Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2018 April ; 2018: 796–799. doi:10.1109/ISBI.2018.8363692.

A METHOD FOR QUANTIFICATION OF CALPONIN EXPRESSION IN MYOEPITHELIAL CELLS IN IMMUNOHISTOCHEMICAL IMAGES OF DUCTAL CARCINOMA IN SITU

Elliot Gray¹, Elizabeth Mitchell², Sonali Jindal², Pepper Schedin², and Young Hwan Chang¹

¹Department of Biomedical Engineering and Computational Biology Program

²Department of Cell, Developmental and Cancer Biology Oregon Health & Science University

Abstract

Ductal carcinoma in situ (DCIS) is breast cancer confined within mammary ducts, surrounded by an intact myoepithelial cell layer that prevents local invasion. A DCIS diagnosis confers increased lifetime risk of developing invasive breast cancer (IBC) and results in surgical excision with radiation, and possibly endocrine- or chemo-therapy. DCIS is known to be over treated, with associated co-morbidities. Biomarkers are needed that delineate patients at low risk of DCIS progression from patients requiring more aggressive treatment. Investigating the role of myoepithelial cell differentiation in barrier function is anticipated to provide insight into DCIS progression and delineate between low and high risk lesions. Here, we develop a high throughput technique to assess loss of myoepithelial differentiation markers. This method facilitates automated analysis of a clinically relevant histopathologic feature, as demonstrated by a high correlation with pathologist annotation ($r = 0.959$), and further, contributes analytical foundations to a multiplexed immunohistochemistry (IHC) approach.

Keywords

DCIS; calponin; invasive breast cancer; feature engineering; myoepithelial cell

1. DCIS AND INVASIVE BREAST CANCER

Incidence of ductal carcinoma in situ (DCIS) increased from 3% in the early 1970's to ~20% of all breast cancers due to widespread use of screening mammography [1]. This year in the U.S., ~50,000 women will be diagnosed with DCIS. DCIS carries an increased lifetime risk of invasive breast cancer (IBC) with long term studies of untreated DCIS showing 30% progression to IBC [2]. Unfortunately, the clinical accuracy to predict which DCIS will remain indolent and which will progress to IBC remains low, leaving an urgent need to discover biomarkers of progression.

The myoepithelial cell is a 'gate keeper' to breast cancer invasion through the mammary duct, exerting tumor suppressive effects by secreting tumor suppressive proteins and by forming a physical barrier that prevents tumor cell escape [3–6]. Variable expression of myoepithelial differentiation markers such as calponin and α -smooth muscle actin (α SMA) are reported in DCIS cases, while physical gaps in the myoepithelium or its loss are used

clinically to define IBC [7]. Further, a rodent model of DCIS revealed a progressive loss in myoepithelial differentiation markers calponin and α SMA prior to transition to IBC, implicating loss of myoepithelial cell differentiation as a precursor to IBC [7]. To date, none of the known myoepithelial cell proteins have been evaluated for their ability to distinguish between indolent and progressing DCIS. In order to assess if loss of myoepithelial differentiation can serve as a biomarker for identifying DCIS lesions at risk for progression, we need an unbiased, high throughput approach to quantify myoepithelial protein expression surrounding DCIS lesions.

2. COMPUTATIONAL HISTOPATHOLOGY

Histopathologic image analysis is a mainstay for quantitative evaluation of tissue structure and in situ protein expression [8]. The fields of histopathology and basic cancer research are increasingly turning to computational image processing techniques to aid in quantification of complex or subtle biological image features [9]. Quantified image features can then be used to predict relevant clinical attributes, such as disease diagnosis. The need for computational assistance is particularly true in the case of multiplexed IHC, where novel staining methods enable 12 or more antibodies to be imaged in one tissue section, substantially increasing pathologist burden [10, 11].

To this end, analysis methods such as Convolutional Neural Networks (CNN), which learn spatial features at multiple scales across an entire image, have been applied to discriminate DCIS from benign lesions in H&E images [12]. CNNs were also used to segment tumor-associated stroma and then to predict DCIS grade or cancer diagnosis [13]. Approaches which compute the morphologic features of cell nuclei, either focusing on epithelial cells, or encompassing all cells in a region of interest, have also found success in predicting DCIS grade or distinguishing between DCIS, benign neoplasia, and normal tissue [12, 14]. Yet, there remains a need for a method that can be applied to current multiplex IHC data to quantify continuity of expression of myoepithelial cell proteins, as focal loss of myoepithelial differentiation may predict increased risk for progression.

