Diagnosing breast cancer by using Raman spectroscopy

Abigail S. Haka*, Karen E. Shafer-Peltier†, Maryann Fitzmaurice‡, Joseph Crowe§, Ramachandra R. Dasari*, and Michael S. Feld*¶

*G. R. Harrison Spectroscopy Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139; †Dade Behring, Newark, DE 19714; ‡University Hospitals of Cleveland, Case Western Reserve University, Cleveland, OH 44106; and §The Cleveland Clinic Foundation, Cleveland, OH 44195

We employ Raman spectroscopy to diagnose benign and malignant lesions in human breast tissue based on chemical composition. In this study, 130 Raman spectra are acquired from *ex vivo* samples of human breast tissue (normal, fibrocystic change, fibroadenoma, and infiltrating carcinoma) from 58 patients. Data are fit by using a linear combination model in which nine basis spectra represent the morphologic and chemical features of breast tissue. The resulting fit coefficients provide insight into the chemical/morphological makeup of the tissue and are used to develop diagnostic algorithms. The fit coefficients for fat and collagen are the key parameters in the resulting diagnostic algorithm, which classifies samples according to their specific pathological diagnoses, attaining 94% sensitivity and 96% specificity for distinguishing cancerous tissues from normal and benign tissues. The excellent results demonstrate that Raman spectroscopy has the potential to be applied *in vivo* to accurately classify breast lesions, thereby reducing the number of excisional breast biopsies that are performed.

spectral diagnosis | optical vibrational disease

In the United States, ≈216,000 new cases of breast cancer are diagnosed each year, and 40,000 women die from the disease (1). Mammography, the most common technique for detecting nonpalpable, highly curable breast cancer, employs x-rays to quantitatively probe density changes in breast tissue. Because these density changes are not uniquely correlated with breast cancer, mammography serves as a screening technique rather than a diagnostic tool. Thus, a lesion found through either clinical breast examination or mammography is always biopsied. Because of current limitations, 70–90% of mammographically detected lesions are found to be benign upon biopsy (2). Breast biopsy is most often performed by surgical excision that removes the entire lesion or by core needle biopsy that removes 5–12 cores of tissue, typically 1 mm in diameter and several centimeters long, to ensure proper sampling (3). The complete diagnostic process, from start to finish, may take months and may include multiple biopsies.

A desire to reduce the number of biopsies performed on benign tissue and the patient trauma, time delay, and high medical costs involved has motivated researchers to explore a variety of minimally invasive optical imaging and spectroscopy techniques to improve breast cancer diagnosis, especially the ability to distinguish benign lesions from malignant ones. These techniques employ visible or near-infrared light, have the potential to provide chemical information, and are less invasive than current diagnostic procedures.

Diffuse optical tomography (DOT) studies the propagation of amplitude-modulated pulses of light through the breast. It is noninvasive and can detect lesions deep within the tissue (4–8). An array of sources and detectors in a measurement cup enables 3D images to be constructed. By using light of different wavelengths, information about scattering and absorption can be extracted to measure oxy- and deoxy-hemoglobin concentrations and the presence of lipids. However, DOT can detect only a limited number of chemicals and is generally of low resolution,

causing small lesions to go undetected. Furthermore, to obtain the sensitivity and specificity required for lesion detection *in vivo*, exogenous agents are often used to improve contrast (9).

Optical spectroscopic techniques are also under investigation for breast cancer diagnosis. Unlike DOT, these techniques sample the tissue locally (≈1 mm³ volume). Light delivery and collection can be accomplished by using optical fibers that can be incorporated into a biopsy needle. As opposed to biopsy, a spectroscopic needle measurement has the advantage of providing immediate diagnosis. Thus, spectroscopy has the potential to reduce both the likelihood of a nondiagnostic needle biopsy (requiring repeat needle or surgical biopsy) and patient anxiety (by eliminating the currently unavoidable wait for pathology diagnosis). Furthermore, with the development of minimally invasive breast cancer therapies, such as radiofrequency ablation, there is the potential for diagnosis and treatment to be performed in a single procedure (10).

Optical techniques that have been applied to breast tissue include fluorescence, reflectance, and Raman spectroscopies. Fluorescence spectroscopy has been applied to *ex vivo* breast tissue, and trends correlating with disease have been observed (11–13). Fluorescence produces relatively large signals. However, the small number of endogenous fluorophores in breast tissue and their broad spectral lineshapes are limiting factors. There are preliminary studies using diffuse reflectance spectroscopy to diagnose breast lesions (14) by monitoring changes in absorption and scattering. Combining fluorescence and diffuse reflectance spectroscopies has shown promise in the breast, as well as high sensitivity and specificity for cancer detection in several other organ systems (15).

Raman spectroscopy can provide detailed chemical information about a tissue sample and thus insight into the chemical changes that accompany breast disease. In contrast to fluorescence, there are a large number of Raman active molecules in breast tissue, and their spectral signatures are sharp and well delineated. The ability to measure several different chemicals is of particular importance in studying breast cancer because of the heterogeneity of the disease. Although Raman spectra provide high information content, the signals are orders of magnitude weaker than fluorescence. However, with careful system design, collection of clinical data in relevant times with safe laser powers can be accomplished. For these reasons, we have investigated Raman spectroscopy as a clinical tool for the diagnosis of a variety of breast pathologies.

