

Journal of Biomedical Optics

BiomedicalOptics.SPIEDigitalLibrary.org

Correlation of breast tissue histology and optical signatures to improve margin assessment techniques

Stephanie Kennedy
Matthew Caldwell
Torre Bydlon
Christine Mulvey
Jenna Mueller
Lee Wilke
William Barry
Nimmi Ramanujam
Joseph Geradts

SPIE.

Stephanie Kennedy, Matthew Caldwell, Torre Bydlon, Christine Mulvey, Jenna Mueller, Lee Wilke, William Barry, Nimmi Ramanujam, Joseph Geradts, "Correlation of breast tissue histology and optical signatures to improve margin assessment techniques," *J. Biomed. Opt.* **21**(6), 066014 (2016), doi: 10.1117/1.JBO.21.6.066014.

Correlation of breast tissue histology and optical signatures to improve margin assessment techniques

Stephanie Kennedy,^a Matthew Caldwell,^a Torre Bydlon,^a Christine Mulvey,^a Jenna Mueller,^a Lee Wilke,^b William Barry,^c Nimmi Ramanujam,^a and Joseph Geradts^{d,*†}

^aDuke University, TOPS Lab, Department of Biomedical Engineering, 136 Hudson Hall, Durham, North Carolina 27708, United States

^bUniversity of Wisconsin Breast Center, Department of Surgery, 600 Highland Avenue, Madison WI 53792, United States

^cDana-Farber Cancer Institute, Department of Biostatistics and Computational Biology, 450 Brookline Avenue, CLS11007, Boston, Massachusetts 02215, United States

^dDuke University Medical Center, Department of Pathology, DUMC3712, 200 Trent Drive, Durham, North Carolina 27710, United States

Abstract. Optical spectroscopy is sensitive to morphological composition and has potential applications in intraoperative margin assessment. Here, we evaluate *ex vivo* breast tissue and corresponding quantified hematoxylin & eosin images to correlate optical scattering signatures to tissue composition stratified by patient characteristics. Adipose sites (213) were characterized by their cell area and density. All other benign and malignant sites (181) were quantified using a grid method to determine composition. The relationships between mean reduced scattering coefficient ($\langle\mu_s'\rangle$), and % adipose, % collagen, % glands, adipocyte cell area, and adipocyte density were investigated. These relationships were further stratified by age, menopausal status, body mass index (BMI), and breast density. We identified a positive correlation between $\langle\mu_s'\rangle$ and % collagen and a negative correlation between $\langle\mu_s'\rangle$ and age and BMI. Increased collagen corresponded to increased $\langle\mu_s'\rangle$ variability. In postmenopausal women, $\langle\mu_s'\rangle$ was similar regardless of fibroglandular content. Contributions from collagen and glands to $\langle\mu_s'\rangle$ were independent and equivalent in benign sites; glands showed a stronger positive correlation than collagen to $\langle\mu_s'\rangle$ in malignant sites. Our data suggest that scattering could differentiate highly scattering malignant from benign tissues in postmenopausal women. The relationship between scattering and tissue composition will support improved scattering models and technologies to enhance intraoperative optical margin assessment. © 2016 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JBO.21.6.066014]

Keywords: breast; margin; scattering; demographics; spectroscopy; histology.

Paper 160139R received Mar. 5, 2016; accepted for publication Jun. 8, 2016; published online Jun. 21, 2016.

1 Introduction

In breast conserving surgery (BCS), the surgeon attempts to excise the cancerous area along with a margin of benign tissue, while conserving as much normal breast volume as possible. Failure to achieve clear margins (as determined by pathology and institutional standards) requires additional surgery. A recent study found that the re-excision (or reoperation) rate varied from 0% to 70%, indicating that there is no reliable intraoperative standard for preventing re-excision.¹ Touch-prep cytology and frozen section analysis have been used to address this need intraoperatively. However, these techniques prolong surgery time (an additional 20 to 40 min), require a trained pathologist to be present, and have technical challenges associated with processing fatty breast tissues. Consequently, these methods are only used at a few medical centers. The current number of patients undergoing BCS is ~270,000/year in the United States, with an annual growth rate of 5.5%.² These statistics provide the motivation for devising innovative solutions to reduce re-excision rates in women undergoing BCS.

Our group has developed a fast and nondestructive quantitative spectral imaging device that measures optical signatures (450 to 600 nm) from which the morphological composition of tissue can be determined.^{3–5} The sensing depth of the device ranges from 0.5 to 2.2 mm,⁵ which satisfies the depth

requirement for effective intraoperative margin detection at Duke University Medical Center and other hospitals. A fast, scalable inverse Monte Carlo model^{6,7} quantitatively determines the wavelength-dependent absorption coefficient (μ_a) and reduced scattering coefficient (μ_s') from the magnitude and shape of the diffuse reflectance spectra. The concentrations of β -carotene (found in adipocytes) and total hemoglobin concentration (found in blood vessels) can be extracted from the absorption (μ_a) spectrum. The μ_s' spectrum is averaged resulting in the mean reduced scattering coefficient ($\langle\mu_s'\rangle$), reflective of the average composition of tissue scatterers. The $\langle\mu_s'\rangle$ has previously been shown to be negatively correlated with the amount of adipose tissue present⁸ and positively correlated with glandular or fibrous content resulting from more cellular and collagen-based structures.⁹

Using this technology, our group optically imaged 88 margins from 70 patients with a sensitivity of 74% and a specificity of 86%. $[\beta\text{-carotene}]/\langle\mu_s'\rangle$ showed the most significant differences between benign margins and margins with tumor involvement, reflecting a decrease in adipose content and an increase in fibroglandular content within malignant tissues.¹⁰ From a related but separate study, the optical signatures from 633 individual sites on the tumor margins resulting from 101 margins (100 patients) were evaluated by pathology and diagnosed as either benign (595) or malignant (38).⁴ The 595 benign sites were divided into adipose (A) ($n = 324$), mixed tissue samples (Mx) ($n = 112$), vessel samples (V) ($n = 64$),

*Address all correspondence to: Joseph Geradts, E-mail: jgeradts@partners.org

†Current address: Brigham and Women's Hospital, Department of Pathology, Boston, Massachusetts 02115, United States

fibroadipose (FA) ($n = 60$), fibroglandular (FG) ($n = 24$), fibrocystic change (FCC) ($n = 4$), fibrous (F) ($n = 6$), and fat necrosis (FN) ($n = 1$). Of these benign samples, 142 were from premenopausal patients and 411 were from postmenopausal patients.⁴ The 38 malignant samples were comprised of invasive ductal carcinoma (IDC) ($n = 22$), ductal carcinoma *in situ* (DCIS) ($n = 12$), lobular carcinoma *in situ* (LCIS) ($n = 3$), and invasive lobular carcinoma (ILC) ($n = 1$). This study examined three levels of optical contrast: (1) malignant versus benign, (2) A versus FA versus FG versus IDC versus DCIS, and (3) benign premenopausal versus postmenopausal.⁴ Results indicated that the site-level optical properties were highly dependent on the composition of the benign breast.⁴ Specifically, scattering increased with fibrous content and decreased with adipose content while $[\beta\text{-carotene}]$ increased with adipose content.⁴ Postmenopausal patients exhibited lower $\langle\mu_s'\rangle$, but higher $[\beta\text{-carotene}]$ than premenopausal patients due to decreased fibrous content and increased adipose content. The $\langle\mu_s'\rangle$ was statistically higher in adipose tissues and lower in fibroglandular tissues compared to that in DCIS and IDC.

