CELL-GRAPH MODELING OF SALIVARY GLAND MORPHOLOGY

Cemal Cagatay Bilgin1, Shayoni Ray2, William P. Daley2, Banu Baydil1, Sharon J. Sequeira2, Bülent Yener1, and Melinda Larsen2

*Funding provided by NIH, R01 DE019244.

Rensselaer Polytechnic Institute, Computer Science Department 110 8th Street, Troy, NY 12180, USA.
University at Albany, State University of New York, Biological Sciences Department, 1400 Washington Avenue, Albany, NY 12222, USA

ABSTRACT

Branching morphogenesis is a developmental process shared by many organs, including the submandibular salivary gland. During morphogenesis, cells within the gland undergo rearrangements to cause changes in the overall tissue morphology. This work presents a methodology based on cell-graphs to quantify these changes in cellular arrangements.

Multiple confocal images of developing salivary gland organ cultures are captured. These cultures are immunostained with a nuclear marker and an epithelial marker to identify epithelial cells as separate from mesenchymal cells. Confocal images are stitched and segmented to identify epithelial and mesenchymal nuclei. Cell-graphs are constructed to model the structural organization of epithelial and mesenchymal cells. Cell-graph metrics are calculated to extract mathematical features that discriminate epithelial vs mesenchymal cells organizations and also distinguish between glands treated with pharmacological inhibitors vs vehicle control.

The results indicate that cell-graph features can be used to both describe and predict the developing salivary gland to provide insights into cellular and physical processes driving morphogenesis.

Index Terms— Salivary Gland, Branching Morphogenesis, Cell-Graphs, Graph-Mining

1. INTRODUCTION

The process of branching morphogenesis is crucial for the development of many organs including the lung, kidney, prostate and mammary glands. The mouse submandibular gland (SMG) is an ideal model system to study branching morphogenesis since its ex-vivo organ culture accurately recapitulates its in-vivo growth characteristics. This process requires invagination of the basement membrane surrounding an epithelial glandular rudiment at specific locations, known as clefts, followed by localized cellular proliferation and outgrowth of the resulting buds. It is the epithelial component of the SMG that undergoes branching morphogenesis, and the behavior of the epithelium is regulated by the surrounding mesenchymal cells. Many specific cell signaling pathways contribute to SMG branching morphogenesis, reviewed in [1].

We recently identified a signaling pathway involving Rho-associated coil-coil kinase (ROCK) that regulates branching morphogenesis by promoting elongation of initiated clefts [2]. We found that ROCK controls a mechanochemical checkpoint to regulate cleft progression both by mediating assembly of the basement membrane component, fibronectin, in clefts and also by controlling cell proliferation.

We present a mathematical model to provide quantitative measurements regarding (i) the cellular changes in the epithelial tissue, and (ii) potential changes that could occur in mesenchymal cells in the presence of a ROCK inhibitor. In this study, we compared the arrangement of cells and their relationships in control SMGs and SMGs treated with an inhibitor of ROCK using the quantitative method of cell-graphs [3, 4].

Our results indicate that structural features computed over the cell-graph are both descriptive and predictive of the developing salivary gland, thus can provide insights into cellular and physical processes driving morphogenesis.

2. METHODOLOGY

2.1. Confocal Imaging of Whole Mount Organ Culture

Ex vivo organ culture: Mouse SMGs were dissected from timed pregnant female mice (strain CD-1, Charles River Laboratories) at embryonic day 13 (4 to 5 buds), with the day of plug discovery designated as E0, following protocols approved by the University at Albany IACUC committee. SMGs were micromanipulated from mandible slices and cultured as described previously [2]. SMG organ cultures were exposed to ROCK inhibitor 140 uM Y27632 (6888000 Calbiochem), or vehicle control for 24 hrs prior to fixation.

Whole-mount immunocytochemistry was performed using SMGs fixed in 4% PFA in phosphate-buffered saline (1XPBS) and 5% sucrose for 20 min at room temperature. SybR Green I (Invitrogen) was used to detect nuclei. Epithelium was detected using an antibody recognizing E-cadherin (BD Biosciences) and a Cy3-conjugated secondary antibody (Jackson ImmunoResearch Lab). SMGs were imaged using a Zeiss LSM510 Meta confocal microscope at 20X (Plan APO/0.75 NA) using the same laser and gain settings for all samples.
Fig. 1. The steps of our methodology are depicted. Images are stitched using the inverse Fourier transform of the phase correlation matrix. The stitched images are then fed to the Otsu thresholding algorithm to find a mask. This mask is used as the input to the active contour method and the epithelial tissue is segmented (c). This final mask is then used to find the epithelial nuclei and mesenchymal nuclei respectively (d) and (e). Using each nucleus as a vertex, we build cell-graphs in our last step (f). Only the zoomed in lower left hand corner of the original image is shown for illustrational purposes.

