|} = 0.$$

In the above definition, the quantity $\lambda$ is the Possion paramter or the 'intensity' (mean number of events per unit area) of the process. For CSR, the number of events in a region $A$ is a Poisson random variable with mean $\lambda|A|$, and the events are uniformly located within $A$. Since the intensity of events is constant at all locations in the study area, the process is said to be homogeneous. Furthermore, if $\lambda$ is independent of location, then the process is stationary. Criterion 2 implies that the probability of a single event in an increasingly smaller area $A$ (adjusted for the area of $A$) is a constant ($\lambda$) independent of the location of region $A$ within the study area of interest. Criterion 3 implies that the probability of two or more events occurring in precisely the same location is zero.

Many statistical tests of CSR are based on the distribution of distances from each event to its nearest neighboring event or the distribution of distances from a randomly selected point to the nearest event. To address the spatial scale, the second-order properties of an observed SPP are often considered. While an intensity function informs on the mean, or first-order properties of a SPP, second-order properties of spatial point processes enable one to depict spatial variation and correlation among events over a wide range of spatial scales. Second-order spatial analysis has already been used in biomedical context [18]. Some key functions include Ripley's $K$ function, variance-stabilized $K$ function ($L$ function) and pair correlation function (PCF).

Ripley's $K$ function is a reduced second moment measure defined as

$$K(h) = \frac{E[\text{number of events within } h \text{ of a random event}]}{\lambda} \tag{2.1}$$

for any nonnegative distance (or spatial lag) $h \geq 0$, where where $\lambda$ is the intensity (mean number of events per unit area) of the spatial point process. Under CSR, $K(h) = \pi h^2$. Hence, $K(h) > \pi h^2$ indicates clustering and regularity otherwise.

For visualization and diagnosis, the variance-stabilizing transformation of the $K$ function, i.e., $L$ function,

$$L(h) = \left\{ \frac{K(h)}{\pi} \right\}^{\frac{1}{2}}. \tag{2.2}$$

Statistical experience shows that the fluctuations of estimated $K$ functions increase with increasing $h$. The root transformation stabilises these fluctuations for both the means and variances. The $L$ function allows a more readily interpretable diagnostic plot because one may plot $h$ versus $L(h) - h$ and compare the resulting curve to its expected value of zero under CSR for all $h$. A nonparametric estimator of $K$ function is given in [20] and the estimate for $L$ function is given by $\hat{L}(h) = \{\hat{K}(h)/\pi\}^{1/2}$, where $K(h)$ is a consistent estimator of $K(h)$.

Obviously, the $K$ function is a cumulative function of $h$ as it measures the expected number of events up to a certain distance. If one is interested in the clustering/regularity observed at a particular distance rather than the cumulative evidence for clustering/regularity observed up to that distance, the pair-correlation function (PCF) $g(h)$ can be used:

$$g(h) = \frac{1}{2\pi h} \frac{dK(h)}{dh}. \tag{2.3}$$

The PCF function provides a scaled measure of the probability of two events occurring at distance $h$ of each other. Since PCF measures spatial correlation at various distances indicating degress of clustering or regularity, it is scale invariant. Under CSR, $g(h) = 1$. $g(h) > 1$ indicates clustering and $g(h) < 1$ means regularity.

For the above second-order functions, Monte Carlo tests based on simulation envelopes are important. Pointwise envelopes are constructed by generating random point patterns according to CSR. In [21], it is shown that envelope-based statistical tests are correct statistical procedures, under appropriate conditions. One can compute envelopes defining percentiles (such as the 5th and 95th percentiles). Such percentiles offer diagnostic implications of spatial scales (distances) at which observed patterns appear to differ from the null hypothesis of CSR. If the envelopes contain the estimated second-order functions, then it can be inferred that the SPP is not significantly deviant from CSR. Moreover, since greater distances produce increasingly inaccurate estimates due to less points and edge effects, we only look at the properties where $h$ is small. However, they do not provide formal statistical inference in the form of a test statistic. Estimation and plotting second-order spatial functions can be performed using the 'spatstat' R package.

### 2.3 Principal Component Analysis and k-Nearest Neighbors

Principal component anlaysis (PCA) concerns explaining the variance–covariance structure of a set of variables through linear combinations of these variables. Its general objectives are dimensionality reduction and interpretation. It frequently serves as intermediate steps in other analyses, such as cluster analysis.

Suppose the random vectors $\mathbf{X}' = [X_1, X_2, ..., X_p]$ have covariance matrix $\mathbf{\Sigma}$. Consider the linear combinations

$$\begin{aligned}
Y_1 &= \mathbf{a}_1'\mathbf{X} = a_{11}X_1 + a_{12}X_2 + ... + a_{1p}X_p,\\
Y_2 &= \mathbf{a}_2'\mathbf{X} = a_{21}X_1 + a_{22}X_2 + ... + a_{2p}X_p,\\
&\vdots\\
Y_p &= \mathbf{a}_p'\mathbf{X} = a_{p1}X_1 + a_{p2}X_2 + ... + a_{pp}X_p, \qquad (2.4)
\end{aligned}$$

It is easy to obtain that

$$\mathrm{Var}(Y_i) = \mathbf{a}_i'\boldsymbol{\Sigma}\mathbf{a}_i \quad i = 1, 2, ..., p, \tag{2.5}$$

$$\mathrm{Cov}(Y_i, Y_j) = \mathbf{a}_i'\boldsymbol{\Sigma}\mathbf{a}_j \quad i, j = 1, 2, ..., p. \tag{2.6}$$

The principal components are the linearly uncorrelated combinations of variables with maximal variances.

The k-Nearest Neighbors algorithm (kNN) is a non-parametric lazy learning method and one of the simplest of machine learning algorithms used for classification and regression [22]. In kNN classification, an object is classified by a majority vote of its neighbors and assigned to the class most frequently occurring among its $k$ nearest neighbors measured by some distance function. Figure 1 illustrates how kNN works for different $k$ values and gives some commonly used distance functions. The best choice of $k$ entirely depends on the observed data and is usually chosen empirically. If $k$ is too small, it will lead to noisy decision boundaries. Conversely, if $k$ is too large, it will lead to over-smoothed boundaries. In practice, the optimal $k$ should be chosen so that only nearby samples are included and the error rate is minimized. If $k = 1$, the object is assigned to the class of the unique nearest neighbor, which is usually called the nearest neighbor algorithm. The performance of the kNN classifier also depends significantly on the distance metric used. For continuous variables, a commonly used distance metric is Euclidean distance. For high dimensional data, dimension reduction is typically performed on the raw data before applying the kNN algorithm on the transformed data in feature space.