It has previously been established that all cellular and extracellular components of breast tissue contribute to scattering but the relative magnitude of each contributing component has not been measured independently.¹¹ Furthermore, given that both fibroglandular and malignant tissue contain glands and collagen, it is important to deconstruct how scattering is influenced by the composition of the benign and malignant areas. In this study, the amount of epithelium (glands), collagen, and adipose tissue was quantified from histological sections of the optically measured sites. Correlations between the scattering parameter $\langle\mu_s'\rangle$ and the histological composition of the tissue stratified by various demographic parameters, including age, body mass index (BMI), breast density, and menopausal status were explored.

The principal goal of our study was to deconstruct sources of scattering as related to breast tissue composition in order to better understand the sources of spectral contrast observed in positive breast tumor margins.

2 Materials and Methods

2.1 Patient Population and Measurement Procedure

An *ex vivo* study to evaluate partial mastectomy specimens with diffuse optical spectroscopy in patients undergoing surgery for breast malignancies was approved by the Institutional Review Board at Duke University as detailed in previous publications.³⁻⁵ The following patient characteristics were collected (if available): mammographic breast density (MBD), menopausal status, neo-adjuvant treatment status (chemotherapy or endocrine therapy), age, BMI, and surgical re-excision status. Based on preoperative mammograms, each patient was assigned an MBD based on a four-point scale indicative of increasing amounts of fibroglandular tissue: 1 (fatty), 2 (scattered fibroglandular), 3 (heterogeneously dense), and 4 (extremely dense). Women were designated as postmenopausal if they had one of the following: either a bilateral salpingo-oophorectomy or lack of a menstrual cycle for more than 1 year.

The surgical procedures, optical measurements, and the histological processing of the tissue are described elsewhere.⁴ A total of 38 malignant sites and 595 benign sites from the previous study⁴ were examined for this analysis. The corresponding hematoxylin & eosin (H&E) slides for each of these sites were requested from the Duke surgical pathology archives. Each H&E slide was imaged, quantified and classified to determine the proportion of glands, collagen, and adipose tissues as well as the size and density of adipocytes in predominantly fatty tissues. This process is shown in Fig. 1.

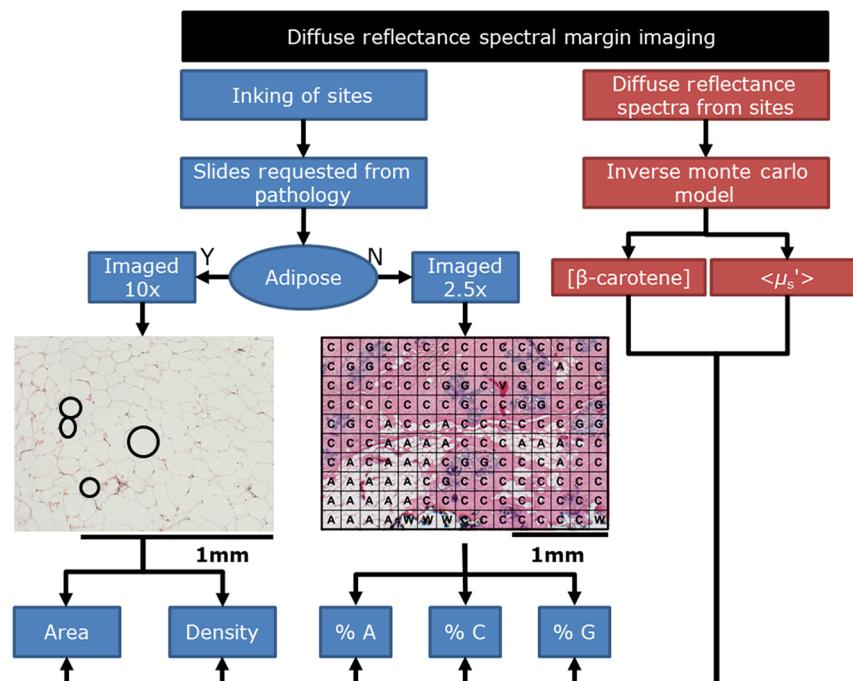


Fig. 1 Flowchart demonstrating the study design. Diffuse reflectance spectra were measured and processed using the Monte Carlo model of reflectance to extract tissue optical properties. The corresponding site was inked and then submitted for pathology. The corresponding $5\ \mu\text{m}$ H&E sections from the optically interrogated sites were imaged and detailed morphological composition was quantified (A, adipose; C, collagen; G, glands).

The 595 benign sites were categorized by the study pathologist (JG) as follows: A, Mx, V, FA, FG, F, FCC, and FN. The 38 malignant sites were similarly categorized as follows: IDC, DCIS, ILC, and LCIS. In order to focus on the underlying sources of scattering from adipocytes, collagen, and glands, we excluded sites that had significant additional components [i.e., V ($n = 64$)] as well as sites that lacked a large enough sample size to draw conclusions regarding trends [i.e., FCC ($n = 4$), FN ($n = 1$), ILC ($n = 1$), and LCIS ($n = 3$)]. When specifically examining the adipose sites, the only features quantified from these sites were adipocyte size and density as shown on the left-hand side of Fig. 1. The black circles within the adipose H&E image are illustrative of the adipocyte features. The remaining benign sites were comprised of two to three of the following components: adipose, collagen, and glands and the fraction of each of these components were quantified as shown on the right side of Fig. 1. Although some of the sites contained vessels or fluid filled cysts; these were accounted for but not included in this analysis. In regions with inflammatory cells, the box was classified according to the predominant component. Inflammatory cells have a range of diameters: 10 to 12 μm in neutrophils, eosinophils, and basophils; 6 to 15 μm in lymphocytes; and $\sim 18 \mu\text{m}$ in monocytes.¹² These inflammatory components mimic densely packed epithelial cells; thus, when inflammatory cells comprised the majority of the 200 $\mu\text{m} \times 200 \mu\text{m}$ area, the tissue section was designated as glandular or nucleated, but this was a rare event. When most of the section contained collagen with a few inflammatory cells, the tissue section was designated as collagen. A total of 526 benign sites (A, Mx, FA, FG, and F) and 34 malignant sites (IDC and DCIS) were included in the imaging protocol. Figure 2 shows the included and excluded sites for this study; the exclusion and inclusion criteria for the imaging analysis are further detailed in Sec. 2.2.1 for adipose sites and Sec. 2.2.2 for nonadipose sites.

