



Co-Occurring Gland Angularity in Localized Subgraphs: Predicting Biochemical Recurrence in Intermediate-Risk Prostate Cancer Patients

George Lee^{1*}, Rachel Sparks², Sahirzeeshan Ali¹, Natalie N. C. Shih³, Michael D. Feldman³, Elaine Spangler⁴, Timothy Rebbeck⁴, John E. Tomaszewski⁵, Anant Madabhushi^{1*}

1 Case Western Reserve University, Department of Biomedical Engineering, Cleveland, Ohio, United States of America, **2** Rutgers, the State University of New Jersey, Department of Biomedical Engineering, Piscataway, New Jersey, United States of America, **3** University of Pennsylvania, Department of Pathology and Laboratory Medicine, Philadelphia, Pennsylvania, United States of America, **4** University of Pennsylvania, Department of Clinical Epidemiology and Biostatistics, Philadelphia, Pennsylvania, United States of America, **5** University at Buffalo, State University of New York, Department of Pathology and Anatomical Sciences, Buffalo, New York, United States of America

Abstract

Quantitative histomorphometry (QH) refers to the application of advanced computational image analysis to reproducibly describe disease appearance on digitized histopathology images. QH thus could serve as an important complementary tool for pathologists in interrogating and interpreting cancer morphology and malignancy. In the US, annually, over 60,000 prostate cancer patients undergo radical prostatectomy treatment. Around 10,000 of these men experience biochemical recurrence within 5 years of surgery, a marker for local or distant disease recurrence. The ability to predict the risk of biochemical recurrence soon after surgery could allow for adjuvant therapies to be prescribed as necessary to improve long term treatment outcomes. The underlying hypothesis with our approach, co-occurring gland angularity (CGA), is that in benign or less aggressive prostate cancer, gland orientations within local neighborhoods are similar to each other but are more chaotically arranged in aggressive disease. By modeling the extent of the disorder, we can differentiate surgically removed prostate tissue sections from (a) benign and malignant regions and (b) more and less aggressive prostate cancer. For a cohort of 40 intermediate-risk (mostly Gleason sum 7) surgically cured prostate cancer patients where half suffered biochemical recurrence, the CGA features were able to predict biochemical recurrence with 73% accuracy. Additionally, for 80 regions of interest chosen from the 40 studies, corresponding to both normal and cancerous cases, the CGA features yielded a 99% accuracy. CGAs were shown to be statistically significantly ($p < 0.05$) better at predicting BCR compared to state-of-the-art QH methods and postoperative prostate cancer nomograms.

Citation: Lee G, Sparks R, Ali S, Shih NNC, Feldman MD, et al. (2014) Co-Occurring Gland Angularity in Localized Subgraphs: Predicting Biochemical Recurrence in Intermediate-Risk Prostate Cancer Patients. PLoS ONE 9(5): e97954. doi:10.1371/journal.pone.0097954

Editor: Zhuang Zuo, UT MD Anderson Cancer Center, United States of America

Received: January 27, 2014; **Accepted:** April 27, 2014; **Published:** May 29, 2014

Copyright: © 2014 Lee et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by Department of Defense (W81XWH-12-1-0171); Department of Defense (W81XWH-11-1-0179); National Cancer Institute of the National Institutes of Health under award numbers R01CA136535-01, R01CA140772-01, and R21CA167811-01; the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award number R43EB015199-01; the National Science Foundation under award number IIP-1248316; and the QED award from the University City Science Center and Rutgers University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: george.lee@case.edu (GL); anant.madabhushi@case.edu (AM)

Introduction

Each year in the United States, nearly 60,000 men diagnosed with prostate cancer (CaP) undergo radical prostatectomy (RP) [1]. In cases for which there is no prior evidence of disease spread, treatment of CaP with RP has generally resulted in favorable long term outcome [2]. However, for 15–40% of RP patients, biochemical recurrence (BCR) occurs within 5 years of surgery [1]. BCR is commonly defined as a detectable persistence of prostate specific antigen (PSA) of at least 0.2 ng/ml and is suggestive of either local or distant recurrence of disease necessitating further treatment [3]. Consequently, it is important to be able to predict the risk of BCR soon after surgery, so that if needed, adjuvant treatments can be initiated.

Gleason scoring [4] is a pathology based grading system based on the visual analysis of glandular and nuclear morphology. Low

Gleason scores have been associated with more favorable longer term prognosis for prostate cancer, while the converse is true for higher Gleason scores [5]. Gleason scoring combines the grade of the most common and second most common patterns within the tissue section, resulting in a Gleason sum ranging from 2 (least aggressive) to 10 (most aggressive). Gleason score is currently regarded as the best biomarker for predicting disease aggressiveness and longer term, post-surgical patient outcome [4]. Unfortunately, post-surgical outcome of prostate cancer patients with intermediate Gleason scores can vary considerably [6]. Some statistical tables suggest a 5-year BCR-free survival rate as low as 43% in men with Gleason sum 7 [5]. Furthermore, Gleason scoring is subject to considerable inter-reviewer variability [7]. Allsbrook et al. [7] reported a kappa-coefficient of 0.4 representing moderate agreement amongst pathologists for grading Gleason score 7 patterns. Therefore, the prognostic value of Gleason

scoring alone for predicting BCR in RP patients with intermediate Gleason scores appears to be limited.

Over the last two decades, many postoperative nomograms have been developed to incorporate additional clinical variables such as tumor stage, pre-operative PSA, or positive surgical margins [5,8–11] in order to predict patient and disease outcome. The Kattan nomogram [8] incorporated these parameters to predict 80 month BCR free survival following radical prostatectomy. Han et al. [5] incorporated Gleason sum, tumor stage, and pre-operative PSA into a series of probability tables, known as the Han Tables. Subsequently, the Stephenson nomogram [9] added the date of surgery as a prognostic variable. The University of California at San Francisco built their own risk score predictor (CAPRA) [10] to separate postoperative CaP patients into low, medium, and high risk categories. Apart from the parameters used in the Kattan nomogram, the CAPRA score also included the percentage of positive biopsy cores into their risk assessment. Hinev et al. [11] performed an independent study advocating the use of the Memorial Sloan Kettering Cancer Center (MS-KCC) nomogram, developed by Kattan and Stephenson, suggesting superior prediction of 5-year BCR compared to the Han tables. The MS-KCC nomogram adds additional variables such as age and time free of cancer. These nomograms represent the state-of-the-art in postoperative CaP prediction of BCR, but still rely heavily on Gleason score, which is derived from pathologist interpretation.

The recent advent of digital whole slide scanners has allowed for high resolution digitization of tissue slides. These digitized slide images can be subsequently subjected to quantitative histomorphometry (QH). A variety of QH tools have been previously employed for describing, classifying, and diagnosing disease patterns from histopathology images [12]. In the context of excised prostate pathology specimens, QH has been utilized successfully in a wide range of applications from cancer detection to prognosis, Monaco et al. developed and employed a markov random field (MRF) algorithm for detection of prostate cancer [13].

Some researchers have also explored the role of image texture in characterizing the appearance of CaP morphology. For the purpose of automated CaP grading, Jafari-Khouzani et al. [14] examined the role of second order image intensity texture features based on co-occurrence matrices. Co-occurrence matrices evaluate the frequency with which two image intensities appear within a pre-defined distance of each other within a neighborhood. A series of second-order statistical features (e.g. entropy) [15] to describe the co-occurrence matrix can then be extracted and serve to describe the local image texture. Other texture features such as first order statistical image intensity and steerable gradient filters (e.g. Gabor filters) [16] have also been used to predict CaP. Texture features, while they have been shown to be useful for characterizing CaP morphology, often suffer from a lack of transparency and interpretability.

Another class of approaches have attempted to explicitly model CaP appearance by interrogating the spatial arrangement of individual nuclei and glands. Tabesh et al. [17] investigated color, texture, and structural morphology to perform automated Gleason scoring in prostate histopathology. In Doyle et al. [16], morphological descriptors such as gland size and perimeter ratio were shown to distinguish benign and malignant histological regions. Veltri et al. [18] investigated nuclear morphology using a descriptor called nuclear roundness variance to predict biochemical recurrence in men with prostate cancer.

Many researchers have also attempted to model QH tissue architecture, via the use of graphs networks to characterize the

spatial arrangement of nuclei and glands [16,19–21]. Christens-Barry et al. used Voronoi- and Delaunay-based graph tessellations to describe tissue architecture in CaP histology [19]. Doyle et al. [16] showed that the Minimum Spanning Trees, in addition to Voronoi, Delaunay features appeared to be strongly correlated to Gleason grade. However, these features are derived from fully connected graphs. This approach suggests that nuclei embedded within stromal and epithelial regions will be connected via these graphs and hence the graph edges will traverse the epithelial stromal interfaces and regions [22]. Hence the features extracted from Voronoi or Delaunay graphs represent the “averaged” attributes of both stromal and epithelial architecture, and thus overlooks the local contributions of stroma and epithelium independently within the graphs. Alternatively, analysis of local subgraphs [20,21,23,24], which unlike global graphs (e.g. Voronoi and Delaunay) that aim to capture a global architectural signature for the tumor, can allow for quantification of local interactions within flexible localized neighborhoods. Gunduz et al. [23] noted a natural clustering of cells and utilized cell graphs to model gliomas and differentiate cancerous, healthy, and non-neoplastic inflamed tissue. Demir et al. [25] and others [20,23,24] have developed a set of graph features to quantify the local cell-graphs. Bilgin et al. [20] similarly extracted features from different types of local cell graphs for classification of breast tissue. Features were extracted from simple, probabilistic, and hierarchical cell-graphs, as well as a hybrid combination of simple and hierarchical approaches. Similarly, Ali et al [21], utilized attributes of probabilistic cell-cluster graphs for differentiating oropharyngeal cancers. These subgraphs offer the advantage of being able to explicitly and independently model spatial architecture of nuclei and glands within the epithelial and stromal regions.

