# Collagen IV and laminin-1 expression in embryonic mouse lens using principal components analysis technique

# Collagen IV and laminin-1 expression in embryonic mouse lens using principal components analysis technique

**Ouafa Sijilmassi[1,2*], José Manuel López Alonso[2], María Del Carmen Barrio Asensio[1], Aurora Del Río Sevilla[1]**

[1]*Universidad Complutense De Madrid, Faculty of Optics and Optometry, Anatomy and Human Embryology Department, Avda Arcos de Jalón, 118, Madrid, Spain, 28037*
[2] *Universidad Complutense De Madrid, Faculty of Optics and Optometry, Optics Department, Avda Arcos de Jalón, 118, Madrid, Spain, 28037.*

[*]*o.sijilmassi@ucm.es*

**Keywords:** Principal Component Analysis, folic acid deficiency, collagen IV, laminin-1, lens, development.

## 1. Introduction.

The lens of the eye is a flawless tissue for the study of numerous essential biological processes due to the fact that not only its structure but also its morphological and biochemical specializations are unique (Bloemendal, 1981). Its metabolism has been the subject of a great interest of research. Generally, the transparency of the lens depends of the highly ordered state of fiber cells and extracellular matrix (ECM). The most important ECM of the ocular lens is found in the lens capsule which is an epithelial cell basement membrane and it is rich in type IV collagen and laminin-1 (Forrester *et al.*, 2015).

Both proteins affects the growth and morphology of the lens eye (Mcdevitt, 2012). Synthesized capsular material, like collagen IV and laminin-1, appears to be produced in the cortical lens fibers, both anterior and posterior, and in lens epithelial cells (Young & Ocumpaugh, 1966, Danysh & Duncan, 2009). We can divide the lens in three important parts (Figure 1):

1. The lens capsule: it is a noncellular basement membrane that completely encloses the lens and serves as a basal attachment site for epithelial cells. In

the embryo, it consists of collagen IV and laminin-1 secreted by the lens epithelial cells and fibers cells (Bloemendal, 1981, Kelley *et al.*, 2002), which facilities the development of the anterior, posterior and equatorial lens capsule (Danysh & Duncan, 2009).

2. The lens epithelium: is a single layer of cells beneath the anterior capsule (Bloemendal, 1981). The lens epithelial cells of the central zone represent a stable population of cells whose numbers slowly reduces with age. Epithelial cells localized just anterior to equator are known as germinative cells; this region (germinative zone) is the part with higher cell proliferation. In this zone, the cells divide constantly. From this region these new fiber cells migrate to the posterior zone to become lens fibers (Bron, 1997). This process occurs in the transitional zone called Bow Region (BR) where the newly elongated and differentiated epithelial cells into fiber can produce ECM proteins, such as collagen IV and laminin-1, in the posterior capsule (Yan *et al.*, 2005).

3. The lens fibers: the lens body consists of tightly packed fibers with very little extracellular space. These fibers are developed constantly throughout life by the elongation of lens epithelial cells near the lens equator (Bron, 1997). Elongated lens fibers still contain most of the organelles present in epithelial cells. Nevertheless, as development continues, degradation of all cytoplasmic organelles inside the fibers cells is essential for the upkeep of lens transparency (Bassnett & Mataic, 1997).

Between the lens epithelium and lens fibers there is a transition or gap region. This region is called "bonding zone" in the rest of the paper.
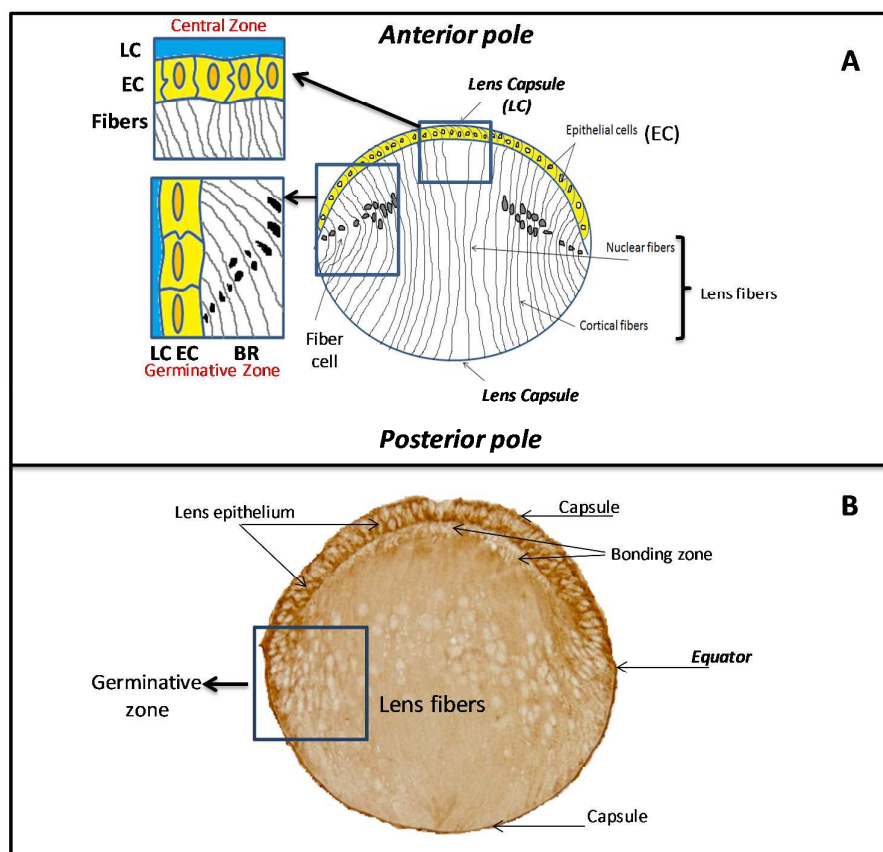
**Figure 1**: A, Schematic drawing of the embryonic lens. B, embryonic mouse lens (14.5 gestational days) labeled with anti-collagen IV: Central epithelial zone (CZ) (top square), Germinative zone lens epithelium(GZ) (bottom square).

The eye is a part of the nervous system. Folic acid (FA), also termed folate, is essential for the development of the nervous system and other associated structures (Tang & Finnell, 2003, Reynolds, 2006). It has a crucial role in the regional regulation of apoptosis and cell differentiation (Tang *et al.*, 2005). Moreover, it is vital for normal cellular growth and differentiation (Hwang *et al.*, 2015). Many of the nervous system abnormalities are accompanied by eye defects such as congenital cataracts and anophthalmia (Glaser *et al.*, 1994). Maternal folate deficiency is related with other fetal congenital malformations: clinical studies have confirmed an association between FA deficiency and various types of defects such as cardiovascular defects (Christensen *et al.*, 2013, Czeizel *et al.*, 2013), growth retardation (Miller *et al.*, 1989), craniofacial

malformations (Piedrahita *et al.*, 1999) in addition to neural tube defect (Czeizel et al., 2013, Bower & Stanley, 1989).

Some studies have been shown that mutations, deficiency or abnormal expression of type IV collagen or laminin-1 result in cataracts and several lens defects (Firtina *et al.*, 2009, Takei *et al.*, 2001, Gross & Dowling, 2005, Semina *et al.*, 2006). These molecules are essential for the correct development of the previously explained lens structures. It is unknown if FA affects the expression of these two molecules, for this reason the present work was performed to investigate if FA deficiency can change the expression of type IV collagen and laminin-1 in mouse lens that could produce different malformations in the lens.

Normally, immunohistochemistry technique is used to evaluate antigen expression in tissues (Taylor & Levenson, 2006). Estimation of the abundance of antigens expression has frequently relied on visual scoring. However, it is known that different observers differ in their interpretation of the same immunohistochemistry section, in some cases the variability of interpretation may be even greater (Taylor, 1994). The quantification of immunostaining by means of digital image analysis allows us to measure accurately protein expression percentages in immunobiological stained tissues (Durán & Arriazu, 2013). Nevertheless, most computer-based techniques for immuno-image analysis techniques have so far had limited applicability owing to several drawbacks including a need for specific software systems, and often need considerable user input (Goedkoop *et al.*, 2005, Matkowskyj *et al.*, 2003). In this study, we are the first to evaluate and to quantify collagen IV and laminin-1 expression in lens tissue on immunohistochemical stained section using an automated image analysis termed the Principal Component Analysis (PCA) as a simple method which can generate high throughput analysis in a cost-effective manner.

For this purpose, we study the expression of type IV collagen and laminin-1 in embryos mice lens whose mothers were fed with different FA deficiency diets. We compare these expressions with the results of a control group. The paper is structured as follows: section 2 explains the materials used, finalizing with the capture of the images. Section 3 is devoted to the explanation of the image processing used to compare the expression of the type collagen IV and laminin-1 among the different groups (Principal Component Analysis – PCA). Results and discussion are exposed in section 4. Finally, a conclusion section is included.

## 2. Materials.

### 2.1 Animals and diets

Eight-week-old C57/BL/6J female mice (Harlan Laboratories, Barcelona, Spain) had free access to either a control (2 mg FA/kg diet) that met FA requirements for mice (COUNCIL). Deficient groups females for two (D2) and eight (D8) weeks were fed with a FA diet with 0 mg / kg diet + 1% succinylsulfathiazole that is a non-absorbable sulfonamide to ensure the absence of endogenous synthesis of FA in the intestine. At 14.5 gestational days (E14.5) the embryos were removed by cesarean section for immunohistochemistry method. Then, they were washed in PBS (Phosphate Buffered Saline) and fixed for 48 h in buffered formol. After fixing, heads were covered in paraffin. Frontal sections were made with a thickness of 5μm and placed on slides.