2.2. Image Stitching

High resolution images of the tissue samples were captured using a Zeiss LSM510 Meta confocal microscope at 20X. With this resolution it is not possible to image the whole gland. To image complete biological specimens with high resolution, images having overlapping regions are captured and stitched together. For stitching, we take the same approach as described in [5]. When tie images, A and B, are related by a translation, the Fourier transform based phase correlation method can be used to find the shift. A correlation matrix \( Q(F(A), F(B)) \) of the phases of the Fourier transform is calculated. The inverse Fourier transform of the phase correlation matrix \( F^{-1}(Q) \) gives the highly correlated regions of the images. There is usually more than one highly correlated region in an image, and the correct shift is found by cross-correlation on the overlapping areas of the input images. A non-linear blending image fusion is also performed to compensate for non-uniformity in the image.

2.3. Image Segmentation

We compare and contrast epithelial tissues and mesenchymal tissues as well as the inhibitor-treated versus untreated tissues. This requires the segmentation of the epithelial region and the mesenchymal region in a given image.

Bud detection is performed in two steps. First, a coarse initial segmentation of the epithelium is performed using the Otsu thresholding algorithm. Then the result of this step is used for the initialization of the active contours without edges technique [6]. In this approach the image \( u_0 \) is assumed to be formed by two regions of approximately piecewise-constant intensities \( u_1^0, u_1^0 \). Let \( C \) be a curve in \( \Omega \), and \( \omega \) denote the region inside the curve \( C \), and \( \Omega \setminus \omega \) the region outside the curve \( C \). Using the area and the length of this curve as regularization terms, Chan and Vese approach introduces the energy functional \( F(\epsilon_1, \epsilon_2, C) \) given as in equation (1), and
solves the problem of minimizing this functional,
\[
F(c_1, c_2, C) = \mu \text{Length}(C) + \nu \text{Area}(\omega) \\
+ \lambda_1 \int_{\omega} |u_0(x, y) - c_1|^2 dx dy \\
+ \lambda_2 \int_{\Omega \setminus \omega} |u_0(x, y) - c_2|^2 dx dy
\]  
over the curves \( C \) and the average intensity values \( c_1, c_2 \) inside and outside the curves, where \( \mu \geq 0, \nu \geq 0, \lambda_{1,2} \) are fixed parameters.

2.4. Cell-Graph Formation

Formally, a graph is represented by \( G = (V, E) \) where \( V \) is the vertex set and \( E \) is a binary relation on \( V \). Our structural modeling considers each cell as a vertex and captures the pairwise distance relationship between these cells. In cell-graph formation step, we build edges between the vertices calculated in the previous step. A relationship between two vertices is hypothesized if these two vertices are touching or close to each other. Biologically, this might mean that these cells are communicating with each other using chemical signals, cell-cell adhesions, etc.

We find the center of mass for each cell and store their \( x, y \) coordinates. We hypothesize a communication by setting a link between two nodes if the euclidean distance between them is less than a threshold that ensures a physical contact between the corresponding cell membranes. The Euclidean distance between two cells is simply given by equation (2)

\[
d(u, v) = \sqrt{(u_x - v_x)^2 + (u_y - v_y)^2},
\]

where \( u_x \) and \( u_y \) are \( x \) and \( y \) coordinates of node \( u \) respectively.

The cell-graph approach has two main advantages over the Delaunay triangulations. First, Delaunay graphs are one single connected component and second, Delaunay graphs are planar. There is no evidence for these two assumptions to hold in a tissue and moreover a Delaunay triangulation is a very special case of a cell-graph.

2.5. Metric Calculation

We extract a rich set of features from the structural representations of the tissues. Our feature set can be grouped into three categories: simple metrics, distance-based metrics and connectedness metrics. The simplest metrics defined on a graph are the number of nodes and number of edges. These correspond to the total number of cells in the tissue, and the total number of cells in communication respectively.

The next set of features extracted from cell-graphs are the distance-based metrics. These metrics quantify how far the nodes are apart from each other. First, the shortest path distances from each and every node to every other node in the graph are calculated. The eccentricity of a node \( u \) is then given as the maximum shortest path distance from node \( u \) to any other node in the graph. Using the eccentricity values for each node, the diameter of the graph is given by the maximum eccentricity. Similarly the minimum eccentricity is called the radius, and the nodes that have eccentricity values equal to the radius are defined as central points. The number of central points in the graph is also another feature.

Finally, we calculate the connectedness and cliquishness metrics which includes the size of the largest connected component where each node is reachable from each other. The ratio of this component to the whole graph size gives the giant connected component ratio. One very common metric used in social networks is the average clustering coefficient. For a vertex, the ratio of the links that vertex’s neighbors have in between to the total number that can possibly exist gives the clustering coefficient. We take the average value of these and assign it as average clustering coefficient. A full list of the features used in this work is given in [3].