Raman spectroscopy is an inelastic scattering process in which photons incident on a sample transfer energy to or from molecular vibrational modes (16). It is a coherent two-photon

---

BIOPHYSICS

process in which a molecule simultaneously absorbs an incident photon and emits a Raman photon, accompanied by its transition from one energy level to another, giving rise to a frequency (i.e., energy) shift of the emitted photon. Because the energy levels are unique for every molecule, Raman spectra are chemical-specific. Individual bands in the Raman spectrum are characteristic of specific molecular motions. Raman spectroscopy is particularly amenable to *in vivo* measurements, because the powers and excitation wavelengths that are used do not affect the tissue and the penetration depth is relatively large (17).

Early studies based on small sample sets observed spectral trends, indicating the promise of Raman spectroscopy for breast cancer diagnosis (18–21). However, this initial work, which relied on peak height ratios for data analysis, was unable to differentiate benign lesions from malignant lesions. Given the intended application, in which there is an *a priori* expectation of nonnormal tissue, accurate distinction of benign and malignant lesions is crucial. In view of the wealth of information available from Raman spectroscopy and the biochemical complexity of breast lesions, a method of analysis that utilizes the entire Raman spectrum, rather than peak height ratios, is necessary to distinguish between benign and malignant lesions.

Our own initial research, based on principal component analysis of tissue Raman spectra, showed that benign and malignant tumors could be differentiated (22). Although principal component analysis utilizes the entire Raman spectrum, it affords little insight into the chemical changes responsible for disease diagnosis. To provide information about the chemical basis for diagnosis and understand the relationship between a tissue's Raman spectrum and its disease state, we developed a spectroscopic model of breast tissue (23). This model fits macroscopic tissue spectra with a linear combination of basis spectra derived from Raman microscopy of various breast tissue morphological structures. These basis spectra represent the epithelial cell cytoplasm, cell nucleus, fat, β-carotene, collagen, calcium hydroxyapatite, calcium oxalate dihydrate, cholesterol-like lipid deposits, and water. This modeling approach is based on the assumptions that the Raman spectrum of a mixture is a linear combination of the spectra of its components and that signal intensity and chemical concentration are linearly related (24). The resulting fit coefficients yield the contribution of each basis spectrum to the macroscopic tissue spectrum, thereby elucidating the chemical/morphological makeup of the lesion. These same morphological changes are routinely used by pathologists to diagnose disease. However, unlike conventional pathology, which is subject to interobserver variation, Raman spectroscopy assesses these changes in an objective, reproducible manner without the need for tissue removal (25, 26).

Here, our Raman spectroscopic model is used to characterize the chemical/morphological composition of a range of *ex vivo* breast tissue specimens and pathologies and to predict the breast tissue disease state. The data demonstrate the efficacy of the model in both diagnosing breast disease and understanding the chemical/morphological variations associated with disease progression. The results indicate that the technique has the potential to be applied *in vivo* to accurately classify breast lesions, thereby reducing the number of excisional breast biopsies that are performed.

Materials and Methods

Tissue Preparation. Breast tissue was obtained from patients who were undergoing surgical breast biopsy reduction mammoplasties and prophylactic mastectomies. Upon removal, the samples were snap-frozen in liquid nitrogen for storage and then passively thawed at room temperature and kept moist with PBS. After spectral acquisition, specimens were marked with India ink to indicate the region sampled, fixed in formalin, routinely processed, paraffin-embedded, cut through the marked locations

in 5-μm-thick sections, and stained with hematoxylin/eosin. The histological slides were examined by an experienced breast pathologist who was blinded to the outcome of the Raman spectroscopy analysis. A total of 130 spectra from 58 patients were examined by using Raman spectroscopy: 49 normal tissues from 25 patients (20 female, 2 male, and 3 unknown; 8 white, 12 black, and 5 unknown) with a mean age of 34.3 years (range of 13–75 years); 50 from benign lesions, 16 with fibrocystic change from 16 patients (10 female, 1 male, and 5 unknown; 5 white, 3 black, and 8 unknown) with a mean age of 40.3 years (range of 13–75 years), 3 from lesions diagnosed as ductal epithelial hyperplasia (DEH) from 2 white female patients with ages of 11 and 49 years, and 31 from fibroadenomas from 6 patients (4 female and 2 unknown; 2 black and 4 unknown) with a mean age of 20.8 years (range of 13–40 years); and 31 from malignant lesions all diagnosed as infiltrating carcinoma (10 ductal, 2 lobular, 1 ductal and lobular, and 1 mammary, not otherwise specified) from 16 patients (8 female and 8 unknown; 5 white, 2 black, and 9 unknown) with a mean age of 57.6 years (range of 46–77 years). Patient information was not available for a sample that was diagnosed as fat necrosis. Because multiple spectra were collected from each patient, some tissue samples are included in both the normal and diseased categories, depending on the pathology underlying the exact region of data collection.