In our application, the following procedure is adopted for classification and validation using kNN. First, for a morphometric variable in a specific flap/zone in each RPE flatmount, the quantiles of all observations are computed from the 0.10 quantile to the 0.90 quantile with an increment of 0.04, so that 21 transformed data points are obtained in each eye for the variable of interest. Store the 21 data points in a row vector and stack the row vectors for all eyes. Hence, our initial data consists of a matrix $\mathbf{A}$ with 21 columns with each row

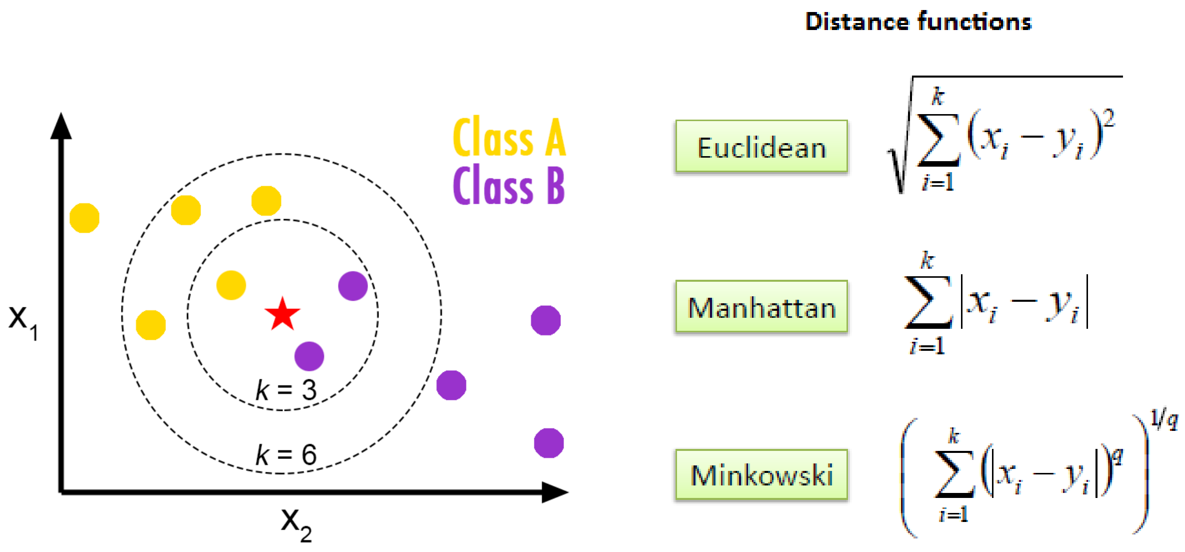


**Figure 2.1** An illustration of the kNN algorithm (right) and some common distance functions (right). Taken from http://bdewilde.github.io/blog/blogger/2012/10/26/classification-of-hand-written-digits-3.

corresponding to the RPE of a mouse. Second, apply PCA to the matrix **A** and choose the first 5 principal components, so that a reduced matrix **B** is obtained with column number 5. Third, the optimal $k$ value is selected using the leave-one-out cross validation method with the matrix **B**. Finally, the rows of matrix **B** are randomly partitioned and 80% of eyes are chosen as the training data for building the classification model in each classification run. The rest 20% of eyes are used for validating the model. The classification and validation procedure is repeated 2000 times and the prediction accuracy is calculated as the mean of the proportions of correctly classified objects in each iteration. For classification of the genotype, which is a binary variable, other quantities such as sensitivity (percentage of rd10 correctly identified as rd10) and specificity (percentage of C57 correctly identified as such) can also be calculated to assess the performance of the kNN classification.

# Chapter 3

# RESULTS AND DISCUSSION

## 3.1 Second-order Spatial Analysis of RPE Cell Patterns

In this section, we report the results from both simulation and analysis of real data of normal and diseased human RPE patterns. Second-order spatial analysis on the simulated data shows the distinct cell patterns between normal and diseased eyes. Cells in normal RPE seets have more homogeneous distributions than those in AMD eyes. The real data analysis of human RPE cell paterns further shows that the second-order spatial properties of the cell centroids provide evidence of the cell deformation and rearrangement in AMD eyes.

### 3.1.1 Analysis of simulated RPE

Figure 3.1 shows the simulated RPE tissues (Courtesy of Dr. Karina Mazzitello) with cell apoptosis occuring in multiple locations as well as the estimated second-order spatial function plots. A sequence of 120 simulated RPE images respresented the tissue pattern evolvement as a result of cell death. We selected and analyzed 5 images at different time points of the simulation. From (A) to (E), it is obvious that the cells around the holes become larger and more stretched at later stages, in comparison with the normal cells at initial stages. If we look at distance around 20 pixels in (F), the SPP formed by cell centers exhibit a pattern of regularity for all images. However, at larger distances, markedly distinct patterns occur. The initial normal cells result in a curvy shape for estimated $L$ function while the patterns of cells in (D) and (E) lose this property and appear more clustered in distances between 40 and 100. This can be explained as follows. As cells stretch and distort, patches of smaller cells are formed and the cell centers are more clustered at certain spatial scales. Similar conclusion can be drawn for the PCF plot. In summary, the clear oscillatory shape in both the variance stabilized $K$ function ($L$ plot) and pair correlation function (PCF) plot for normal eyes results from the well-ordered nearly hexagonal cell packing. With AMD disease, the cells become more ellyptical in shape, resulting in the irregularities in cell center distances.

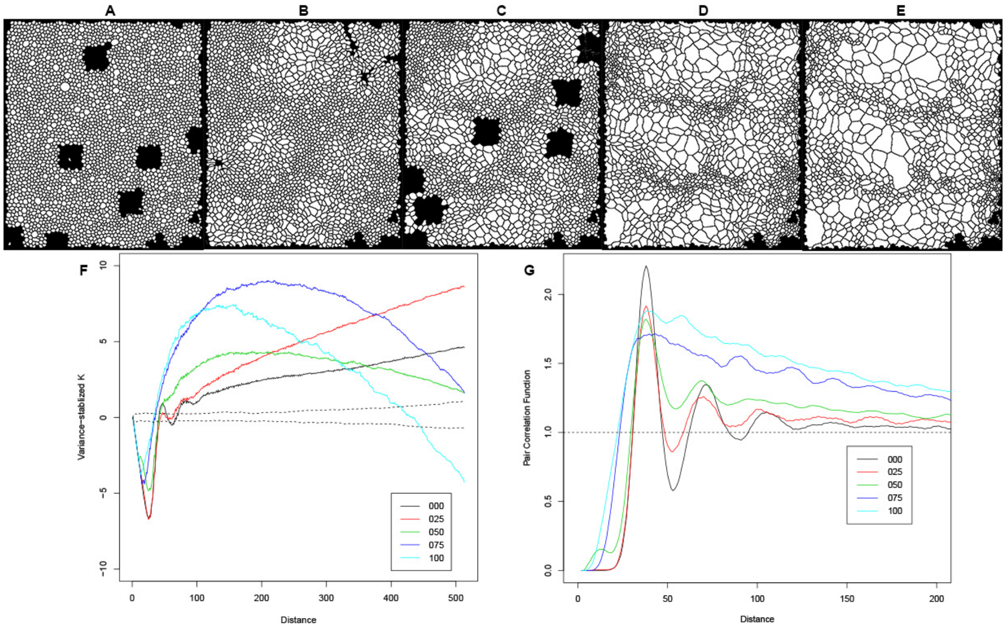
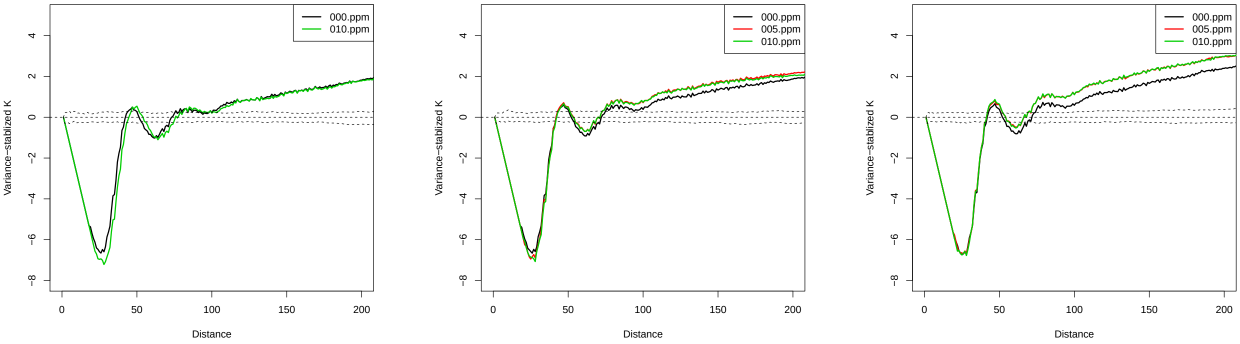