In this paper, we describe a new QH methodology which aims to utilize the directionality of glands and associate disorder in gland orientations to predict the degree of malignancy and subsequently the risk of post-surgical biochemical recurrence in CaP patients. The hypothesis is that normal benign glands align themselves with respect to the fibromuscular stroma, and thus display a coherent directionality. Malignant prostate glands, however, lose their capability to orient themselves and display no preferred directionality. Additionally with increasing degree of malignancy and disease aggressiveness, the coherence of gland orientations within localized regions is completely disrupted. In other words, the entropy (which captures disorder) in gland orientations tends to increase as a function of malignancy.

The CGA features aim to capture the directional information in localized gland networks in excised histopathology sections to characterize differences in gland orientation between (a) malignant and benign regions and (b) CaP patients who do and do not experience biochemical recurrence following RP. The CGA methodology comprises the following main steps.

For CGAs, a segmentation algorithm is first employed to individually segment gland boundaries from digitized pathology sections. To each gland, we ascribe an angle that reflects the dominant orientation of the gland based off the major axis as shown in Figure 1(a). A subgraph is then constructed to link together glands proximal to each other into a gland network as illustrated in Figure 1(b). The subgraphs, unlike the graphs for Voronoi, Delaunay and minimum spanning trees that have been previously used to characterize global glandular architecture [26], allows for characterization of local gland arrangement. Use of local subgraphs prevent graph edges from traversing heterogeneous tissue regions such as stroma and epithelium. The co-occurrence matrix, previously used to characterize image intensity textures, is used to capture second-order statistics of gland

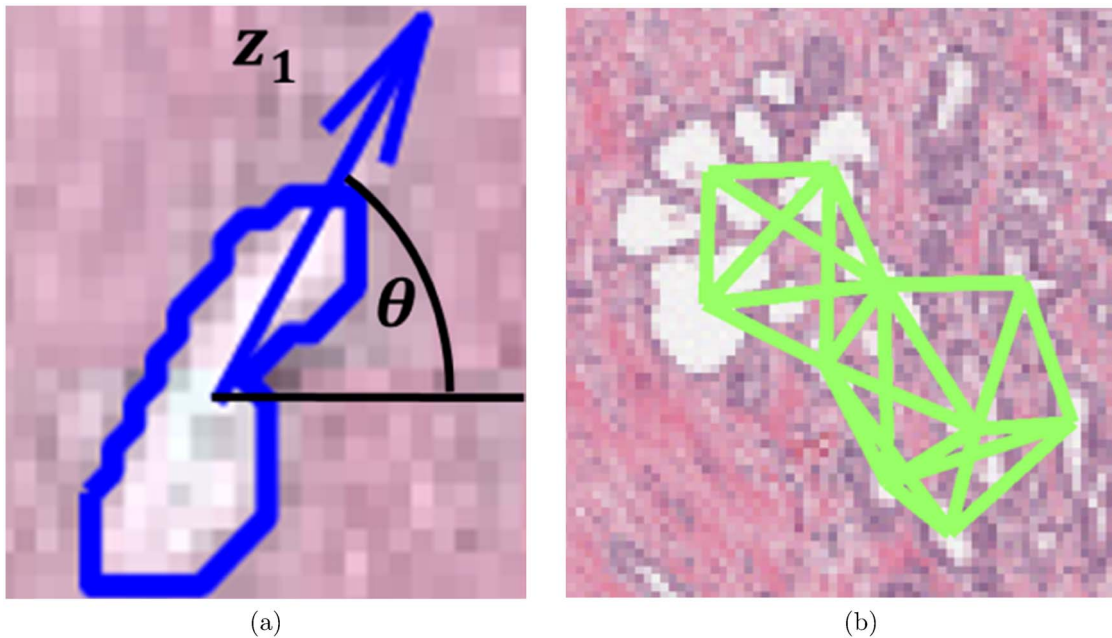


Figure 1. Gland characteristics of interest for calculating CGA features. (a) For each gland, the angle between the major axis of the gland (z_1) and the x-axis is calculated. (b) Subgraphs connect the centroids of neighboring glands into locally connected gland networks. doi:10.1371/journal.pone.0097954.g001

orientations within each gland network in the image. Hence each co-occurrence matrix captures the frequency with which orientations of two glands proximal to each other co-occur. Co-occurrence features such as entropy are extracted from the co-occurrence matrix associated with each gland network to capture the degree to which proximal gland orientations are similar or divergent to each other. Hence a neighborhood with a high entropy value would reflect a high degree of disorder among gland orientations while a low entropy value reflects that the gland angles appear to be aligned roughly in the same direction.

Given that we expect to see glandular angle disorder in (a) malignant versus benign regions and (b) biochemical recurrence cases versus non-recurrence cases, second-order statistical angular features like entropy represent a novel, reproducible, and interpretable way to characterize disease appearance on histopathology. Unlike first order statistics of angles, the co-occurring gland angular features are able to implicitly capture the cyclical properties of gland orientation. The use of local subgraphs generated by a probabilistic decaying function help define local gland networks within which the CGA features can be extracted and analyzed.

In this work, we demonstrate the utility of CGA features for the following classification tasks: 1) differentiating cancerous and non-cancerous prostate tissue regions, and 2) distinguishing CaP patients with and without biochemical recurrence following radical prostatectomy.

Figure 2 shows two representative studies: a biochemical recurrence (BCR) and a non-biochemical recurrence (NR) case. For the BCR case, we can see greater disorder in the gland orientation illustrated via the vector plot in Figure 2(f). The angle-based colormap for BCR characterizes the disorder in BCR cases, as evidenced by the a large spectrum of colors, each color representing a different orientation. Conversely, for the NR case, (Figure 2(n)), the colormap shows a smaller range of colors, suggesting less variance in the gland directionality. The gland directionality differences are also reflected in Figures 2(d), (l) via

the angular co-occurrence matrix. The brightness of the off-diagonal elements of the matrix reflect greater co-occurrences of differentially oriented gland angles for the BCR case (Figure 2(d)) compared to the NR case (Figure 2(l)). These differences in the angular co-occurrence matrices are detected by the second order statistics, as Figures 2(h), (p) illustrate different color patterns based on the value of the statistics in each subgraph.

It can also be observed in Figures 2(c), (k) that subgraphs capture local gland neighborhoods. Since the subgraphs are localized and limited to the epithelial regions alone, the contributions from within the stromal regions are minimized. The CGAs therefore provide a compact, interpretable and quantitative representation of gland architecture and prostate cancer morphology which can be employed to distinguish (a) cancer from benign regions and (b) BCR from NR cases.

The remainder of this paper is structured as follows. We first introduce the theory and methodology for **CGAs**. **Materials and Methods** outlines the process of obtaining the study cohorts and provides details for workflow and comparative methodologies used in this study. **Experimental Results and Discussion** provides specific instances in which we test our CGA methodology. Lastly, **Concluding Remarks** discusses our overall contributions and future work.

Quantitative Histomorphometry via Co-occurring Gland Angularity (CGA)

Notation

We define an image scene as $\mathcal{I} = (c, f)$, where the image scene \mathcal{I} is described by a spatial grid C of locations $c \in C$, each of which are associated with a unique intensity value $f(c)$. For intensity images $f(c) \in \mathbb{R}^1$ and for color images $f(c) \in \mathbb{R}^3$. We define a sub-region, $\mathcal{R} \in C$, within the scene, where a subgraph $\mathcal{G}(\mathcal{R})$ can be defined.

Each \mathcal{R} is comprised of a number of glands γ_i , which are represented as nodes, $\gamma_i \in \mathcal{R}$, where $i \in \{1, \dots, n_g\}$, where n_g is the