The mice were kept in the animal house of the School of Medicine at the Universidad Complutense of Madrid at 22±2°C, with light-dark cycles of 12/12 h. The animal facilities and the experimental protocol used in the studies reported herein were reviewed and ethically approved by the Animal Experimentation Committee of the Universidad Complutense of Madrid.

*2.2 Immunohistochemistry*

For immunohistochemistry, sections were then incubated for 2hrs at room temperature with either 1:200 policlonal rabbit IgG anti-mouse laminin (Sigma-Aldrich) or 1:200 policlonal rabbit IgG anti-human collagen IV (ICN Biomedical Inc., Aurora, OH). Labeling was developed using the Rabbit/Mouse EnVision[TM] Peroxidase System, a peroxidase-conjugated dextran polymer (Dako Corp., Carpinteria, CA), and 3,30-diaminobenzidine (DAB kit) as chromogen (Dako Corp.) (Martínez-Sanz *et al.*, 2008). All slides were analyzed with a Leica DMRB microscope and subsequently photographed by a Leica DFC 320 digital camera with x10 magnification and a pixel size of 3.45μm x 3.45μm.

## 3. Principal Component Analysis (PCA) method applied to lens images.

The objective of the present work is the comparison among lens images groups (control, D2, and D8). The main goal is to test the expression level of collagen IV and laminin-1. The images are prepared as follows:

1. Nine images of the whole eye are selected for each group.

2. For each image, a Region of Interest (ROI) containing the lens, is selected. Inside this ROI a computer mask is applied to produce a circular image of the lens (pixels that are within the circle will be analyzed, outsiders would have a zero value) with the same size for all the images. The radius of the mask is chosen in order to fit the lens images and to be applicable to all of them. This mask is implemented in a MATLAB code (Toolbox, 2015). At this stage, we have nine images of the lens in each group with the same size (nine images labeled with collagen IV and nine with laminin-1). All images are converted to a grayscale and their sizes are 601x665 pixels.

*3.1 PCA Background. Application to immunohistochemistry images*

*3.1.1 Motivation of the method*

The principal components analysis (PCA) is a well-known statistical technique originally used to deal with large volumes of data. The purpose of it is to reduce the dimensionality of data trying to find the most relevant factors of the data that allow us to discard some of them as not relevant and thus reduce the volume of data to be treated (Donald, 1990). Indeed, as such, the method has been applied to the treatment of images in different contexts. In relation to immunohistochemistry images there have been several types of uses:

1. A large number of proteins or compounds can be expressed in a tissue and it is not possible to label it with all of them. The PCA method is used to determine which proteins are most representative of all the variability of the data and the usual process is to mark the tissues only with them and not with the others. This reduces the number of samples to be used and the types of labeling (Laas *et al.*, 2017, Gry *et al.*, 2010, Sobhani *et al.*, 2011). In other cases, the immunohistochemical images are used to illustrate the morphological patterns characteristic of the expression of the molecular markers (Moro *et al.*, 2016).

2. Enhancement of the immunohistochemistry imaging contrast by performing the PCA analysis on the RGB image taken with a color CCD camera, together with automatic classification algorithms or Machine Learning (Loukas *et al.*, 2003, Khorshed, 2013, Sarnecki *et al.*, 2016, Komura & Ishikawa, 2017, Van Eycke *et al.*, 2017).

3. The degree of marking of an immunohistochemical image is usually done through a subjective numerical "scoring" that assigns a "score" to each image.

The PCA method is usually used on the data of these numerical values(Ocak *et al.*, 2014). This is usually the most common use of PCA.

Immunohistochemical staining is a method used to evaluate the expression of antigens in tissues (Taylor & Levenson, 2006). The interpretation of this type of labeling is based on the microanatomic distribution of the staining, the proportion of positive cells and the intensity of immunostaining (Seidal *et al.*, 2001). However, it is not always easy to interpret. Most immunohistochemical studies evaluate changes in protein expression based on a visual estimate, a difficult task that requires a lot of time. In the opinion of many studies, this type of quantification of protein expression has a very subjective approach characterized by great intra- and inter-observer variability (Becker, 1993, Taylor, 1994, Rexhepaj *et al.*, 2008).

To minimize the limitations of this technique, such as the time invested and subjectivity in the interpretation of the results, we have implemented an automated system based on the processing of immunohistochemical images in order to be an alternative tool to quantify the expression of proteins and their distribution in the tissues. In this paper, with the application of the principal component analysis (PCA) we intend to accurately measure the percentage of expression of collagen IV and laminin-1 in some ocular structures, as well as regionalize it. In the second instance, with this procedure we want to improve intra and interobserver variability and thus create an objective method capable of reducing interobserver variability in the evaluation of tissues.

*3.1.2 PCA and image representation*

Principal Component Analysis (PCA) technique allows us to study the variance of a set of images from the scatter plot of their pixels. Figure 2 shows, as an example, this scatter plot, for the case of having only three images. We restrict the example to three frames because it is the maximum number to be represented with a three-dimensional 3D plot in which the three orthogonal directions are associated with the frames. The signal given by each pixel is represented by a point in a 3D diagram whose coordinates are the three different values of that pixel in the three images, after subtracting the mean value of each image to each pixel coordinate. When all the pixels are located in this diagram, a cloud of dots is obtained; each dot is one pixel. For a complete description of the method, see (López-Alonso *et al.*, 2002).
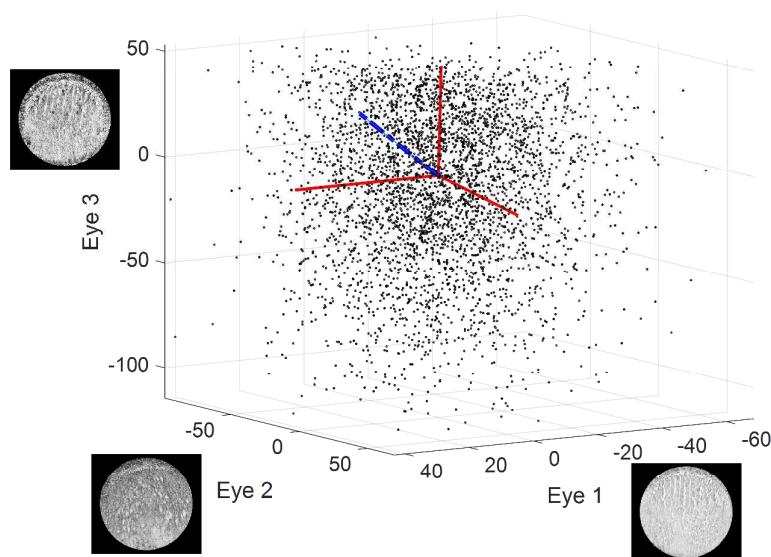


**Figure 2:** Scatter plot for a set of three images

The starting point of the principal components method is the calculation of the covariance matrix **S**, between the images. One of the properties of this matrix is that it allows us to calculate what is the variance $V_u$ of the data along an *u* direction in the scatter plot like (Donald, 1990):

$$V_u = u^T S u \qquad (1)$$

where $u$ is a column vector of module unit in the considered direction. The eigenvectors $(e_\alpha)$ of the matrix $S$ are normalized vectors in mutually orthogonal directions that maximize the variance $V_u$ in those directions. These maximized variances are the eigenvalues $(\lambda_\alpha)$ of $S$. As an example, such directions have been represented in the particular case of Figure 2 (red lines). In addition, each of the directions of the scatter plot can be associated with an image $Y_u$ , linear combination of the starting images with zero mean $I_k$ by:

$$Y_u = \sum_{k=1}^{N} u(k) I_k \qquad (2)$$

where N is the number of images. In the case of taking the addresses of the $e_\alpha$ eigenvectors, those images, $Y_\alpha$, are the images corresponding to the principal components. Because PCA orthogonality, the total variance of the images is the sum of the variance associated with each direction, that is, the sum of the eigenvalues, so that each of them can be given a relative weight in the variance total given by:

$$\Omega_\alpha = \frac{\lambda_\alpha}{\sum_{\alpha=1}^{N} \lambda_\alpha} \qquad (3)$$

On the other hand, the average image is in the direction $(\frac{1}{N}, \frac{1}{N}, .... ^{N)} ..., \frac{1}{N})$ corresponding to the normalized direction $u_m = 1/\sqrt{N}(1,1, .... .^{N)} ...,1)$. Therefore, we can calculate the variance of the data in that direction as $V_m = u_m^T S u_m$. The relative weight in the data of the average image can be calculated as:

$$\Omega_m = \frac{V_m}{\sum_{\alpha=1}^{N} \lambda_\alpha} = \frac{V_m}{V_T} \qquad (4)$$

being $V_T$ the total variance. As an example, this middle direction is drawn in figure 2 (dash line in blue).

We can compare the decomposition in PCA with simply taking the mean image through three considerations:

1.- Since the extraction of the main components is from greater to lesser representation in the total variance, we can compare $\Omega_1$ with $\Omega_m$. By construction $\Omega_1$ will always be greater than $\Omega_m$ since it is the percentage of variance in the direction of maximum dispersion. Hence, if we must choose an image as representative of the whole, it should be that of the first main component. The difference between $\Omega_1$ and $\Omega_m$ will give us an idea of the error we make when taking the average image as the most representative.