3. IMMUNOHISTOCHEMICAL METHODS

Formalin fixed paraffin embedded (FFPE) tissue of DCIS from women diagnosed with DCIS only (no evidence of IBC) were obtained through the OHSU Knight Biobank in full accordance of the OHSU IRB guidelines (OHSU IRB 4918).

Tissue sections were stained with a multiplexed IHC method, using antibodies to calponin, α SMA, and cytokeratin 18 (CK18) [10]. Hematoxylin staining was used to allow single cell image registration and color images were deconvolved with ImageJ as described previously [10]. DCIS-associated myoepithelium (n=6 DCIS cases, 27 DCIS lesions) were assessed for percent calponin expression (i.e. continuity), using visual assessment of the IHC signal by three independent investigators including a trained pathologist (S.J.). The annotated values were averaged and then compared to the computational method described here.

4. SPATIAL ANALYSIS OF MYOEPITHELIUM

In cross section of a normal mammary duct, the myoepithelium forms a closed, continuous boundary surrounding a single layer of luminal epithelium, whereas in DCIS, the myoepithelium encloses multiple layers of neoplastic cells. Here we detected tumor cells by CK18, delineated the myoepithelium with α SMA, and evaluated myoepithelial cell differentiation using calponin (Fig 1a). We developed a procedure for analysis of the spatial distribution of calponin expression in the myoepithelium, specifically looking for areas lacking calponin expression.

4.1 Definition of Continuity of Expression

While assessing myoepithelial-layer integrity is routine clinical practice for pathologists to delineate DCIS from IBC, markers to distinguish indolent from progressing DCIS are lacking. We hypothesize that myoepithelial cell differentiation markers may predict risk of DCIS progression, and thus developed a method to assess continuity of the myoepithelial differentiation protein calponin as a test case. We utilized cases of DCIS where the myoepithelial layer was deemed intact. An image key was created by a pathologist (S.J.) to visually assess the percentage of the myoepithelium where calponin is expressed. For example, if 50% of DCIS myoepithelial cells are lacking calponin, then the lesion would be annotated with a continuity of 50%.

4.2. Segmentation

To reduce reliance on visual pathologic review, we developed several ways to estimate calponin continuity using segmentation and geometric feature calculation. Our segmentation procedure uses simple thresholding by Otsu's method [15]. The result is three binary images comprised from α SMA, CK18, and calponin IHC images (Fig 1b). We refer to these individual IHC-derived images as "masks". These masks are then used to construct two "composite masks" described below, and which are used for geometric quantification.

The first composite mask, called the "DCIS mask" (Fig. 1d), is formed using the union of all three simple masks. DCIS lesions stained with our three markers tend to be significantly larger than other connected objects in such a binary image, such as noise, myofibroblasts, and blood vessels. We used morphological operations such as dilation and erosion to refine the DCIS mask by smoothing edges and removing small objects. Each connected object (connected is defined here to mean pixelwise 8-connectivity) in the DCIS mask was assigned a unique ID, which was used to match the pathologist's annotation to our geometric calculations.

The second composite mask, called the "calponin mask" (Fig. 2a) was found by taking the intersection of the simple calponin mask with the DCIS mask. We refined it by setting pixels equal to zero that were further than 60 pixels (30 μ m) inside the DCIS mask boundary. For method development, the selection of 60 pixels was empirically determined to be the maximum thickness of myoepithelia (in our data set). We set this constraint to omit calponin staining that can be found within the tumor center both specifically in papillary type DCIS and non-specifically in necrotic regions, which we found to sometimes confound analysis of

the relevant specific staining of myoepithelial cells. Thus, each DCIS mask is associated with a corresponding calponin mask, which contains the region of positive calponin staining for a single myoepithelium.