The mean ages and age ranges for the patients in this study reflect the natural age incidence of each lesion (27). The peak age of incidence for stromal fibroplasia, the predominant manifestation of fibrocystic change encountered in this study, is in the fourth and fifth decades (30s and 40s). That of fibroadenoma is much earlier [the third decade (20s)], and fibroadenomas account for the majority of breast lesions requiring biopsy in that age range. The peak age of incidence for infiltrating carcinoma is the sixth decade (50s).

Raman Spectroscopic Measurements. Data were acquired by using a Raman system described in refs. 28 and 29. The excitation spot is $\approx$100 μm in diameter, and light diffusion in the tissue results in a sampled volume of $\approx$1 mm^3. Raman spectra were acquired with a 10- to 30-s integration time, depending on signal intensity, and a spectral resolution of 8 cm^{-1}. The average laser excitation power varied between 100 and 150 mW. The fluences used in this study are safe for clinical investigations (30). No tissue damage was observed, either grossly or upon histological review.

Data Processing. Data processing was preformed as described in refs. 23 and 28. Model fitting was performed by using a linear combination of basis spectra with a nonnegativity constraint. The contribution of each basis spectrum, obtained from the model described above, to the breast tissue specimens was acquired by normalizing the fit coefficients (excluding water, because it is applied exogenously) such that they sum to one. To determine the error in our fit coefficients, we used a χ^2 analysis (31). χ^2 analysis is a well known method for calculating the goodness of a fit as well as the error associated with model fitting. The error bars (one SD), shown in Fig. 3, are generated from this analysis. The Raman spectra in each diagnostic group have different signal-to-noise ratios; thus, mean errors are reported for each pathology. Fitting errors for the two diagnostic model components, fat and collagen, are 0.010 and 0.006 for normals, 0.040 and 0.022 for fibrocystic change, 0.012 and 0.006 for fibroadenoma, and 0.034 and 0.016 for infiltrating carcinoma, respectively. Errors are slightly larger for fat than for collagen because the Raman spectrum of fat has more similarity to other model components than that of collagen.

Logistic regression, a discriminate analysis technique, was used to correlate the normalized fit coefficients with the diagnostic categories (32) for all combinations of the eight morphological components in the model. A likelihood ratio test was used
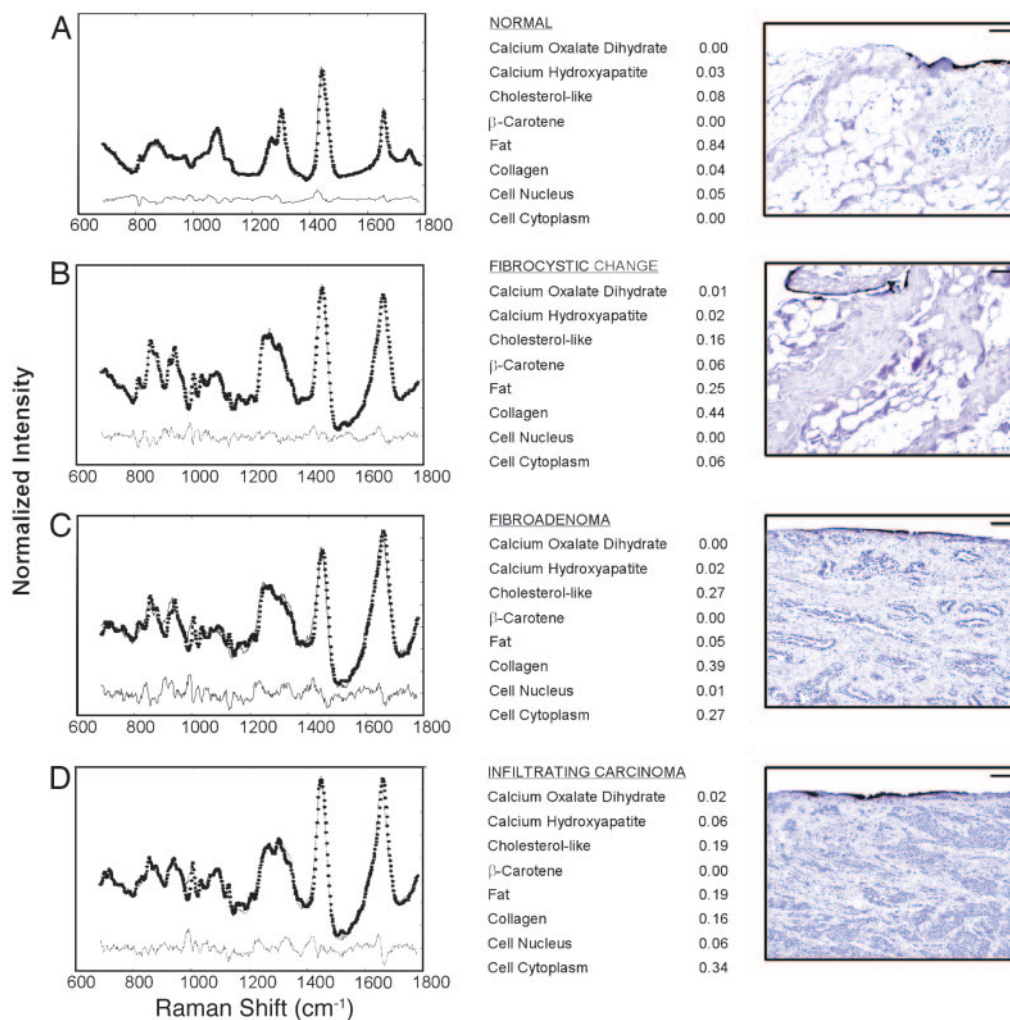
Fig. 1. Model fits. Normalized Raman spectra (solid lines), model fits (dotted lines), residuals (shown below), fit coefficients, and images from hematoxylin/eosin (H&E)-stained sections used to make the histopathologic diagnosis for normal breast tissue (*A*), fibrocystic change (*B*), fibroadenoma (*C*), and infiltrating carcinoma (*D*). The India ink used to record the region of spectral examination is seen as a black line on the tissue surface in the H&E images. (Scale bars, 100 μm.)