2.2 Imaging and Quantification of Detailed Morphological Composition of Breast Tissue Histological Sections

The criteria for inclusion of the sites in this analysis were as follows: (1) the measurement ink could be clearly identified, (2) the imaged area corresponded with the pathologist's diagnosis, and (3) the tissue was intact throughout the area of interest.

2.2.1 Adipocyte imaging and quantification

The inclusion criteria were met for 213 adipose sites. The image of each adipose histological section was acquired with a Zeiss Axio Imager upright microscope, 10 \times ocular with a color camera, and a 10 \times objective. The images were acquired using a halogen light source, a QImaging MicroPublisher 5.0 MP color camera, and MetaMorph 7.6.5, which was used to adjust the acquisition time and RGB gain. The field of view (FOV) was 1 mm \times 1.3 mm with a resolution of 1.04 μm .

An algorithm developed by our group was used to determine adipocyte area (size) and density on the 213 adipose H&E images. Specifically, the green channel of the RGB images was used in the algorithm as it provided a convenient method of separating the primarily pink and blue stained tissue from white fat. All images were preprocessed with a two-dimensional implementation of an edge-preserving bilateral filter. Subsequently, the MATLAB[®] implementation of the Canny edge detector was used to extract the outlines of the adipocytes. The interior of each outlined shape was measured to obtain cell area and the number of shapes was counted to provide an estimate of cell density. Empirically determined cell-area thresholds of 129.3 and 22,569 μm^2 (diameters approximately between 12 and 170 μm) were used to limit the counted results to those with a high probability of being an adipocyte. These thresholds correspond in the range of prior literature estimates of median cell diameter.^{13–16}

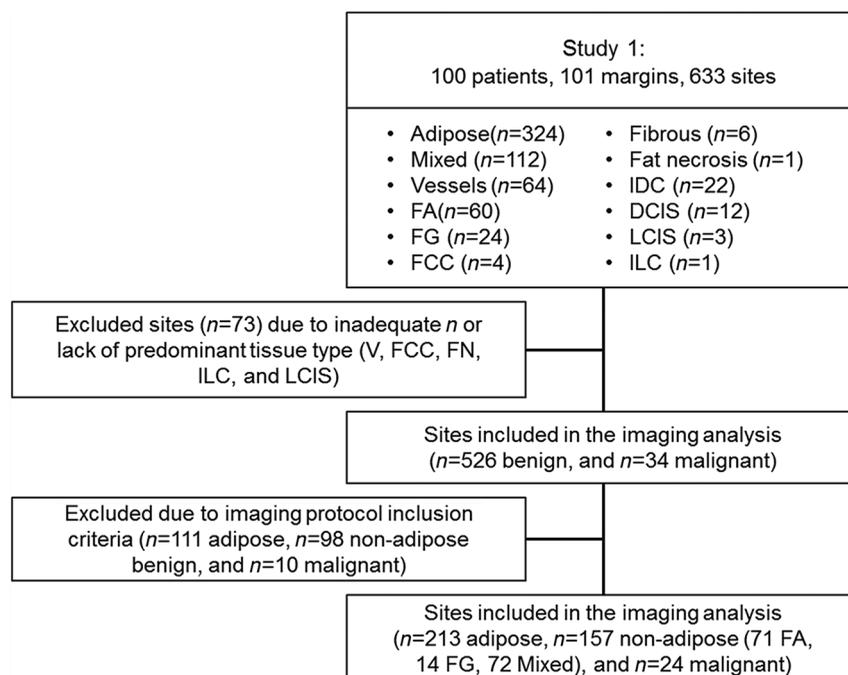


Fig. 2 REMARK diagram illustrating the excluded and retained sites for study investigation.

2.2.2 Imaging and quantification of fibroadipose, fibroglandular, mixed, and malignant tissues

Precise pathological coregistration was not possible due to loss of the ink marker in 14 malignant sites. The inclusion criteria were not met for 98 nonadipose benign sites. This resulted in a total of 157 nonadipose benign sites and 24 malignant sites that were entered into the detailed image analysis process.

The nonadipocyte images were acquired with a 10× ocular, 10% neutral density filter, and a 2.5× objective. The resulting FOV for the 2.5× objective was ~5.4 mm × 4 mm with a resolution of 2.0812 μm. Due to the large FOV, a shading correction image was obtained and subtracted from the images to account for illumination variations. The orange measurement ink on each histological section, made at the time of the optical spectroscopic measurements, was identified and oriented parallel to either the long or short axis of the microscope's imaging field. As described previously, each diffuse reflectance channel was 3 mm in diameter with a sensing depth of ~0.5 to 2.2 mm. When images were acquired, two scale bars were drawn on the image to represent the 2-mm depth and the 3-mm surface footprint along the area with ink denoting the measured site. Grids of 200 μm encompassing the 3 mm × 2 mm area were overlaid on each image (~150 boxes per image/site). These 157 benign and 24 malignant images were then shown to raters for additional classification.

The nonadipocyte images were classified and diagnosed in two ways. First, the initial diagnosis determined by the pathologist describing the predominant tissue type (as described in the previous study⁴) was retained for each image of the corresponding H&E slide. This diagnosis was also used to verify correct spatial imaging and separation into benign and malignant sites. Second, five raters were polled to further separate the pathologist's categorization to determine the predominant tissue types in each of the 181 images. Three of the five raters were graduate students with experience and training on viewing and evaluating basic histological and pathological images; the remaining two raters were trained to recognize adipocytes, collagen, and nuclei. A kappa statistic was used to ensure agreement across raters. The purpose of the raters' classification was to use an unbiased population to identify which of the following components were present in an image: collagen (C), adipose (A), and glands (G). Some images contained a combination of only two of these three components, while other images contained all three components. The 181 images included malignant sites but only the pathologist's classification was used to define malignant and benign. The images were classified as (1) collagen and adipose, (2) collagen and glands, or (3) mixed. The mode of the five raters' classification for the benign sites resulted in the following tissue classifications: FA-adipose and collagen; FG-collagen and glands; Mix-adipose, collagen, and glands; Mal-malignant sites. Using ImageJ and Excel, either a 10 × 15 or 15 × 10 grid of 200 μm × 200 μm boxes, corresponding to the 3 mm × 2 mm area, was overlaid on each image. Each of the 150 boxes was diagnosed as A (adipose), C (collagen), G (glands), V (vessels), F (fluid), or W (white space). V and F were not evaluated. Quantified histological parameters were calculated as shown below. This expression allowed an investigation into $\langle \mu_s' \rangle$ as it related to the % adipose, % collagen, and % glands within an H&E image that accounts for white space: %Tissue X = (#of boxes of Tissue X) / (Total # of boxes - # boxes of Tissue W). The relationships between histological variables and $\langle \mu_s' \rangle$ were analyzed using

Pearson's correlation coefficient, while Wilcoxon rank-sum tests were used to evaluate the relationships between menopausal status and $\langle \mu_s' \rangle$ by comparing the median scattering values across four independent populations (premenopausal with low collagen, postmenopausal with low collagen, premenopausal with high collagen, and postmenopausal with high collagen). Menopausal status was identified by the patient clinical records, while low and high collagen was defined by having a % collagen about the median value observed in the data (see Sec. 3.2). The data were plotted using box plots to provide a visualization of the summary statistics. The tops and bottoms of each "box" corresponded to the 25th and 75th percentiles of the samples, respectively, with the interquartile range covering the distance. The sample median and skewness of the data were represented with a horizontal line across each box. The observations within 1.5 times the interquartile range away from the top or bottom were depicted with whiskers. Observations beyond the whisker length were marked as outliers. Notches displayed the variability of the median between samples, where box plots whose notches did not overlap had different medians at the 5% significance level.

