

Journal of Biomedical Optics

SPIEDigitalLibrary.org/jbo

Analysis of second-harmonic-generation microscopy in a mouse model of ovarian carcinoma

Jennifer M. Watson
Photini F. Rice
Samuel L. Marion
Molly A. Brewer
John R. Davis
Jeffrey J. Rodriguez
Urs Utzinger
Patricia B. Hoyer
Jennifer K. Barton

Analysis of second-harmonic-generation microscopy in a mouse model of ovarian carcinoma

Jennifer M. Watson,^a Photini F. Rice,^a Samuel L. Marion,^b Molly A. Brewer,^c John R. Davis,^d Jeffrey J. Rodriguez,^e Urs Utzinger,^a Patricia B. Hoyer,^b and Jennifer K. Barton^a

^aUniversity of Arizona, Biomedical Engineering, 1657 E. Helen Street, Building 240, P.O. Box 210240, Tucson, Arizona 85721

^bUniversity of Arizona, Department of Physiology, Basic Sciences, Room 4122, 1501 N. Campbell Avenue, P.O. Box 245051, Tucson, Arizona 85724-5051

^cUniversity of Connecticut Health Center, Carole and Ray Neag Comprehensive Cancer Center, Division of Gynecologic Oncology, 263 Farmington Avenue, MC 1614, Farmington, Connecticut 06030-2875

^dUniversity of Arizona, Department of Pathology, Tucson, Arizona

^eUniversity of Arizona, Electrical/Computer Engineering, P.O. Box 210104, Tucson, Arizona

Abstract. Second-harmonic-generation (SHG) imaging of mouse ovaries *ex vivo* was used to detect collagen structure changes accompanying ovarian cancer development. Dosing with 4-vinylcyclohexene diepoxide and 7,12-dimethylbenz[a]anthracene resulted in histologically confirmed cases of normal, benign abnormality, dysplasia, and carcinoma. Parameters for each SHG image were calculated using the Fourier transform matrix and gray-level co-occurrence matrix (GLCM). Cancer versus normal and cancer versus all other diagnoses showed the greatest separation using the parameters derived from power in the highest-frequency region and GLCM energy. Mixed effects models showed that these parameters were significantly different between cancer and normal ($P < 0.008$). Images were classified with a support vector machine, using 25% of the data for training and 75% for testing. Utilizing all images with signal greater than the noise level, cancer versus not-cancer specimens were classified with 81.2% sensitivity and 80.0% specificity, and cancer versus normal specimens were classified with 77.8% sensitivity and 79.3% specificity. Utilizing only images with greater than of 75% of the field of view containing signal improved sensitivity and specificity for cancer versus normal to 81.5% and 81.1%. These results suggest that using SHG to visualize collagen structure in ovaries could help with early cancer detection. © 2012 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: 10.1117/1.JBO.17.7.076002]

Keywords: collagen; ovarian cancer; multiphoton; two-photon; imaging; 7,12-dimethylbenz[a]anthracene.

Paper 12129 received Feb. 22, 2012; revised manuscript received Apr. 25, 2012; accepted for publication May 18, 2012; published online Jul. 4, 2012.

1 Introduction

In the U.S. alone there are more than 20,000 new cases of ovarian cancer each year and more than 15,000 deaths per year. The most common type of ovarian malignancy is derived from epithelial cells and is more likely to occur in postmenopausal women. Mortality rates are high because an effective screening test does not currently exist. Only 15% of ovarian cancers are found before metastasis has occurred. If ovarian cancer is found and treated before metastasis, the 5-year survival rate is 94% (versus 28% for metastatic disease).¹ Thus, an early diagnostic test to detect premalignant changes would save many lives.

Current methods of screening consist of measuring serum levels of cancer antigen 125 (CA-125) and transvaginal ultrasound.^{2,3} Serum levels of CA-125 are often elevated in women with ovarian cancer.^{2,4-7} However, CA-125 levels are influenced by conditions other than ovarian cancer such as other cancers, lung disease, liver cirrhosis, hysterectomy, obesity, and smoking habits.⁸⁻¹² Furthermore, CA-125 is usually elevated only in patients with stage II-IV cancer, not in patients with borderline tumors or stage I ovarian cancer.¹³

Transvaginal ultrasound can be used to visualize both ovaries and evaluate the size of lesions to determine the extent of tumor

growth and metastasis.^{14,15} However, tumor morphology and vascular perfusion, as seen by ultrasound, are not enough to identify abnormalities in ovaries of normal volume or distinguish between benign and malignant tumors.^{16,17} Furthermore, CA-125 combined with ultrasound does not decrease the number of mortalities resulting from ovarian cancer.^{18,19}

Owing to the low performance of CA-125 and ultrasound in detection of early cancers, women who are at high risk may be advised to undergo a prophylactic salpingo-oophorectomy (removal of the ovaries and fallopian tubes). Whereas this procedure is highly effective at reducing cancer risk, removal of the ovaries is known to increase morbidity and mortality.²⁰⁻²⁴

Optical methods that have been investigated for detection of ovarian cancer include spectroscopy, optical coherence tomography (OCT), confocal microscopy, photoacoustic imaging (PAI), and multiphoton microscopy (MPM). Several studies have been performed utilizing these modalities. However, most comparisons are between normal and advanced cancer because women rarely present with early-stage ovarian cancer.

Reflectance and fluorescence spectroscopy can differentiate normal and neoplastic ovarian tissue with good sensitivity and specificity.²⁵⁻²⁹ Limitations of spectroscopy include shallow depth of penetration and low spatial resolution. OCT images of the ovary show details of tissue microstructure such as surface epithelium, follicles, cysts, collagen bundles, and vessels.^{30,31}

Address all correspondence to: Jennifer K. Barton, University of Arizona, Biomedical Engineering, 1657 E. Helen Street, Building 240, P.O. Box 210240, Tucson, Arizona 85721. Tel: +1 520-621-4116; Fax: +1 520-626-4824; E-mail: barton@u.arizona.edu

Furthermore, differences between normal and abnormal/neoplastic ovarian tissue are seen, such as epithelial inclusions, invaginations, and differences in attenuation.^{27,28,30,32,33} However, OCT does not have the resolution necessary to visualize dysplasia. Confocal microscopy produces subcellular-resolution images that can be used to identify cancer occurring on the surface of the ovary, but the depth of imaging is limited.^{29,34,35} PAI has the largest depth of imaging (2 to 3 cm) and, owing to differences in absorption properties, can visualize large structures such as corpora lutea, follicles, and blood vessels.^{31,36} Likewise, malignant and normal ovaries in postmenopausal women can be distinguished by their different absorption properties.³⁷ However, PAI has relatively low resolution and may be confounded by benign conditions with high vascularity or hemorrhage, or early-stage cancers without significant vascularity changes.