to determine which fit coefficients were significant for diagnosis and what probability thresholds, based on these fit coefficients, correctly classified the most samples. To cross-validate our algorithm, we used a leave-one-out cross-validation analysis, which enabled recycling of the data to produce a more robust diagnostic algorithm. Using maximum likelihood estimation, we determined the probability that a breast sample is normal, fibrocystic change, fibroadenoma, or invasive carcinoma. Receiver operating characteristic (ROC) curves were generated by changing the probability threshold for assigning a classification. The ROC curves in Fig. 4 display unique values of sensitivity and specificity.

Results and Discussion

Model Fits. To understand the relationship between a tissue sample's Raman spectrum and its disease state, we examined the contribution of each model basis spectrum to spectra acquired from a variety of pathologies. Model fits to Raman spectra acquired from normal, benign, and malignant samples of breast tissue are shown in Fig. 1, with corresponding images from hematoxylin/eosin-stained sections used to make the histopathologic diagnosis. The difference between the measured spectrum and the model fit, the residual, is shown below each spectrum. The lack of significant structure in the residuals

demonstrates that the model accounts for the majority of the spectroscopic features observed and describes the data well. The fit coefficients, also displayed in Fig. 1, represent the amount that each model basis spectrum must be weighted to recreate the tissue spectrum, thereby providing insight into the chemical/morphological makeup of the tissue. Fit coefficients are a function of both the concentration of a particular model component and its Raman scattering cross section (which indicates the strength of the signal at unit concentration).

Four of the 130 spectra in our data set were excluded from diagnostic algorithm development because of an insufficient number of samples: three from benign samples diagnosed as DEH and one from a benign sample diagnosed as fat necrosis. The sample diagnosed as fat necrosis exhibited an intense fluorescent background, most likely due to lipids oxidized during tissue necrosis that, with a larger data set, may prove to be diagnostic of fat necrosis.

Spectral Fit Coefficients and Tissue Morphology. The fit coefficients, given by the model and normalized to sum to one, represent contributions of chemicals and morphological features to the macroscopic tissue spectrum. Fig. 2 shows histograms of the average fit coefficients for normal breast tissue, fibrocystic change, fibroadenoma, and infiltrating carcinoma. In Fig. 2*A*,
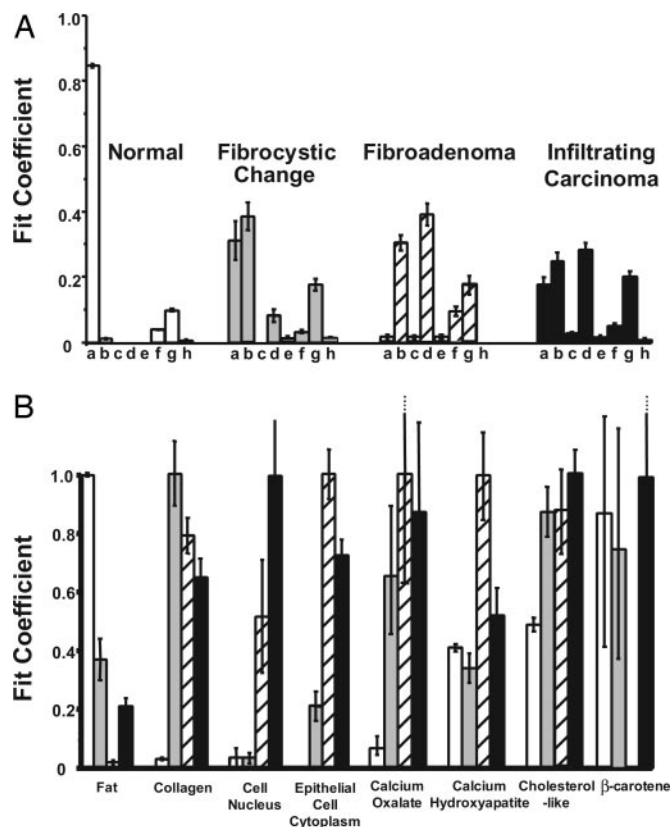
Fig. 2. Histograms displaying the average composition of samples diagnosed as normal (white), fibrocystic change (gray), fibroadenoma (striped), and infiltrating carcinoma (black). The one-SD confidence intervals are shown for each model component. (*A*) Data are grouped according to pathological diagnosis: fat (a), collagen (b), cell nucleus (c), epithelial cell cytoplasm (d), calcium oxalate (e), calcium hydroxyapatite (f), cholesterol-like (g), and β-carotene (h). (*B*) Data are clustered by model component, and each histogram cluster is normalized to the largest of the four values.