3 Results

3.1 Patient Characteristics

The patient characteristics in terms of age, BMI, BD, menopausal status, and chemotherapy status are shown in Table 1. The average age and BMI of the population were 58.5 years and 30.2, respectively. The investigated population represented all four breast density categories with a higher prevalence of MBD-2 and MBD-3. The population studied was skewed toward postmenopausal and chemo-naïve patients.

3.2 Representative Data and Summary Variables

Representative wavelength-dependent μ_a and μ_s' spectra and histological images are shown in Fig. 3. This figure demonstrates that adipose tissue contributed to increased absorption [Fig. 3(a)] and fibroglandular tissue contributed to increased scattering [Fig. 3(b)]. Additionally, both collagen and glands played a role in determining tissue scattering, while adipose sites exhibited a baseline scattering signature. The corresponding H&E stained histology images for the fibroglandular (FG), fibroadipose (FA), and adipose (A) sites are shown in Figs. 3(c)–3(e).

Table 1 Patient characteristics from the investigated population.

Age (mean, range) years	58.5, 30 to 87
BMI (mean, range)	30.2, 18.3 to 49.2
Breast density	
MBD-1	17%
MBD-2	34%
MBD-3	38%
MBD-4	11%
Pre versus postmenopausal (%)	27% versus 73%
Naïve versus neoadjuvant therapy (%)	81% versus 19%

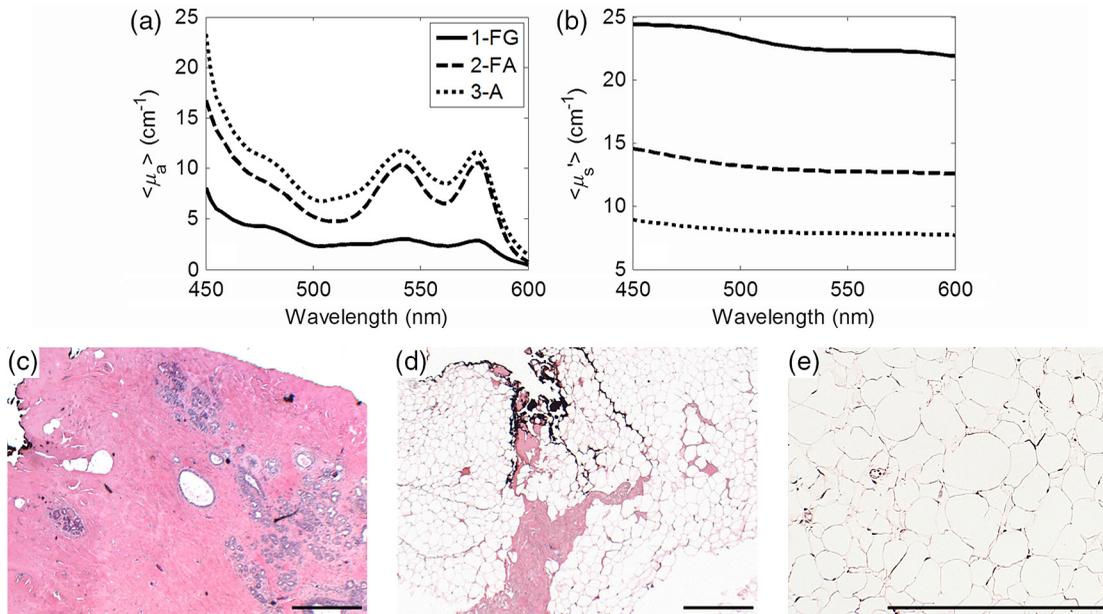


Fig. 3 (a) The corresponding mean absorption coefficient ($\langle \mu_a \rangle$) and (b) the mean reduced scattering coefficient ($\langle \mu_s' \rangle$) for fibroglandular (FG), fibroadipose (FA), and adipose (A). (C–E) The corresponding histology images for the fibroglandular (FG), fibroadipose (FA), and adipose (A). This figure demonstrates that adipose tissue contributes to increased absorption and fibroglandular tissue contributes to increased scattering. The scale bars represent 500 μm . The adipose H&E was imaged at 100 \times while the FG and FA images were acquired at 25 \times .

Table 2 shows the mean and standard deviation for the optical and histological variables as a function of tissue type. A Kruskal–Wallis test (one-way ANOVA on ranks) demonstrated there is a statistically significant difference in the distribution of $\langle \mu_s' \rangle$ among the five tissue types shown in Table 2 (chi-squared = 44.612, df = 4, p -value = 4.787×10^{-9}).

The inter-rater agreement was $\kappa = 0.69$ indicating a high degree of agreement across raters. The 394 sites imaged were taken from 93 patients. Of the 394 total sites, 370 were benign and 24 were malignant. The latter included 15 foci of IDC and 9 foci of DCIS. The benign sites did not include specific lesions such as FCC, fibroadenomas or papillomas. The highest scattering was seen in the FG sites followed by Mal, Mix, FA, and A sites. [β -carotene] was highest in Mal sites followed by A, FA, Mix, and FG sites. Although [β -carotene] was the highest in Mal, the values were not statistically significantly different

than those observed in the adipose sites ($p = 0.26$). The percent adipose tissue decreased in the following order: A, FA, Mix, Mal, and FG, respectively. The percent collagen decreased as follows: FG, Mal (lower than FG due to inclusion of DCIS and IDC), Mix, FA, and A. The percent glandular tissue was highest in Mal followed by FG and Mix; FA sites had minimal glandular tissue if any.

3.3 Patient Characteristics in Sites with Low and High Percentages of Collagen

Table 3 shows the breakdown according to sites with either a low or high percentage of collagen using the median value (44%) of all sites as the cutoff. There was an even distribution of number of sites and patients with high and low percentage collagen. Sites were evenly distributed between MBD-2 and

Table 2 Sample sizes, mean values, and standard deviations for the optical and histological variables by tissue type. The mean and range of values demonstrate that FA has the highest % adipose, FG has the highest quantified collagen and μ_s' , and malignant (IDC and DCIS) sites have the highest quantified glandular parameters.