4.3. Automatic Finding of Myoepithelium Discontinuities

To enable continuity evaluation of calponin expression within the myoepithelial cell layer, we constructed line segments along the shortest path between disconnected objects in the calponin mask. In our data, DCIS lesions are frequently non-convex in shape (e.g. Fig. 1). Calponin masks also tend to have a high variability in the size, shape, and number of objects, which precludes the approach of fully connecting all objects, or the use of convex hull or Matlab's Alpha Shape to connect only objects along the outside boundary. We also found that a calponin mask can have a single discontinuity, yet consist of a single fully connected object with one hole, that we could not distinguish mathematically from a perfectly continuous mask. We solved these problems in part by using the Simple Linear Iterative Clustering (SLIC) superpixel method to partition the calponin mask [16].

SLIC has two parameters that govern the size and compactness, respectively, of each superpixel. The size parameter governs the average area of superpixels, and high compactness produces more regular, square-shaped superpixels. Superpixels of size comparable to that of a cell, or about 10 μ m wide, and relatively high compactness of 100, were found empirically to give superior results compared to other parameter choices. By partitioning the calponin mask this way, we hoped to detect a single discontinuity in an otherwise topologically connected object, and also to provide a basis for constraining a search for discontinuities to neighboring regions of the mask.

To define local, or neighboring image regions, we first found the Voronoi tessellation of the set of centroids of all superpixels (Fig. 2b) [17]. This returns a one-to-one mapping of superpixels to Voronoi cells. We designated all superpixels whose Voronoi cells were touching as neighbors and as candidates for constructing a connecting line segment between them. It can also be stated that superpixels were called neighbors if they were connected by an edge in the Delaunay triangulation, which is the dual of the Voronoi tessellation [17].

Finally, we constructed the set of all “discontinuity” line segments by connecting a subset of neighboring superpixels along the shortest path between their respective perimeters. We connected two neighboring superpixels if they were not touching, and if their Voronoi cells were bridged by the DCIS mask boundary (pink line in Fig. 2b). We define the set L of superpixel pairs satisfying these properties as follows. Let $B(t)$ be an ordering of the pixels in the DCIS mask boundary, such that $B(t)$ is touching $B(t+1)$ for all $t=1, \dots, p-1$ where p is the number of pixels in the DCIS mask $B(1)$ is touching $B(p)$. This is the precise output of the Matlab function “bwboundaries” on the DCIS mask. We denote the and write: ID of the Voronoi cell containing the pixel $B(t)$ as V_t and write:

$$L = \left\{ \left(V_t, V_{\text{mod}_p(t+1)} \right) \middle| V_t \neq V_{\text{mod}_p(t+1)}, 1 \leq t \leq p \right\} \quad (1)$$

We use the modulo function $\text{mod}_p(t+1)$ to denote the fact that pixels $B(1)$ and $B(p)$ are touching (8-connected). A final refinement is performed by discarding all of the line segments that do not intersect the outer perimeter of the resulting connected calponin mask (purple curve in Fig. 2c), and the refined set of “discontinuity” line segments is denoted $\{d_i\}$.

4.4. Estimating Per-Lesion Continuity of Calponin Expression

A simple and robust method to quantify the continuity of calponin in the myoepithelium is to compare the area that is stained to a reference area, or area where the myoepithelium “ought to be” stained. We defined

$$\text{naive continuity} = 100 * \frac{\text{Area}_{\text{calponin}}}{\text{Area}_{\text{myoepithelium}}} \quad (2)$$

Here the numerator is the area of the calponin mask for the singular lesion in question, and the denominator is the area of the region between the DCIS mask boundary for the same lesion and the erosion of that boundary by 60 pixels (30 μm) (Fig. 2a). We found that the naïve estimate of continuity had a good correlation with the pathologist’s annotations (Fig. 3 left), and provide it as a benchmark. We utilized the constructed discontinuity line segments described in section (3.3) for a refined estimate of per-lesion continuity:

$$\text{continuity} = 100 * \frac{P - \sum_{i=1}^n d_i}{P} \quad (3)$$

where P is an estimate of the perimeter of the myoepithelium, and $\{d_i\}$ is the set of lengths of the n constructed line segments. The line segments $\{d_i\}$ added to the calponin mask create an (8-)connected object with a hole for which inner and outer perimeters P^O and P^I are measurable (Fig. 2c). Note that in the case of an object with more than one hole, we compute P^I as the inner perimeter of the largest hole. We made the approximation $P \cong (P^I + P^O)/2$, and finally computed

$$\text{continuity} = 100 * \left(1 - \frac{2 \sum_{i=1}^n d_i}{P^I + P^O} \right) \quad (4)$$