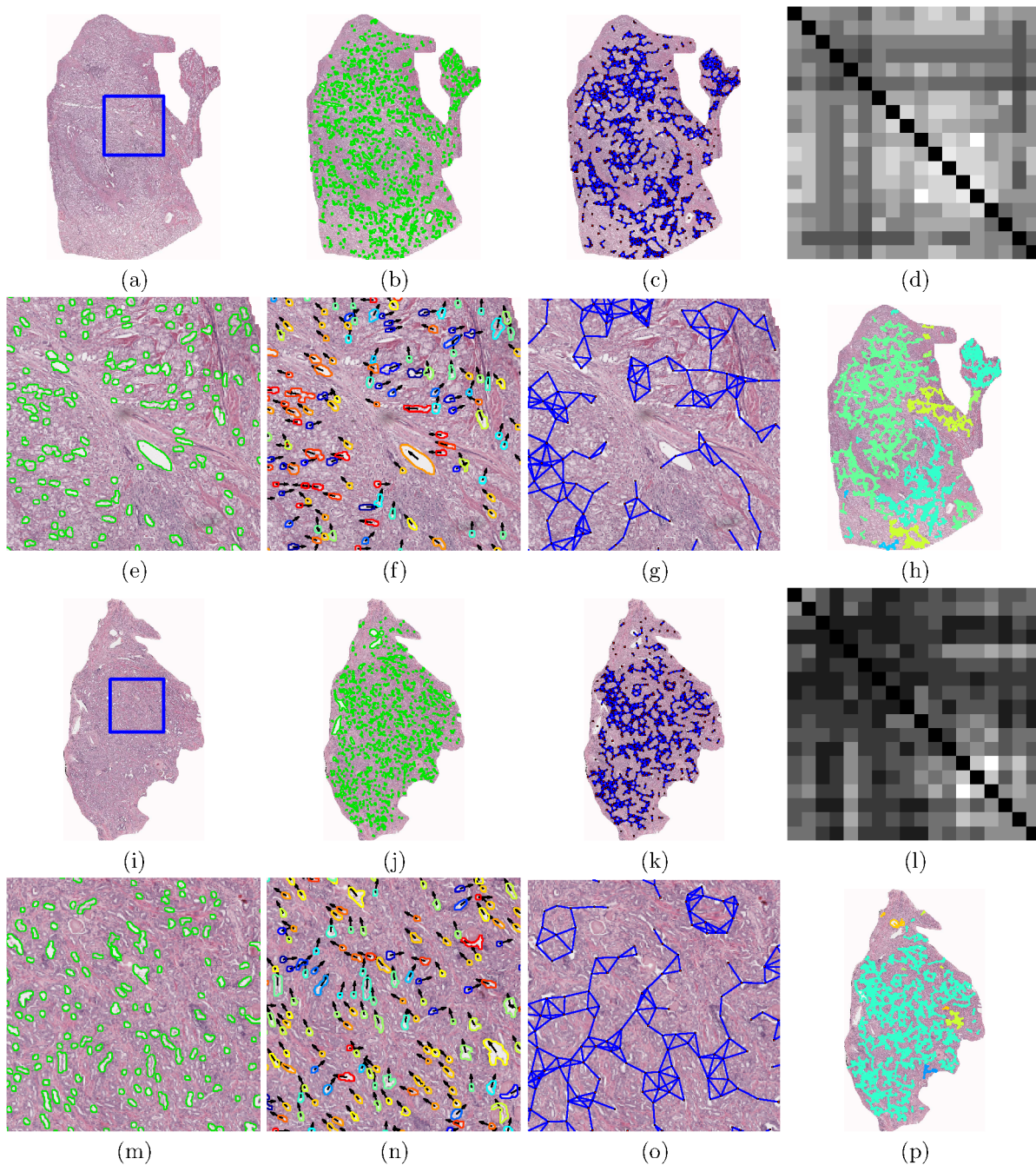


Figure 2. Annotated histological CaP regions ((a) and (i)) pertaining to a BCR (a)–(h) and a NR (i)–(p) case study, respectively. (b), (j) Automated gland segmentation of gland boundaries. (c), (k) Subgraphs showing connections between neighboring glands. An enlarged view of the boxed region in (a) and (i) respectively, illustrates (e), (m) segmented glands, (f), (n) gland angles, and (g), (o) gland network subgraphs. (f), (n) Arrows denote the directionality of each gland. Boundary colors (blue to red) correspond to angles $\theta \in [0^\circ, 180^\circ]$. (g), (o) Localized gland networks define the region of each angular co-occurrence matrix. (d), (l) Summed angular co-occurrence matrices denote the frequency with which two glands of two directionalities co-occur across all neighborhoods (white elements reflect greater co-occurrence). Diagonal co-occurrence values have been omitted to provide better contrast compared to the off-diagonal components. (h), (p) Colormap of the gland subgraphs correspond to the intensity average in each neighborhood.

doi:10.1371/journal.pone.0097954.g002

number of glands in \mathcal{R} . We can also define $B(\gamma_i) \subset \mathcal{R}$ as the set of boundary points associated with gland $\gamma_i \in \mathcal{R}$.

Hence we can formally define $\mathcal{G}(\mathcal{R}) = \{V, E\}$ where V represents the set of glands $\gamma_i \in \mathcal{R}$ and E is the set of edges that connect γ_i to other adjacent glands within \mathcal{R} . Each $\mathcal{G}(\mathcal{R})$ can then be represented via an attribute vector of CGA features $\mathbf{F}(\mathcal{R})$. A

classifier Φ can then be trained to identify any \mathcal{R} as belonging to one of two classes $\{+1, -1\}$. In this work, the classifier Φ will be trained to distinguish each \mathcal{R} as (a) malignant or benign or (b) BCR or not.

Table 1. Description of 13 CGA features.

CGA Feature (Θ)	Description	Additional Notation	
Θ_1	Contrast energy	$\sum_n \{n^2 \mathcal{P}(n)\}$	$\mathcal{P}(n) = \sum_{a,b, a-b =n} \mathcal{M}(a,b)$
Θ_2	Contrast inverse moment	$\sum_{a,b} \left\{ \frac{1}{1+(a-b)^2} \mathcal{M}(a,b) \right\}$	$\mathcal{Q}(k) = \sum_{a,b,a+b=k} \mathcal{M}(a,b)$
Θ_3	Contrast average	$\sum_{a,b, a-b =n} \{ a-b ^2 \mathcal{P}(n)\}$	$p_x = \sum_b \mathcal{M}(a,b)$
Θ_4	Contrast variance	$\sum_{a,b, a-b =n} \{ a-b - \Theta_3\}^2 \mathcal{P}(n)$	$p_y = \sum_a \mathcal{M}(a,b)$
Θ_5	Contrast entropy	$\sum_n \{-\mathcal{P}(n) \log(\mathcal{P}(n))\}$	$HX = -\sum_{a,b} \{p_x \log(p_x)\}$
Θ_6	Intensity average	$\sum_k \{k \mathcal{Q}(k)\}$	$HY = -\sum_{a,b} \{p_y \log(p_y)\}$
Θ_7	Intensity variance	$\sum_k \{(k - \Theta_6)^2 \mathcal{Q}(k)\}$	$HXY =$
Θ_8	Intensity entropy	$\sum_k \{-\mathcal{Q}(k) \log(\mathcal{Q}(k))\}$	$-\sum_{a,b} \{\mathcal{M}(a,b) \log(\mathcal{M}(a,b))\}$
Θ_9	Entropy	$\sum_{a,b} \{-\mathcal{M}(a,b) \log(\mathcal{M}(a,b))\}$	$HXY1 =$
Θ_{10}	Energy	$\sum_{a,b} \{\mathcal{M}(a,b)^2\}$	$-\sum_{a,b} \{\mathcal{M}(a,b) \log(p_x(a)p_y(b))\}$
Θ_{11}	Correlation	$\sum_{a,b} \left\{ \frac{(p_x - \mu_a)(p_y - \mu_b) \mathcal{M}(a,b)}{\sigma_a \sigma_b} \right\}$	$HXY2 =$
Θ_{12}	Information measure 1	$\frac{HXY - HXY1}{\max(HX, HY)}$	$-\sum_{a,b} \{p_x(a)p_y(b) \log(p_x(a)p_y(b))\}$
Θ_{13}	Information measure 2	$(1 - e^{-2(HXY2 - HXY)})^{1/2}$	

doi:10.1371/journal.pone.0097954.t001

Calculating Gland Angles

To determine the directionality for each gland $\gamma_i \in \mathcal{R}$, $i \in \{1, 2, \dots, n_g\}$, we perform principal component analysis [27] on a set of boundary points $B(\gamma_i)$ to obtain the principal components $Z^i = [z_1^i, z_2^i]$. The first principal component z_1^i describes the directionality of gland γ_i in the form of the major axis, along which the greatest variance occurs within $B(\gamma_i)$. The principal axis z_1^i is expressed as vector, of a single directionality, and is defined as

$$z_1^i = \langle z_1^{i,x}, z_1^{i,y} \rangle, \tag{1}$$

where $z_1^{i,x}$ represents the direction in the x -direction, and $z_1^{i,y}$ represents the direction in the y -direction. z_1^i is subsequently converted to an angle $\bar{\theta}(\gamma_i) \in [0^\circ, 180^\circ]$ calculated counterclockwise from the vector $\langle 1, 0 \rangle$ by

$$\bar{\theta}(\gamma_i) = \frac{180^\circ}{\pi} \arctan\left(\frac{z_1^{i,y}}{z_1^{i,x}}\right). \tag{2}$$

A depiction of the process for estimating gland orientations is shown in Figure 1(a).

Defining Local Subgraphs on Glands

Pairwise spatial relationships between glands are defined via sparsified graphs. For the subgraph $\mathcal{G}(\mathcal{R})$ defined on region \mathcal{R} , the individual edges can be defined between all pairs of $\{\gamma_i, \gamma_j\} \in \mathcal{R}$ via a probabilistic decaying function described in Gunduz et al. [20,23].

$$E = \{(i, j) : r < \mathbf{D}(i, j)^{-\alpha}, \forall \gamma_i, \gamma_j \in V\}, \tag{3}$$

where $\mathbf{D}(i, j)$ represents the Euclidean distance between γ_i and γ_j . $\alpha \geq 0$ controls the density of the graph, where α approaching 0 represents a high probability of connecting nodes while α

approaching ∞ represents a low probability. $r \in [0, 1]$ is an empirically determined edge threshold. An example of a resulting glandular subgraph network is shown in Figure 1(b).