Tissue type	n	$\langle \mu_s' \rangle$	[β -carotene]	% Adipose	% Collagen	% Glands
Benign	370	8.3 ± 4.0	20.2 ± 11.8	46.4 ± 29.4	47.3 ± 25.6	5.7 ± 7.2
A	213	7.3 ± 3.1	21.5 ± 11	NA	NA	NA
FA	71	8.5 ± 3.9	20.9 ± 14.7	64.7 ± 23.0	34.5 ± 23.0	0.6 ± 1.1
FG	14	13.0 ± 7	13.6 ± 6.6	2.7 ± 3.0	81.4 ± 8.7	14.3 ± 8.8
Mix	72	10.3 ± 4.5	16.9 ± 10.5	36.9 ± 24.0	53.4 ± 21.8	9.1 ± 6.9
Malignant	24	10.7 ± 4.3	28.1 ± 27.4	35.0 ± 30.5	41.7 ± 24.3	21.7 ± 19.6

Table 3 Summary table for the nonadipose H&E images and corresponding spectra. The data are separated to show the number of data points above and below the median % collagen. The summary shows the number of sites, MBD, age, menopausal status, and tissue type. Sites with a high percentage of collagen were mostly from MBD of 2 or higher. FG sites had at least 56% or higher collagen.

	Collagen	
	<44%	44% to 100%
Patients	52	71
Sites	91	90
Breast density		
MBD1	18	5
MBD2	32	29
MBD3	38	34
MBD4	3	21
Unknown MBD	0	1
Age	61 ± 10.5	54.2 ± 9.0
BMI	29.4 ± 8.0	27.7 ± 7.5
Menopausal status		
Pre	12	42
Post	77	46
Unknown	2	2
Tissue type		
FA	50	21
FG	0	14
Mix	27	45
Mal	14	10

MBD-3. Sites from MBD-4 exhibited a higher fraction of collagen while sites from MBD-1 exhibited lower fractions of collagen, as expected. The sites with low percentage of collagen were from older patients ($p < 0.00005$). Sites with a low percentage of collagen were also weighted toward postmenopausal patients consistent with increased age. Sites with a higher fraction of collagen were evenly split between pre- and postmenopausal patients; FG sites only fell into the higher percent collagen category but FA, Mix, and Mal all exhibited high and low collagen morphologies. The cancer stage and invasive grade populated both high and low collagen categories and was not included (data not shown).

3.4 Scattering versus Collagen in Benign Sites

The relationship between $\langle \mu_s^t \rangle$ and percent collagen was examined; a positive correlation in benign sites is shown in Fig. 4. $\langle \mu_s^t \rangle$ exhibited greater variability in sites with a higher percentage of collagen as shown in Fig. 4(a). Using the adipose sites, scattering was shown to have no significant correlation to adipose tissue morphology as shown in Figs. 4(b) and 4(c).

3.5 Effect of Patient Characteristics and Percentage Collagen on Scattering

In order to investigate the relationship between the patient characteristics and morphology with respect to $\langle \mu_s^t \rangle$, the importance of each variable was analyzed after adjusting for percent collagen. The 213 adipose images were not included in this analysis. The correlations were examined by high and low percent collagen to determine whether the relationships changed with collagen level for benign sites (FA, FG, and Mix) and all sites (FA, FG, Mix, and Mal). The results of the correlations, linear fits, and interactions are shown in Table 4. $\langle \mu_s^t \rangle$ was negatively correlated to age in benign ($r = -0.32$, $p = 4.7 \times 10^{-5}$) and all ($r = -0.32$, $p = 1.4 \times 10^{-5}$) sites and this correlation varied significantly by the collagen level ($r = -0.40$ in high % collagen versus $r = -0.13$ in low % collagen sites, $p = 0.01$). $\langle \mu_s^t \rangle$ was negatively correlated to BMI in benign ($r = -0.32$, $p = 4 \times 10^{-5}$) and all sites ($r = -0.31$, $p = 2.8 \times 10^{-5}$) but this relationship did not vary with collagen level. Age was found to vary with percentage collagen, resulting

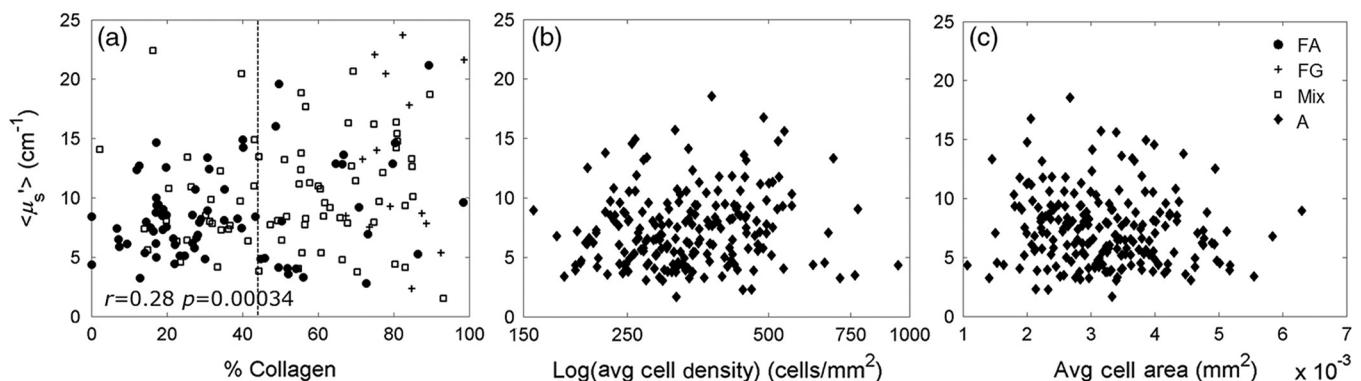


Fig. 4 The relationship between scattering and the % collagen in benign sites. $\langle \mu_s^t \rangle$ is expected to be related to collagen and glands (nuclei and subnuclear components). When considering benign sites (FA, FG, and Mix), (a) $\langle \mu_s^t \rangle$ shows a positive correlation to the % collagen. The sites with collagen greater than 44% exhibit greater variability in $\langle \mu_s^t \rangle$ that cannot be explained by the % adipose tissue. $\langle \mu_s^t \rangle$ shows no correlation to (b) the log of the average adipocyte cell density or (c) the average adipocyte cell area.

Table 4 Correlations of Age, BMI, and MBD to high and low % collagen. Each characteristic was fit as an independent variable with $\langle\mu_s'\rangle$ as the dependent variable using Pearson's correlations and linear fits to establish relationships. Age and BMI were found to be independently associated with $\langle\mu_s'\rangle$. $\langle\mu_s'\rangle$ decreased with both age and BMI but exhibited a stronger negative relationship to age in sites with a high percentage of collagen. MBD was positively correlated with $\langle\mu_s'\rangle$ but not independent of age and BMI.

	Benign		All sites	
	r	p	r	p
Age	-0.32	4.7×10^{-5}	-0.32	1.4×10^{-5}
High % collagen	-0.40	0.01	-0.37	0.02
Low % collagen	-0.13		-0.18	
BMI	-0.32	4.0×10^{-5}	-0.31	2.8×10^{-5}
High % collagen	Ø	NS	Ø	NS
Low % collagen	Ø		Ø	
MBD	0.22	0.01	0.21	4.6×10^{-3}

in a stronger negative correlation to $\langle\mu_s'\rangle$ in sites with a high percentage of collagen. Age was positively correlated to BMI ($r = 0.36$) but BMI remained significant in describing $\langle\mu_s'\rangle$ after correcting for age. BMI did not have a dependence upon percentage collagen yet remained significant after adjusting for age by percent collagen.