**Figure 3.1** Simulated RPE tissue undergoing repeated clustered damage (black area: regions of cell apoptosis) and recovery from a normal RPE pattern.(A-E). F: plots of estimated variance stabilized K function; G: plots of estimated pair correction function. The numbers in the legends correspond to time-steps of the selected images in the simulated sequence.

We also simulated RPE cell sheets from the same normal condition, but different amounts of cells were removed in a single cluster. We removed 10 cells, 50 cells and 100 cells, respectively, and obtained 3 sequences of RPE sheets recovering following the cell removal, each consisting of 10 images. Similar to the above procedures, we selected images at the initial stages and later stages from each sequence and plotted the $L$ functions, as shown in Figure 3.2. With less cells removed, the point patterns of cell centers do not change noticeably. However, when more cells are removed, the $L$ function increase drastically, suggesting that cells are more clustered at a large spatial distance. Comparing (b) and (c) in Figure

3.2, we find greater seperation in the $L$ functions of cell center spatial patterns in the initial and later stages when 100 cells are removed. This small simulation study indicates that with more severe damage to the RPE sheets, more profound changes in the second-order properties of the cell center SPP will be observed. It shows that the second-order spatial analysis is helpful in discriminating the effects of normal aging and AMD disease progression on the RPE sheets.



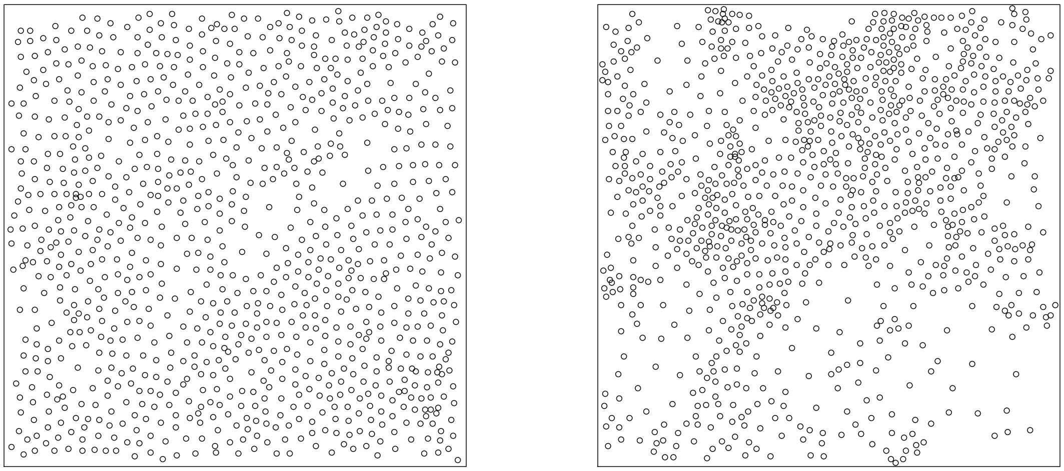
(a) RPE with 10 cells killed (b) RPE with 50 cells killed (b) RPE with 100 cells killed

**Figure 3.2** Estimated $L$ plots for simulated RPE sheets with different amount of cell loss. (a) 10 cells are killed; (b) 50 cells are killed; (c) 100 cells are killed

### 3.1.2 Analysis of human RPE

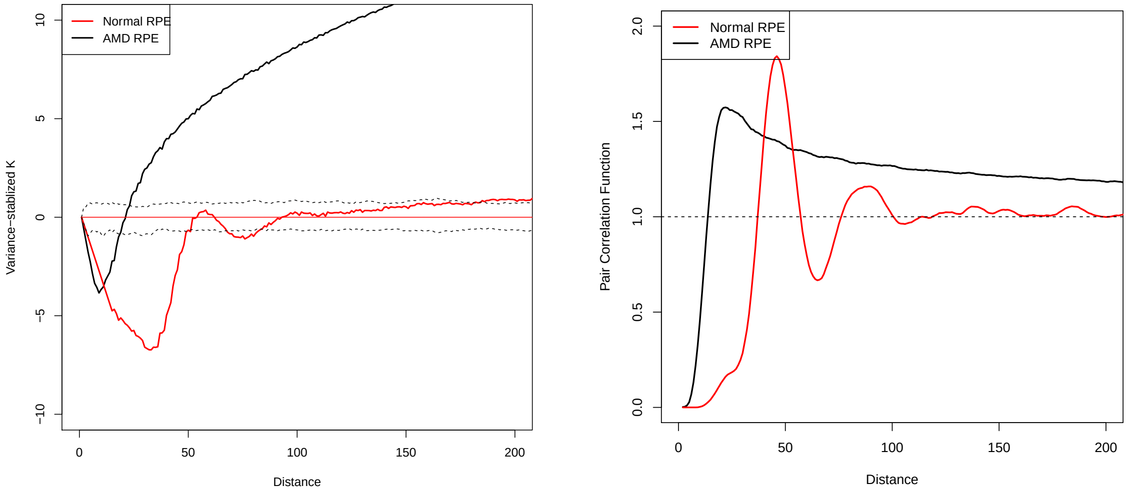
We further performed second-order spatial analysis on both normal and diseased human RPE data. ImageJ was used for segmentation and data extraction. Second-order spatial analysis on human RPE images shows a clear change of cell distribution between normal and AMD eyes. Figure 3.3 shows that the oscillations in the $L$ plot and PCF plot for normal RPE gradually disappear as hexagonal cells stretch and distort. The AMD-like RPE patterns show an increased clustering between distance 50 and 100 (in pixels), indicative of the disordered RPE pattern and close cell packing in AMD. The big dip at distance less than 50 pixels shows that the normal RPE cells are more regular compared to CSR, corresponding to the regular cell-cell distance. The dip also exist for the AMD RPE, but the spatial distance

$h$ decreases to around 10 pixels, which indicates that as a result of cell stretching in diseased RPE, the minimal cell-cell distance decreases. Nonetheless, nearly all of the estimated $L$



(a) Normal human RPE of age 80

(b) AMD RPE of age 80



(c) Estimated $\hat{L}$ Plots

(d) Estimated PCF Plots

**Figure 3.3** Comparison of the second-order spatial properties in normal and diseased RPE in an individual of age 80. (a) Cell center locations for normal RPE; (b) Cell center locations for diseased RPE; (c) Estimated $L$ plots; (d) Estimated pair correlation function plots

function for AMD RPE lies above that of the normal RPE, which implies more clustered cell distribution in AMD RPE sheets. The PCF plots can be interpreted similar to the $L$ plots. The highest peaks suggest that at a distance less than 50 pixels, the SPP of cell centers are clustered than CSR for both normal and diseased RPE. The distance moves to the left for AMD RPE, suggesting again a reduced cell-cell distance. These findings are consistent with the simulation study in the preceeding subsection. The second-order spatial analysis works reasonably well in reflecting the changes in RPE cell morphology as a result of AMD disease. The simulation study provide further insights into the pathophysiology of AMD.