Gland Angularity Co-occurrence Matrices

The objects of interest for calculating CGA features are given by a discretization of the angles $\bar{\theta}(\gamma_i)$, such that $\theta(\gamma_i) = \omega \times \text{ceil}\left(\frac{\bar{\theta}}{\omega}\right)$, where ω is an empirically derived discretization factor. Larger ω provide less specificity for counting co-occurring gland angles and smaller ω may not express co-occurring angles within the individual neighborhoods. The optimal ω was chosen based off a 3-fold randomized cross-validation of parameters $\omega \in \{5, \dots, 45\}$. ω was set to be 10 in this work, allowing for angles to be discretized at every 10 degrees.

Neighbors defined by the local subgraphs \mathcal{G} , allow us to define neighborhoods for each γ_i . For each $\gamma_i \in V_u$, we define a neighborhood \mathcal{N}_i , to include all $\gamma_j \in V$ where a path between γ_i and γ_j , $i, j \in \{1, 2, \dots, n_g\}$ exists via E in the graph \mathcal{G} .

An $N \times N$ angular co-occurrence matrix \mathcal{M} subsequently captures gland angle pairs, $\theta(\gamma_u)$ and $\theta(\gamma_v)$, where $u, v \in \{1, \dots, U\}$ and U is the number of glands in \mathcal{N}_i , which co-occur within each neighborhood \mathcal{N}_i . This can therefore be expressed in the following way.

$$\mathcal{M}_{\mathcal{N}_i}(a, b) = \sum_{\gamma_i: \gamma_j} \sum_{a, b=1}^N \begin{cases} 1, & \text{if } \theta(\gamma_i) = a \text{ and } \theta(\gamma_j) = b \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $N = \frac{180}{\omega}$, the number of discrete angular bins. An example of an angular co-occurrence matrix is shown in Figures 2(d) and (l).

Second Order Gland Angle Statistics

We subsequently extract second order statistical features Θ (Contrast energy, Contrast inverse moment, Contrast average,

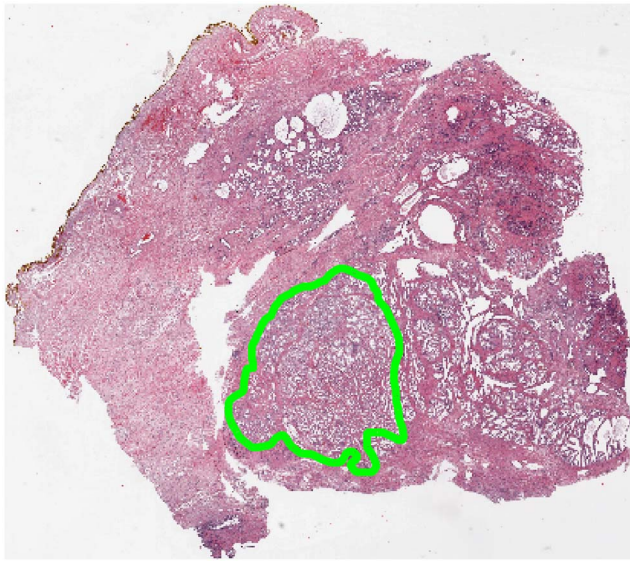


Figure 3. Workflow for building a CGA-based classifier. (a) Gland segmentation is performed on a region of interest. CGA methodology (highlighted within the dashed lines) leverages the gland segmentation to compute CGA features. (b) Angle calculation and (c) Subgraph computation is performed on the segmented image. (d) Angular co-occurrence matrix aggregates co-occurring gland angles within localized gland networks. (e) Mean, standard deviation and range of second order statistics (shown via differentially colored gland networks) create a set of CGA features for the region. (f) A CGA-based classifier can then be built using the features obtained from (e) to distinguish the categories of interest (either cancer versus benign regions or BCR versus NR).

doi:10.1371/journal.pone.0097954.g003

Contrast variance, Contrast entropy, Intensity average, Intensity variance, Intensity entropy, Entropy, Energy, Correlation, and two measures of information) from each angular co-occurrence

matrix $\mathcal{M}_{\mathcal{N}_i}(a,b)$. Formulations of these second order features Θ are described in Table 1. Figure 2 illustrates the visualization of the mean intensity measure of each \mathcal{N}_i on a digitized histopathology image.

Materials and Methods

Ethics Statement

Patients included in the study were obtained from independent sources. Cohort A was collected by Drs. Tomaszewski and Feldman obtained from IRB study “Analysis of Genetic Changes in Genitourinary Cancers” Protocol #707863. Cohort B was part the Score prostate project, run by Dr. Rebbeck, and approved by IRB study UPCC 13808 “Molecular epidemiology of prostate cancer” Protocol #36142. All IRB approval was obtained from the University of Pennsylvania, where the patient data was collected. Written consent was obtained for all patients for long term follow up. De-identified digital pathology samples and biochemical recurrence data used for this study was derived from data collected under informed consent. Since de-identified data was used, IRB consent was not required.

Data Acquisition and Data Description

The datasets (obtained from the Hospital at the University of Pennsylvania) were comprised of 40 CaP patients who had undergone RP treatment. These studies were selected from a much larger cohort of over 3000 cases archived at the Hospital at the University of Pennsylvania. The cases were chosen to have an equal split of cases with BCR and NR following RP. Additionally, the search was limited to just Gleason scores 6–8 and pathologic stage pT2 and pT3.

For all CaP patients, following RP, the excised prostate was sectioned, stained with hematoxylin and eosin (H&E), and digitized at a resolution of $0.5 \mu\text{m}$ per pixel or 20x magnification using an Aperio® whole slide scanner. For each digitized image, CaP regions were annotated by a pathologist, as shown in Figure 3. 56 cancer regions were annotated across 40 patients, 28 from BCR

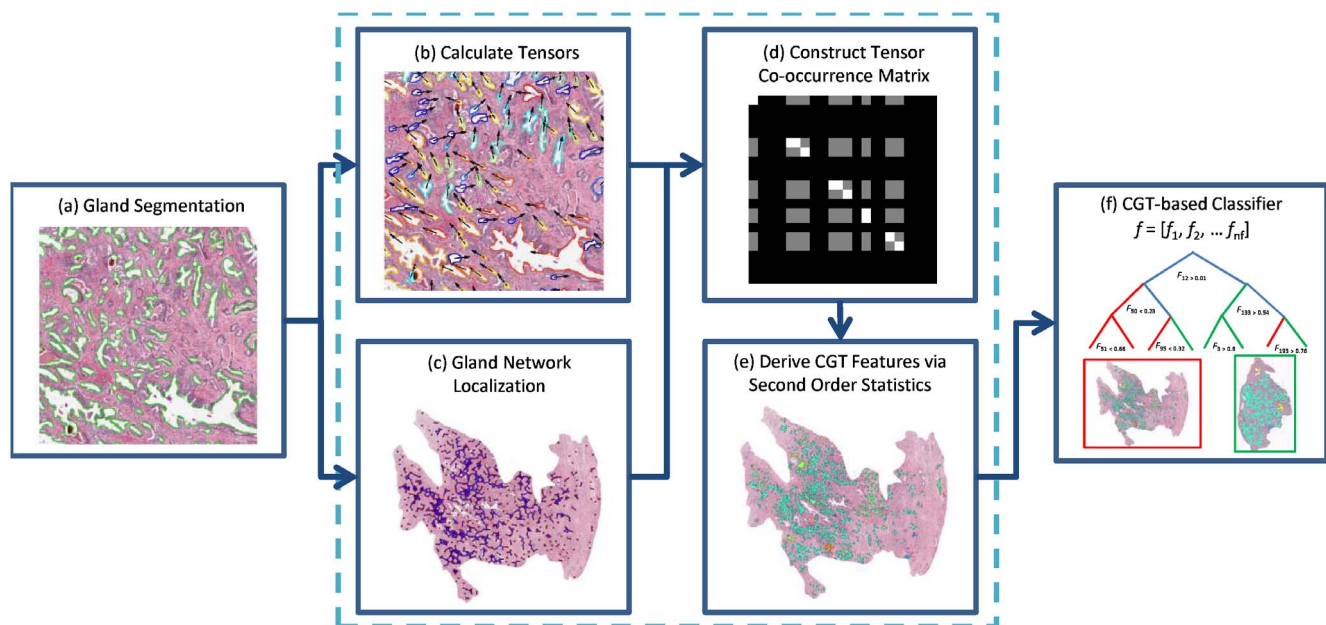


Figure 4. Annotation of a region of interest (shown in green) on prostate histopathology is performed by a pathologist. In this study, quantitative histomorphometric analysis is performed only in these regions. doi:10.1371/journal.pone.0097954.g004

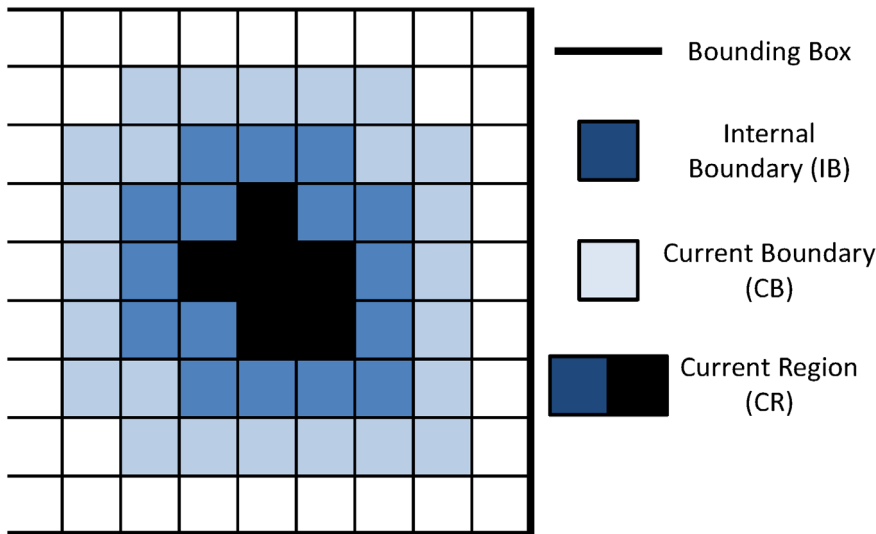


Figure 5. Schematic for region growing.
doi:10.1371/journal.pone.0097954.g005

patients and 28 from NR patients. Since no men without documented and biopsy confirmed prostate cancer undergo radical prostatectomy, there were no regions in this study which did not originate from a patient without cancer. Instead, 24 control regions were culled out from non-cancerous regions of the excised prostate for these patients.