2.- Therefore, for the mean of the images to be the most representative of all of them, their direction should coincide with that of the first eigenvector. We can therefore calculate the angle between these two directions $\theta$ as a second indicator.

3.- Moreover, the difference of $\Omega_1$ and $\Omega_m$ with the 100% gives us an idea that could be information in the data not well expressed by a single representative image.

*3.2 PCA protocol*

From the previous background, the fundamental idea of PCA is to form new variables (images in our case) uncorrelated among them, PCs, which are linear composites of the original. The coefficients of the linear combination are chosen to maximize the variance of PC's. Ordered so that the first few conserve most of the variation present in all of the original variables (Jolliffe, 2002, Jackson, 2005). That is the reason why PCA was first considered as a dimensional reduction technique. It could reduce the original variables into a smaller number of variables, termed Principal Component (PC), which lead to a simpler description of the data (Koch, 2013).

The starting point is the covariance matrix among images, $\mathbf{S}$. The coefficients of the linear combinations used to calculate PC's are the eigenvectors of $\mathbf{S}$ and its eigenvalues ($\lambda_\alpha$) are the variance of the PC's. The Principal components, $Y_\alpha$, are expressed as linear combinations of the original images with zero mean, $I_k$, by means of:

$$Y_\alpha = \sum_{k=1}^{N} e_\alpha(k)I_k \leftrightarrow I_k = \sum_{\alpha=1}^{N} e_\alpha(k)Y_\alpha \qquad (5)$$

where $e_\alpha(k)$ is the $k$-component of the $\alpha$-eigenvector and N is the number of images analyzed (nine per group). As is shown in equation 5, the original images can be expressed as linear combinations of the PC's too. The percentage of the variance of the data (images) explained by the α-PC is given by $\Omega_\alpha$ defined in equation 3. Equations 3 and 5 express the possibility to reconstruct the images only taken into account relevant PC's or groups of PC's considered important by any reason. This process is normally called rectification. The relevant PC's can be identified by different methods. Most them select those ones representing a huge amount of variance. The authors have developed other one to automatically group different PC's. The method is based on the study of the uncertainties in the variance of PC's, that is, $\lambda_\alpha$. The uncertainty in $\lambda_\alpha$ depends on the high order statistics of the images (four order cumulants) (López-Alonso et al., 2002). In any case, the method allows reconstructing the set of images by selecting any principal component or groups of them considered as relevant. This allows studying all the variance of the images and the importance of each one of the spatial structures revealed by each principal component over them. The general scheme of the procedure is shown in figure 3.
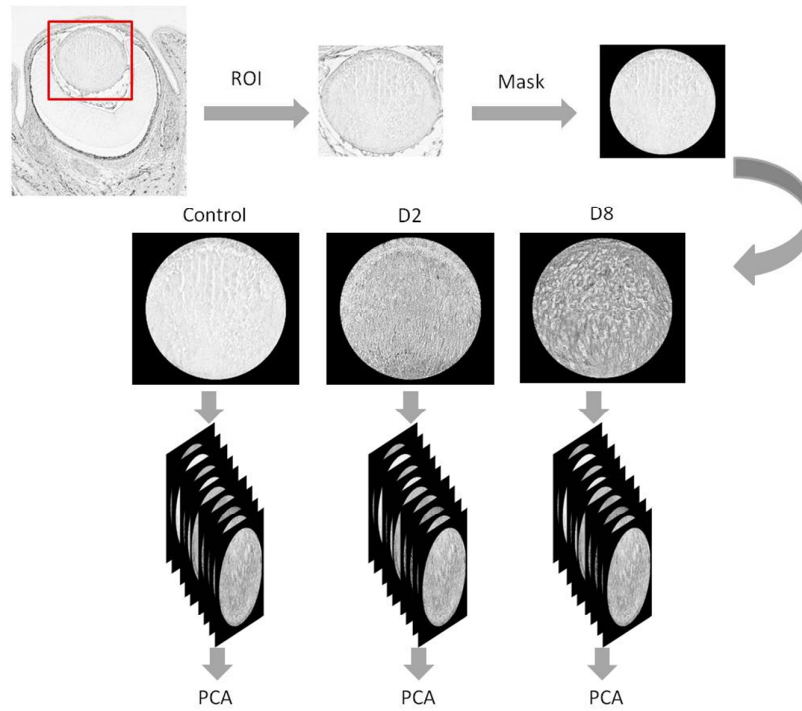
**Figure 3:** Scheme of the lens images selection. A computer mask is applied to the different images of the whole eye in each group (control, D2 and D8). Three groups of nine lens images are then performed to study them statistically by means of Principal Component Analysis

### 3.3 Mahalanobis distance

The principal components are linear combinations of the starting images that have zero covariance between them. The images of each group come from the same underlying probability distribution, so the images of the main components tend to be distributed together according to a multinormal such as:

$$f\big(Y(p)\big) = \frac{1}{(2\pi)^{N/2}|S_Y|^{1/2}} exp\left\{-\frac{1}{2} \; Y(p)^T S_Y^{-1} Y(p)\right\} \qquad (6)$$

Where $Y(p)$ is the vector formed by the coordinates of the pixel $p$ at the base of the eigenvectors of $\mathbf{S}$ (base of principal components). The previous expression allows to calculate a probability to each one of the pixels of the group of images. The exponent that appears in equation 6 is called Mahalanobis distance and has interesting statistical properties (Donald, 1990). In fact, in can be used as such to compare probabilities

among pixels since the probability dependence on pixel only appears in the exponential of equation 6 (Mahalanobis distance).

On the other hand, the transformation of main components can be interpreted geometrically in the scatter plot of the pixels of the images as a translation of the origin to the average of each image followed by a rigid rotation to the new coordinate system given by the eigenvectors of S (see Figure 2). Hence, the Mahalanobis distance can also be written in the base of the original images as,

$$D^2 = (\mathbf{I_k(p)} - \overline{\mathbf{I_k}})^{\mathbf{T}} \mathbf{S}^{-1} (\mathbf{I_k(p)} - \overline{\mathbf{I_k}}) \qquad (7)$$

where $\mathbf{I_k(p)} = [I_1(p),\ldots\ldots,I_k(p)]^T$, being $I_{i=1\ldots k}(p)$ the value of the pixel "$p$" in the image "i" for each group, and $\overline{\mathbf{I_k}}$ is the mean value of the image corresponding to the eye "k" in each group. This distance has been used to detect anomalous pixel values to classify them as different from the rest. In some cases they can be identified as "bad pixels" (López-Alonso & Alda, 2002) or with relevant structures inside the data set (López-Alonso & Alda, 2004). In our case we use the Mahalanobis distance to detect those relevant image pixels whose distances are higher than a certain threshold in probability. These pixels tend to be correlated with significant structures in the data set (López-Alonso & Alda, 2004).

*3.4 Example of application*

As an example, the results for the control group labeled with collagen IV are shown in figure 4. The first PC comprises the 94.1% of the variance. The second and third one´s represent, 1.8% and 1.1% respectively. The remaining ones comprise the 3% of the total variance. In figure 4 are shown PC's #1,2,7 in a gray scale adjusted for each one in

14

order to enhance their spatial structure. With this method, the different spatial structures

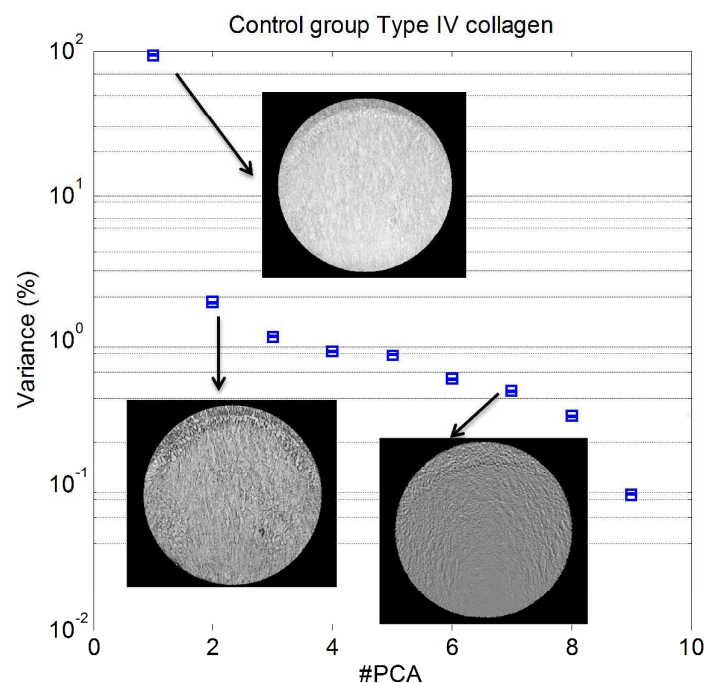of the image are shown in decreasing order of importance.