MPM can achieve submicron resolution, comparable to that of a confocal microscope. Owing to the decrease in scattering at longer wavelengths, MPM using near-infrared light has the ability to image hundreds of microns deeper than confocal microscopy using ultraviolet or visible light. Also, less out-of-focus fluorescence emission is generated, due to the nonlinear properties of the multiphoton process, potentially increasing depth of imaging and improving resolution. The resolution limit of MPM depends on the illumination point spread function. The depth limit of MPM depends on the pulse energy, tissue attenuation (from absorption and scattering), and ratio of collected signal to background signal.³⁸ In MPM, femtosecond pulsed laser light is focused onto the tissue with high numerical aperture optics, resulting in high instantaneous power density in a small volume of tissue. In two-photon excited fluorescence (TPEF), two photons are simultaneously absorbed by a fluorophore and then emitted as one photon at a higher frequency than the incident light. In second harmonic generation (SHG), phase matching of photons in noncentrosymmetric structures results in a scattering event in which two photons are combined into one photon at twice the frequency of the incident light. Remitted light from TPEF and SHG are separated using bandpass filters and collected with photomultiplier tubes. The laser beam is scanned throughout the tissue volume to create a 3-D image set. Because the probability of the multiphoton process is very low outside the small focal volume, very fine sectioning capacity is possible. MPM can be used to see changes in endogenous cellular fluorescence and collagen structure as a result of ovarian cancer.³⁹⁻⁴¹ SHG shows that normal ovaries have thin collagen fibers organized in a net-like structure, whereas malignant ovaries have a denser, wavy collagen structure, possibly resulting from recruitment of activated fibroblasts to the outer rim of the tumor.^{39,41-43} Furthermore, the collagen structure of normal low-risk and normal high-risk postmenopausal ovaries is slightly different.³⁹ SHG may offer a useful balance of sensitivity, resolution, and depth of imaging.

Imaging ovarian tissue *in vivo* or surgical samples *ex vivo* can provide useful indication of the difference between normal and cancerous ovaries, but it is difficult to ascertain what changes preceding cancer development can be visualized. Women usually present with advanced disease, and the etiology of ovarian cancer is poorly understood. Mouse models of ovarian cancer may provide insight into ovarian cancer development. We utilized a cancer model in mice that had undergone early ovarian failure because most ovarian cancers arise in postmenopausal women.⁴⁴ Thus, a follicle-deplete, ovary-intact animal closely approximates the natural human progression through the events

of perimenopause and the postmenopausal stage. 4-Vinylcyclohexene diepoxide (VCD) has been found to induce premature ovarian failure in mice and rats by accelerating the process of atresia in ovarian small pre-antral follicles.⁴⁵ Previous studies in mice demonstrate that VCD-induced follicle loss can cause depletion of the smallest pre-antral follicles within 15 days of daily dosing and complete ovarian failure within 46 days of the onset of dosing.⁴⁶ As a result, the mouse retains little residual ovarian tissue. The model has been developed by treating mice with VCD to induce ovarian failure and subsequently exposing the ovary to a known carcinogen, 7,12-dimethylbenz[*a*]anthracene (DMBA), to induce ovarian cancer. The VCD/DMBA model develops a variety of benign and malignant tumors.⁴⁷

Our overall goal is to develop an imaging method that can determine with certainty whether a woman's ovaries are normal or ovarian cancer is developing. With such a method, high-risk women could undergo a laparoscopic diagnostic test to determine if their ovaries are healthy to avoid, or prolong time to, oophorectomy. In this study, we utilized SHG microscopy to examine micron-scale collagen structure in normal, atypical, and cancerous mouse ovaries. By examining alterations in collagen structure, we may ultimately be able to identify the changes that precede ovarian cancer. Differences in SHG microscopy images of the diagnostic categories were examined by eye and quantified with numerical parameters relating to image frequency content and second-order gray-level statistics. Further, a classification scheme was developed using a support vector machine.

2 Methods

2.1 Animals

All experiments were performed per NIH guidelines, and protocols were approved by the University of Arizona Institutional Animal Care and Use Committee. Female B6C3F1 mice (age 28 days, Harlan, Dublin, VA) were housed in microisolators per NIH guidelines and allowed a 7-day acclimation period before initiating the experiment. Fifty-two 28-day-old mice received intraperitoneal (IP) injections of VCD, 160 mg/kg/day in sesame oil, daily for 20 days, or received sesame oil vehicle only as control. Four months after the end of IP dosing, animals received a single injection of DMBA, 50 μ g in 5 μ L sesame oil, or 5 μ L sesame oil vehicle for controls, under the bursa of the right ovary. Sterile surgical method was used to expose the ovarian bursa for subbursal injection. Prior to surgery animals were anesthetized by IP injection of 2% Avertin at 0.015 mL per gram body weight. The left ovary was not injected. Therefore, there were four experimental groups: both VCD and DMBA exposed, only VCD exposed, only DMBA exposed, and neither VCD nor DMBA exposed. Ovaries were harvested at 5 or 7 months after subbursal injection with DMBA and immediately imaged. Time from ovary excision to completion of imaging was less than 1 h.

2.2 Imaging

Imaging was performed with a single-beam multiphoton microscope (TrimScope, LaVision BioTec, Bielefeld, Germany) using a titanium:sapphire laser light source (Chameleon Ultra2, Coherent, UK) coupled to the scanner unit, with a pulse width of 120 fs in the sample. The laser intensity was adjusted with an electro-optical modulator (EOM 350-80, Conoptics, USA). Simultaneous SHG and TPEF image data were recorded

through non-descanning reverse detection using triple detector port equipped with Gallium Arsenide (H7422A-40, Hamamatsu, Hamamatsu City, Japan) and bialkali sensors (H6780-01 and H6780-20, Hamamatsu). For this study, only the SHG image data were analyzed. The excitation wavelength was set to 780 nm, and a bandpass filter FF01-377/50 (Semrock) and a dichroic mirror Di01-R405-25 × 36 (Chroma) were used to collect light from SHG. Power on the sample was set to 20 mW. Pixel dwell time was 4.61 μ s, and three-line summing was used. Images were taken at 10- μ m depth increments from the surface of the tissue to 60 to 100 μ m depth. All images had a 400- by 400- μ m field of view and contained 993 by 993 or 1021 by 1021 pixels with 14-bit grayscale resolution.