CGA Extraction Workflow

As previously described in **Notation**, each region \mathcal{R} is characterized by a CGA features vector \mathbf{F} . The \mathbf{F} is used to train a machine learning classifier Φ to distinguish between (a) cancerous from benign regions and (b) BCR from NR patients. The procedure for extracting \mathbf{F} and training Φ is described below and summarized by Figure 4.

Identification of glandular boundaries. The detection and segmentation of gland boundaries is limited to only those regions manually annotated by the pathologist on the digitized histopathology sections. An automatic region-growing based prostate gland segmentation algorithm [13] is used to detect and segment glandular boundaries on the histological image as illustrated in Figure 5. Monaco et al. [13] previously was able to accurately identify prostate cancer regions at 93% accuracy via the gland segmentation procedure described below. Segmentation is performed using the luminance channel in CIELAB color space. In

the luminance channel, glands appear as contiguous, high intensity pixel regions bordered by sharp edges as boundaries. To identify glands, the luminance image is convolved with a Gaussian kernel at multiple scales $\sigma_g \in \{0.025, 0.05, 0.1, 0.2\}$ mm to account for multiple gland sizes. The peaks (maxima) of the resulting smoothed luminance images are used as seeds for a region growing procedure briefly outlined below.

1. A $12\sigma_g \times 12\sigma_g$ bounding box is initialized around each initial seed pixel, which represents the current region (CR), with 8-connected pixels surrounding CR, denoted as the current boundary (CB).
2. Next, the pixel in CB with the highest intensity is removed from CB and incorporated into CR. The 8 surrounding pixels of this new CR pixel, which are not already in CR, are incorporated into CB.
3. The boundary strength is identified at each iteration as shown in Figure 5. We define the internal boundary (IB) as all CR pixels adjacent to CB. Boundary strength is defined as the mean intensity of the pixels in IB minus the mean intensity of the pixels in CB.
4. Steps 2 and 3 are repeated until the algorithm attempts to add a pixel outside the bounding box.

Table 2. Overview of clinical datasets employed in this study.

	Clinical Variables	Cohort A (20)	Cohort B (20)	Combined Cohort (40)
Pathological Gleason Score	3+3	4 (20%)	1 (5%)	5 (12.5%)
	3+4	7 (35%)	17 (85%)	24 (60%)
	4+3	7 (35%)	2 (10%)	9 (22.5%)
	3+5	1 (5%)	- (-)	1 (2.5%)
	4+4	1 (5%)	- (-)	1 (2.5%)
Pathologic Stage	pT2	8 (40%)	12 (60%)	20 (50%)
	pT3a	9 (45%)	6 (30%)	15 (37.5%)
	pT3b	3 (15%)	2 (10%)	5 (12.5%)

doi:10.1371/journal.pone.0097954.t002

Table 3. Summary of Quantitative Histomorphometric (QH) features to compare against CGA features as well as the number of features used to characterize each feature type.

Feature Type (QH)	Description	#
Gland Morphology	Area Ratio, Distance Ratio, Standard Deviation of Distance, Variance of Distance, Distance Ratio, Perimeter Ratio, Smoothness, Invariant Moment 1–7, Fractal Dimension, Fourier Descriptor 1–10 (Mean, Std. Dev, Median, Min/Max of each)	100
Voronoi Diagram	Polygon area, perimeter, chord length: mean, std. dev., min/max ratio, disorder	12
Delaunay Triangulation	Triangle side length, area: mean, std. dev., min/max ratio, disorder	8
Minimum Spanning Tree	Edge length: mean, std. dev., min/max ratio, disorder	4
Glandular Density	Density of glands, distance to nearest gland	24
Co-occurrence Texture	Contrast energy, Contrast inverse moment, Contrast average, Contrast variance, Contrast entropy, Intensity average, Intensity variance, Intensity entropy, Entropy, Energy, Correlation, two measures of information: mean, std. dev.	26

doi:10.1371/journal.pone.0097954.t003

5. The optimal region is defined as region CR at the iteration where maximum boundary strength was achieved.

Overlapping regions are subsequently resolved by removing the region with the lowest boundary strength. An example of our results can be seen in Figure 4(b).

CGA feature extraction. Based on the gland segmentation detailed in the section, **Identification of Glandular Boundaries**, CGA features are calculated as described in the previous section, **Quantitative Histomorphometry via Co-occurring Gland Angularity (CGA)**. The optimal parameters were chosen based off a 3-fold randomized cross-validation procedure for $\omega \in \{5, 10, \dots, 45\}$ and $\alpha, r \in \{0.5, \dots, 1\}$. The best combination was found to be $\omega = 10$, $\alpha = 0.5$ and $r = 0.2$, which was used to build the angle co-occurrence matrix for all cases in these experiments. Mean, standard deviation, and range of the CGA features described in the section **Second Order Gland Angle Statistics** are calculated. This yields **F** which is a set of 39 CGA features.

Building a CGA-based classifier. For each classification task (see **Experimental Results and Discussion**), a training set of positive and negative categories were compiled from the set of 40 cases (see Table 2). The training set was used to train a classifier Φ in conjunction with **F** to distinguish between the categories of interest. The trained classifier Φ was then used to

assign each region \mathcal{R} or study \mathcal{I} into classes $\{+1, -1\}$ based on the classification tasks (a) distinguishing cancer from benign or (b) BCR from NR patients. 3-fold randomized cross-validation was used to train and evaluate classifier robustness. This involved randomly splitting the entire dataset into 3 equally sized sets with 2 subsets used for Φ training and 1 subset used for independent evaluation. This procedure was repeated 100 times. In all our experiments a random forest classifier (a bootstrapped aggregation of multiple decision tree classifiers) was used. We refer the reader to [28] for additional details on the random forest classifier.

Comparative Methodologies

In order to compare the performance of the CGA features for the different classification tasks described in **Experimental Results and Discussion**, we explicitly modeled and evaluated a number of other state of the art (a) QH features and (b) nomograms, described below.

Quantitative histomorphometric attributes. *Gland Morphology (M)*: Morphological descriptors [16] are extracted from the segmented glandular boundaries obtained in textbfIdentification of Glandular Boundaries Statistics such as the area ratio, perimeter ratio, and distance ratio are derived from the gland boundary information and the mean, standard deviation, median, and the ratio between the minimum and maximum values are calculated across all glands [16]. These features are summarized in Table 3.

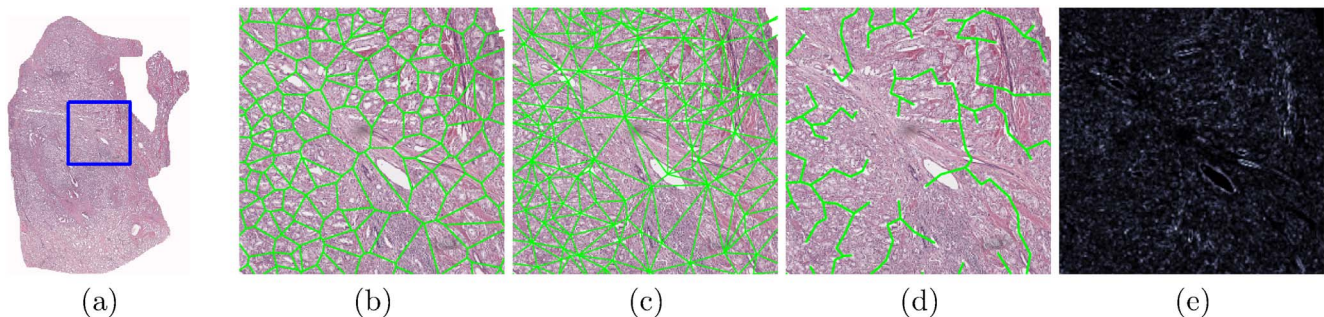


Figure 6. Examples of quantitative histomorphometric features for comparing against CGA features. QH features are extracted from (a) an annotated region on a digitized prostate histology slide following radical prostatectomy. Graphs for (b) Voronoi, (c) Delaunay, and (d) Minimum Spanning Trees as well as (e) a texture image feature are shown from the area denoted by a blue box in (a).

doi:10.1371/journal.pone.0097954.g006

Table 4. Mean and Standard Deviation of (a) ϕ^{Acc} and (b) ϕ^{AUC} for QH features in distinguishing cancer from non-cancer regions over 100 runs of randomized 3-fold cross validation with a random forest classifier.