**Figure 4:** Principal Component structure for the control group. The PCA #1,2,7 are shown close to their eigenvalues.

The probability distribution function (pdf) and cumulative distribution function (cdf) of

$D^2$ for the control group labeled with collagen IV is shown in figure 5. A threshold in

distance is selected as the one representing the 10% higher distances (cdf higher than

0.9). In the case of control group (figure 5) this threshold is 27.07. This threshold has

been selected in order to be able to visualize more clearly the relevant high distance

area. A very high threshold selects almost single points instead of zones, while a minor

one does not differ much from the total Mahalanobis distance map. Hence, an

intermediate level of 90% has been chosen. In the experience of the authors this is

usually an adequate threshold. The result can be seen in figure 5, where the total

Mahalanobis distance map is represented next to the pixels whose distance is above

10% higher. It can be seen that the selected pixels represent the anterior part of the lens

mainly. This part is the more labeled zone normally (see figure 1.B). For the sake of

clarity, these types of images are called "threshold distance images" (th-distance images) in the rest of the present work.
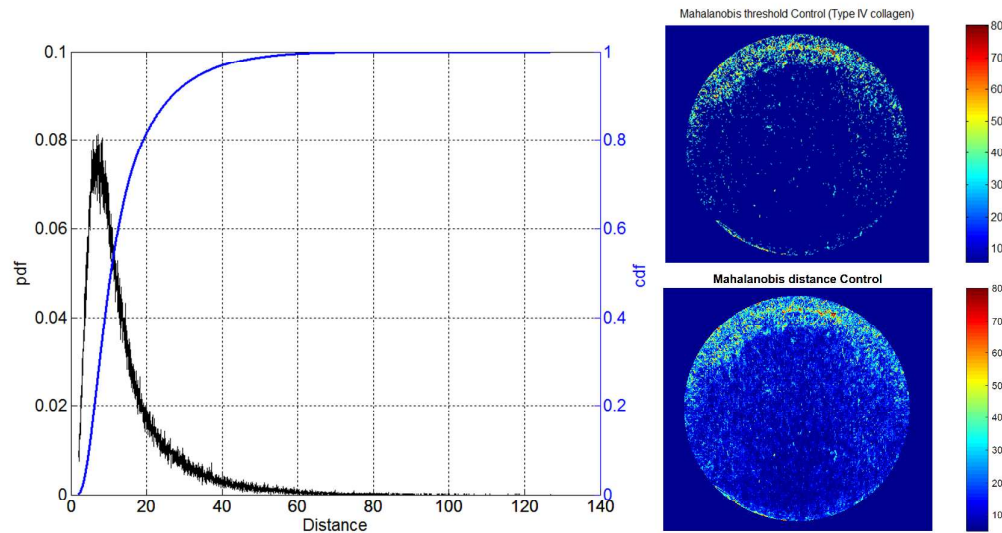


**Figure 5:** Probability distribution function (pdf, left axis) and cumulative distribution function (cdf, right axis) for the control lens images labeled with anti-collagen IV. The images show the pixels with a Mahalanobis distance higher than the threshold used (27.07 for a cdf higher than 0.9) (up image) and the total Mahalanobis distance map (down image).

## 4. Application of the method and results

### 4.1 Mahalanobis distance results.

The previous Mahalanobis distance method is applied to the three groups of lens images (control, D2 and D8) labeled with Type IV collagen or laminin-1 respectively. The results are shown in figure 6. As can be seen, cdf functions are slightly different in case of collagen IV and more similar in case of laminin-1 cdf's. This difference between the collagen IV and laminin-1 is more evident when comparing the th-distance images. In case of collagen IV, the number of selected pixels decreases in the anterior lens zone while spreading into the lens in the D2 and D8 groups. In case of laminin-1, the differences among groups are not so evident. The results of this analysis show the idea

that the spatial expression of collagen IV and laminin-1 change from control to D2 and D8 groups. The next section of the paper is devoted to the study of the spatial expression of these molecules applying PCA method.
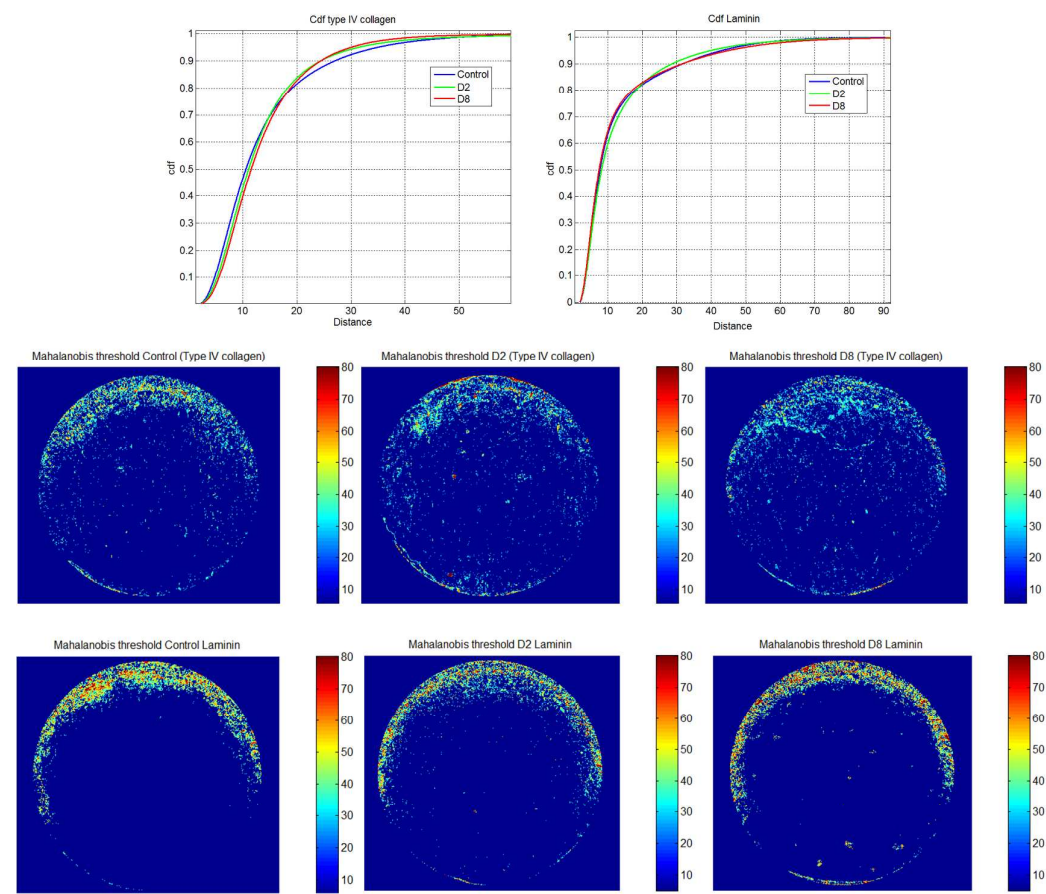


**Figure 6:** Cumulative distribution functions (cdf) for the Mahalanobis distance for control, D2 and D8 groups labeled with anti-collagen IV (left) or anti-laminin-1 (right). The images show the pixels with a Mahalanobis distance higher than the threshold. The distance threshold is selected as the one having cumulative probability higher than 0.9 for the control group and it´s the same for D2 and D8 groups.

## 4.2 Spatial expression of collagen IV and laminin-1

The PCA method is applied to the control, D2, and D8 groups for collagen IV and laminin-1. Percentages of data variance explained by each PC are shown in figure 7. In both cases (collagen IV and laminin-1), and for all groups, there are a first PC comprising a large amount of variance (higher than 94%). In case of collagen IV, the rest of control group PC´s represent a higher variance than the ones for D2 and D8

groups. Contrariwise, in case of laminin-1, for D2 group the #2,3 PC´s are like the control group, while the rest of PC´s exhibit a behavior like collagen IV (higher values for the control group). For this reason, we group the PCA's into processes as follows: in the case of collagen IV the first processes (PC1) represents the greater variability in the data, and we call the grouping of the rest components as 'Rest of Processes', i.e. from PC2 to PC9.

In case of the laminin-1, we divided them in a first processes (PC1) that represents the greater variability, process two (PC2) (which comprises PC´s 2 and 3), and finally `Rest of processes´ that represents the grouping of the rest components, from PC 4 to 9. The percentage of variance represented by the different processes is given in table 1. This percentage is calculated as $\sum_{\alpha \in L} \Omega_{\alpha}$ , where L is the set of PCs indexed in the process.
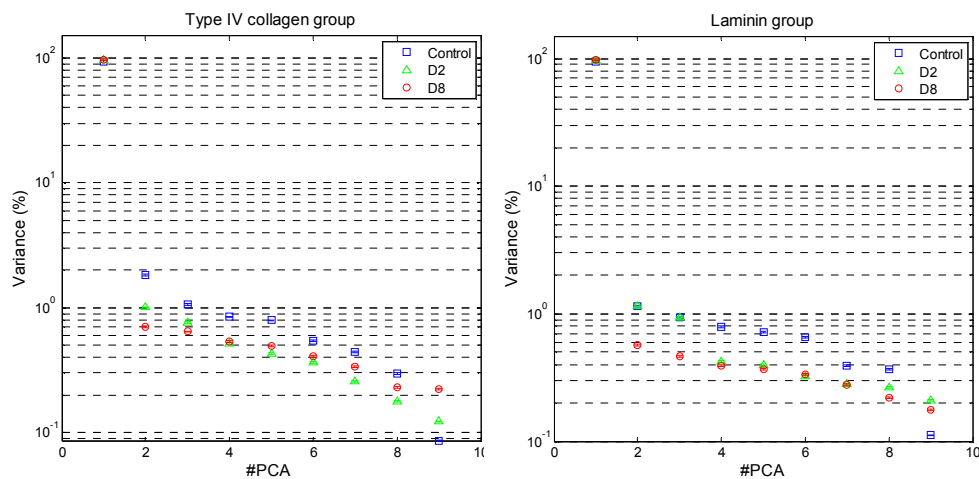