2.3 Histology and Pathological Evaluation

After imaging, ovaries were fixed in Bouin's solution for 2 to 4 h, transferred to 70% ethanol, dehydrated, embedded in paraffin blocks, and sectioned at 5 μ m thickness. Orientation was carefully maintained from explant to imaging, fixation, paraffin embedding, and sectioning, by maintaining anatomical orientation at explant and placing the ovary face up on filter paper indicating medial-lateral and superior-inferior locations. Histology sections were taken perpendicular to the area imaged, allowing a cross-sectional view of the imaged edge. Every 20th section was mounted and stained with hematoxylin and eosin. All histologic specimens were evaluated by a pathologist and a gynecologic oncologist with veterinary training. Any ovary with suspected tumor had additional sections immunostained with cytokeratin (anti-cytokeratin 18 antibody [E431-1] and rabbit polyclonal to wide-spectrum cytokeratin, Abcam Inc., Cambridge, MA), per the manufacturer's recommended protocol, to determine if the tumor was of epithelial origin. The specimens were diagnosed per pathologic findings into the following seven categories: normal, DMBA-effect, tubular adenoma, tubular adenoma with areas of focal dysplasia, granulosa cell tumor, Sertoli-Leydig cell tumor, or adenocarcinoma. Normal ovaries were those which contained only healthy tissue or changes consistent with a normal aging process. DMBA-effect was a benign abnormality, caused by DMBA exposure, characterized by epithelial cell proliferation, degenerating follicles, degenerating corpora lutea, and highly active steroidogenic cells. Tubular adenoma was a benign epithelial tumor of glandular origin characterized by cells organized in tubules. The limited number of granulosa cell and Sertoli-Leydig cell tumors seen precluded their inclusion in the image analysis. Adenocarcinoma, a malignant tumor arising from the epithelial cells of glandular tissue, is the most common form of ovarian cancer in women.

2.4 Analysis and Classification

Images were analyzed by eye and characteristic features were identified. On the basis of visual examination, it was expected that computation of spatial frequency content and standard gray-level co-occurrence matrix (GLCM) parameters might capture the variations in collagen fiber thickness and periodicity seen by eye, subsequently enabling automatic classification of images into correct diagnostic groups. All images were preprocessed by resampling to 1024 by 1024 pixels using bilinear interpolation. All image processing and analysis was performed in MATLAB (R2011a, Mathworks).⁴⁸

From the fifty-two animals included in the study, 92 specimens were imaged and 59 specimens were included in the

analysis. Twelve specimens were not imaged due to instrument or investigator error. Thirty-one of the specimens were not included in the analysis because they did not contain ovary or the ovary was entirely covered by fat and/or connective tissue in the area imaged. Two other specimens were excluded from the analysis because they were the only example from a unique diagnosis (granulosa cell tumor and Sertoli-Leydig cell tumor). Images were excluded from the analysis if the average gray level was less than 1% above the noise floor of the imaging system, if they contained artifacts from fur or other debris, or if the imaged area was not ovary, as verified by histology.

After exclusion of unusable ovaries and images, the following data were available for analysis: normal (25 ovaries, 315 images), DMBA-effect (11 ovaries, 115 images), tubular adenoma (10 ovaries, 94 images), tubular adenoma with dysplasia (9 ovaries, 54 images), and adenocarcinoma (4 ovaries, 55 images).

2.5 Fourier Analysis

The two-dimensional discrete Fourier transform was computed for each image using the standard FFT algorithm. The images contained primarily low-frequency content with some high-frequency noise. To remove noise and analyze the lower-frequency content, a frequency cutoff was determined by evaluating the fiber size by eye. The smallest collagen fiber width was approximately 6 pixels, which equates to a spatial frequency of one-sixth the maximum, so the frequency range used in the analysis was limited to the lowest sixth of spatial frequencies. This lower frequency range was divided into three equal-width circular bands: low-, middle-, and high-frequency bands. The power in each band was computed and normalized to the total power in the three bands. The DC value was excluded from the lowest-frequency region.

2.6 Gray-Level Co-Occurrence Matrix Analysis

GLCM analysis is a widely used texture analysis method developed by Haralik et al.⁴⁹ The GLCM is formed by counting the number of occurrences of a gray level adjacent to another gray level, at a specified pixel distance and direction. The result is a matrix with rows and columns representing gray levels and elements containing the probability of the gray-level co-occurrence. A separate matrix can be generated for each pixel separation and each direction. Symmetric GLCMs were computed using 64 gray-levels with the gray-level limits being the minimum and maximum gray-levels in the image—that is, gray-level values in each image were linearly scaled such that the highest gray-level in the image became 64 and the lowest gray-level in the image became 1. For a 1024- by 1024-pixel image, a practical upper bound on pixel separation was 50 pixels. Separations of 1 to 50 in 1-pixel increments were used for this study. Because collagen fiber orientation was not consistent from ovary to ovary, parameters for four orientations (0 deg, 45 deg, 90 deg, and 135 deg) were computed and averaged.

From each GLCM, four parameters (contrast, correlation, energy, and homogeneity) that capture essential image characteristics were computed as follows:

$$\text{Contrast: } \sum_{i,j} |i - j|^2 p(i, j), \quad (1)$$

$$\text{Correlation: } \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)[p(i, j)]}{\sigma_i \sigma_j}, \quad (2)$$

$$\text{Energy: } \sum_{i,j} p(i, j)^2, \quad (3)$$

$$\text{Homogeneity: } \sum_{i,j} \frac{p(i, j)}{1 + |i - j|}, \quad (4)$$

where $p(i, j)$ is the probability of gray level i occurring next to gray level j , $\mu_i = \sum_i \sum_j i \cdot p(i, j)$, $\mu_j = \sum_i \sum_j j \cdot p(i, j)$, $\sigma_i = \sum_i \sum_j (i - \mu_i)^2 \cdot p(i, j)$, and $\sigma_j = \sum_i \sum_j (j - \mu_j)^2 \cdot p(i, j)$. All parameters have a maximum value of 1 and a minimum value of 0 or -1.^{48,49} There are many other parameters that can be computed from the GLCM. To minimize the computation time required, a few parameters with the most potential were selected.

High contrast occurs when an image has a high number of pixel pairs with large differences in gray level occurring at the specified separation and orientation. High correlation occurs in images with periodic features. Energy (or angular second moment) is highest in images with uniform gray level or uniform gray-level differences at the specified separation and lower for those with more variation in gray levels. Finally, homogeneity (or inverse difference moment) is highest in images with pixels of the same or similar gray levels at the specified separation and orientation.

Combining the Fourier and GLCM parameters resulted in a total of 203 parameters to describe each single image. The first three parameters were the power in the low-, middle-, and high-frequency bands. The rest of the parameters are GLCM contrast, correlation, energy, and homogeneity for 1-, 2-, 3-, . . . , 50-pixel separations.