	Gland Morphology	Voronoi	Delaunay	Minimum Spanning Tree	Gland Density	Texture	CGA
ϕ^{Acc}	96.80±1.21%	90.14±2.28%	92.04±1.33%	94.23±1.72%	96.26±1.45%	96.08±0.87%	99.06±0.76%
p -value	3.2183e-28	2.1923e-35	1.0692e-35	7.0137e-35	8.3012e-31	4.6573e-35	-
ϕ^{AUC}	0.9816±0.0087	0.9126±0.0239	0.9252±0.0192	0.9488±0.0195	0.9629±0.0152	0.9459±0.0058	0.9951±0.0077
p -value	7.1411e-23	2.2052e-34	2.2058e-34	3.9028e-34	7.4428e-31	2.1993e-34	-

Associated Wilcoxon Rank Sum Test p -values for ϕ^{Acc} and ϕ^{AUC} of the QH and CGA features.
doi:10.1371/journal.pone.0097954.t004

Table 5. Mean and Standard Deviation of (a) ϕ^{Acc} and (b) ϕ^{AUC} for QH features in distinguishing BCR from non-recurrence patients over 100 runs of randomized 3-fold cross validation with a random forest classifier.

	Gland Morphology	Voronoi	Delaunay	Minimum Spanning Tree	Gland Density	Texture	CGA
ϕ^{Acc}	59.20±4.07%	64.47±4.72%	64.82±4.66%	64.62±3.95%	64.60±3.60%	68.80±4.15%	73.18±5.02%
p -value	5.418e-10	1.2562e-08	1.1547e-10	1.7115e-11	8.4514e-10	6.9014e-28	-
ϕ^{AUC}	0.5817±0.0567	0.5600±0.0472	0.5657±0.0414	0.5594±0.0430	0.5550±0.0429	0.6017±0.0568	0.6693±0.0711
p -value	8.4424e-08	3.6478e-06	1.453e-07	4.0656e-09	4.9759e-11	0.012387	-

Associated Wilcoxon Rank Sum Test p -values for ϕ^{Acc} and ϕ^{AUC} of the QH and CGA features.
doi:10.1371/journal.pone.0097954.t005

Table 6. Mean and Standard Deviation of (a) ϕ^{Acc} and (b) ϕ^{AUC} for predicting BCR over 100 runs of randomized 3-fold cross validation via Random Forest classifiers.

CaP Predictor	Kattan	Stephenson	CAPRA	MS-KCC	CGA
ϕ^{Acc}	58.30 ± 4.83%	54.75 ± 4.34%	52.90 ± 3.19%	59.45 ± 5.41%	85.75 ± 4.89%
<i>p</i> -value	4.2558e-35	3.3989e-35	1.3793e-35	5.2764e-35	-
ϕ^{AUC}	0.6246 ± 0.0929	0.7199 ± 0.0935	0.6579 ± 0.0906	0.6182 ± 0.0764	0.7959 ± 0.0591
<i>p</i> -value	5.1173e-25	6.0125e-10	2.1506e-21	7.2366e-29	-

Associated Wilcoxon Rank Sum Test *p*-values for ϕ^{Acc} and ϕ^{AUC} of each risk assessment tool compared to CGA features for predicting BCR are shown.
doi:10.1371/journal.pone.0097954.t006

Voronoi Diagram (V): Voronoi diagrams divide each region \mathcal{R} into non-overlapping polygons, each associated with a gland γ_i , where each edge bisects two neighboring gland centroids γ_i and γ_j , where $i, j \in \{1, 2, \dots, n_g\}$. An example of the Voronoi Diagram constructed on top of a digitized pathology image with gland centers serving as nodes/vertices is shown in Figure 6(b). Statistics (See Table 3) such as the area, perimeter and chord length are recorded for each polygon and the average, standard deviation, disorder, and the ratio between the minimum and maximum values are calculated across all polygons in the image.

Delaunay Triangulation (D): Delaunay triangulation divides the image \mathcal{R}_i into triangles whose edges connect the gland centroids γ_i and γ_j with γ_i and γ_j serving as the graph vertices. An example of the Delaunay triangulation graph is shown in Figure 6(c). Delaunay triangulation is related to the Voronoi diagram in that for each polygon in the Voronoi diagram, there is an accompanying edge which connects its gland centroid γ_i with γ_j of an adjacent polygon. Edge length and area are computed for each triangle and the mean, standard deviation, minimum to maximum ratio, and disorder statistics are calculated across all triangles (see Table 3).

Minimum Spanning Tree (MST): MSTs are another graph representation where all gland centroids γ_i are connected with a minimum total edge weight defined as $\arg\min_E \sum_{i,j} E_{ij} \times \mathbf{D}(i,j)$, where $\mathbf{D}(i,j)$ is the Euclidean distance between γ_i and γ_j . An example is shown in Figure 6(d). The average, standard deviation, disorder, and minimum to maximum ratio statistics are calculated across all edges in the graph (see Table 3).

Gland Density (GD): GD features encompass two types of features. The first set of features denotes the number of γ_j which lie within a 10, 20, 30, 40, and 50 pixel radius of each γ_i . The second set of features denoted the distance between each γ_i and its 3, 5, and 7 nearest centroids γ_j . The average, standard deviation, and disorder for each of these features are computed across all glands (see Table 3).

Co-occurrence Textures (T): Second order co-occurrence features are calculated from a symmetric co-occurrence matrix which aggregates the frequency with which two pixel intensities co-occur within a pre-determined spatial distance around each pixel $c \in \mathcal{R}$. The size of the co-occurrence matrix is determined by the

maximum possible intensity value in the image, which for 8-bit images is $2^8 = 256$. A spatial distance of 1 pixel was chosen for our experiments. For each $c \in \mathcal{R}$, the second order co-occurrence features described in [15] are computed from the co-occurrence matrix. The mean and standard deviation across all $c \in \mathcal{R}$ are used to build a single texture feature descriptor for each \mathcal{R} .

Risk assessment nomogram and scoring systems. *Kattan Nomogram (K):* The Kattan nomogram [8] was one of the earliest prediction tools to be developed for predicting biochemical failure following radical prostatectomy. Clinical predictors for the Kattan nomogram include 1) Pre-operative PSA, 2) Gleason Sum, 3) Primary Gleason score, 4) Surgical Margins, 5) Prostate Capsular Invasion, 6) Seminal Vesicle Invasion (SVI), and 7) Lymph Node Involvement. A raw score s , $0 \leq s \leq 300$ (higher score reflects higher risk of BCR) is derived from these predictors and risk for each patient is assessed in terms of a probability of being BCR free for a particular time interval following surgery.

Stephenson Nomogram (S): The Stephenson nomogram [9] was developed along with Michael Kattan to incorporate the year of surgical intervention for predicting BCR. Based on 1) year of Radical Prostatectomy, 2) surgical margins, 3) extraprostatic extension (EPE), 4) seminal vesicle invasion (SVI), 5) lymph node involvement, 6) primary gleason grade, 7) secondary gleason grade, and 8) pre-operative PSA. A raw score s , $0 \leq s \leq 240$, (higher score pertains to higher risk of BCR) is derived from these clinical features. Risk for each patient is assessed in terms of a score related to the probability of being BCR free for a particular time interval following surgery.

Cancer of the Prostate Risk Assessment (CAPRA): The University of California in San Francisco (UCSF) CAPRA test [10], is based on overall score $s \in \{0, 1, \dots, 10\}$, where 10 represents the highest risk of BCR. Clinical predictors for CAPRA include 1) Age, 2) Pre-operative PSA, 3) Primary Gleason, 4) Secondary Gleason scores, 5) Tumor Stage, and 6) Percent Positive Biopsy Cores.

Memorial Sloan-Kettering Cancer Center (MS-KCC) Nomogram (MSK): One of the most popular nomograms with contributions from Kattan and Stephenson is the MS-KCC nomogram [8,9]. The MS-KCC nomogram incorporates 1) Pre-Treatment PSA, 2) Age, 3) Primary Gleason Grade, 4) Secondary Gleason Grade, 5) Gleason Sum, 6) Year of Prostatectomy, 7) Months Free of Cancer, 8) Surgical Margins, 9) Extra Capsular Extension, 10)

Table 7. Independent validation of ϕ^{AUC} for predicting BCR from non-recurrence CaP patients.

CaP Predictor	Kattan	Stephenson	CAPRA	MS-KCC	CGA
ϕ^{AUC}	0.5500	0.5600	0.6050	0.5850	0.7600

doi:10.1371/journal.pone.0097954.t007

Table 8. Logrank test p -values for comparison of Kaplan-Meier survival curves of Cohort B stratified into BCR and NR groups by CGA features and CaP risk assessment tools.

CaP Predictor	Kattan	Stephenson	CAPRA	MS-KCC	CGA
p -value	0.1846	0.2934	0.4215	0.0596	0.0016

The most significant p -value is shown in **bold**.
doi:10.1371/journal.pone.0097954.t008

Seminal Vesicle Involvement, and 11) Lymph Node Involvement. Risk for each patient is assessed in terms of a score related to the probability of being BCR free for a particular time interval following surgery.

Classification accuracy. The predictive value of each classifier, obtained with the CGA features and comparative methodologies, is evaluated via classification accuracy ϕ^{Acc} and via the area under the receiver operating characteristic curve (AUC) ϕ^{AUC} . The Wilcoxon Rank Sum Test [29] was utilized to determine statistical significance between the classification performance of the QH features listed in Table 3 and the CGA features.