5. RESULT AND DISCUSSION

We found that the naïve method (Eq. 2) achieves a high correlation ($R = 0.791$) with the mean visually annotated continuity of calponin (i.e. pathologist assessment). Abdallah *et al.* recently reported computational measurements of the sensitivity of calponin, P63, CD10, and αSMA in demonstrating the myoepithelium of DCIS lesions, with results comparable to our “naïve method” [18].

The proposed method (Eq. 4) achieved a Pearson's-correlation of $R = 0.959$ with pathologist's annotation. We also observed some discordance between pathologist and our method's assessments. One outlier (Fig. 3 right) was annotated at 10% continuity by pathologist assessment, but found to be 50% continuous by our computational method. This sample had faint staining and low signal-to-noise ratio, so it is likely that the discrepancy resulted from the choice of threshold by Otsu's method, which might be corrected using sample intensity normalization and flat threshold choice [19], or a more advanced segmentation method.

One drawback of our method is the reliance on segmentation of the DCIS mask boundary. Central necrosis or tissue-loss during the staining process were found to be the primary confounders of segmentation, although we note that cases like this would be difficult to analyze traditionally as well. Segmentation results must be either checked visually (as we did here), or filtered for artifacts. Future work to ensure the robustness of segmentation, for example by filtering out images with severe tissue loss during imaging, should be done to take full advantage of our method.

This study describes a computational method to assess continuity of calponin in mammary myoepithelium in cases of DCIS. This method, closely matches what can be detected by the trained human eye but in significantly reduced time and in the absence of observer bias. This spatial-structural analysis of continuity of myoepithelial differentiation proteins combined with multiplex IHC technique promises the capacity for integration of myoepithelial, immune cell and tumor intrinsic factors, which may help predict risk of DCIS progression to invasive disease.

Acknowledgements:

This research was supported by a grant from Susan G. Komen® PDF17480342 (E.M), and the Oregon Health and Science University (OHSU) Center for Spatial Systems Biomedicine.

6. REFERENCES

- [1]. Jansen SA, "Biology of DCIS and Progression to Invasive Disease," in *Breast Cancer Biology for the Radiation Oncologist*: Springer, 2012, pp. 35–48.
- [2]. Collins LC, Tamimi RM, Baer HJ, Connolly JL, Colditz GA, and Schnitt SJ, "Outcome of patients with ductal carcinoma in situ untreated after diagnostic biopsy," *Cancer*, vol. 103, no. 9, pp. 1778–1784, 2005. [PubMed: 15770688]
- [3]. Adriance MC, Inman JL, Petersen OW, and Bissell MJ, "Myoepithelial cells: good fences make good neighbors," *Breast Cancer Research*, vol. 7, no. 5, p. 190, 2005. [PubMed: 16168137]
- [4]. Pandey PR, Saidou J, and Watabe K, "Role of myoepithelial cells in breast tumor progression," *Frontiers in bioscience: a journal and virtual library*, vol. 15, p. 226, 2010.
- [5]. Polyak K and Hu M, "Do myoepithelial cells hold the key for breast tumor progression?," *Journal of mammary gland biology and neoplasia*, vol. 10, no. 3, pp. 231–247, 2005. [PubMed: 16807803]
- [6]. Gudjonsson T, Adriance MC, Sternlicht MD, Petersen OW, and Bissell MJ, "Myoepithelial cells: their origin and function in breast morphogenesis and neoplasia," *Journal of mammary gland biology and neoplasia*, vol. 10, no. 3, pp. 261–272, 2005. [PubMed: 16807805]
- [7]. Russell TD et al., "Myoepithelial cell differentiation markers in ductal carcinoma in situ progression," *The American journal of pathology*, vol. 185, no. 11, pp. 3076–3089, 2015. [PubMed: 26343330]