2.7 Image Classification

The five diagnoses were used to separate the images for classification. Carcinoma was compared to normal, and carcinoma was compared to all other images (noncarcinoma). Also, benign tubular adenoma was compared to tubular adenoma with dysplasia. Image classification was performed using a support vector machine (SVM) algorithm. The SVM is a machine learning classifier for separating two classes.⁵⁰ MATLAB default settings were used, with sequential minimal optimization for finding the hyperplane and either a linear or quadratic kernel to map the data into kernel space. The classification function uses the equation

$$c = \sum_i a_i k(s_i, x) + b, \quad (5)$$

where s_i are the support vectors, a_i are the weights (slope), b is the bias (intercept), and k is a kernel function, which was chosen to be linear (dot product) or quadratic. If $c \geq 0$, then x is classified as a member of the first group; otherwise x is classified as a member of the second group.⁴⁹

Training was performed on 25% of the data selected at random. Training was performed 100 times, with replacement of training images to the image pool before each subsequent training iteration. The performance of the classifier was determined by finding the average and standard deviation of the area under the receiver operator characteristic (ROC) curve for the 100 training sets. The ROC curve is a plot of the true-positive rate versus the false-positive rate (i.e., sensitivity versus 1—specificity) and is calculated using the true classes (specified by the user) and the output classes from the SVM at various bias values. A larger area under the ROC indicates better

classifier performance.^{51,52} Sensitivity and specificity were determined by selecting the point on the ROC curve where sensitivity and specificity were approximately equal.

The minimum number of parameters for optimal binary classification of each group pair was found using sequential forward selection (SFS).⁵³ For SFS, the area under the ROC curve was evaluated for each of the 203 parameters using 100 training sets (all images in the training set have known diagnoses). The best single parameter was then combined with each of the remaining 202 parameters, and 100 training sets were performed to find the highest performing pair of parameters. The best two parameters were then combined with the remaining 201 parameters, and 100 training sets were performed to find the highest performing trio of parameters. This process was repeated until the optimal set of parameters was identified. An additional parameter was kept only if the additional parameter increased the performance by greater than one standard deviation from the previous performance. Differences in optimal parameters were checked for statistical significance using a linear mixed-effects model, to account for multiple images from each ovary. The mixed-effects model included a random intercept and a robust sandwich estimator to estimate the covariance matrix.

To verify that the properly trained classifier performs better than random guessing, the classifier was challenged with a random grouping test, in which training was performed as above, but classes were assigned randomly instead of correctly. The random assignments were given in the same proportion as true image classes.

Many images contained large regions without signal. Images with high and low signal area were trained on and tested separately to see if higher signal area resulted in better sensitivity and specificity. Images from normal and carcinoma diagnoses were placed into the following groups based on visual inspection: images having greater than 75% of the field of view (FOV) containing signal (19 carcinoma and 54 normal), images with greater than 25% of the FOV containing signal (40 carcinoma and 119 normal), and images with less than 25% of the FOV containing signal (8 carcinoma and 12 normal). The parameters that were found to be best for normal versus carcinoma (when training on all images) were used for training, and 100 iterations were performed.

3 Results

Images representing the general appearance of each diagnosis are shown in Fig. 1. Images of normal ovaries have thin, straight collagen fibers that weave in all directions around the many different-sized follicles. Images from DMBA-effect ovaries have various-sized fibers with small voids in some areas. Tubular adenoma ovary images have primarily thin collagen fibers with distinctive fuzzy dots. Images from ovaries with tubular adenoma and tubular adenoma with dysplasia are similar, but more variation in collagen fiber thickness can be seen in tubular adenoma with dysplasia. The carcinoma ovary images tend to have thick, wavy collagen fibers that are ordered in the same direction, close together, and in thick bands not covering the entire field of view.

The SFS resulted in selection of two or three image analysis parameters for each binary classification. Using more than three parameters never resulted in improved performance. Carcinoma was best differentiated from normal alone, and from all other diagnoses combined, using two parameters: power in the highest frequency region (PHF) and GLCM energy with 38-pixel separation. Tubular adenoma was best differentiated from

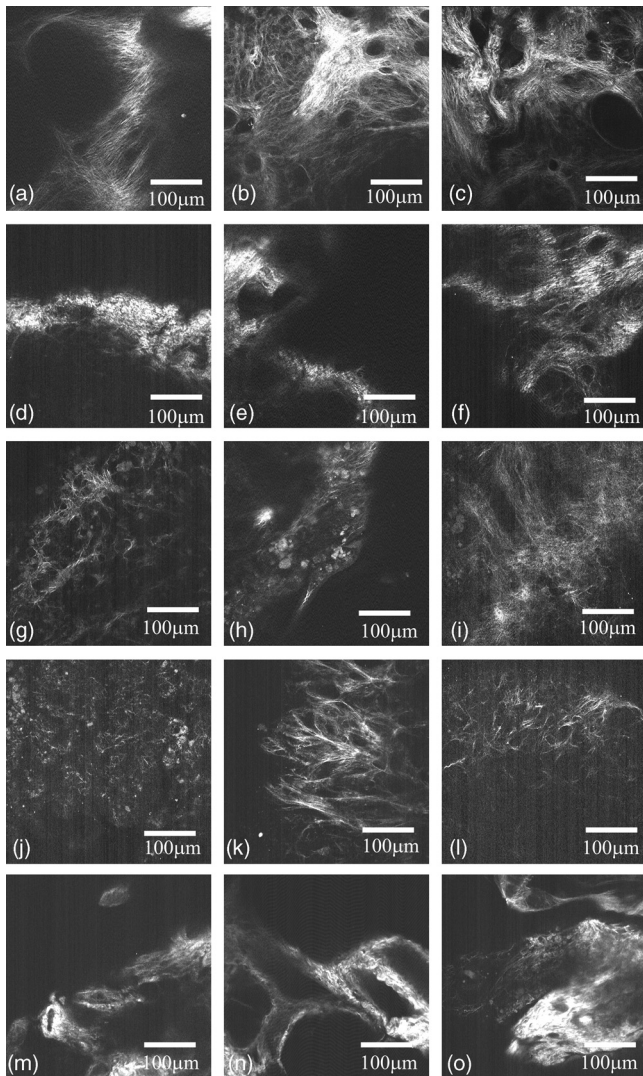


Fig. 1 Images from three ovaries of each diagnosis: normal (a–c), DMBA-effect (d–f), tubular adenoma (g–i), tubular adenoma with dysplasia (j–l), and carcinoma (m–o).

tubular adenoma with dysplasia using three parameters: GLCM contrast at 17- and 25-pixel separation and GLCM energy at 22-pixel separation. The average values of these five parameters are shown in Figs. 2 and 3. There was a statistically significant difference between carcinoma and all other diagnoses for PHF ($P < 0.0150$) (Fig. 2). There was a statistically significant difference between cancer and all other diagnoses, except tubular adenoma with dysplasia, for GLCM energy with 38-pixel separation ($P \leq 0.0025$) (Fig. 2). There was a statistically significant difference between tubular adenoma and tubular adenoma with dysplasia for GLCM contrast at 17- and 25-pixel separation and GLCM energy at 22-pixel separation ($P \leq 0.0015$) (Fig. 3).