3.6 Effects of Patient Characteristics and Adipocyte Morphology on Scattering

To confirm that adipocytes did not contribute to the positive correlation between collagen and $\langle\mu_s'\rangle$, the relationship between $\langle\mu_s'\rangle$ with adipocyte size and log (adipocyte density) as well as age and BMI, stratified by MBD was examined and is shown in Table 5 for the 213 adipose samples. This variability was not dependent upon the size and density of the adipocytes, nor age and BMI ($p > 0.05$). Adipose tissues, however, did contribute to a baseline scattering value. After adjusting the morphology components by breast density, a stronger relationship between average cell area (ACA) and $\langle\mu_s'\rangle$ was seen in sites from high breast density patients ($r = -0.29$, $p = 0.026$), but the relationship to cell density remained nonsignificant. Analyzing the relationships between scattering and characteristics associated with MBD shows that $\langle\mu_s'\rangle$ is not correlated to patient characteristics or adipocyte morphology in adipose sites ($n = 213$, $p > 0.05$).

3.7 Scattering, Collagen Density and Menopausal Status

To determine how these scattering differences might affect optical contrast in patients of different breast compositions, scattering was compared across sites with low and high percentages of collagen as well as pre- and postmenopausal status as a binary surrogate for age. Menopausal status was used because it was highly positively correlated with age ($r = 0.75$). Scattering was found to be significantly lower in benign sites with low

Table 5 Analyzing the relationships between scattering and characteristics associated with MBD shows $\langle\mu_s'\rangle$ is not correlated to patient characteristics or adipocyte morphology in adipose sites; $n = 213$, all $p > 0.05$. ACA is significantly larger in sites from high breast density patients ($r = -0.29$) versus low breast density patients ($r = 0.02$). Ø indicates no correlation; NS, not significant.

	$\langle\mu_s'\rangle$ (cm^{-1})	
	r	p
Age (years)	Ø	NS
MBD1-2	Ø	NS
MBD3-4	Ø	
BMI (kg/m^2)	Ø	NS
MBD1-2	Ø	NS
MBD3-4	Ø	
ACA (mm^2)	Ø	NS
MBD1-2	0.02	0.026
MBD3-4	-0.29	
Log (ACD) (cells/mm^2)	Ø	NS
MBD1-2	Ø	NS
MBD3-4	Ø	

percentage of collagen in postmenopausal women compared to benign sites with a high percentage of collagen in premenopausal women (Fig. 5, unadj. $p = 5.3 \times 10^{-6}$) as would be expected. However, unexpectedly, benign sites with a high percentage of collagen in postmenopausal women were not statistically different from those with low percentage collagen in either pre- or postmenopausal women. These sites from

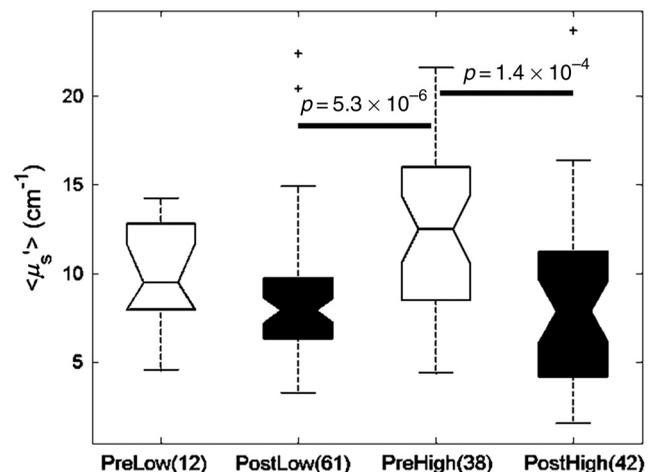


Fig. 5 Box plots demonstrating menopausal status analysis with respect to high and low % collagen. Sites from postmenopausal patients with low and high % collagens exhibit the lowest scattering values. High % collagen sites from premenopausal patients exhibit the highest scattering values. The difference in $\langle\mu_s'\rangle$ values for malignant and benign sites is greater in postmenopausal patients.

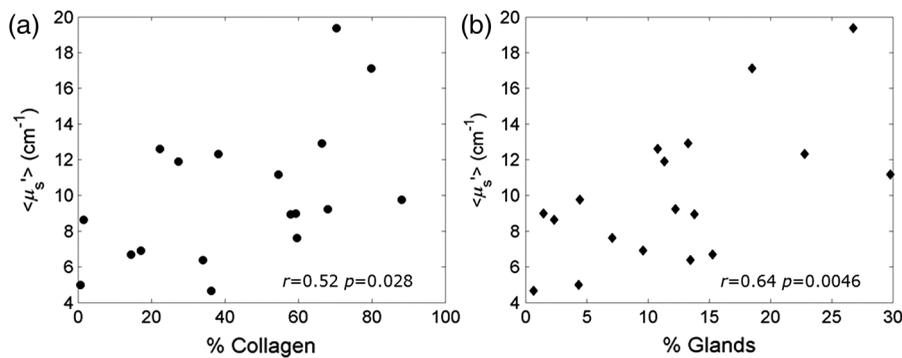


Fig. 6 Statistical modeling results show that both % collagen and % glands independently relate to scattering in a positive direction with no relationship to one another. The relationship between scattering and % collagen (a) or % glands (b) in malignant sites demonstrates a slightly stronger dependence upon glands.

postmenopausal women with high percentage of collagen were statistically lower than that of sites from premenopausal women with a high percentage of collagen (Fig. 5, unadj. $p = 1.4 \times 10^{-4}$).

3.8 Relative Contributions of Collagen versus Glands to Scattering

Lastly, the relationship between percent collagen and percent glands in benign, malignant, and all sites was analyzed to determine the relative contributions of stroma versus epithelium to $\langle \mu_s' \rangle$. The relationships among $\langle \mu_s' \rangle$, the percent collagen, and percent glands are shown in Fig. 6 in order to determine the contribution of collagen and glands to $\langle \mu_s' \rangle$, respectively.

The linear models (Table 6) identified that both percent collagen and percent glands were independently associated with $\langle \mu_s' \rangle$ in all sites and sites with glands (FG, Mix, and Mal). The analysis of collagen versus glands was narrowed to investigate sites with glands between 0% and 35% (the dynamic range of the data). The linear model reflected an equivalent relationship to scattering from glands and the collagen in benign sites ($r = 0.18$ versus $r = 0.17$). However, as shown in Table 6, the malignant sites had a slightly stronger positive relationship of percent glands ($r = 0.64$, $p = 0.005$) to $\langle \mu_s' \rangle$ compared to the collagen content and $\langle \mu_s' \rangle$ ($r = 0.52$, $p = 0.03$).