data are grouped according to pathologic diagnosis; in Fig. 2*B*, data are clustered by model component. In Fig. 2*B*, each histogram cluster is normalized to the largest of the four values to highlight the relative changes between pathologies. The one-SD confidence intervals are also displayed in Fig. 2 to illustrate the variation in fit coefficients of a particular model component between samples with the same pathology.

Examination of the fit coefficients for each pathology provides insight into the chemical changes associated with breast disease. The fit coefficients of normal breast tissue indicate that it is primarily composed of fat. Normal breast tissue contains both glandular and adipose tissues (33). Glandular tissue consists of ducts lined by epithelial cells and a supportive collagenous extracellullar matrix. Adipose tissue is primarily composed of adipocytes (cells containing large amounts of cytoplasmic fat), although small quantities of extracellular matrix are present. Overall, ducts represent only a small volume of the tissue, and thus our model accurately characterizes normal breast tissue as predominately composed of fat with small contributions from collagen. Contributions from fat are particularly prominent because adipose tissue has a large Raman scattering cross section relative to most other model components. The model does not show a contribution from epithelial cells to normal breast tissue. However, the lack of an epithelial cell contribution does not mean that the normal samples do not contain epithelial cells; rather, it means that they are not present in sufficient quantities

or with a strong enough Raman scattering cross section to appreciably contribute to the macroscopic Raman spectrum.

The fit coefficients of the breast lesions indicate a markedly different chemical/morphological composition than that of normal breast tissue. First, the amount of collagen increases in all abnormal breast tissues. This finding is consistent with known breast pathology, because lesion formation is often accompanied by fibrosis, a scarring process characterized by an increased stromal component, and thus by both proliferation of fibroblasts and accumulation of collagen. The relative increase in collagen is most pronounced in fibrocystic change, a benign condition that can manifest itself as fibrosis, adenosis (an increase in the number of ducts), or cyst formation (dilation of ducts with fluid). Each of these changes can occur with or without the others. In the fibrocystic lesions examined in the present study, increases in the fit coefficients of the collagen, epithelial cell cytoplasm, and cholesterol-like basis spectra replace the large contribution from fat in normal breast tissue.

Fibroadenoma is a benign tumor of a different lineage than all other lesions in this study (34). It is most closely related to phylloides tumors, the malignant counterpart of which is not carcinoma but cystosarcoma phylloides, in which the stroma, rather than the epithelium, is malignant. Lesions diagnosed as fibroadenoma show an increased contribution from collagen due to fibroblast proliferation and accumulation of collagen that results in expansion of the stroma (35). They also show an increased contribution from both the cell nucleus and epithelial cell cytoplasm basis spectra as a consequence of the number of fibroblasts and epithelial cells present.

Similar to fibroadenoma, lesions diagnosed as infiltrating carcinoma show an increased contribution from collagen, in this case due to fibroblast proliferation in response to stromal invasion by the malignant epithelial cells (35). The fit coefficients of such lesions also display an increase in the amount of epithelial cell cytoplasm and cell nucleus. Both infiltrating carcinomas and fibroadenomas exhibit large increases in the number of cells relative to other lesions. Enlargement of cell nuclei is a hallmark of cancer, and thus a higher nuclear-to-cytoplasm (N/C) ratio is a diagnostic criterion routinely used by pathologists (36, 37). In our studies, the spectroscopic parameter characterizing the N/C ratio is obtained by dividing the fit coefficient of the cell nucleus basis spectrum by that of the epithelial cell cytoplasm basis spectrum. Fibrocystic change and fibroadenoma have mean N/C parameters of 0.01 and 0.04, respectively, whereas infiltrating carcinoma has a much higher mean N/C parameter of 0.08. Although clear trends are seen in the mean values, there is significant variability of the N/C parameter within pathologies.

Further differences between lesions diagnosed as fibroadenoma and infiltrating carcinoma exist in the amount of fat present. Samples diagnosed as fibroadenoma have less fat than those diagnosed as infiltrating carcinoma, as can be seen in Fig. 2, because fibroadenoma is an expansile lesion that grows by pushing the fatty breast tissue aside. Infiltrating carcinoma, in contrast, infiltrates in between the fat cells, so some adipocytes are retained within the carcinoma.

Other breast pathologies are much less common and were not represented in the study specimens. Three samples were diagnosed as DEH, a proliferation of the ductal epithelium that confers an increased risk for breast cancer (38). The mean N/C parameter of these three lesions is 0.05, intermediate between that of normal breast and infiltrating carcinoma, indicating the potential for detecting precancerous changes in the breast by using Raman spectroscopy.