Visualizing these data in another manner, Fig. 4 shows a plot of PHF versus GLCM energy at 38-pixel separation for carcinoma and normal diagnoses. Carcinoma values appear to fall on the outer border of the normal values. Figure 5 shows a plot of GLCM contrast at 17- and 25-pixel separations for tubular adenoma and tubular adenoma with dysplasia diagnoses. The tubular adenoma with dysplasia values appear to fall on a line with similar slope but higher intercept than tubular adenoma.

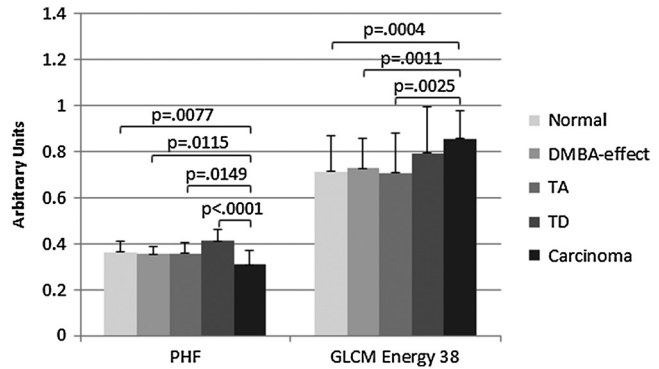


Fig. 2 Average values for parameters used to separate carcinoma from normal and carcinoma from noncarcinoma. TA = tubular adenoma, TD = tubular adenoma with dysplasia. Only positive standard deviations are shown.

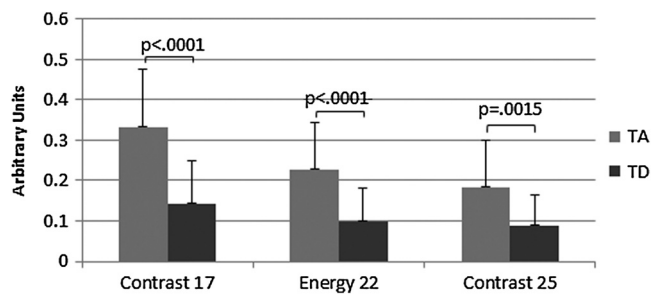


Fig. 3 Average values for parameters used to separate tubular adenoma (TA) from tubular adenoma with dysplasia (TD). Only positive standard deviations are shown.

The area under the ROC curve (AUC) was always greater than 0.78 for correct class and less than 0.61 for random class, showing that all training groups that were trained with the correct class performed better than training groups that were trained with random class assignments (25% of data in training, 75% testing, 100 iterations). Also, quadratic kernel resulted in a performance similar to that of linear kernel (AUC 0.83 for linear and AUC 0.81 for quadratic) for carcinoma versus normal and carcinoma versus noncarcinoma.

When training on 25% and testing on the remaining 75%, the classifier showed better than 74% average sensitivity and

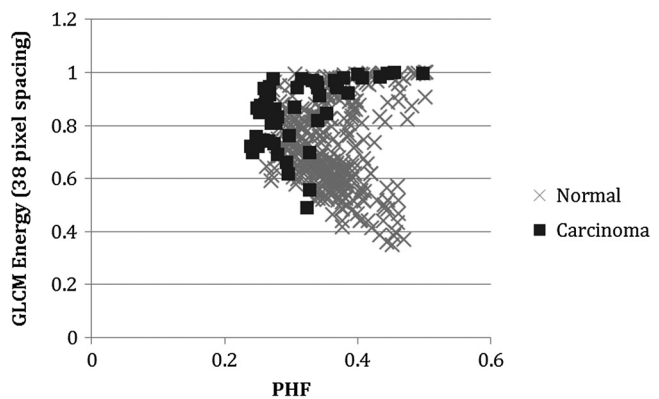


Fig. 4 Plot of power in the high-frequency region versus GLCM energy for carcinoma and normal diagnoses.

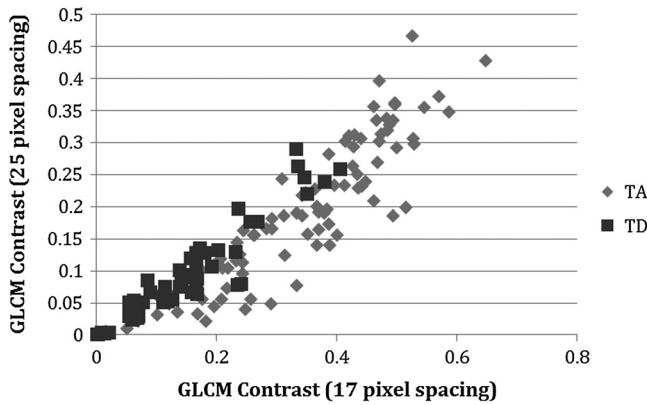


Fig. 5 Plot of contrast with 17-pixel separation vs. contrast with 25-pixel separation for tubular adenoma (TA) and tubular adenoma with dysplasia (TD).

specificity results for all groups. The quadratic kernel resulted in higher sensitivity and specificity than the linear kernel for carcinoma versus normal (77.8% and 79.3%, compared to 75.4% and 76.1%, respectively). The same result was seen for carcinoma versus noncarcinoma (81.2% and 80.0%, compared to 76.6% and 75.4%). The highest sensitivity and specificity achieved for each group are shown in Table 1.

For carcinoma versus normal, images with signal in less than 25% of the FOV had low sensitivity and specificity. Using images with signal in greater than 25% of the FOV produced sensitivity and specificity slightly better than using all images. Using images with signal in greater than 75% of the FOV produced the highest sensitivity and specificity (Table 2).

4 Discussion

The VCD/DMBA-treated animal model successfully led to ovarian adenocarcinoma, the most commonly occurring ovarian malignancy in women. Therefore, the collagen structure changes seen in this model may be similar to structural changes seen in human ovarian cancer. Unfortunately, the incidence of

adenocarcinoma was only 35% in the VCD/DMBA group (and 0% in the control groups). Thus, the number of carcinoma specimens was very small compared to the number of specimens from other diagnoses. Combined with relatively large variation in the appearance of carcinoma regions, the analysis results presented here must be considered preliminary.