Classification Accuracy (ϕ^{Acc}) measures the ability of a classifier to correctly predict class label for \mathcal{R} or \mathcal{I} within the independent testing set and is calculated as

$$\phi^{Acc} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{5}$$

where TP , TN , FP , and FN refer to the true positive, true negative, false positive, and false negative detection rates associated with Φ respectively.

Receiver operating characteristic. ϕ^{AUC} represents an overall measure of predictive value for classifier Φ , independent

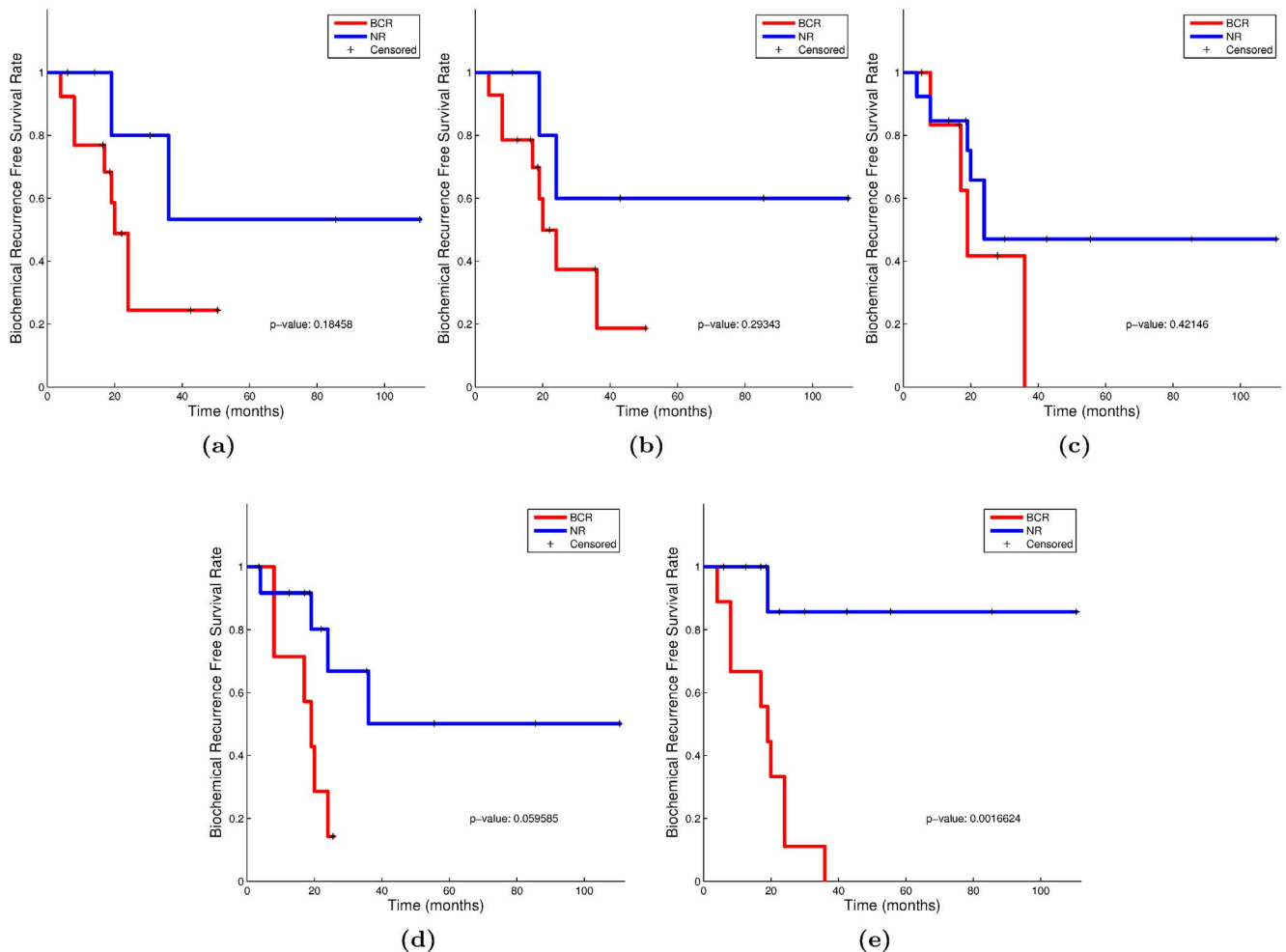


Figure 7. Comparison of Kaplan-Meier BCR-free survival curves differentiated via (a) Kattan nomogram, (b) Stephenson nomogram, (c) UCSF-CAPRA, (d) MS-KCC nomogram, and (e) CGA classifiers on an independent 20 patient cohort. Lower p -values are indicative of better predictors of BCR.
doi:10.1371/journal.pone.0097954.g007

of the decision threshold. The receiver operative characteristic (ROC) curve is constructed by computing $1 - \phi^{Spec}$ and ϕ^{Sens} at each decision threshold where ϕ^{Spec} and ϕ^{Sens} are defined as

$$\phi^{Sens} = \frac{TP}{TP + FN}, \quad (6)$$

and

$$\phi^{Spec} = \frac{TN}{TN + FP}. \quad (7)$$

The AUC represents the area under the ROC curve, where an AUC of 1 reflects perfect classification and an AUC of 0.5 suggests that the classifier is performing no better than random guessing.

Kaplan-meier analysis. Kaplan-Meier analysis [30] is used to compare the BCR-free survival time between positive and negative control groups. In this study, the two groups are determined by a predictor Φ . When plotted onto time versus BCR-free survival rate, the BCR free survival rate of the group will decrease at the time when a patient develops BCR. Thus, we expect the curve for the set of patients predicted to have BCR to drop quickly while the set of patients predicted as NR should remain BCR-free with the corresponding Kaplan-Meier curve not dropping off. The quantitative difference between the survival outcome can be determined via the logrank test [31]. The non-parametric test yields a p -value, where lower p -values denote greater significance between the survival distributions.

Experimental Results and Discussion

We compare the performance of the CGA features versus the other QH (see Table 3) and risk assessment tools (see **Risk Assessment Nomogram and Scoring Systems**) in the context of the following experiments.

1. Distinguishing cancerous versus non-cancerous regions.
2. Predicting biochemical recurrence versus non-recurrence in CaP patients following RP.
3. Comparing BCR prediction of CaP patients following RP via CGAs against QH and risk assessment tools in terms of ROC and Kaplan-Meier analysis.

Each experiment and accompanying results are described in detail below and were conducted in accordance with the CGA extraction workflow described in the previous section.

Experiment 1: Distinguishing Cancerous versus Non-cancerous Regions

80 regions were annotated by expert pathologists pertaining to 56 cancerous regions and 24 non-cancer regions. We compare the efficacy of CGA features with the QH features described in Table 3 for the purpose of differentiating cancerous regions from non-cancerous regions. For each set of CGA or QH features, a classifier was trained to distinguish between the cancerous and non-cancerous regions as previously described in **Building a CGA-based classifier**.

Mean and standard deviation for ϕ^{Acc} and ϕ^{AUC} for the different QH and CGA features are shown in Table 4. CGA shows statistically significant ($p < 0.05$) improvement in terms of ϕ^{Acc} and

ϕ^{AUC} compared to all QH features. In fact the CGA features yield a near perfect classification performance for this particular task.

Experiment 2: Identifying CaP Patients with Biochemical Recurrence following Surgery

For each of 40 patients, the largest annotated cancer region for each of the 40 patients was selected. We compare the efficacy of CGA features with other state of the art QH features (Table 3) for differentiating patients who will develop BCR from those who will not, following RP. Similar to the procedure described in Experiment 1, the CGA and QH features were used to train classifiers to distinguish the BCR and NR cases over 100 runs of 3-fold cross-validation.

As illustrated in Table 5, CGAs outperformed each of the 6 other QH features in predicting BCR in 40 CaP patients in terms of ϕ^{AUC} and ϕ^{Acc} . These results were statistically significant ($p < 0.05$). Not only do these results suggest that the CGA features were able to outperform other QH features, they also seem to suggest that in more aggressive CaP, disorder in glandular orientations appears to progressively increase.

Experiment 3: Comparing CGA Features versus Risk Assessment Tools

The data was collected from two independent sources: Cohort A from the department of Pathology and Laboratory Medicine at the University of Pennsylvania and Cohort B from the department of Clinical Epidemiology and Biostatistics at the University of Pennsylvania. Of the 40 patients, only 20 patients in Cohort B had associated clinical variables for nomogram prediction of BCR. A further breakdown of the data is summarized in Table 2. We utilize this independent data collection to perform an independent study with a larger predicted training set compared to the 3-fold cross-validation performed in the other experiments.

Similar to the CGA classifier, a nomogram based classifier is also constructed based off the corresponding risk scores, s . For the results in Table 6, training and evaluation of these classifiers was performed as described in Experiments 1 and 2 within Cohort B.

CGAs outperformed 4 state-of-the-art prostate cancer nomograms, demonstrating 85% classification accuracy compared to 59% accuracy for the MS-KCC nomogram, which had the second highest ϕ^{Acc} (Table 6). CGAs showed statistically significant improvement in ϕ^{AUC} over all nomograms. Perhaps the most significant message in the results shown in Table 6 is that the CGA features in the absence of clinical variables PSA, Gleason score, stage, etc. were still able to outperform the 4 nomograms.

To increase the size of the testing set to 20 patients, we performed a study using two independent cohorts. Analysis via the calculation of the Receiver Operating Characteristic (ROC) curve is used to determine the overall performance across all classification thresholds of each classifier Φ .