The histograms in Fig. 2 exhibit relatively little contribution from the two types of calcifications found in breast tissue, calcium hydroxyapatite and calcium oxalate dihydrate. Because of their diagnostic importance, microcalcifications in fresh breast tissue are not typically made available for scientific research, and
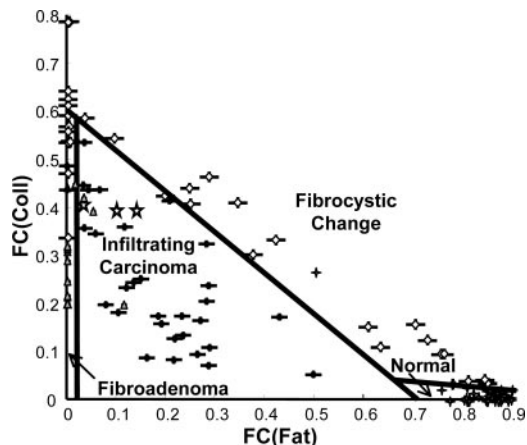
Fig. 3. Scatter plot displaying the fat and collagen content for all pathologies encountered in this study. Error bars (one SD) are indicated. Several of the error bars are smaller than the symbols used to denote each data point. Normal, gray stars; fibrocystic change, diamonds; fibroadenoma, triangles; infiltrating carcinoma, squares; DEH, white stars.

Fig. 4. ROC curves illustrating the ability of Raman spectroscopy to separate lesions diagnosed as infiltrating carcinoma from benign and normal breast tissues. The ROC curve of two indistinguishable populations, represented by the dashed line, is included for comparison.

thus our study did not include lesions containing microcalcifications. However, in a parallel study, we have shown that the Raman spectra of microcalcifications contain significant diagnostic information (39). As this technique moves into a clinical setting, microcalcifications will be encountered more frequently and in greater abundance, which should provide additional diagnostic information and increase diagnostic accuracy.

Diagnostic Algorithm. The fit coefficients not only provide insight into the composition of the tissue but also are used to develop diagnostic algorithms. A diagnostic algorithm was developed that examines all of the data simultaneously and is, to our knowledge, the first spectral-based algorithm to separate breast tissues according to specific pathological diagnoses.

The fit coefficients corresponding to the fat and collagen basis spectra were the key diagnostic parameters in differentiating pathologies. Samples were initially divided into two groups based on their collagen [FC(Coll)] and fat [FC(Fat)] content. One group contained infiltrating carcinomas and fibroadenomas; the second contained normal tissue and fibrocystic lesions. The equation for the decision line, drawn by logistic regression, is $FC(Coll) = -0.85FC(Fat) + 0.60$. The resultant two clusters were then further subdivided by using logistic regression. Again, the diagnostic power of all eight model components was assessed. Fibrocystic change and normal tissue were separated based on their fat and collagen contents, with the decision line given by $FC(Coll) = -0.06FC(Fat) + 0.10$. Fibroadenoma and infiltrating carcinoma were separated based solely on their fat content. The decision line for this separation is $FC(Fat) = 0.02$.

Fig. 3 displays a scatter plot of FC(Coll) and FC(Fat) for all pathologies encountered in this study, as well as the decision lines that separate samples according to diagnoses. Table 1
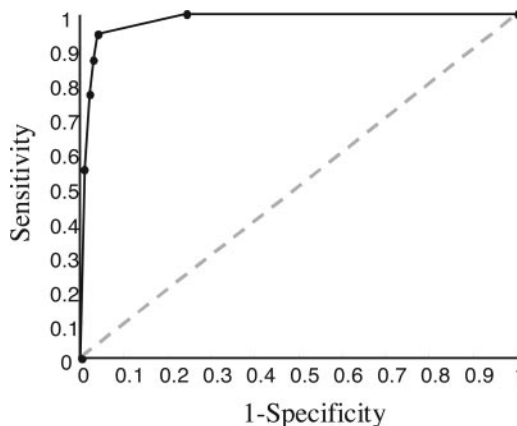
compares the pathologic diagnosis with that of the Raman diagnostic algorithm for our data set. The algorithm yields a sensitivity of 94% (29/31), specificity of 96% (91/95), and overall accuracy of 86% (108/126) for detecting infiltrating carcinoma. Also shown in Fig. 3 are the fat and collagen fit coefficients for the three samples diagnosed as DEH. In future applications, this diagnostic algorithm could be used in a prospective manner, and the fit coefficients of collagen and fat simply plotted to determine where they fall in the diagnostic plane.

Fig. 4 displays a ROC curve that illustrates the ability of Raman spectroscopy to separate lesions diagnosed as infiltrating carcinoma from benign and normal breast tissues. A ROC curve illustrates the tradeoff between sensitivity and specificity by plotting the true-positive rate against the false-positive rate for the different possible probability thresholds of a diagnostic test. The closer the curve comes to the 45-degree diagonal, shown as a dashed line, the less accurate the diagnostic test. The ROC curve clearly illustrates the ability of Raman spectroscopy to accurately diagnose breast cancer and demonstrates how the diagnostic scheme can be adjusted to obtain the desired degree of sensitivity at the cost of specificity.