4 Discussion

In this article, we demonstrate that scattering is positively correlated to % collagen, which is consistent with previous findings.¹⁷ In sites with a high percentage of collagen

Table 6 Examining the dynamic range of the data includes restricting sites to % glands <40%. Here, the % glands and % collagen were found to independently contribute to $\langle \mu_s' \rangle$ in both benign and malignant sites. $\langle \mu_s' \rangle$ increased with both collagen and glands; $\langle \mu_s' \rangle$ exhibited a stronger positive relationship to glands in malignant sites.

$\langle \mu_s' \rangle$ (cm ⁻¹)	Benign (n = 85)		Malignant (n = 18)		All (103)	
	R	p	r	P	r	P
% Glands	0.18	0.09	0.64	0.005	0.25	0.01
% Collagen	0.17	0.13	0.52	0.03	0.23	0.02

(% collagen $\geq 44\%$), scattering exhibited higher variability compared to sites with <44% collagen. We demonstrated that $\langle \mu_s' \rangle$ variability was not dependent upon adipocyte morphology (cell area and density), BMI, or age. Although cell area was negatively correlated to scattering in high breast density women, this was not replicated in low breast density women. In addition, the cell density parameter also lacked a relationship to scattering and breast density. Thus, the adipose component was assumed to have minimal contribution to the variability or linear trends observed between $\langle \mu_s' \rangle$ and tissue morphology from adipose tissues, leading to the conclusion that the trends stemmed primarily from the collagen and glandular contributions. These findings allowed us to test the same relationships in FA, FG, Mix, and Mal tissues with $\langle \mu_s' \rangle$ decreasing with age, BMI, and MBD. However, MBD did not provide additional information not already described by either age or BMI.

Interestingly, in women with high collagen content, there was a relatively higher negative correlation between $\langle \mu_s' \rangle$ and age. The lowest scattering was seen in postmenopausal sites. Studies have shown that breast cancer risk increases with both age and breast density.^{18–23} Because age is strongly related to menopausal status, we hypothesize that hormonal changes and structural changes to collagen are tied to the decreasing trend in scattering in older, postmenopausal women. These hypotheses will require further studies and validation. Scattering is influenced by the scatterer size, the density of scatterers, and the refractive index mismatch between the scatterer and surrounding medium.²⁴ Changes in the rate of collagen turnover or changes in diameters could affect the number and size of scatterers, respectively. Decreased permeability and swelling observed with age²⁵ would be indicative of dehydration or possible changes in the refractive index mismatch. Median scattering levels are similar across low and high fibroglandular contents in postmenopausal women. These results indicate scattering could provide improved contrast in postmenopausal patients compared to premenopausal patients as scattering signal from malignant tissue would be highlighted due to the decreased background scattering from normal tissue.

In this study, we also investigated the stromal (collagen) and epithelial (glands) contributions to scattering. Previously, we had found that both the fibro-connective (collagen) and glandular portions were positively correlated with $\langle \mu_s' \rangle$ but the analysis was not performed separately on benign and malignant sites independently, nor were we able to identify the stronger contributor.¹⁷ Separating the comparison by diagnosis was an

important step that allowed a better understanding of the contribution of glands to $\langle\mu'_s\rangle$ and how this relationship changed in malignant sites. Our data demonstrates that collagen and glands have equal and independent contributions to scattering in benign sites; in malignant sites, however, the contribution from glands was stronger compared to the collagen contribution to scattering. Glands seem to be a stronger contributor to scattering in malignant tissues. In malignant tissues, the glandular component often occupies a larger area of the tissue; the cells/nuclei generally are larger and more variable in malignant tissues, which are among the features used by pathologists to arrive at a cancer diagnosis.²⁶ This is an important potential source of contrast in malignant and fibroglandular tissues where collagen densities may be similar, but where glandular contributions are expected to be significantly different due to the presence of proliferating cancer cells in malignant tissues. In the event a site is positive and contains a large glandular component, this DRS technology will highlight the morphological difference compared to benign sites. However, in the case of malignant sites that may contain small amounts of DCIS, or large contributions of fat, the optical scattering signature will not be as distinct from benign tissues.

It is important to revisit our current scattering model based on Mie theory given the contributions of the various constituents to scattering in this study. New scattering models, which capitalize upon size distributions of particles or techniques using polarization, could be used to more effectively separate collagen from glandular components. Bartek et al.²⁷ used electron microscopy to estimate subcellular particle sizes for the purpose of predicting scattering in breast tissue; they identified particle sizes ranging from 10 to 500 nm with two distributions centered around 20 to 25 nm and 110 to 230 nm, respectively, but found that Mie theory with a spherical treatment exceeded expected scattering values based on clinical data. Thus, different scattering models may be employed to describe differences in size distributions, which can further help separate collagen from glandular contributions to scattering.

It is important to note a potential source of error in the histological quantification that cannot currently be corrected. Due to the large specimen volume, the quantification was performed on a 5- μm paraffin section from a 3-mm block of formalin fixed tissue, while the DRS technology probes the entire 3 mm thickness. If other slices from the 3-mm block were quantified, the breakdown of tissue components might vary. In addition, the inked site was often larger than 3 mm making perfect coregistration difficult and requiring some approximations. However, the rough estimates of tissue composition show correlations to spectral signatures that could not occur by chance. In fact, if an automated higher resolution method of quantification were implemented, these correlations might increase in strength. In addition, the malignant sites were comprised of both invasive and *in situ* carcinomas of different grades and stages; to address the effect of tumor characteristics, larger sample sizes would need to be investigated to make this comparison.

A finding that was previously observed in a prior study⁴ and further confirmed here was that malignant cells had higher average $[\beta\text{-carotene}]$ than pure adipose sites; however, there was a wide range in both groups such that the values were not statistically significantly different across malignant versus adipose sites. In addition, the malignant sample size is small versus the adipose sample size. Of note, the malignant sites investigated in this study demonstrated a wide range of % adipose composition

($35 \pm 30.5\%$). Previous research done by Brown et al. found that $[\beta\text{-carotene}]$ was complex even within adipose tissues, and that varying breast density and adipocyte cell size contributed to variations in $[\beta\text{-carotene}]$.¹⁰ There also is published evidence that adipocytes associated with neoplastic cells are biochemically and functionally distinct.^{28–31}

5 Conclusions

Through this stepwise analysis, we confirmed that $\langle\mu'_s\rangle$ is positively correlated to collagen and negatively correlated to age and BMI. $\langle\mu'_s\rangle$ exhibited increased variability with collagen level, which was not dependent on the adipose morphology. A stronger negative correlation between age and $\langle\mu'_s\rangle$ was seen in sites with a high percentage of collagen as compared to sites with a significantly lower percentage of collagen. Scattering was found to be lowest in postmenopausal women regardless of the percentage of collagen indicating that menopausal status or age could be an important demographic factor for algorithms based on scattering contrast for breast cancer diagnosis. Collagen and glands have equivalent yet independent contributions to $\langle\mu'_s\rangle$ in benign sites. In malignant sites, glands demonstrated a stronger positive correlation than collagen to $\langle\mu'_s\rangle$. This information can be leveraged to improve the interpretation of scattering contrast and to better distinguish between fibroglandular and malignant sites, which can ultimately improve intraoperative margin assessment.