**Figure 7:** Scree plot for the eigenvalues for all three groups (control, D2 and D8) expressed in % of variance. Collagen IV (Left) and laminin-1 (Right)

18

Table 1 The percentage of variance represented by the sum of the various eigenvalues in % for the three groups. D2 and D8 are the deficient groups.

| Collagen IV | | | | | | Laminin-1 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Control | | D2 | | D8 | | Control | | | D2 | | | D8 | | |
| PC 1 % | PC 2-9 % | PC 1 % | PC 2-9 % | PC 1 % | PC 2-9 % | PC 1 % | PC 2-3 % | PC 4-9 % | PC 1 % | PC 2-3 % | PC 4-9 % | PC 1 % | PC 2-3 % | PC 4-9 % |
| 94 | 6 | 96 | 4 | 96 | 4 | 95 | 2 | 3 | 96 | 2 | 2 | 97 | 1 | 2 |

In order to study the difference in the spatial expression of collagen IV and laminin-1 between control and deficit groups, an Labeling index, $O_l^{P_i}$ ($l$={collagen, laminin}; $P_i$ ={PC1, PC2, and RP}) is defined as:

$$O_l^{P_i} = \frac{\overline{R(l\_I_{Def}^{P_i})} - \overline{R(l\_I_C^{P_i})}}{\overline{l\_I_C}} \times 100; \qquad (8)$$

where $\overline{R(l\_I_{Def}^{P_i})}$ is the mean value of the rectification respect to the $P_i$ process of the deficit group images (D2 or D8) labeled with $l$_molecules, $\overline{R(l\_I_C^{P_i})}$ is the same, but applied to the control group, and $\overline{l\_I_C}$ is the mean value of the control group for the $l$-molecule. This index indicates the spatial expression of the analyzed molecule in percentage respect to the control group. Moreover, it analyses the expression due to each $P_i$ separately (PC1 and Rest in case of collagen IV. PC1, PC2 and Rest in case of laminin-1).

The result in case of laminin-1 PC2 process (rectification with PC´s #2,3) for D2 group is shown in figure 8.
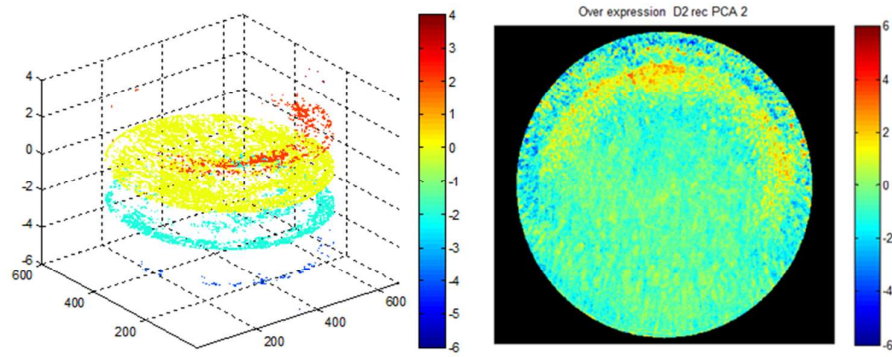
**Figure 8:** Labeling index of laminin-1 in case of D2 group and process PC2 (rectification with PC´s #2,3).

Left, frontal view. Right, contour level graph.

The results of labeling index for collagen IV and laminin-1 are shown in figures 9 and 10 respectively. In both cases, FA deficiency affect severely expression levels. The images are analyzed in the next subsections. Zones where we appreciate changes are: the lens epithelial cells (central and germinative zones), the bonding zone between lens fibers and epithelial cells, the Bow region (BR) and finally lens fibers (see figure 1).
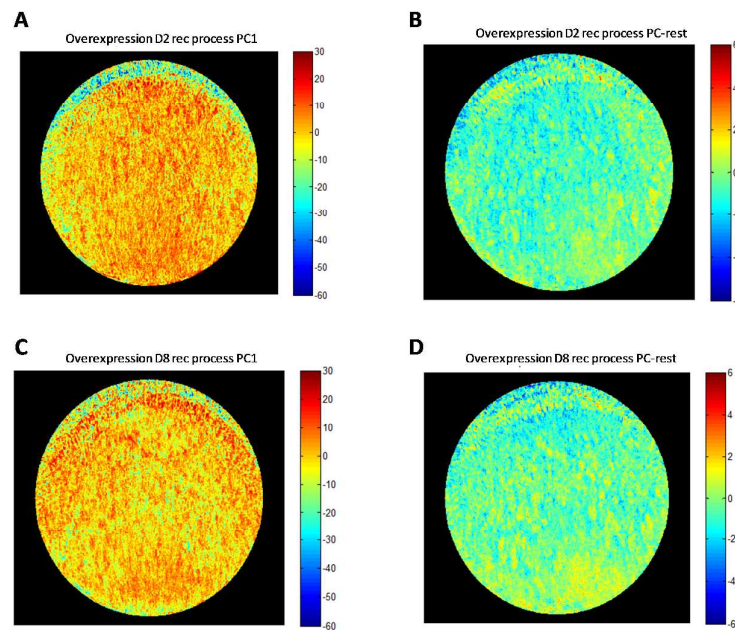


**Figure 9:** Labeling index of collagen IV for D2 deficit group (up row) and D8 deficit (bottom row). The collagen IV expression for the images reconstructed with PC1 process is shown in left images (A, C, scale from -60% to +30%) while collagen IV expression for the Rest process is shown in right images (B, D scale from -6% to 6%)
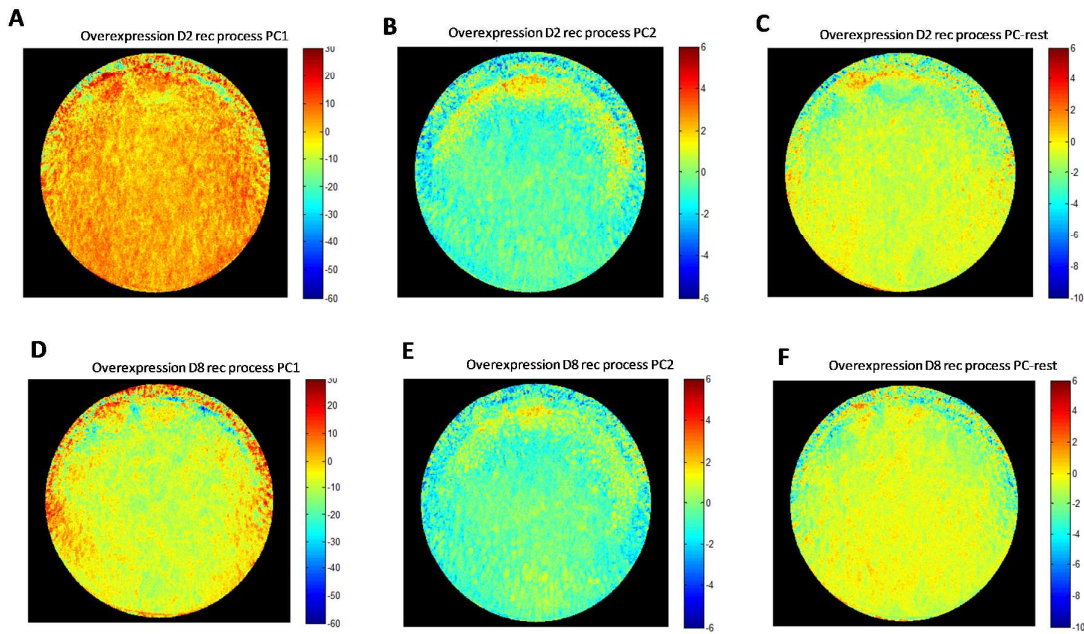
**Figure 10:** Labeling index of laminin-1 for D2 deficit group (up row) and D8 deficit group (bottom row). The laminin1 expression for the images reconstructed with PC1 process is shown in left column (A, D, scale from -60% to +30%). The laminin-1 expression for the images reconstructed with PC2 process is shown in the middle column (B, F scale from -6% to 6%). Overexpression for the images reconstructed with Rest process is shown in the right column (C, F scale from -10% to 6%)

Comparison when taking only the mean image of each group (Control, D2 and D8) as representative is made with the parameters $\Omega_1, \Omega_m$ and $\theta$ defined previously. The results of these parameters appear in the table 2 for collagen IV and laminin-1 and for Control, D2 and D8 group of images.

**Table 2:** Variance percentage values represented by PC1 and the mean image. Angle between the representative direction of the mean and the first eigenvector for collagen IV and laminin-1 in the different groups.

| | Collagen IV | | | Laminin-1 | | |
|---|---|---|---|---|---|---|
| | Control | D2 | D8 | Control | D2 | D8 |
| $\Omega_1$ % | 91.76 | 95.18 | 92.78 | 94.88 | 96.03 | 97.20 |
| $\Omega_m$ % | 91.03 | 94.95 | 91.77 | 91.51 | 94.99 | 96.07 |
| $\theta^o$ | 5.15 | 2.81 | 6.03 | 10.91 | 5.98 | 6.20 |

As can be seen, neither the mean nor the PC1 represent the entire structure of the data. In our case there is approximately between 5%-10% variance in the images not explained by them. Furthermore, this unexplained variance has spatial structure in the image since the angle between the direction of the mean and PC1 does not match either. Particularly striking is the case of the laminin-1 control group and the change that occurs with the D2 and D8 groups.