For the results in Table 7, we perform the classification on Cohort B by creating a classifier Φ trained from Cohort A. For the Random Forest classifier, each prediction is given a fuzzy decision value \hat{p} between -1 and 1 . Nomograms do not require further training beyond the original fitting done in the original study from which the nomogram was developed. Each nomogram is designed to predict BCR risk based on a score s . By setting decision thresholds at different \hat{p} and s for Φ respectively, we obtained sensitivity and specificity scores at each threshold. The area under the ROC curve (ϕ^{AUC}) is subsequently calculated for each Φ to compare their performance in predicting BCR within Cohort B.

CGA features demonstrate a clear improvement over 4 state-of-the-art nomograms, showing an AUC of 0.76 compared to AUCs

near 0.5 for the comparative nomograms. The weak performance on this independent cohort suggests the difficulty that these nomograms may have in predicting the risk of BCR for men with intermediate-risk Gleason scores and suggests that the addition of QH features could improve upon the current nomogram standards.

This training and testing procedure was repeated for Kaplan-Meier analysis of 20 patients within Cohort B. Kaplan-Meier curves demonstrate the difference in BCR free survival time associated with each risk assessment tool. The logrank test was used to determine a p -value associated with the difference between the survival curves. A lower p -value indicates greater differences in the BCR-free survival between the predicted BCR and NR groups.

In Figure 7, we show Kaplan-Meier survival curves based on the predicted BCR and NR groups of each Φ . Based on the logrank test (shown in Table 8), patients were best differentiated via the CGA features, with a p -value of 0.0016 compared to 0.0596 for the MS-KCC nomogram, which had the next lowest p -value. The CGA features represent the only feature set which show statistically significant differentiation ($p < 0.05$) in the survival outcomes of its predicted patient cohorts.

Concluding Remarks

In this paper, we present a novel set of quantitative histomorphometric features, co-occurring gland angles (CGAs), calculated on local subgraphs. CGAs represent a novel combination of subgraphs, gland angles, and angular co-occurrence matrices to quantify the local disorder in the gland angles on histopathology. For a cohort of 40 patients with Gleason scores 6–8 and treated

with radical prostatectomy, we found CGA features demonstrated a statistically significant ($p < 0.05$) improvement in classification accuracy in distinguishing (a) cancer from benign regions and (b) biochemical recurrence from non-recurrence patients following surgery compared to 6 other state of the art QH features. Furthermore, we found CGAs to outperform 4 state-of-the-art postoperative nomograms for predicting BCR in CaP patients. Kaplan-Meier analysis of the independent cohort of prostate cancer patients with intermediate-risk pathological Gleason scores (via the logrank test) demonstrated that only the CGAs showed a statistically significant ($p < 0.05$) difference in predicting the survival distributions.

We do however acknowledge that our work did have its limitations. Firstly, even though randomized cross-validation strategies were employed to remove bias in the classifier, it would be optimal to validate our classifier on an independent testing set. Additionally, we also wish to control for the clinical variables used in this study (grade, stage, PSA levels, etc.) to further remove potential noise in the study cohorts. We hope to address these limitations in future work. An immediate next step will be to evaluate the performance of the CGA features in predicting aggressive disease from needle core biopsies alone as opposed to prostatectomy specimens.

Author Contributions

Conceived and designed the experiments: GL. Performed the experiments: GL. Analyzed the data: GL. Contributed reagents/materials/analysis tools: NS MD ES TR JT. Wrote the manuscript: GL. Key contributors for methodological idea conception: RS SA AM. Contributed code for implementation of the methodology: RS SA.

References

- Trock BJ, Han M, Freedland SJ, Humphreys EB, DeWeese TL, et al. (2008) Prostate cancer-specific survival following salvage radiotherapy vs observation in men with biochemical recurrence after radical prostatectomy. *JAMA* 299: 2760–2769.
- Pinto F, Prayer-Galetti T, Gardiman M, Sacco E, Ciaccia M, et al. (2006) Clinical and pathological characteristics of patients presenting with biochemical progression after radical retropubic prostatectomy for pathologically organ-confined prostate cancer. *Urol Int* 76: 202–208.
- D'Amico AV, Chen MH, Roehl KA, Catalona WJ (2004) Preoperative psa velocity and the risk of death from prostate cancer after radical prostatectomy. *N Engl J Med* 351: 125–135.
- Epstein JI (2010) An update of the gleason grading system. *Journal Of Urology (the)* 183: 433.
- Han M, Partin AW, Zahurak M, Piantadosi S, Epstein JI, et al. (2003) Biochemical (prostate specific antigen) recurrence probability following radical prostatectomy for clinically localized prostate cancer. *J Urol* 169: 517–523.
- Stark JR, Perner S, Stampfer MJ, Sinnott JA, Finn S, et al. (2009) Gleason score and lethal prostate cancer: does 3+4 = 4+3? *J Clin Oncol* 27: 3459–3464.
- Allsbrook W Jr, Mangold KA, Johnson MH, Lane RB, Lane CG, et al. (2001) Interobserver reproducibility of gleason grading of prostatic carcinoma: general pathologist. *Hum Pathol* 32: 81–88.
- Kattan MW, Wheeler TM, Scardino PT (1999) Postoperative nomogram for disease recurrence after radical prostatectomy for prostate cancer. *Journal of Clinical Oncology* 17: 1499–1499.
- Stephenson AJ, Scardino PT, Eastham JA, Bianco FJ, Dotan ZA, et al. (2005) Postoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy. *Journal of Clinical Oncology* 23: 7005–7012.
- Cooperberg MR, Freedland SJ, Pasta DJ, Elkin EP, Presti JC, et al. (2006) Multiinstitutional validation of the ucsf cancer of the prostate risk assessment for prediction of recurrence after radical prostatectomy. *Cancer* 107: 2384–2391.
- Hinev A, Anakievski D, Kolev N, Marianovski V, Hadjiev V (2011) Validation of pre-and postoperative nomograms used to predict the pathological stage and prostate cancer recurrence after radical prostatectomy: a multi-institutional study. *Journal of BU ON: official journal of the Balkan Union of Oncology* 16: 316.
- Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, et al. (2009) Histopathological image analysis: a review. *IEEE Rev Biomed Eng* 2: 147–171.
- Monaco JP, Tomaszewski JE, Feldman MD, Hagemann I, Moradi M, et al. (2010) High-throughput detection of prostate cancer in histological sections using probabilistic pairwise markov models. *Med Image Anal* 14: 617–629.
- Jafari-Khouzani K, Soltanian-Zadeh H (2003) Multiwavelet grading of pathological images of prostate. *IEEE Trans on Biomedical Engineering* 50: 697–704.
- Haralick R, Shanmugam K, Dinstein I (1973) Textural features for image classification. *IEE Transactions on Systems, Man and Cybernetics* 3: 610–621.
- Doyle S, Feldman MD, Shih N, Tomaszewski J, Madabhushi A (2012) Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer. *BMC bioinformatics* 13: 282.
- Tabesh A, Teverovskiy M, Pang HY, Kumar VP, Verbel D, et al. (2007) Multifeature prostate cancer diagnosis and gleason grading of histological images. *IEEE Trans Med Imaging* 26: 1366–1378.
- Veltri RW, Isharwal S, Miller MC, Epstein JI, Partin AW (2010) Nuclear roundness variance predicts prostate cancer progression, metastasis, and death: A prospective evaluation with up to 25 years of follow-up after radical prostatectomy. *Prostate* 70: 1333–1339.
- Christens-Barry W, Partin A (1997) Quantitative grading of tissue and nuclei in prostate cancer for prognosis prediction. *Johns Hopkins Apl Technical Digest* 18: 226–233.
- Bilgin C, Demir C, Nagi C, Yener B (2007) Cell-graph mining for breast tissue modeling and classification. *Conf Proc IEEE Eng Med Biol Soc* 2007: 5311–5314.
- Ali S, Madabhushi A (2012) An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. *IEEE Transactions on Medical Imaging* 31: 1448–1460.
- Ali S, Veltri R, Epstein JI, Christudass C, Madabhushi A (2013) Cell cluster graph for prediction of biochemical recurrence in prostate cancer patients from tissue microarrays. In: *Proc. SPIE 8676, Medical Imaging, Digital Pathology*.
- Gunduz C, Yener B, Gultekin SH (2004) The cell graphs of cancer. *Bioinformatics* 20 Suppl 1: i145–i151.
- Gunduz-Demir C (2007) Mathematical modeling of the malignancy of cancer using graph evolution. *Math Biosci* 209: 514–527.
- Demir C, Gultekin SH, Yener B (2005) Augmented cell-graphs for automated cancer diagnosis. *Bioinformatics* 21 Suppl 2: ii7–i12.
- Lee G, Sparks R, Ali S, Madabhushi A, Feldman MD, et al. (2013) Co-occurring gland tensors in localized cluster graphs: Quantitative histomorphometry for predicting biochemical recurrence for intermediate grade prostate cancer. In:

- Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on. IEEE, 113–116.
27. Hotelling H (1933) Analysis of a complex of statistical variables into principal components. *Journal of educational psychology* 24: 417.
 28. Breiman L (2001) Random forests. *Machine learning* 45: 5–32.
 29. Fagerland MW (2012) t-tests, non-parametric tests, and large studies a paradox of statistical practice? *BMC Medical Research Methodology* 12: 78.
 30. Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. *Journal of the American statistical association* 53: 457–481.
 31. Mantel N (1966) Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* 50: 163–170.