Effect of Age on the Diagnostic Algorithm. The female breast undergoes substantial biochemical alterations at menopause. Specifically, a large amount of collagen is replaced by fat, resulting in breast tissue that is less dense on mammography. Because our diagnostic algorithm is based on fat and collagen content as assessed by Raman spectroscopy, we investigated trends in these two parameters as a function of patient age. It is notable that although the normal samples span a wide age range, our technique characterizes them as predominately composed of fat because the Raman scattering cross section is much higher for fat than for collagen. Thus, we observe significant contributions from collagen only when

Table 1. Comparison of the pathologic diagnosis with that of the Raman diagnostic algorithm

	Pathology diagnosis			
Raman diagnosis	Normal (49 spectra)	Fibrocystic change (31 spectra)	Fibroadenoma (15 spectra)	Infiltrating carcinoma (31 spectra)
Normal	45	1	0	0
Fibrocystic change	4	22	0	0
Fibroadenoma	0	7	12	2
Infiltrating carcinoma	0	1	3	29

it is present in large quantities (for instance, in the setting of dense stromal fibrosis). To rigorously confirm that our diagnostic algorithm is not influenced by patient age, we examined the correlation between the spectroscopic parameter characterizing the fat-to-collagen ratio, obtained by dividing FC(Fat) by FC(Coll), and age. Fibroadenoma was excluded from this analysis because it is a juvenile disease and has a much different age range than the other pathologies in our study (27). We found a correlation coefficient of -0.140 for the spectroscopic parameter characterizing the fat-to-collagen ratio and age. We also examined the relationship between the fat-to-collagen ratio and age within individual pathologies, obtaining correlation coefficients of -0.104, 0.137, and -0.025 for invasive carcinoma, fibrocystic change, and normals, respectively. Thus, we do not observe age-dependent trends in the fat and collagen contents of our data. Our results indicate that Raman spectroscopy is much less sensitive to breast density and menopausal status than many other optical techniques.

Implications. We have demonstrated the ability of Raman spectroscopy to diagnose benign and malignant breast lesions with high sensitivity and specificity *ex vivo* in a laboratory setting. These diagnoses are based on chemical/morphological changes that are known to accompany breast disease. These excellent *ex vivo* results are a necessary step toward data acquisition in hospital settings, both *ex vivo* directly following biopsy and *in vivo* during breast surgery. These preclinical and clinical studies are made possible by the recent development of a Raman optical fiber probe designed for medical applications (40). The probe is optimized to collect high signal-to-noise ratio data from tissue in clinically relevant times (1 s).