Given that the main components are a linear combination of the images and, by the orthogonality of the eigenvectors (López-Alonso et al., 2002), we can reconstruct the starting images from them using equations 5. In this way we can reconstruct the rest of the variance not explained by PC1 maintaining its spatial structure for study. This possibility is not given if we only take the average (mean image) as representative because, implicitly, we take the rest of the variance as noise or something to be ignored. We believe that this is one of the great advantages of the PCA method. It is noteworthy that the method tells us that there is structure in 10% of the data, which is very difficult to appreciate at a glance or through a subjective method, which is the usual way to evaluate this type of microscopy images.

In order to compare with the mean image of each group a similar index is calculated as:

$$O_l^m = \frac{\overline{I_{Def}} - \overline{I_C}}{\overline{I_C}} \times 100 \tag{9}$$

where $\overline{I_{Def}}$ is the mean of the deficit group (D2 or D8) and $\overline{I_C}$ the mean of the control group. For the results of table 1 the result of this index should be like the index of PC1. The comparison images for collagen IV and laminin-1, groups D2 and D8 are shown in the following figures 11 and 12:
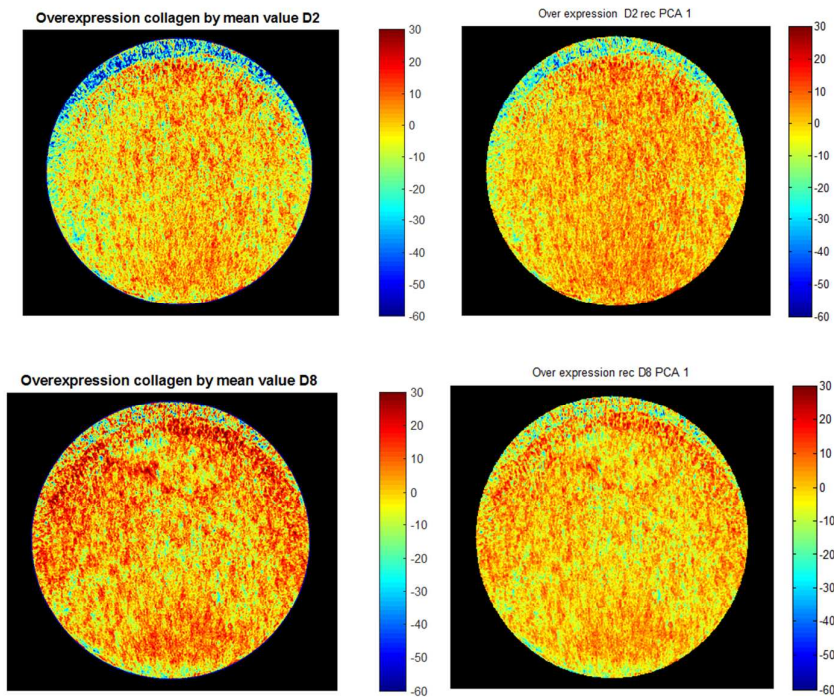
**Figure 11:** *Collagen IV labeling index for groups D2 (up row) and D8 (bottom row). Left, mean images for each group, compared to control group, used. Right, PCA method for the PC1 index.*



**Figure 12:** Laminin-1 labeling index for groups D2 (up row) and D8 (bottom row). Left, mean images for each group, compared to control group, used. Right, PCA method for the PC1 index.
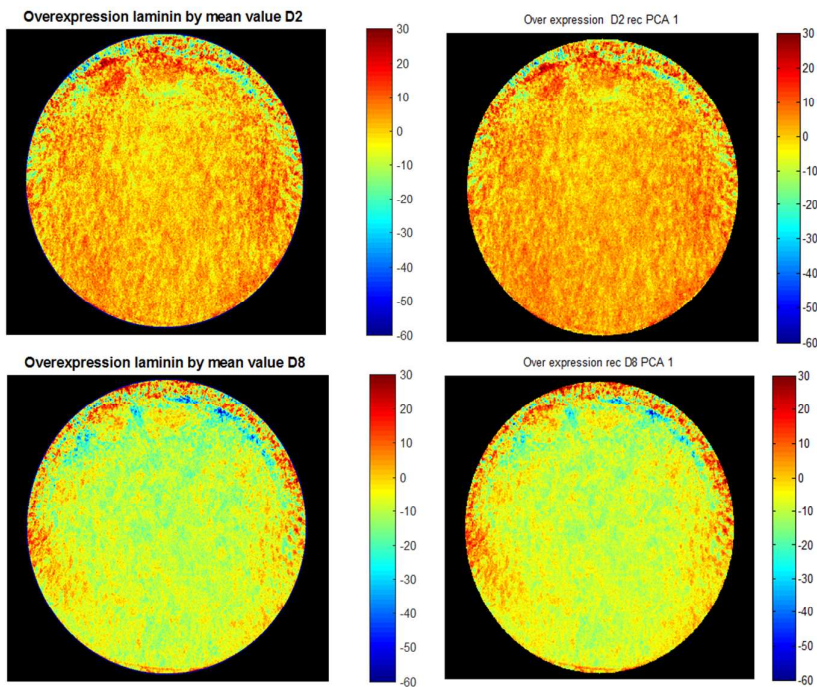
As expected from table 2, the results are different comparing the index using the average of the images with the index using the PC1 reconstruction.

The true power of the principal components method is observed in figures 9 and 10 where the image of the previous index is given for other main components other than PC1. This is especially interesting in the case of laminin-1 where it can be distinguished that there is structure in the junction and epithelial zone with possible biological relevance and with an overexpression of around 3% compared to the comparison with the control group. This would be very difficult to detect by means of a visual evaluation of the immunohistochemical image, which is one of the great advantages of the PCA method. In addition, the method allows quantifying and regionalizing the marking of molecules in an automatic way, eliminating in a simple way the subjective scoring that usually takes place in these images. Finally, we must remember that PCA method allows to the reconstruct all the images with the required principal components. That is, to calculate the index of equation 8, the images of each group are reconstructed with, for example, PC1. We obtain 9 images, each one corresponding to a different lens. In the index we take the mean value of the reconstruction as representative of the group. Anyway, it is possible to examine anyone of these nine reconstructed images in order to see the relevance of PC 1 in each one of the individual lens. This cannot be done if we take only the mean value of images.

In the following sections we will describe the changes found regarding the control group for the two molecules (collagen IV and laminin-1). They are analyzed distinguishing between the different parts of the lens. From our knowledge this work is the first to analyze the impact of FA deficiency on these molecules and their relationship with biological structures like the lens. Getting to know what is the molecular mechanism by which these structural changes take place would require new experiments and tests from a biological point of view. The present work should be considered only as an indicator of where these changes occur and their possible

relationship with problems in the development of the lens. The following sections of the results and discussion are devoted to these possible problems and impacts.

## 5. Discussion

*5.1 Expression of collagen IV.*

As shown in the previous sections for the case of collagen images, the method selects the images reconstructed with PC1 and the rest of the principal components (PC1 and 'Rest' processes) as different from the control images group. In this section we are going to discuss the results observed for each process along with its possible relationship with biological structures. It is important to note that the method automatically selects structures giving clues about their possible origin and possible future tests to be performed.

### 5.1.1  Reconstruction with PC1 process

The reconstruction of images with the PC1 process exhibits an overexpression of collagen IV around 10%-20% in lens fibers zone respect to the control. This overexpression is a little higher in the bonding and posterior fiber zone. The increase of overexpression in the bonding zone is more evident in case of D8 group when comparing with D2 group. Contrariwise, there is a decrease of 30%-40% in the lens epithelial cells, slightly more pronounced in D2 group than in D8 group (See figure 9 A and C).

By E14 (gestational day 14), in a normal mouse development, the apical surfaces of the epithelial cells are in contact with the major part of the apical one of subjacent lens fiber cells, but the bonding is not fully completed yet. Firtina et al (Firtina et al., 2009) related the overexpression of collagen IV in this area with a loss of contact between fiber cells and epithelial cells resulting in the formation of gap between the fiber and

epithelial cells. In our histological sections this loss of contact in bonding zone is also detected and coincides with an increase in the labeling index in this region.

Moreover, overexpressing collagen IV that we have observed in mouse lens fibers has been also observed by Firtina et al. which attribute to the retention of collagen IV within the fiber cells and appears to induce cataract (Firtina et al., 2009).

In addition, we observe a decrease of expression collagen IV in all epithelial cells. It is known that FA deficiency causes decrease in cell proliferation (Craciunescu *et al.*, 2004). This lessening proliferation could result in a reduction of collagen synthesis. An abnormality of collagen synthesis may be decreases lens epithelial cell proliferation too. FA deficiency is attributed to the impaired collagen synthesis in the skin (Hautvast & Barnes, 1974). Altered collagen synthesis it has been shown to produce vacuoles in lens and irregular lens capsule (Gould *et al.*, 2007)-(Van Agtmael *et al.*, 2005). Vacuoles are clearly visible in some of our histological sections.

### 5.1.2   *Reconstruction with 'Rest' process*

Finally, the reconstruction for the Rest Process exhibits a slightly increase of 2% in the bonding and in fiber in the posterior zone in both cases (D2 and D8) (See figure 9 B and D), and again, as in PC1 case, a decrease in the all the epithelial zones. Given the orthogonality of the processes given by the principal components, these changes do not occur exactly at the same points as for the P1 process, suggesting a different mechanism of under or over-expression. The method can therefore be used as a guide for the design of tests or experiments that seek to find these different mechanisms since, in itself, it only gives quantitative information.