## 3.2 Classification of Mouse Genotype and Age through Cell Morphometry

We applied kNN to the mouse RPE morphometric data for classification of the genotypes and ages of RPE in different spatial regions. From macula outward, there are 4 zones for each RPE sheet, labelled zone 1, zone 2, zone 3 and zone 4, respectively. Further, we considered combining adjacent zones so that we have 3 zones corresponding roughly to macula, mid-periphery and far periphery regions. Table 3 shows the four cases with and without combining zones. The classification of genotype was performed in three age groups, which is different from previous studies. Similarly, the classification of age was conducted in different genotypes. The separation of age groups or genotype groups allows us to reveal more information for each genotype/age combination.

**Table 3.1** Different cases of zone combination

| Case | Description |
|---|---|
| I | All 4 zones remain the same |
| II | zone 1 and 2 as **zone 1**, zone 3 as **zone 2**, zone 4 as **zone 3** |
| III | **zone 1**, zone 2 and 3 as **zone 2**, zone 4 as **zone 3** |
| IV | **zone 1**, **zone 2**, zone 3 and zone 4 as **zone 3** |

The classification of genotype/age was further performed in each zone/flap. The accuracy of classification was plotted against the zone/flap and the results are reported in Figures 3.4-3.13. In all succeeding figures, the error bars represent the 95% confidence intervals.

### 3.2.1 Classification of genotype by zone in different age groups

Figures 3.4-3.7 show the classification results of genotype by zone in different age groups for all cases. In Figure 3.4, the area and perimeter variables give nearly linearly decreasing accuracy with increasing zone numbers in age group 1. Other variables do not give consistent results. However, most variables have higher classification accuracy in zones 1 and 2. This means in young mice, the region closer to the macula is more informative of distinguishing the genotypes. Notice that nearly all classification accuracies are lower than 80%. Hence, the difference between normal and diseased RPE is really not so remarkable at early stage of the disease. In age group 2, the pattern changes abrutly. Area and perimeter again have the best performance, with accuracies close to 1. Solidity, extent and form factor all have the lowest accuracy in zone 2 but higher accuracy in zone 1 and zone 4. Eccentricity and major/minor ratio have the same pattern with highest accuracy in zone 3. At this stage, the RPE cells in AMD individuals begin to undergo more change in cell sizes while maintaining relatively stable cell shapes. Cell shapes distort from hexagonal shape more drastically in macula and peripheral regions of the RPE. A shift in the patterns also occurs in age group 3. It can be found that in age group 3 the zones with larger numbers have higher accuracy for most shape variables. This suggests that as mice grow older, the RPE cells undergo even more distortion in the periphery regions than in macula region. Note also that overall accuracy goes up.

For Cases II, III and IV, Figures 3.5-3.7 give consistent results similar to the findings in Figure 3.4. In the young age group, since the differences between normal and AMD RPE are negligible, no obvious patterns are found and the classification accuracy is relatively low. In the medium age group, cell sizes and cell shapes begin to change. Both area and perimeter perform similarly well in all zones, suggesting cell sizes are changing globally with age, bringing differences in RPE morphology among normal and AMD individuals. Shape variables have better classification of the genotype in zones 1 and 3. This is a direct indication that cell shapes in macula and peripheral regions go through more drastic change as AMD progresses. In the old populations, the peripheral region is more indicative of whether an individual is normal or diseased.

Our results provide some insights into the study of RPE morphology. We suggest that for rd10 RPE cells, instead of measuring the whole RPE flatmount, researchers can directly measure cell size in macula or far-peripheral region for medium aged individuals, and obtain cell shape measurements in the far-peripheral regions for old individuals.

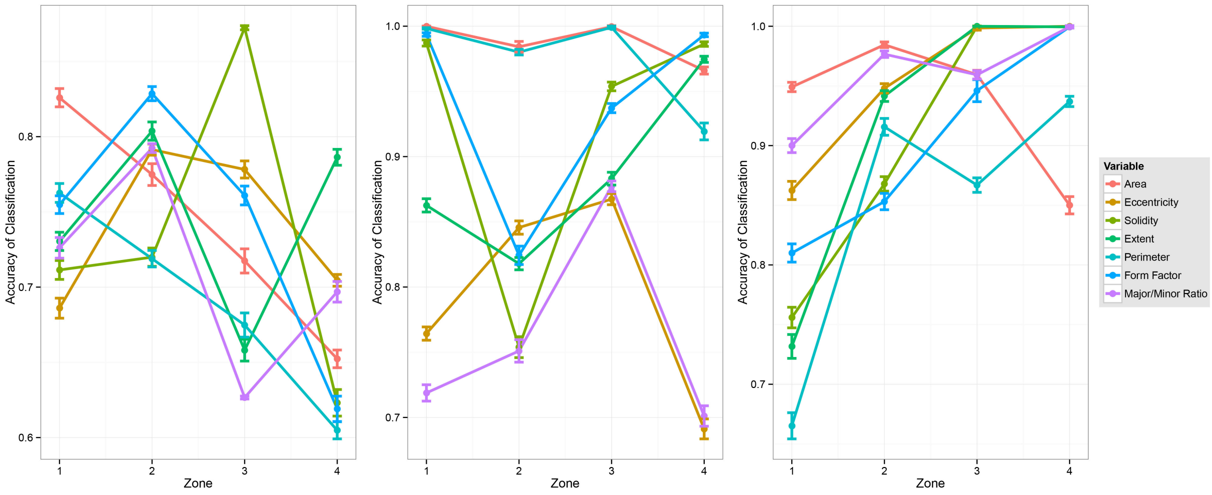


**Figure 3.4** Classification of genotype by zone in three age groups for Case I. Figures from left to right correspond to age group 1, age group 2 and age group 3, respectively.