1. American Cancer Society (2005) *Breast Cancer Facts and Figures 2004*. Available at www.cancer.org/docroot/STT/stt_0_2004.asp. Accessed Aug. 5, 2005.
2. Johnson, J., Dalton, R., Wester, S., Landercasper, J. & Lambert, P. (1999) *Arch. Surg. (Chicago)* **134,** 712–715.
3. Liberman, L., Dershaw, D. D., Rosen, P. P, Abramson, A. F., Deutch, B. M. & Hann, L. E. (1994) *Radiology* **192,** 793–795.
4. Fantini, S., Walker, S. A., Franceschini, M. A., Kaschke, M., Schlag, P. M. & Moesta, K. T. (1998) *Appl. Opt.* **37,** 1982–1989.
5. Hebden, J., Veenstra, H., Dehghani, H., Hillman, E., Schweiger, M., Arridge, S. & Delpy, D. (2001) *Appl. Opt.* **40,** 3278–3287.
6. Quaresima, V., Matcher, S. J. & Ferrari, M. (1998) *Photochem. Photobiol.* **67,** 4–14.
7. Tromberg, B. J., Coquoz, O., Fishkin, J. B., Pham, T., Anderson, E. R., Butler, J., Cahn, M., Gross, J. D., Venugopalan, V. & Pham, D. (1997) *Philos. Trans. R. Soc. London B* **352,** 661–668.
8. Shah, N., Cerussi, A., Eker, C., Espinoza, J., Butler, J., Fishkin, J., Hornung, R. & Tromberg, B. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 4420–4425.
9. Ntziachristos, V., Yodh, A. G., Schnall, M. & Chance, B. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 2767–2772.
10. Fornage, B. D., Sneige, N., Ross, M. I., Mirza, A. N., Kuerer, H. M., Edeiken, B. S., Ames, F. C., Newman, L. A., Babiera, G. V. & Singletary, S. E. (2004) *Radiology* **231,** 215–224.
11. Gupta, P. K., Majumder, S. K. & Uppal, A. (1997) *Lasers Surg. Med.* **21,** 417–422.
12. Majumder, S. K., Gupta, P. K., Jain, B. & Uppal, A. (1998) *Lasers Life Sci.* **8,** 249–264.
13. Yang, Y., Katz, A., Celmer, E. J., Zurawska-Szczepaniak, M. & Alfano, R. R. (1997) *Photochem. Photobiol.* **66,** 518–522.
14. Bigio, I. J., Bown, S. G., Briggs, G., Kelley, C., Lakhani, S., Pickard, D., Ripley, P. M, Rose, I. G. & Saunders, C. (2000) *J. Biomed. Opt.* **5,** 221–228.
15. Yang, Y., Katz, A., Celmer, E. J., Zurawska-Szczepaniak, M. & Alfano, R. R. (1996) *Lasers Life Sci.* **7,** 115–127.
16. Raman, C. V. & Krishnan, K. S. (1928) *Nature* **121,** 501–502.
17. Hanlon, E. B., Manoharan, R., Koo, T. W., Shafer, K. E., Motz, J. T., Fitzmaurice, M., Kramer, J. R., Itzkan, I., Dasari, R. R. & Feld, M. S. (2000) *Phys. Med. Biol.* **45,** R1–R59.
18. Alfano, R. R., Liu, C. H., Sha, W. L., Zhu, H. R., Akins, D. L., Cleary, J., Prudente, R. & Cellmer, E. (1991) *Lasers Life Sci.* **4,** 23–28.
19. Frank, C. J., McCreery, R. L. & Redd, D. C. B. (1995) *Anal. Chem.* **67,** 777–783.
20. Frank, C. J., Redd, D. C. B., Gansler, T. S. & McCreery, R. L. (1994) *Anal. Chem.* **66,** 319–326.
21. Redd, D. C., Feng, Z. C., Yue, K. T. & Gansler, T. S. (1993) *Appl. Spectrosc.* **47,** 787–791.
22. Manoharan, R., Shafer, K., Perelman, L., Wu, J., Chen, K., Deinum, G., Fitzmaurice, M., Myles, J., Crowe, J., Dasari, R. & Feld, M. (1998) *Photochem. Photobiol.* **67,** 15–22.
23. Shafer-Peltier, K. E., Haka, A. S., Fitzmaurice, M., Crowe, J., Dasari, R. R. & Feld, M. S. (2002) *J. Raman Spectrosc.* **33,** 552–563.
24. Buschman, H. P., Deinum, G., Motz, J. T., Fitzmaurice, M., Kramer, J. R., van der Laarse, A., Bruschke, A. V. & Feld, M. S. (2001) *Cardiovasc. Pathol.* **10,** 69–82.
25. Elston, C. W., Sloane, J. P., Amendoeira, I., Apostolikas, N., Bellocq, J. P., Bianchi, S., Boecker, W., Bussolati, G., Coleman, D., Connolly, C. E., *et al.* (2000) *Eur. J. Cancer* **36,** 1769–1772.
26. Tsuda, H., Akiyama, F., Kurosumi, M., Sakamoto, G. & Watanabe, T. (1999) *Jpn. J. Clin. Oncol.* **29,** 413–420.
27. Rosen, P. P. & Oberman, H. A. (1992) *Atlas of Tumor Pathology: Tumors of the Mammary Gland* (Armed Forces Institute of Pathology, Washington, D.C.), 3rd Series, pp. 101, 157.
28. Shafer-Peltier, K. E., Haka, A. S., Motz, J. T., Fitzmaurice, M., Dasari, R. R. & Feld, M. S. (2002). *J. Cell. Biochem.* **87,** S125–S137.
29. Brennan, J. F. B., Wang, Y., Dasari, R. R. & Feld, M. S. (1997) *Appl. Spectrosc.* **51,** 201–208.
30. Motz, J. T. (2003) Ph.D. thesis (Massachusetts Institute of Technology, Cambridge, MA).
31. Kenney, J. F. & Keeping, E. S. (1951) *Mathematics of Statistics* (Van Nostrand, Princeton), Part 2, 2nd Ed.
32. Shama, S. (1996) *Applied Multivariate Techniques* (Wiley, New York).
33. Rosen, P. P. (1997) *Breast Pathology* (Lippincott–Raven, Philadelphia).
34. Houssami, N., Cheung, M. N. & Dixon, J. M. (2001) *Med. J. Aust.* **174,** 185–188.
35. Cotran, R. S., Kumar, V. & Robbins, S. L. (1989) *Robbins Pathologic Basis of Disease* (Saunders, Philadelphia), 4th Ed., pp. 1189, 1195.
36. Elston, C. W. & Ellis, I. O. (1991) *Histopathology* **19,** 403–410.
37. Hoda, S. A. & Rosen, P. P. (2002) *Am. J. Clin. Pathol.* **118,** 101–108.
38. Page, D. L. & Dupont, W. D. (1990) *Cancer* **66,** 1326–1335.
39. Haka, A. S., Shafer-Peltier, K. E., Fitzmaurice, M., Crowe, J., Dasari, R. R. & Feld, M. S. (2002) *Cancer Res.* **62,** 5375–5380.
40. Motz, J. T., Hunter, M., Galindo, L. H., Gardecki, J. A., Kramer, J. R., Dasari, R. R. & Feld, M. S. (2003) *Appl. Opt.* **43,** 542–554.