- [8]. Zaha DC, "Significance of immunohistochemistry in breast cancer," World journal of clinical oncology, vol. 5, no. 3, p. 382, 2014. [PubMed: 25114853]
- [9]. Menter DG et al., "Computerized image analysis of Ki-67 in ductal breast carcinoma in situ," Analytical and quantitative cytology and histology, vol. 23, no. 3, pp. 218–228, 2001. [PubMed: 11444192]
- [10]. Tsujikawa T et al., "Quantitative Multiplex Immunohistochemistry Reveals Myeloid-Inflamed Tumor-Immune Complexity Associated with Poor Prognosis," Cell Reports, vol. 19, no. 1, pp. 203–217, 2017. [PubMed: 28380359]
- [11]. Stack EC, Wang C, Roman KA, and Hoyt CC, "Multiplexed immunohistochemistry, imaging, and quantitation: a review, with an assessment of Tyramide signal amplification, multispectral imaging and multiplex analysis," Methods, vol. 70, no. 1, pp. 46–58, 2014. [PubMed: 25242720]
- [12]. Radiya-Dixit E, Zhu D, and Beck AH, "Automated Classification of Benign and Malignant Proliferative Breast Lesions," Scientific reports, vol. 7, no. 1, p. 9900, 2017. [PubMed: 28852119]
- [13]. Bejnordi BE et al., "Deep learning-based assessment of tumor-associated stroma for diagnosing breast cancer in histopathology images," arXiv preprint arXiv:1702.05803, 2017.
- [14]. Yamamoto Y et al., "Quantitative diagnosis of breast tumors by morphometric classification of microenvironmental myoepithelial cells using a machine learning approach," Scientific Reports, vol. 7, 2017.
- [15]. Otsu N, "A threshold selection method from gray-level histograms," IEEE transactions on systems, man, and cybernetics, vol. 9, no. 1, pp. 62–66, 1979.
- [16]. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, and Süsstrunk S, "SLIC superpixels compared to state-of-the-art superpixel methods," IEEE transactions on pattern analysis and machine intelligence, vol. 34, no. 11, pp. 2274–2282, 2012. [PubMed: 22641706]
- [17]. Liebling TM and Pournin L, "Voronoi diagrams and Delaunay triangulations: Ubiquitous siamese twins," Documenta Mathematica, ISMP, pp. 419–431, 2012.
- [18]. Abdallah DM MEDN, "Comparative Immunohistochemical Study of P63, SMA, CD10 and Calponin in Distinguishing in situ from Invasive Breast Carcinoma," J Mol Biomark Diagn, vol. 8, no. 342, 2017.
- [19]. Khan AM, Rajpoot N, Treanor D, and Magee D, "A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution," IEEE Transactions on Biomedical Engineering, vol. 61, no. 6, pp. 1729–1738, 2014. [PubMed: 24845283]

Figure 1.

Human DCIS lesion multiplex IHC stained to delineate myoepithelial cell layer (α SMA+) from tumor cells (CK18+), as well as assess myoepithelial cell differentiation (calponin). Otsu's method is applied to color-deconvoluted, multiplexed, IHC images (a) producing simple binary masks (b). The union of the masks is found (c) and processed with mathematical morphology to produce a segmentation of each DCIS lesion, including the α SMA- and calponin-positive myoepithelium (d).

Figure 2.

Several intermediate steps of our image analysis method. **(a)** The grayscale image of calponin staining is shown outlined in blue by the result of “calponin mask” segmentation. The outer orange curve shows the boundary of the “DCIS mask” for this lesion. The inner orange boundary shows the morphological erosion of the DCIS mask. The *naïve continuity* is the area outlined in blue divided by the area between the orange curves. **(b)** calponin mask is shown in brown with SLIC superpixel partition. Voronoi cells computed for the superpixel centroids are drawn in light blue, and the DCIS mask is drawn in dark blue and outlined in pink. Connections made across calponin mask gaps are in yellow. **(c)** The grayscale image of calponin is shown again, with estimated outer (P^O , purple) and inner (P^I teal) connected calponin mask perimeters shown. Scale bar 50 μ m.

Figure 3:
Correlation between analytical calponin continuity estimation and pathologist's annotation for Naïve estimate and the proposed method.