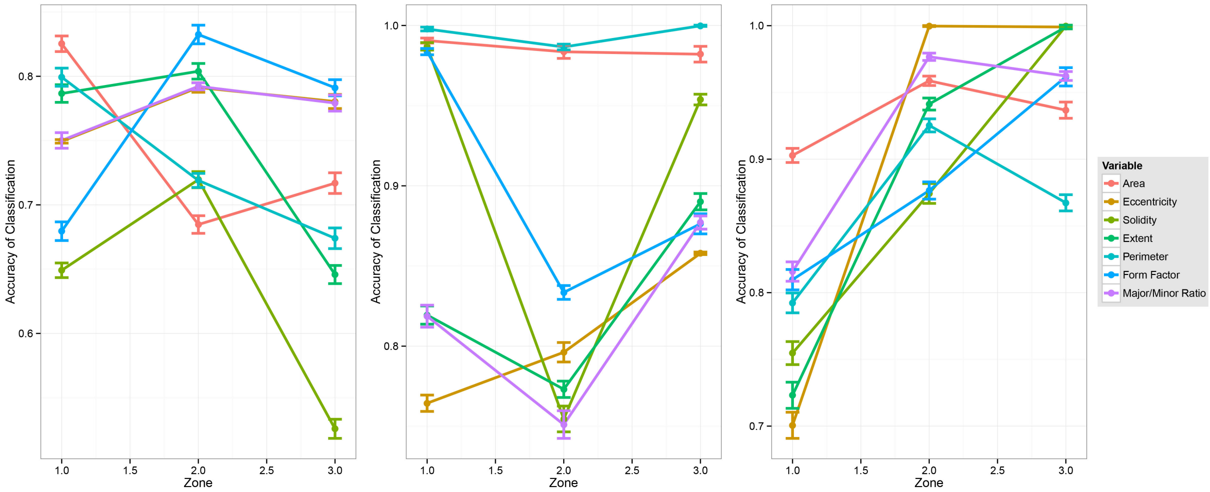


**Figure 3.5** Classification of genotype by zone in three age groups for Case II. Figures from left to right correspond to age group 1, age group 2 and age group 3, respectively.

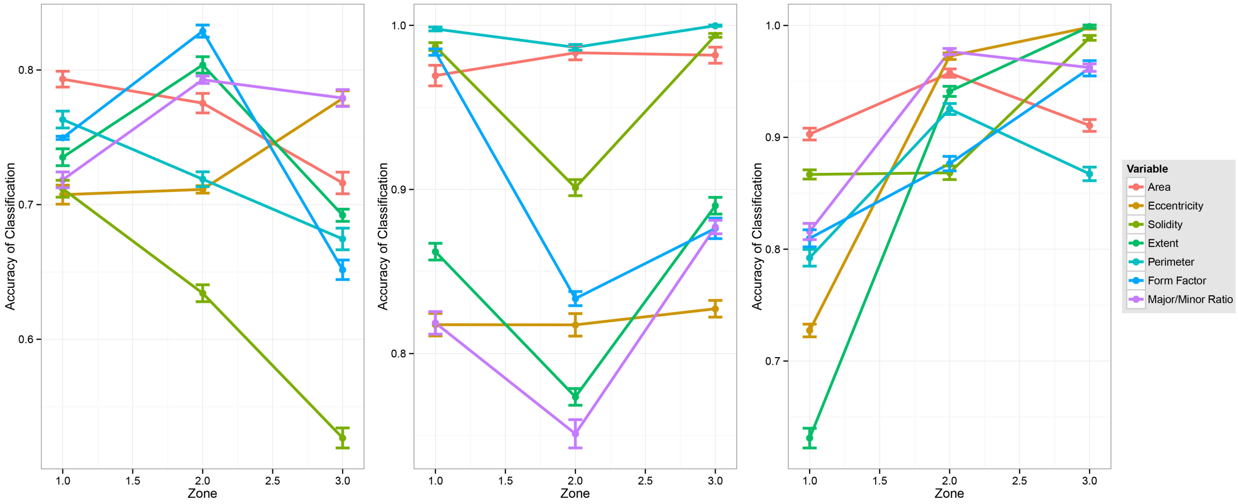


**Figure 3.6** Classification of genotype by zone in three age groups for Case III. Figures from left to right correspond to age group 1, age group 2 and age group 3, respectively.

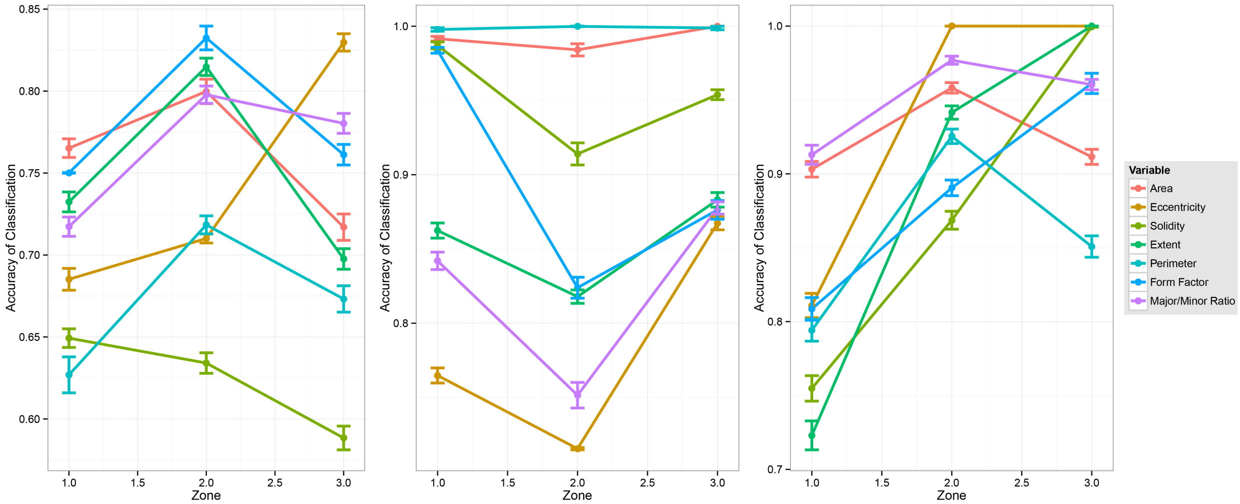


**Figure 3.7** Classification of genotype by zone in three age groups for Case IV. Figures from left to right correspond to age group 1, age group 2 and age group 3, respectively.

### 3.2.2 Classification of age by zone in different genotype groups

Figures 3.8-3.11 present the results for the classification of age by zone in different genotype groups. In Figure 3.8, we observe that for normal RPE, that is, the C57BL/6J genotype, all variables yield good classification of the age except eccentricity, extent and major/minor ratio. No conspicuous difference exists in zones. In rd10 RPE, the outer peripheral regions are best at classifying the age groups. These observations remain the same for Cases II, III and IV. The interpretations and implications are as follows. In normal individuals, aging brings global change to the RPE morphology whereas in diseased individuals, aging results in more profound change in RPE morphology in the peripheral region.