Both VCD and DMBA caused atrophy of ovarian tissue and occasionally adhesions, making the ovaries difficult to find and separate from fat and other tissue. This difficulty resulted in exclusion of many samples that were histologically confirmed as nonovarian tissue in the area imaged. Also, VCD and DMBA caused effects in the ovary that are not common in women. VCD caused the development of benign tubular adenomas. Tubular adenomas are rare in the human ovary, so the findings related to this, although possibly useful in a very small population, would not directly translate to changes seen in women at high risk for adenocarcinoma. Occasionally, the tubular adenomas also developed focal areas of dysplasia. Because dysplasia often precedes cancer, identification of collagen changes due to dysplasia would be very useful for detecting precancerous changes in human ovary, but it is unclear if the collagen morphology changes accompanying tubular adenoma with dysplasia would be the same as collagen morphology changes occurring during dysplasia in the absence of tubular adenoma.

DMBA caused a condition we labeled “DMBA-effect.” The DMBA-effect included changes associated with highly active steroidogenic cells, degenerating follicles, degenerating corpora lutea, and a proliferative epithelial layer. The entire effect is not seen in humans, but proliferation of the epithelial layer is a risk factor for ovarian cancer, so the collagen changes that were seen in the DMBA-effect may translate to humans if the change in collagen morphology is related to epithelial proliferation. Average parameter values for DMBA-effect images frequently fell between values for normal and carcinoma, suggesting the possibility that collagen changes may be due to proliferation of the epithelial layer; however, the differences in parameter values for DMBA-effect and normal were not statistically significant.

Training results revealed that power in the highest-frequency band and GLCM energy with 38-pixel separation were the most

Table 1 Testing results for 100 iterations.

Groups tested	Sensitivity, % (SD)		Specificity, % (SD)	
Carcinoma versus normal	77.8	(11.3)	79.2	(6.8)
Carcinoma versus noncarcinoma	81.2	(11.1)	80.0	(5.0)
TA versus TD	80.2	(3.8)	82.7	(4.6)

Note: TA = tubular adenoma, TD = tubular adenoma with dysplasia.

Table 2 Testing results for carcinoma versus normal.

<25% Signal area		All images		>25% Signal area		>75% Signal area	
Sensitivity, % (SD)	Specificity, % (SD)	Sensitivity, % (SD)	Specificity, % (SD)	Sensitivity, % (SD)	Specificity, % (SD)	Sensitivity, % (SD)	Specificity, % (SD)
62.0 (32.4)	66.7 (14.4)	77.8 (11.3)	79.2 (6.8)	79.5 (15.0)	81.3 (9.7)	81.5 (12.7)	81.1 (9.7)

useful parameters for separation of both carcinoma versus normal and carcinoma versus noncarcinoma images. It is expected that the frequency content of the images would vary from group to group because the collagen fiber width is visually different. In the carcinoma images, collagen fibers appear to be much thicker than collagen fibers in tubular adenoma or normal images, which will translate to lower relative power in the high-frequency region. It is also expected that GLCM energy would change for different collagen morphologies, because the collagen width and spacing affects the transition of gray levels across the image. Normal has lower average energy at 38-pixel separation because the fibers are thinner, causing more variation in gray levels when moving across the image.

The increase in performance for carcinoma versus noncarcinoma over carcinoma versus normal is likely due to carcinoma having greater separation from other diagnoses than from normal. For example, carcinoma and tubular adenoma with dysplasia have a larger separation in the average value for power in the high-frequency region than carcinoma and normal. Difficulty in distinguishing the carcinoma and normal categories is likely due to the wide variation in collagen structure within normal ovaries, leading to a variation in computed image features. Normal ovaries have considerable image variation because the collagen structure depends on the structures it surrounds—stromal cells, follicles, or scar tissue from a new corpus luteum will all look different.

For simplicity, we initially intended to use only a linear kernel for classification. However, owing to the parabolic shape of the plot of the parameters (Fig. 4) for carcinoma and normal, a quadratic kernel was also tested. Training performance of the quadratic kernel was slightly lower (AUC lower) than the linear kernel, but the quadratic kernel gave better sensitivity and specificity results for both carcinoma versus normal and carcinoma versus noncarcinoma. There was an increase in sensitivity and specificity despite a decrease in AUC because the portion of the ROC curve near equal sensitivity and specificity for the quadratic kernel was above the ROC curve for the linear kernel at the selected point. If another point on the ROC curve was selected, the sensitivity and specificity may be greater for the linear kernel than for the quadratic kernel. These results suggest that the ideal line for separation is neither linear nor quadratic.

Because of varying surface topology, clefts, and invaginations in the ovaries, as well as variations in ovary composition, there were frequently large portions of the image FOV that did not contain signal. Sensitivity and specificity results were similar to random when only images with less than 25% of the FOV containing signal were used. The sensitivity and specificity results were improved when only images containing signal in greater than 25% of the FOV (rather than all images) were analyzed. Restricting analysis to only images with greater than 75% of the FOV containing signal resulted in incremental improvement, but also required the exclusion of many images. These results indicate that images used for detection of cancer require signal in at least one-quarter of the FOV. In this study, the presence of images with a small percentage of the FOV containing signal was mainly due to the small radius of curvature of the mouse ovaries, which, when combined with the water-immersion objective, precluded a flat field of view. The radius of curvature of a human ovary would be much larger, and a clinical implementation of the system would likely be a contact probe, allowing the physician to obtain an image of a flat surface. When imaging a flat surface, there is likely to be signal in

the entire FOV, reducing the number of images that would need to be excluded from the analysis.

Tubular adenoma and tubular adenoma with dysplasia were separated with the largest sensitivity and specificity of the three groups tested. For tubular adenoma versus tubular adenoma with dysplasia, training results showed that the most useful parameters were GLCM contrast at 17- and 25-pixel separation and GLCM energy at 22-pixel separation. The excellent performance when using energy and contrast at different pixel separations is likely related to the periodicity of fiber spacing in the images, which was generally between 5 and 50 pixels. These results indicate that tubular adenoma images contain a pattern that repeats at approximately 22-pixel separation and larger differences in gray levels at 17- and 25-pixel separations than tubular adenoma with dysplasia. Images of tubular adenoma and tubular adenoma with dysplasia had a very distinct visual appearance, with the presence of dots intermixed with numerous thin fibers. It is unclear what tissue constituent is responsible for these dots, and more investigation is required to find the cause of this signal. Clinically, distinguishing a benign condition from dysplasia would be useful, although as stated, the incidence of tubular adenoma in women is small.