Acknowledgments

We would like to thank the Department of Pathology at Duke for providing us access to the H&E slides corresponding to all of the data we have collected. Also, we would like to thank the Duke Comprehensive Cancer Center and the Duke Translational Research Institute for their generous support. The DUMC surgical pathology staff performed the processing of the partial mastectomy specimens. This work was supported by the Department of Defense award W81XWH-05-1-036, National Institute of Health award 1R41CA128160-01, and was supported in part by Duke University's CTSA grant 1 UL1 RR024128-01 from NCRR/NIH.

References

1. L. E. McCahill et al., "Variability in reexcision following breast conservation surgery," *JAMA* **307**(5), 467–475 (2012).
2. "US surgical procedure volumes," Medtech Insight Report #A607, Life Science Intelligence, Huntington Beach, California (2009).
3. L. G. Wilke et al., "Rapid noninvasive optical imaging of tissue composition in breast tumor margins," *Am. J. Surg.* **198**(4), 566–574 (2009).
4. S. Kennedy et al., "Optical breast cancer margin assessment: an observational study of the effects of tissue heterogeneity on optical contrast," *Breast Cancer Res.* **12**(6), R91 (2010).
5. T. M. Bydlon et al., "Performance metrics of an optical spectral imaging system for intra-operative assessment of breast tumor margins," *Opt. Express* **18**(8), 8058–8076 (2010).
6. G. M. Palmer and N. Ramanujam, "Monte Carlo-based inverse model for calculating tissue optical properties. Part I: theory and validation on synthetic phantoms," *Appl. Opt.* **45**(5), 1062–1071 (2006).
7. G. M. Palmer et al., "Monte Carlo-based inverse model for calculating tissue optical properties. Part II: application to breast cancer diagnosis," *Appl. Opt.* **45**(5), 1072–1078 (2006).
8. C. Zhu et al., "Diagnosis of breast cancer using diffuse reflectance spectroscopy: comparison of a Monte Carlo versus partial least squares analysis based feature extraction technique," *Lasers Surg. Med.* **38**(7), 714–724 (2006).

9. S. Srinivasan et al., "Interpreting hemoglobin and water concentration, oxygen saturation, and scattering measured in vivo by near-infrared breast tomography," *Proc. Natl. Acad. Sci. U. S. A.* **100**(21), 12349–12354 (2003).
10. J. Q. Brown et al., "Optical spectral surveillance of breast tissue landscapes for detection of residual disease in breast tumor margins," *PLoS One* **8**(7), e69906 (2013).
11. S. Thomsen and D. Tatman, "Physiological and pathological factors of human breast disease that can influence optical diagnosis," *Ann. N.Y. Acad. Sci.* **838**, 171–193 (1998).
12. M. H. Ross and W. Pawlina, *Histology: A Text and Atlas with Correlated Cell and Molecular Biology*, 5th ed., Lippincott Williams & Wilkins, Philadelphia, Pennsylvania (2006).
13. M. Jernas et al., "Separation of human adipocytes by size: hypertrophic fat cells display distinct gene expression," *FASEB J.* **20**(9), 1540–1542 (2006).
14. H. C. Chen and R. V. Farese Jr., "Determination of adipocyte size by computer image analysis," *J. Lipid Res.* **43**(6), 986–989 (2002).
15. M. Ashwell et al., "Human fat cell sizing—a quick, simple method," *J. Lipid Res.* **17**(2), 190–192 (1976).
16. T. Bjornheden et al., "Computerized determination of adipocyte size," *Obes. Res.* **12**(1), 95–105 (2004).
17. C. Zhu et al., "Diagnosis of breast cancer using fluorescence and diffuse reflectance spectroscopy: a Monte-Carlo-model-based approach," *J. Biomed. Opt.* **13**(3), 034015 (2008).
18. B. Pogue et al., "Characterization of hemoglobin, water, and NIR scattering in breast tissue: analysis of intersubject variability and menstrual cycle changes," *J. Biomed. Opt.* **9**(3), 541–552 (2004).
19. A. E. Cerussi et al., "Sources of absorption and scattering contrast for near-infrared optical mammography," *Acad. Radiol.* **8**(3), 211–218 (2001).
20. N. Shah et al., "Noninvasive functional optical spectroscopy of human breast tissue," *Proc. Natl. Acad. Sci. U. S. A.* **98**(8), 4420–4425 (2001).
21. T. Li et al., "The association of measured breast tissue characteristics with mammographic density and other risk factors for breast cancer," *Cancer Epidemiol. Biomarkers Prev.* **14**(2), 343–349 (2005).
22. M. R. Bani et al., "Factors correlating with reexcision after breast-conserving therapy," *Eur. J. Surg. Oncol.* **35**(1), 32–37 (2009).
23. V. Assi et al., "Clinical and epidemiological issues in mammographic density," *Nat. Rev. Clin. Oncol.* **9**(1), 33–40 (2012).
24. S. L. Jacques, "Optical properties of biological tissues: a review," *Phys. Med. Biol.* **58**(11), R37–R61 (2013).
25. A. J. Bailey, R. G. Paul, and L. Knott, "Mechanisms of maturation and ageing of collagen," *Mech. Ageing Dev.* **106**(1–2), 1–56 (1998).
26. J. L. Connolly et al., *Role of the Surgical Pathologist in the Diagnosis and Management of the Cancer Patient*, D. W. Kufe et al., ed., BC Decker, Holland-Frei Cancer Medicine, Hamilton, Ontario (2003).
27. M. Bartek et al., "Estimation of subcellular particle size histograms with electron microscopy for prediction of optical scattering in breast tissue," *J. Biomed. Opt.* **11**(6), 064007 (2006).
28. M. N. Duong et al., "Adipose cells promote resistance of breast cancer cells to trastuzumab-mediated antibody-dependent cellular cytotoxicity," *Breast Cancer Res.* **17**, 57 (2015).
29. Y. Y. Wang et al., "Adipose tissue and breast epithelial cells: a dangerous dynamic duo in breast cancer," *Cancer Lett.* **324**(2), 142–151 (2012).
30. B. Dirat et al., "Cancer-associated adipocytes exhibit an activated phenotype and contribute to breast cancer invasion," *Cancer Res.* **71**(7), 2455–2465 (2011).
31. H. G. Moon et al., "Molecular characteristics of cancer-associated adipocytes in breast cancer," AACR ed., in *AACR 106th Annual Meeting 2015*, 18–22 April 2015 (2015).

Stephanie Kennedy is currently a senior consultant working for ICON Plc in the Commercialisation and Outcomes division. She received her BBME from Catholic University of America in 2006 and her PhD in biomedical engineering from Duke University in 2012. She has authored and coauthored several papers in the field of biomedical optics in conjunction with the Tissue Optical Spectroscopy Lab at Duke University.

Biographies for the other authors are not available.