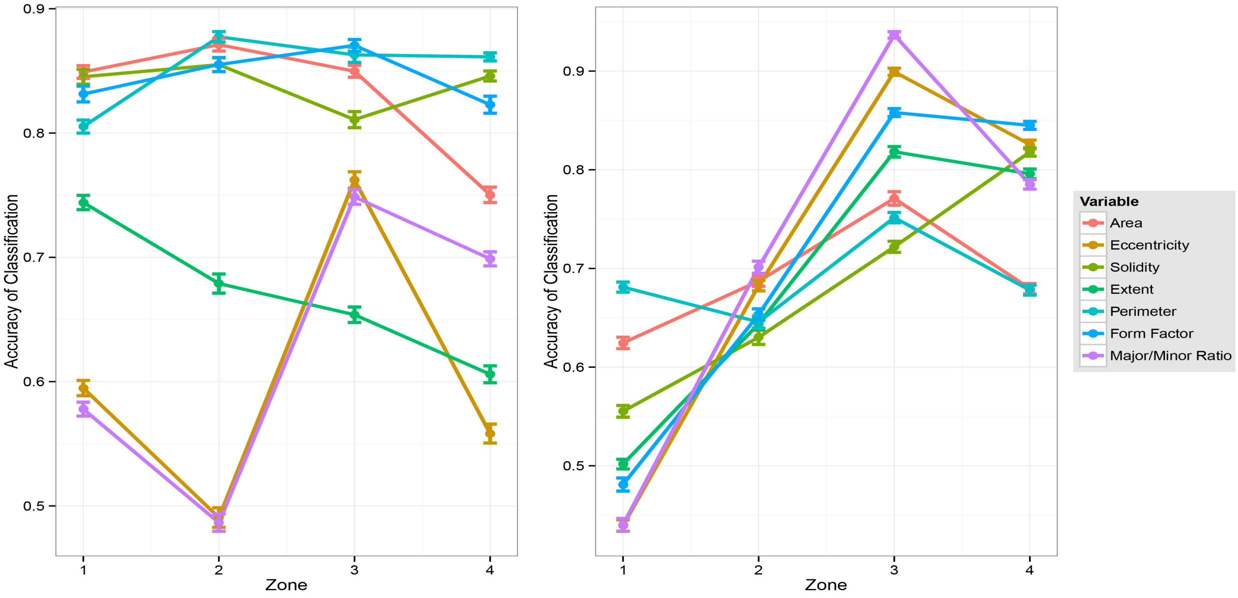


**Figure 3.8** Classification of age by zone in two genotype groups for Case I. Left: C57BL/6J; right: rd10.

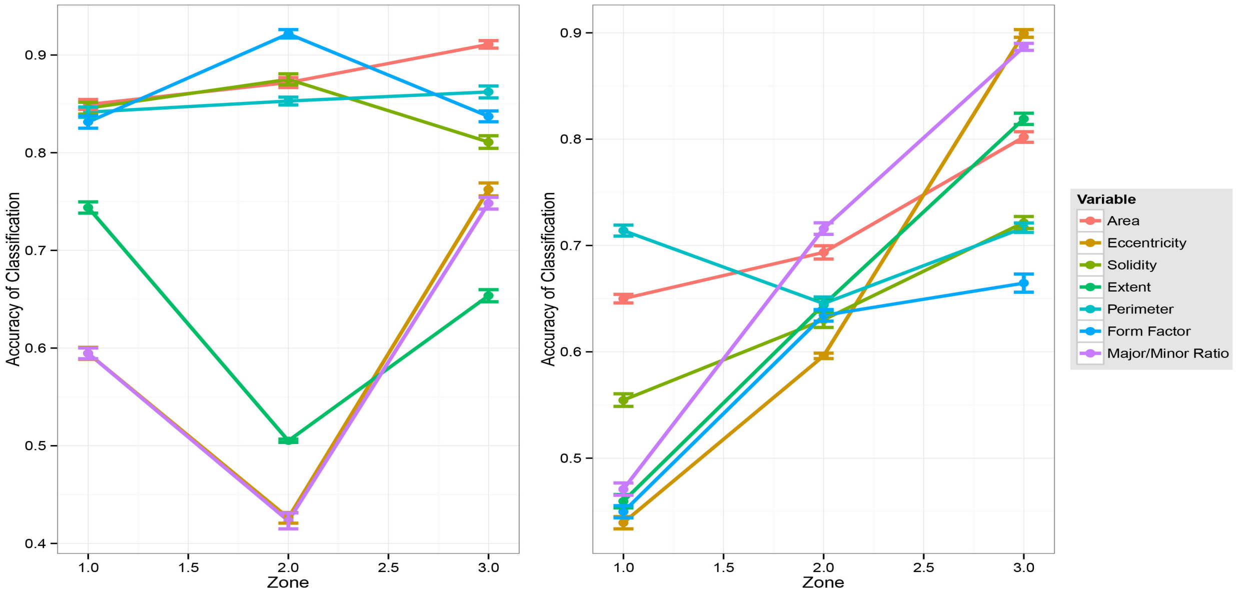


**Figure 3.9** Classification of age by zone in two genotype groups for Case II. Left: C57BL/6J; right: rd10.

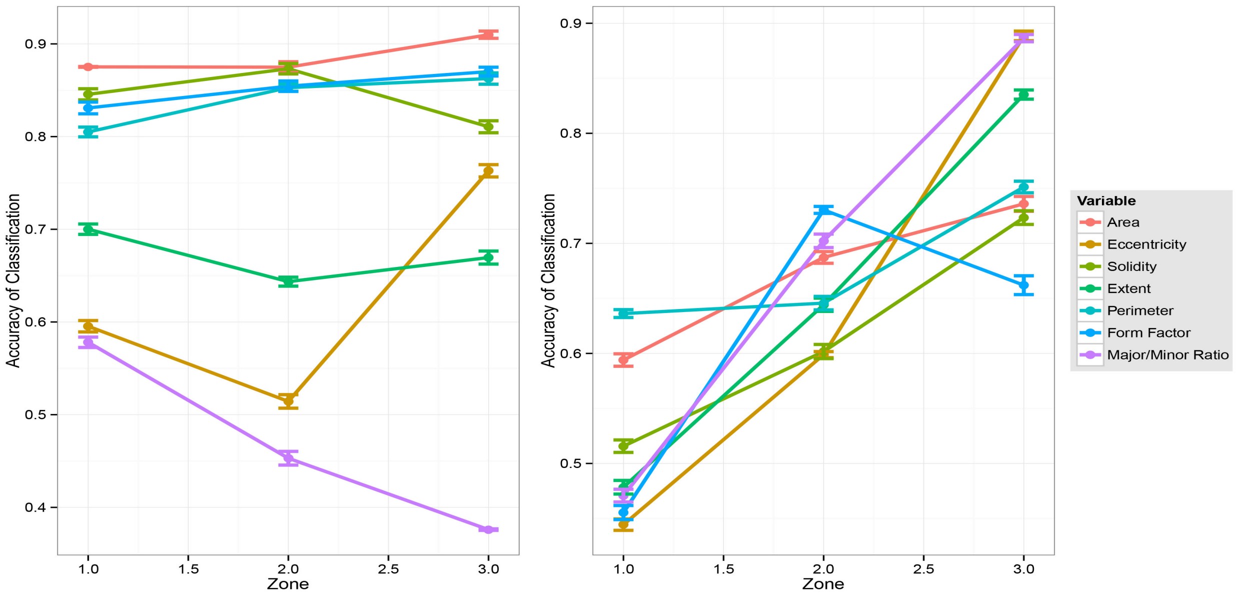


**Figure 3.10** Classification of age by zone in two genotype groups for Case III. Left: C57BL/6J; right: rd10.

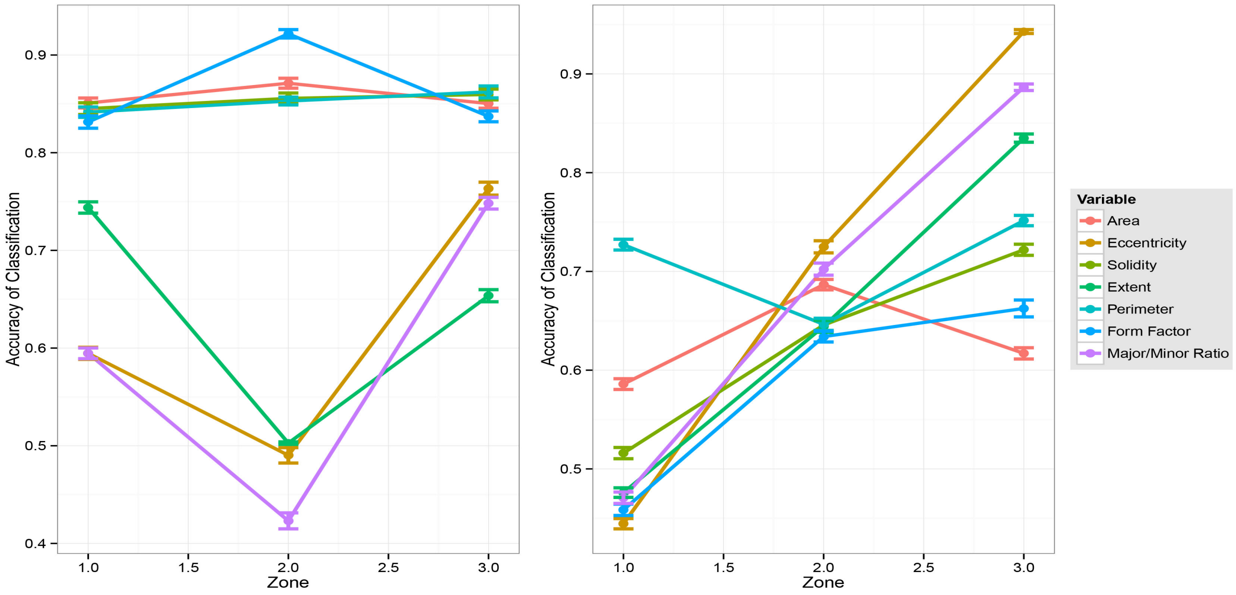


**Figure 3.11** Classification of age by zone in two genotype groups for Case IV. Left: C57BL/6J; right: rd10.

### 3.2.3 Classification of genotype by flap in different age groups

The classification results of genotype by flap in different age groups are shown in Figure 3.12. We do not observe patterned differences among flaps in all of the