The high sensitivity to dysplastic changes is particularly exciting because the histologically verified region of dysplasia was typically deep in the ovary, not in the superficial volume imaged. This finding suggests that there may be a field effect influencing the surrounding collagen structure for hundreds of micrometers. A field effect has been noted in light-scattering spectroscopy of the colon.⁵⁴ As the number of dysplastic ovaries in this study was small, however, a firm conclusion cannot be drawn.

Classifier performance for every group was higher with true diagnosis assignments than with random diagnosis assignments. This result proves that the classifier is able to separate better than random chance.

Twenty to fifty percent of BRCA-positive (high-risk) patients develop ovarian cancer.⁵⁵ If 1000 BRCA-positive patients underwent prophylactic oophorectomy, 500 to 800 of these procedures would be extraneous, causing unnecessary morbidity and mortality. If 1000 patients were tested before oophorectomy using a diagnostic test with 80% sensitivity and specificity, as we have shown, then only 100 to 160 would be unnecessary oophorectomies. The trade-off is that 100 to 160 women with cancer would not be caught by the test. By changing the bias value on the ROC curve, 100% sensitivity and 51% specificity can be selected. Using the corresponding curve for separation, the diagnostic test would successfully detect all cancers and result in unnecessary oophorectomy of 245 to 392 patients. Compared to the current method of prophylactic ovary removal, this diagnostic test would cut the number of unnecessary oophorectomies in half, greatly reducing unnecessary morbidity and mortality.

The analysis and classification methods used here provided a simple way to separate two classes at a time. A more sophisticated classifier would enable multiple diagnoses, and would incorporate data from a full stack or multiple stacks of images to capture the variation in ovary morphology. The differences in collagen structure between categories are often very subtle, and there is significant variation within a group, particularly in the case of normal. Tumors are often much more uniform throughout—this is seen most obviously in the tubular adenoma, which look very similar from slice to slice and ovary to ovary.

Additionally, the dynamic changes in collagen structure may make the same diagnosis appear different depending on the extent of disease. In the future, we plan to examine whether combining SHG images with TPEF images and/or optical coherence tomography images can improve classifier performance. We also plan to analyze images obtained *in vivo* at multiple time points during an animal's lifetime to better evaluate changes during early disease development. The *in vivo* imaging series will allow us to select ovaries with disease at the end of the study and look back in time to determine changes in collagen structure during early disease. With better understanding of early disease, we hope to develop an optical diagnostic test for ovarian cancer to use in women who are candidates for prophylactic oophorectomy. The ideal diagnostic system would be implemented minimally invasively using a micro-endoscope, like a falloposcope, that could image both the fallopian tubes and ovaries. Such a test could potentially reduce the number of unnecessary salpingo-oophorectomies in high-risk women and thus improve their quality of life.

Acknowledgments

This study was sponsored in part by National Institutes of Health, National Cancer Institute research grant R01 CA119200 and the University of Arizona Cancer Center Support Grant (CCSG—CA023074).