26

*5.2 Expression of laminin-1*

As mentioned in section 4.2, for the case of laminin-1, PCs 2-3 behave as in the control group in case of D2 group. For the D8 group the behavior is different, showing a lower variance than control group (see figure 7). For the 'Rest' of principal components, their weight is lower than in the control group, for both D2 and D8. For this reason, we group the PCAs into processes in the following way. A first process (PC1) containing the greatest variability in the data. A second process (PC2) that contains the PCs 2 and 3 (difference of behavior in D2 and D8) and, finally, a final process (RP) containing the reconstruction of the data with the PC's from 3 to 9.

*5.2.1 Reconstruction with PC1 process*

In case of PC1 the overexpression is higher (~ 20%) in all epithelial cells, bonding, and BR zone for both D2 and D8 groups (Figures 10 A and D). The increases of laminin-1 observed in lens epithelial cell, especially in the germinative and transitional zones, is more pronounced in D8. It has been shown that laminin promotes growth, adhesion and protein synthesis in epithelial cells in vivo (Sugrue & Hay, 1981). In fact, it is considered potent regulator of epithelial cell behavior (De Arcangelis *et al.*, 2001). Laminin-1 is essential for the formation of epithelial structures, it facilitates that cells acquire apicobasal polarity and adopt an epithelial character (Ivanovitch *et al.*, 2013). Hence, expressing higher levels of laminin-1 in epithelial cells compared with control maybe will have a remarkable impact on the function of the lens epithelial cells. Yan et al. (Yan et al., 2005) have been shown that laminin-1 was expressed abundantly in lens epithelial cells. In the first time, the consequence of this increase of laminin-1 expression disrupted the normal intracellular signaling as well as cell differentiation, polarization and adhesion; all are required for the maintenance of epithelial cells variability and identity and for normal lens fibers formation. An abnormal deposition of

laminin-1 may causes disorganization of ECM proteins in lens capsule which can influences lens epithelial homeostasis (Weaver *et al.*, 2006).

Likewise, D2 group presents an overexpression (between 5% and 10%) in lens fibers. Nevertheless, D8 group present opposite results: an underexpression (~20%) in all lens fiber. Decreased laminin-1 expression in the lens fiber could lead to a variety of lens defects with severe lens degeneration. Previous studies of laminin-1 mutant embryos have suggested that the lens epithelium and fiber cells did not differentiate normally (Semina et al., 2006).

### 5.2.2 Reconstruction with PC2 process (PC's #2,3) and Rest process (PC's #4-9)

The second process (PC2) suggests an increase of laminin-1 expression (~3%) in the bonding zone and a decrease in some epithelial zones (~ 4%) in both groups (D2 and D8) (Figures 10 B and E). This could be a relevant fact that, to our knowledge, has not been described in the literature, perhaps because the overexpression is small compared to the masking produced in the image not treated with the PCA method. This fact opens the way for studies to discover the origin of this behavior of laminin in the bonding zone.

There is no relevant difference for the PC-Rest process in D2 and D8 groups respect to control (Figures 10 C and F). Most of the image exhibits a 0% of overexpression while higher expressions are localized mainly in points not marking clearly a zone.

It is interesting to highlight the fact that the general behavior of collagen IV is opposite to laminin-1: collagen IV it tends to increase its expression in fiber lens zone while decreasing in epithelial cells zone. Relation between the cross-expression of Collagen IV and laminin-1 has been reported previously. For example, collagen IV aggregates were seen in the lens fibers of laminin-1 mutant's zebrafish (Pathania *et al.*, 2014).

## 6. Conclusions

Immunohistochemistry section staining is not always easy to interpret. Our present work focuses on the spatial localization and quantification of two ECM molecules, collagen IV and laminin-1, in the embryonic mouse lens applying principal component analysis technique.

It is important to remember that reduction of the income of folate produces a marked impairment in collagen IV and laminin-1 expression. These abnormalities in collagen IV and laminin-1 expression in embryos lens could influence on epithelial cells adhesion and fiber cells differentiation resulting in cataracts formation due to lens disorganization.

As far as we know, at the present time no one has studied the effect of FA deficiency on laminin-1 and type IV collagen in ocular tissues. The direct analysis of microscopic images immunolabeled with laminin-1 or collagen IV provide us with a quantitative idea of the infra/over expression of both molecules in ocular lens. Therefore, we analyze the images from a statistically point of view using the PCA analysis and comparing a control group with a D2 and D8 group. This technique permits to reconstruct the original images taking into account different spatial structures (processes) to analyze independently. These processes could be associated to different spatial organization of collagen IV and laminin-1.

Our results show that there is a variation of both molecules distribution in the FA deficient embryos. In case of type IV collagen, it tends to be overexpressed especially in the equatorial zone, posterior zone and bonding zone. In the epithelial zone there is a decrease in its expression comparing with the control group. In case of laminin1 the results suggest the opposite: an increase of expression respect to control in the epithelial and germinative zone while decreasing in fiber zone.

In conclusion, our automated image analysis technique is based on processing images from a microscope, with this method it can measure a wide variety of biological tissues quantitatively, rapidly and reliably. This image analysis tool has commonly been based on mathematical model with the purpose of interpreting information contained in image data and extract quantitative measures with ease. On other hand, immunohistochemistry staining is an important tool for evaluate, locate, and quantify protein expression. Taken together, automated immunohistochemistry image analysis, can help evaluating and comparing the staining intensity and percentages of over/under-expression in different biological tissues which can facilitate reliably distinguish protein level differences in any given tissue. These technique, in combination, improve accuracy and increase the ease of the analysis to determinate expression levels in immunohistologically stained sections.

In our case, the method cannot be used to assure the biological mechanism under such findings: it only suggests possible changes in the way collagen IV and laminin-1 organize inside the lens. Further research should be made in this direction.

References:

Bassnett, S. & Mataic, D. (1997) Chromatin degradation in differentiating fiber cells of the eye lens. *The Journal of cell biology,* **137,** 37-49.

Becker, R. L. (1993) Standardization and quality control of quantitative microscopy in pathology. *Journal of Cellular Biochemistry,* **53,** 199-204.

Bloemendal, H. (1981) *Molecular and cellular biology of the eye lens,* John Wiley & Sons.

Bower, C. & Stanley, F. J. (1989) Dietary folate as a risk factor for neural-tube defects: evidence from a case-control study in Western Australia. *The Medical Journal of Australia,* **150,** 613-619.

Bron, A. J. (1997) Wolff's Anatomy of the Eye and Orbit.

Christensen, K. E., Zada, Y. F., Rohlicek, C. V., Andelfinger, G. U., Michaud, J. L., Bigras, J.-L., Richter, A., Dubé, M.-P. & Rozen, R. (2013) Risk of congenital heart defects is influenced by genetic variation in folate metabolism. *Cardiology in the Young,* **23,** 89-98.

COUNCIL, N. R. COMMITTEE ON ANIMAL NUTRITION. SUBCOMMITTEE ON POULTRY NUTRITION (1994) Nutrient Requirements of Poultry. Washington: National Academy Press.

Craciunescu, C. N., Brown, E. C., Mar, M. H., Albright, C. D., Nadeau, M. R. & Zeisel, S. H. (2004) Folic acid deficiency during late gestation decreases progenitor cell proliferation and increases apoptosis in fetal mouse brain. *Journal of Nutrition,* **134,** 162-166.

Czeizel, A. E., Dudás, I., Vereczkey, A. & Bánhidy, F. (2013) Folate deficiency and folic acid supplementation: the prevention of neural-tube defects and congenital heart defects. *Nutrients,* **5,** 4760-4775.

Danysh, B. P. & Duncan, M. K. (2009) The lens capsule. *Experimental eye research,* **88,** 151-164.

De Arcangelis, A., Lefebvre, O., Méchine☐Neuville, A., Arnold, C., Klein, A., Rémy, L., Kedinger, M. & Simon☐Assmann, P. (2001) Overexpression of laminin α1 chain in colonic cancer cells induces an increase in tumor growth. *International journal of cancer,* **94,** 44-53.

Donald, M. (1990) Multivariate statistical methods. McGraw Hill, New York.

Durán, E. & Arriazu, R. (2013) Quantification of Protein Expression on an Immunofluorescence Section by Using the MetaMorph Image Analysis System. *Universal Journal of Applied Science,* **1,** 86-94.

Firtina, Z., Danysh, B. P., Bai, X., Gould, D. B., Kobayashi, T. & Duncan, M. K. (2009) Abnormal expression of collagen IV in lens activates unfolded protein response resulting in cataract. *Journal of Biological Chemistry,* **284,** 35872-35884.

Forrester, J. V., Dick, A. D., McMenamin, P. G., Roberts, F. & Pearlman, E. (2015) *The eye: basic sciences in practice,* Elsevier Health Sciences.

Glaser, T., Jepeal, L., Edwards, J. G., Young, S. R., Favor, J. & Maas, R. L. (1994) PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. *Nature genetics,* **7,** 463-471.

Goedkoop, A. Y., De Rie, M. A., Teunissen, M. B., Picavet, D. I., Van Der Hall, P. O., Bos, J. D., Tak, P. P. & Kraan, M. C. (2005) Digital image analysis for the evaluation of the inflammatory infiltrate in psoriasis. *Archives of dermatological research,* **297,** 51.

Gould, D. B., Marchant, J. K., Savinova, O. V., Smith, R. S. & John, S. W. (2007) Col4a1 mutation causes endoplasmic reticulum stress and genetically modifiable ocular dysgenesis. *Human molecular genetics,* **16,** 798-807.