age groups, suggesting no orientational differences in where aging affects the cell size or shape. We do observe, similar to previous findings, that in the medium age group, area and perimeter perform better in discriminating the disease versus normal RPE and that in the old age group, cell shape measurements are good indicators of rd10 genotype.

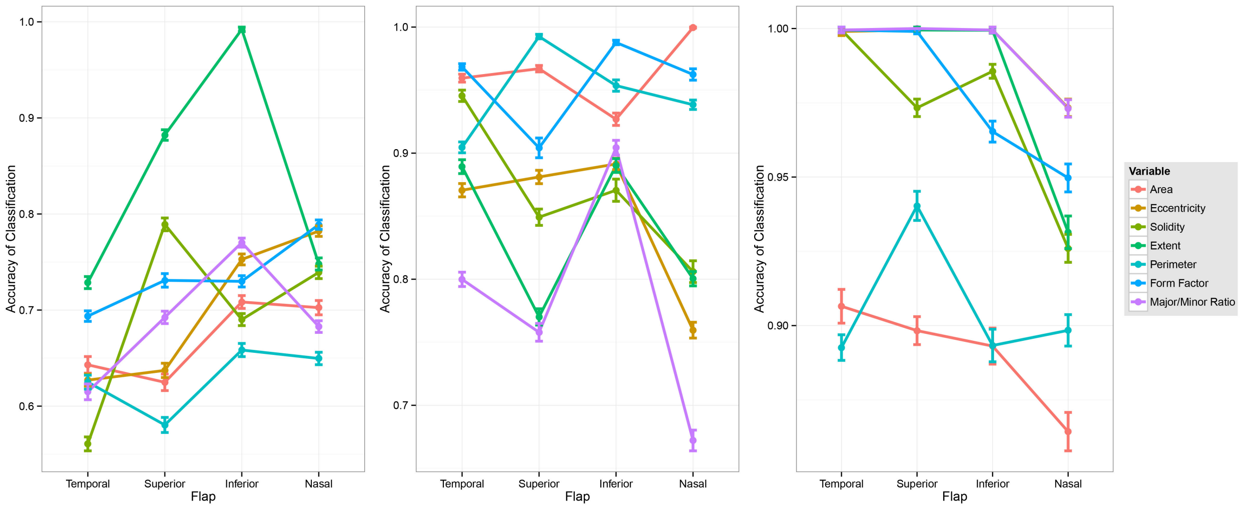


**Figure 3.12** Classification of genotype by flap in three age groups. Figures from left to right correspond to age group 1, age group 2 and age group 3, respectively.

### 3.2.4 Classification of age by flap in different genotype groups

Figure 3.14 shows the classification of age by flap in twp genotypes. We cannot make any conclusion with regard to which flap classifies age the best or what variables perform the best as no obvious trends are found.