References

1. *Cancer Facts and Figures 2011*, American Cancer Society www.cancer.org (2011).
2. R. C. Bast, Jr. et al., "New tumor markers: Ca125 and beyond," *Int. J. Gynecol. Cancer* **15**(Suppl 3), 274–281 (2005).
3. U. Menon et al., "Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)," *Lancet Oncol.* **10**(4), 327–340 (2009).
4. A. P. Singh et al., "Clinical potential of mucins in diagnosis, prognosis, and therapy of ovarian cancer," *Lancet Oncol.* **9**(11), 1076–1085 (2008).
5. M. Donach et al., "Combined use of biomarkers for detection of ovarian cancer in high-risk women," *Tumour Biol.* **31**(3), 209–215 (2010).
6. C. H. Clarke et al., "Proteomic biomarkers apolipoprotein A1, truncated transthyretin and connective tissue activating protein III enhance the sensitivity of CA125 for detecting early stage epithelial ovarian cancer," *Gynecol. Oncol.* **122**(3), 548–553 (2011).
7. C. S. Zhu et al., "A framework for evaluating biomarkers for early detection: validation of biomarker panels for ovarian cancer," *Cancer Prev. Res. (Phila.)* **4**(3), 375–383 (2011).
8. S. Milman, K. D. Whitney, and N. Fleischer, "Metastatic medullary thyroid cancer presenting with elevated levels of CA 19-9 and CA 125," *Thyroid* **21**(8), 913–916 (2011).
9. R. Molina et al., "Tumor markers (CEA, CA 125, CYFRA 21-1, SCC and NSE) in patients with non-small cell lung cancer as an aid in histological diagnosis and prognosis. Comparison with the main clinical and pathological prognostic factors," *Tumour Biol.* **24**(4), 209–218 (2003).
10. C. C. Johnson et al., "The epidemiology of CA-125 in women without evidence of ovarian cancer in the prostate, lung, colorectal and ovarian cancer (PLCO) screening trial," *Gynecol. Oncol.* **110**(3), 383–389 (2008).
11. A. S. Dehaghani et al., "Factors influencing serum concentration of CA125 and CA15-3 in Iranian healthy postmenopausal women," *Pathol. Oncol. Res.* **13**(4), 360–364 (2007).
12. D. K. Pauler et al., "Factors influencing serum CA125II levels in healthy postmenopausal women," *Cancer Epidemiol. Biomarkers Prev.* **10**(5), 489–493 (2001).
13. J. Tcherkassova et al., "Combination of CA125 and RECAF biomarkers for early detection of ovarian cancer," *Tumour Biol.* **32**(4), 831–838 (2011).
14. L. S. Cohen et al., "Three-dimensional power Doppler ultrasound improves the diagnostic accuracy for ovarian cancer prediction," *Gynecol. Oncol.* **82**(1), 40–48 (2001).
15. D. A. Fishman et al., "The role of ultrasound evaluation in the detection of early-stage epithelial ovarian cancer," *Am. J. Obstet. Gynecol.* **192**(4), 1214–1222 (2005).
16. J. R. van Nagell, Jr. et al., "Ovarian cancer screening with annual transvaginal sonography: findings of 25,000 women screened," *Cancer* **109**(9), 1887–1896 (2007).
17. I. Ohel et al., "Three-dimensional power Doppler ultrasound in ovarian cancer and its correlation with histology," *Arch. Gynecol. Obstet.* **281**(5), 919–925 (2010).
18. S. S. Buys et al., "Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial," *JAMA* **305**(22), 2295–2303 (2011).
19. L. J. Havrilesky et al., "Development of an ovarian cancer screening decision model that incorporates disease heterogeneity: implications for potential mortality reduction," *Cancer* **117**(3), 545–553 (2011).
20. M. Salhab, S. Bismohun, and K. Mokbel, "Risk-reducing strategies for women carrying BRCA1/2 mutations with a focus on prophylactic surgery," *BMC Womens Health* **10**(28) (2010).
21. L. I. Batista et al., "Coordinated prophylactic surgical management for women with hereditary breast-ovarian cancer syndrome," *BMC Cancer* **8**, 101 (2008).
22. W. A. Rocca et al., "Premature menopause or early menopause and risk of ischemic stroke," *Menopause: J. N. Am. Menopause Soc.* **19**(3), 272–277 (2011).
23. J. Fanning, E. Yacoub, and R. Hojat, "Laparoscopic-assisted cytoreduction for primary advanced ovarian cancer: Success, morbidity and survival," *Gynecol. Oncol.* **123**(1), 47–49 (2011).
24. L. T. Shuster et al., "Prophylactic oophorectomy in premenopausal women and long-term health," *Menopause Int.* **14**(3), 111–116 (2008).
25. U. Utzinger et al., "Reflectance spectroscopy for *in vivo* characterization of ovarian tissue," *Lasers Surg. Med.* **28**(1), 56–66 (2001).
26. V. Untereiner et al., "Optical diagnosis of peritoneal metastases by infrared microscopic imaging," *Anal. Bioanal. Chem.* **393**(6–7), 1619–1627 (2009).
27. L. P. Hariri et al., "Simultaneous optical coherence tomography and laser induced fluorescence imaging in rat model of ovarian carcinogenesis," *Cancer Biol. Ther.* **10**(5), 438–447 (2010).
28. E. M. Kanter et al., "Dual modality imaging of a novel rat model of ovarian carcinogenesis," *J. Biomed. Opt.* **11**(4), 041123 (2006).
29. M. A. Brewer et al., "Imaging of the ovary," *Technol. Cancer Res. Treat.* **3**(6), 617–627 (2004).
30. L. P. Hariri et al., "Laparoscopic optical coherence tomography imaging of human ovarian cancer," *Gynecol. Oncol.* **114**(2), 188–194 (2009).
31. Y. Yang et al., "Integrated optical coherence tomography, ultrasound and photoacoustic imaging for ovarian tissue characterization," *Biomed. Opt. Express* **2**(9), 2551–2561 (2011).
32. Y. Yang et al., "Optical scattering coefficient estimated by optical coherence tomography correlates with collagen content in ovarian tissue," *J. Biomed. Opt.* **16**(9), 090504 (2011).
33. Y. Yang et al., "Potential role of a hybrid intraoperative probe based on OCT and positron detection for ovarian cancer detection and characterization," *Biomed. Opt. Express* **2**(7), 1918–1930 (2011).
34. S. F. Elahi et al., "Longitudinal molecular imaging with single cell resolution of disseminated ovarian cancer in mice with a LED-based confocal microendoscope," *Mol. Imag. Biol.* **13**(6), 1157–1162 (2011).
35. A. A. Tanbakuchi et al., "Clinical confocal microlaparoscope for real-time *in vivo* optical biopsies," *J. Biomed. Opt.* **14**(4), 044030 (2009).
36. A. Aguirre et al., "Coregistered three-dimensional ultrasound and photoacoustic imaging system for ovarian tissue characterization," *J. Biomed. Opt.* **14**(5), 054014 (2009).
37. A. Aguirre et al., "Potential role of coregistered photoacoustic and ultrasound imaging in ovarian cancer detection and characterization," *Transl. Oncol.* **4**(1), 29–37 (2011).
38. N. J. Durr et al., "Maximum imaging depth of two-photon autofluorescence microscopy in epithelial tissues," *J. Biomed. Opt.* **16**(2), 026008 (2011).
39. N. D. Kirkpatrick, M. A. Brewer, and U. Utzinger, "Endogenous optical biomarkers of ovarian cancer evaluated with multiphoton microscopy," *Cancer Epidemiol. Biomarkers Prev.* **16**(10), 2048–2057 (2007).

40. R. M. Williams et al., "Strategies for high-resolution imaging of epithelial ovarian cancer by laparoscopic nonlinear microscopy," *Transl. Oncol.* **3**(3), 181–194 (2010).
41. O. Nadiarnykh et al., "Alterations of the extracellular matrix in ovarian cancer studied by second harmonic generation imaging microscopy," *BMC Cancer* **10**(94), 1–14 (2010).
42. J. Adur et al., "Recognition of serous ovarian tumors in human samples by multimodal nonlinear optical microscopy," *J. Biomed. Opt.* **16**(9), 096017 (2011).
43. D. Granot et al., "In vivo imaging of the systemic recruitment of fibroblasts to the angiogenic rim of ovarian carcinoma tumors," *Cancer Res.* **67**(19), 9180–9189 (2007).
44. M. R. McLemore et al., "Epidemiologic and genetic factors associated with ovarian cancer," *Cancer Nurs.* **32**(4), 281–290 (2009).
45. L. N. Springer et al., "Involvement of apoptosis in 4-vinylcyclohexene diepoxide-induced ototoxicity in rats," *Toxicol. Appl. Pharmacol.* **139**(2), 394–401 (1996).
46. L. P. Mayer et al., "The follicle-depleted mouse ovary produces androgen," *Biol. Reprod.* **71**(1), 130–138 (2004).
47. Z. R. Craig et al., "7,12-Dimethylbenz[a]anthracene induces Sertoli-Leydig cell tumors in the follicle-depleted ovaries of mice treated with 4-vinylcyclohexene diepoxide," *Comp. Med.* **60**(1), 10–17 (2010).
48. Matlab R2011b Documentation. <http://www.mathworks.com> (2011).
49. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern. SMC* **3**(6), 610–621 (1973).
50. C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.* **20**(3), 273–297 (1995).
51. N. A. Obuchowski, "Receiver operating characteristic curves and their use in radiology," *Radiology* **229**(1), 3–8 (2003).
52. M. H. Zweig and G. Campbell, "Receiver-operating characteristic plots: a fundamental evaluation tool in clinical medicine," *Clin. Chem* **39**(4), 561–577 (1993).
53. A. W. Whitney, "A direct method of nonparametric measurement selection," *IEEE Trans. Comput.* **20**(9), 1100–1103 (1971).
54. H. K. Roy et al., "Spectroscopic microvascular blood detection from the endoscopically normal colonic mucosa: biomarker for neoplasia risk," *Gastroenterology* **135**(4), 1069–1078 (2008).
55. M. C. King, J. H. Marks, and J. B. Mandell, and New York Breast Cancer Study Group, "Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2," *Science* **302**(5645), 643–646 (2003).