Gross, J. & Dowling, J. (2005) Laminin Beta 1 and Laminin Gamma 1 Are Essential for Normal Lens and Retinal Development in Zebrafish. *Investigative Ophthalmology & Visual Science,* **46,** 2441-2441.

Gry, M., Oksvold, P., Pontén, F. & Uhlén, M. (2010) Tissue-specific protein expression in human cells, tissues and organs. *Journal of Proteomics & Bioinformatics,* **3,** 286-293.

Hautvast, J. & Barnes, M. (1974) Collagen metabolism in folic acid deficiency. *British Journal of Nutrition,* **32,** 457-469.

Hwang, S. Y., Kang, Y. J., Sung, B., Kim, M., Kim, D. H., Lee, Y., Yoo, M., Kim, C. M., Chung, H. Y. & Kim, N. D. (2015) Folic acid promotes the myogenic differentiation of C2C12 murine myoblasts through the Akt signaling pathway. *International journal of molecular medicine,* **36,** 1073-1080.

Ivanovitch, K., Cavodeassi, F. & Wilson, S. W. (2013) Precocious acquisition of neuroepithelial character in the eye field underlies the onset of eye morphogenesis. *Developmental cell,* **27,** 293-305.

Jackson, J. E. (2005) *A user's guide to principal components,* John Wiley & Sons.

Jolliffe, I. (2002) *Principal component analysis,* Wiley Online Library.

Kelley, P. B., Sado, Y. & Duncan, M. K. (2002) Collagen IV in the developing lens capsule. *Matrix biology,* **21,** 415-423.

Khorshed, R. A. (2013) A cell level automated approach for quantifying antibody staining in immunohistochemistry images. A structural approach for quantifying antibody staining in colonic cancer spheroid images by integrating image processing and machine learning towards the implementation of computer aided scoring of cancer markers. University of Bradford.

Koch, I. (2013) *Analysis of multivariate and high-dimensional data,* Cambridge University Press.

Komura, D. & Ishikawa, S. (2017) Machine learning methods for histopathological image analysis. *arXiv preprint arXiv:1709.00786.*

Laas, E., Ballester, M., Cortez, A., Graesslin, O. & Daraï, E. (2017) Unsupervised Clustering of Immunohistochemical Markers to Define High-Risk Endometrial Cancer. *Pathology & Oncology Research,* 1-9.

Loukas, C. G., Wilson, G. D., Vojnovic, B. & Linney, A. (2003) An image analysis□based approach for automated counting of cancer cell nuclei in tissue sections. *Cytometry part A,* **55,** 30-42.

López-Alonso, J. M. & Alda, J. (2002) Bad pixel identification by means of principal components analysis. *Optical engineering,* **41,** 2152-2158.

López-Alonso, J. M. & Alda, J. (2004) Characterization of artifacts in fully digital image-acquisition systems: Application to web cameras. *Optical Engineering,* **43,** 257-266.

López-Alonso, J. M., Alda, J. & Bernabéu, E. (2002) Principal-component characterization of noise for infrared images. *Applied optics,* **41,** 320-331.

Martínez-Sanz, E., Del Río, A., Barrio, C., Murillo, J., Maldonado, E., Garcillán, B., Amorós, M., Fuerte, T., Fernández, Á. & Trinidad, E. (2008) Alteration of medial-edge epithelium cell adhesion in two Tgf-β3 null mouse strains. *Differentiation,* **76,** 417-430.

Matkowskyj, K. A., Cox, R., Jensen, R. T. & Benya, R. V. (2003) Quantitative immunohistochemistry by measuring cumulative signal strength accurately measures receptor number. *Journal of Histochemistry & Cytochemistry,* **51,** 205-214.

Mcdevitt, D. (2012) *Cell biology of the eye,* Elsevier.

Miller, P. N., Pratten, M. K. & Beck, F. (1989) Growth of 9.5□day rat embryos in folic□acid□deficient serum. *Teratology,* **39,** 375-385.

Moro, C. F., Fernandez-Woodbridge, A., D'souza, M. A., Zhang, Q., Bozoky, B., Kandaswamy, S. V., Catalano, P., Heuchel, R., Shtembari, S. & Del Chiaro, M. (2016) Immunohistochemical typing of adenocarcinomas of the pancreatobiliary system improves diagnosis and prognostic stratification. *PloS one,* **11,** e0166067.

Ocak, S., Friedman, D. B., Chen, H., Ausborn, J. A., Hassanein, M., Detry, B., Weynand, B., Aboubakar, F., Pilette, C. & Sibille, Y. (2014) Discovery of new membrane-associated proteins overexpressed in small-cell lung cancer. *Journal of Thoracic Oncology,* **9,** 324-336.

Pathania, M., Semina, E. V. & Duncan, M. K. (2014) Lens extrusion from Laminin alpha 1 mutant zebrafish. *The Scientific World Journal,* **2014**.

Piedrahita, J. A., Oetama, B., Bennett, G. D., Van Waes, J., Kamen, B. A., Richardson, J., Lacey, S. W., Anderson, R. G. & Finnell, R. H. (1999) Mice lacking the folic acid-binding protein Folbp1 are defective in early embryonic development. *Nature genetics,* **23,** 228-232.

Rexhepaj, E., Brennan, D. J., Holloway, P., Kay, E. W., McCann, A. H., Landberg, G., Duffy, M. J., Jirstrom, K. & Gallagher, W. M. (2008) Novel image analysis approach for quantifying expression of nuclear proteins assessed by immunohistochemistry: application to measurement of oestrogen and progesterone receptor levels in breast cancer. *Breast Cancer Research,* **10,** R89.

Reynolds, E. (2006) Vitamin B12, folic acid, and the nervous system. *The lancet neurology,* **5,** 949-960.

Sarnecki, J. S., Burns, K. H., Wood, L. D., Waters, K. M., Hruban, R. H., Wirtz, D. & Wu, P.-H. (2016) A robust nonlinear tissue-component discrimination method for computational pathology. *Laboratory Investigation,* **96,** 450.

Seidal, T., Balaton, A. J. & Battifora, H. (2001) Interpretation and quantification of immunostains. *The American journal of surgical pathology,* **25,** 1204-1207.

Semina, E. V., Bosenko, D. V., Zinkevich, N. C., Soules, K. A., Hyde, D. R., Vihtelic, T. S., Willer, G. B., Gregg, R. G. & Link, B. A. (2006) Mutations in laminin alpha 1 result in complex, lens-independent ocular phenotypes in zebrafish. *Developmental biology,* **299,** 63-77.

Sobhani, I., Tap, J., Roudot-Thoraval, F., Roperch, J. P., Letulle, S., Langella, P., Corthier, G., Van Nhieu, J. T. & Furet, J. P. (2011) Microbial dysbiosis in colorectal cancer (CRC) patients. *PloS one,* **6,** e16393.

Sugrue, S. P. & Hay, E. D. (1981) Response of basal epithelial cell surface and Cytoskeleton to solubilized extracellular matrix molecules. *The Journal of Cell Biology,* **91,** 45-54.

Takei, K., Furuya, A., Hommura, S. & Yamaguchi, N. (2001) Ultrastructural fragility and type IV collagen abnormality of the anterior lens capsules in a patient with Alport syndrome. *Japanese journal of ophthalmology,* **45,** 103-104.

Tang, L. S. & Finnell, R. H. (2003) Neural and orofacial defects in Folbp1 knockout mice. *Birth Defects Research Part A: Clinical and Molecular Teratology,* **67,** 209-218.

Tang, L. S., Santillano, D. R., Wlodarczyk, B. J., Miranda, R. C. & Finnell, R. H. (2005) Role of Folbp1 in the regional regulation of apoptosis and cell proliferation in the developing neural tube and craniofacies. In: *American Journal of Medical Genetics Part C: Seminars in Medical Genetics.* Wiley Online Library.

Taylor, C. & Levenson, R. (2006) Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment II. *Histopathology,* **49,** 411-424.

Taylor, C. R. (1994) An exaltation of experts: concerted efforts in the standardization of immunohistochemistry. *Human pathology,* **25,** 2-11.

Toolbox, O. (2015) MatLab, r2015b edition. September.

Van Agtmael, T., Schlötzer-Schrehardt, U., McKie, L., Brownstein, D. G., Lee, A. W., Cross, S. H., Sado, Y., Mullins, J. J., Pöschl, E. & Jackson, I. J. (2005) Dominant mutations of Col4a1 result in basement membrane defects which lead to anterior segment dysgenesis and glomerulopathy. *Human molecular genetics,* **14,** 3161-3168.

Van Eycke, Y.-R., Allard, J., Salmon, I., Debeir, O. & Decaestecker, C. (2017) Image processing in digital pathology: an opportunity to solve inter-batch variability of immunohistochemical staining. *Scientific Reports,* **7,** 42964.

Weaver, M. S., Sage, E. H. & Yan, Q. (2006) Absence of SPARC in lens epithelial cells results in altered adhesion and extracellular matrix production in vitro. *Journal of cellular biochemistry,* **97,** 423-432.

Yan, Q., Perdue, N., Blake, D. & Sage, E. H. (2005) Absence of SPARC in murine lens epithelium leads to increased deposition of laminin-1 in lens capsule. *Investigative ophthalmology & visual science,* **46,** 4652-4660.

Young, R. W. & Ocumpaugh, D. E. (1966) Autoradiographic studies on the growth and development of the lens capsule in the rat. *Invest Ophthalmol,* **5,** 583-589.