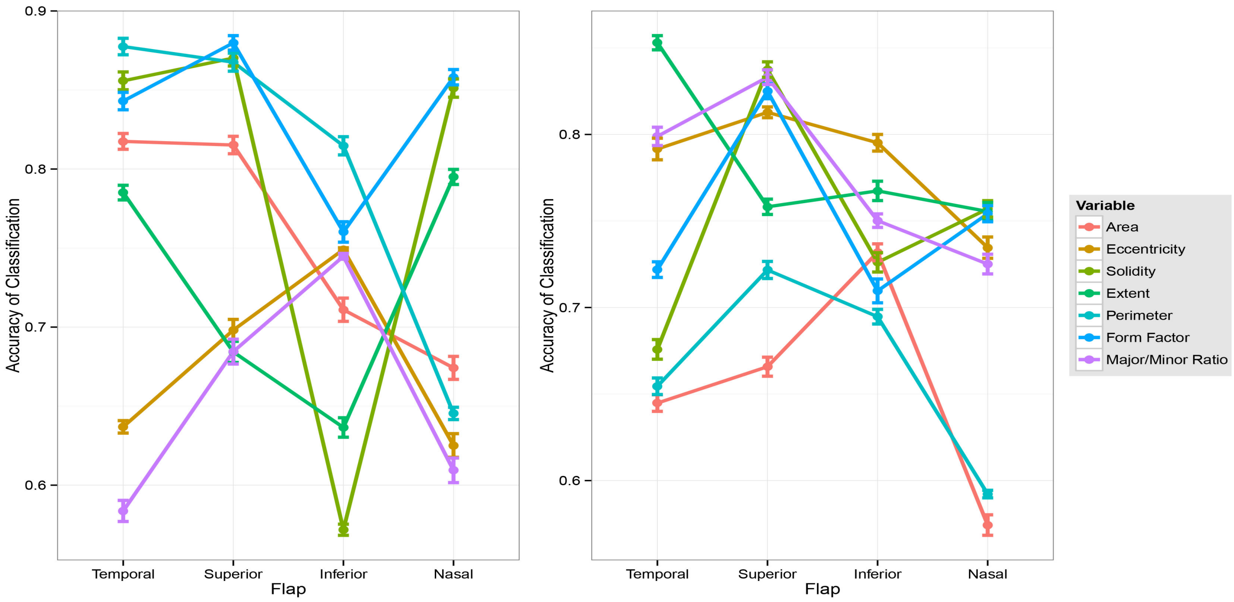


**Figure 3.13** Classification of age by flap in two genotype groups. Left: C57BL/6J; right: rd10.

# Chapter 4

# CONCLUSION AND FUTURE WORK

## 4.1 Conclusion

In this thesis, we quantitatively analyzed RPE morphometry from the spatial aspect. Second-order spatial analysis reveals distinct patterns of cell packing in normal and diseased RPE sheets. Cells are more clustered in diseased RPE sheets, indicative of cell reorganization brought about by AMD disease. The more severe the damage is to the RPE sheet, the more clustered the cell center point patterns will be. kNN algorithm provides a powerful tool in classifying the genotype and age of normal and diseased RPE in different spatial regions of the RPE sheet. In medium age groups, cell size measurements perform the best in classifying the disease status. In old individuals, peripheral regions have higher classification accuracy for area and shape variables, indicating that AMD leads to more cell shape deformation in periphery of RPE. The study does not reveal significant patterns in flaps. Our study provides some quantitative tools for analyzing spatially RPE morphometry, which shall be helpful for investogators to obtain RPE data efficiently and for clinicians to diagnose the disease.

## 4.2 Limitations and Future Work

It is recognized that this thesis is only a small effort towards unveiling the fundamental knowledge for AMD and RPE. Several limitations are present and continual research are required to validate what has been found in this study, to formalize the procedures and to develop new methods. For example, for second-order spatial analysis of human RPE, researchers may be concerned with when the RPE cell point patterns in AMD individuals begin to deviate from those of normal ones. Hence, one may collect more RPE flatmounts (preferably non-invasively) at different ages and explore the aging effects seperately in both normal and diseased RPE. Besides, it is known that two cell centers can not lie too close to each other. Hence, the hard-core process may be better suited to model the cell center spatial process. For analysis of real human RPE sheets, where to select the cutboxes and how large the cutboxes should be might influence the results substantially. The whole RPE gives use complete information about the characteristics of each individual. In reality, for the purpose of spatial analysis, we could only select cutboxes from certain regions since some areas are messy and require accurate image segmentation techniques. Several issues are also noticed in RPE morphometric analysis. The classification study can be coupled with other types of study to mutually corroborate on the findings. For example, researchers need to investigate and identify the underlying causes of changes in the periphery compared to the macular region. In addition, other classification methods might also be used, including support vector machines and linear discrimination analysis. Theoretical studies for the data structure might be helpful for formulating rigorous models and procedures.

# REFERENCES

[1] Boyer, M. M., Poulsen, G. L., Nork, T. M. (2000). Relative contributions of the neurosensory retina and retinal pigment epithelium to macular hypofluorescence. *Arch Ophthalmol.* 118(1):27-31.

[2] Strauss, O. (2005). The retinal pigment epithelium in visual function. *Physiol Rev* 85:845-881.

[3] Klein R. , Cruickshanks K. J., Nash S. D., et al. (2010). The prevalence of age-related macular degeneration and associated risk factors. *Arch Ophthalmol* 128, 750-758.

[4] Friedman D. S., O'Colmain B. J., Munoz B., et al. (2004). Prevalence of age-related macular degeneration in the United States. *Arch Ophthalmol* 122, 564-572.

[5] Lim, L. S., et al. (2012). Age-related macular degeneration. *The Lancet* 379, 1728-1738.

[6] de Jong P. T. (2006). Age-related macular degeneration. *N Engl J Med* 355, 1474-1485.

[7] Khandhadia, Sam, Jocelyn Cherry, and Andrew John Lotery. (2012). Age-related macular degeneration. *Neurodegenerative Diseases*. Springer, 15-36.

[8] Mehta, S. (2015). Age-Related Macular Degeneration. *Primary Care* 42 (3): 377-391.

[9] Rizzolo, Lawrence J. (2007). Development and role of tight junctions in the retinal pigment epithelium. *International Review of Cytology* 258, 195-234.

[10] Thompson, D. A.W. (1996). In: Bonner J. ed. *On Growth and Form.* Cambridge University Press, 88-131.

[11] Boatright, J. H., Dalal, N., Chrenek, M. A., Gardner, C., Ziesel, A., Jiang, Y. and Nickerson, J. M. (2015). Methodologies for analysis of patterning in the mouse RPE sheet. *Molecular Vision* 21, 40.

[12] Chrenek, M. A., et al. (2012). Analysis of the RPE sheet in the rd10 retinal degeneration model. *Retinal Degenerative Diseases*, 641-647.

[13] Chrenek, M. A., Dalal, N., Gardner, C., Grossniklauss, H., Jiang, Y., Boatright, J. H. and Nickerson, J. M. (2012). Analysis of the RPE sheet in the rd10 retinal degeneration model. *Adv Exp Med Biol.* 723: 641-647.

[14] Jiang, Y., et al. (2013). Functional Principal Component Analysis Reveals Discriminating Categories of Retinal Pigment Epithelial Morphology in Mice. *Invest Ophthalmol Vis Sci.* 54:7274-7283.

[15] Jiang, Y., et al. (2014) Analysis of Mouse RPE Sheet Morphology Gives Discriminatory Categories. *Adv Exp Med Biol.* 801: 601-607.

[16] Rashid, A., et al. (2016). RPE Cell and Sheet Properties in Normal and Diseased Eyes. *Adv Exp Med Biol.* 854: 757-763.

[17] Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns.* Academic Press.

[18] Illian J, Penttinen A., Stoyan H., Stoyan D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns.* Wiley: Chichester.

[19] Waller L. A. and Gotway C. A. (2004). *Applied Spatial Statistics for Public Health Data.* Wiley: Hoboken, NJ.

[20] Waller, L. A., Särkkä, A., Olsbo, V., Myllymäki, M., Panoutsopoulou, I. G., Kennedy, W. R. and WendelschaferCrabb, G. (2011). Second-order spatial analysis of epidermal nerve fibers. *Statistics in Medicine* 30(23), 2827-2841.

[21] Baddeley, A., et al. (2014). On tests of spatial pattern based on simulation envelopes. *Ecological Monographs* 84(3), 477-489.

[22] Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* 46 (3): 175-185.