2.6. Cell-Graph Mining

We use support vector machines to learn the topological difference between treated and untreated samples of epithelial and mesenchymal tissues. SVM projects the data into a higher dimension and finds an optimal separating hyperplane between data points such that the data points of different classes fall on the opposite sides of this hyperplane. The parameters of the optimal separating hyperplane are given by the quadratic programming optimization problem which minimizes \( g \), where \( g(u, \xi) = \frac{1}{2} \|w\|^2 + M \sum_{i=1}^{N} \xi_i \) with linear equality and inequality constraints. Reformulating this minimization problem, the SVM problem can be defined as in equation (3),

\[
\min: W(\alpha) = - \sum_{i=1}^{N} \alpha_i + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_i y_j \alpha_i \alpha_j k(x_i, x_j)
\]

\text{s.t.: } \sum_{i=1}^{N} y_i \alpha_i = 0, \forall i : 0 \leq \alpha_i \leq M

where \( \alpha_i \) is a Lagrange multiplier that corresponds to the sample \( x_i, y_i \) is the corresponding class label for \( x_i \) and \( k(., .) \) is a kernel function. The use of kernel functions is a very crucial feature of SVMs. Using this kernel function, input data is transformed into a new space and the optimal separating hyperplane is found in this new space. Using nonlinear kernel functions, SVMs can also separate nonlinear datasets. We have used the radial basis kernel, also called Gaussian kernel, defined as \( k(x_i, y_i) = e^{-\gamma \|x_i - y_i\|^2} \).

3. EXPERIMENTS AND RESULTS

The dataset contains 10 inhibitor-treated and 10 untreated tissue samples, where each sample has two regions, namely ep-
ithelium and mesenchyme. We solve five classification problems on this dataset to compare and contrast the structural differences between the treated and untreated epithelial and mesenchymal tissue samples.

First, we perform the epithelial versus mesenchymal tissue comparison, in untreated tissue samples only, to find the topological differences of mesenchymal and epithelial graphs that coexist in the same tissue. We then perform the same comparison for treated tissue samples. Similarly, this ensures that epithelial graphs are compared to mesenchymal graphs in the same tissue sample. This first set of experiments do not give insight to whether the treated and untreated examples are distinguishable from each other.

We also consider the structural differences between the treated and untreated samples. Thus, we compare (i) only the epithelial graphs of the treated and untreated samples, and (ii) only the mesenchymal tissues of treated and untreated samples.

Finally, we consider the gland as a whole (rather than comparing the epithelial and mesenchymal tissues separately) and represent it by the union of epithelial and mesenchymal graph features. The classification problem is solved to distinguish treated glands from untreated ones.

<table>
<thead>
<tr>
<th>Link Thr</th>
<th>Epi-Mes in Tr</th>
<th>Epi-Mes in Ut</th>
<th>Tr-Ut in Mes</th>
<th>Tr-Ut in Epi</th>
<th>Tr-Ut in Gland</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>100</td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>95</td>
<td>95</td>
<td>90</td>
<td>60</td>
<td>85</td>
</tr>
</tbody>
</table>

Table 1. Cell-Graph mining training accuracy is given. Modeling the structural organization of epithelial cells is sensitive to the link threshold since the epithelial compartment includes cells that are in different functional states (i.e., undergoing differentiation into both acinar and ductal cells), thus heterogeneity within the epithelial branching.

<table>
<thead>
<tr>
<th>Link Thr</th>
<th>Epi-Mes in Tr</th>
<th>Epi-Mes in Ut</th>
<th>Tr-Ut in Mes</th>
<th>Tr-Ut in Epi</th>
<th>Tr-Ut in Gland</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>95</td>
<td>95</td>
<td>85</td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td>3</td>
<td>70</td>
<td>95</td>
<td>85</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>80</td>
<td>65</td>
<td>65</td>
<td>70</td>
<td>60</td>
</tr>
</tbody>
</table>

Table 2. Cell-Graph mining accuracy using leave-one-out testing method is given. Sensitivity to the link threshold is also observed due to a different structure-function relationship within epithelial compartment and heterogeneity within the epithelial branching which may require a subclassification problem to be solved for higher accuracy. The accuracy can be further increased by increasing the cardinality of the training and testing sets.

The training and the leave-one-out testing accuracy of cell-graph mining method are summarized in tables 1 and 2 respectively. Table 2 shows that for link lengths of 3 units of distance, cell-graphs can achieve 85% accuracy for the treated-untreated comparison in mesenchymal tissue samples and 80% accuracy for the treated and untreated epithelial graph comparison, respectively.

We remark that the epithelial compartment includes cells that are undergoing differentiation into both acinar and ductal cells, which is not detectable using the E-cadherin marker we have used. Due to the variances in growth of different SMGs, it may be necessary to solve a sub-classification problem to distinguish cellular sub-compartments within the epithelium for more accurate quantification.

4. CONCLUSION

We show that cell-graph mining can be used to classify epithelium vs mesenchymal tissues in untreated- and Y-27632 inhibitor-treated embryonic SMG tissues from single confocal images. More importantly, we can classify epithelium as untreated- or treated-, mesenchyme as untreated- or treated-, and whole tissue (epithelium and mesenchyme) as untreated- or inhibitor-treated using the same methods. From these results we will extract features that will provide insight into the action of the ROCK inhibitor Y27632 on salivary gland development and will be used to construct a predictive temporal mathematical model of branching morphogenesis.

5. REFERENCES


