

AD _____

Award Number: W81XWH-05-1-0380

TITLE: Fluorescence and Diffuse Reflectance Spectroscopy for Breast Cancer Diagnosis during Core Needle Biopsy

PRINCIPAL INVESTIGATOR: Changfang Zhu Ph.D.

CONTRACTING ORGANIZATION: University of Wisconsin
Madison, WI 53706

REPORT DATE: September 2007

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 1-SEP-2007		2. REPORT TYPE Annual Summary		3. DATES COVERED (From - To) 31 AUG 2006 - 30 AUG 2007	
4. TITLE AND SUBTITLE Fluorescence and Diffuse Reflectance Spectroscopy for Breast Cancer Diagnosis during Core Needle Biopsy				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER W81XWH-05-1-0380	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Changfang Zhu Ph.D E-Mail: czhu@cae.wisc.edu				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) ORGANIZATION: University of Wisconsin Madison, WI 53706				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The goal of this project is to explore the potential of using tissue fluorescence and diffuse reflectance spectroscopy for breast cancer detection during a core needle breast biopsy. Both empirically based and model based approaches have been explored for the extraction of diagnostically useful information from the tissue spectra, as well as the diagnosis of breast cancer based on these extracted features. For the sample set investigated in our study, both intrinsic tissue properties extracted using model based approach and spectral features extracted using empirical spectral analysis could provide sensitivity and specificity of up to 89% for discriminating breast malignancy. These outcomes provided the foundation for using the optical sensor based on tissue fluorescence and diffuse reflectance spectroscopy as an adjunct diagnostic tool, which has the potential to provide guidance for core needle breast biopsy.					
15. SUBJECT TERMS Fluorescence, Diffuse reflectance, Spectroscopy, Breast Cancer, Diagnosis, Biopsy					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 56	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (include area code)

Table of Contents

Introduction.....	4
Body.....	4
Key Research Accomplishments.....	7
Reportable Outcomes.....	8
Conclusions.....	9
References.....	9
Appendices.....	A-1

Introduction

Over the past 50 years, breast cancer has become a major health problem for women. Currently finding breast cancer early and treating it are the most important strategies to fight this disease. Core needle biopsy has been increasingly used in the breast cancer care for the early diagnosis of this disease. However, the caveat of this minimally invasive procedure is the limited sampling accuracy. The long-term goal of this research has been to develop an optical sensor based on tissue fluorescence and diffuse reflectance spectroscopy, which has the potential to serve as adjunct diagnostic tool to guide the biopsy thus improving the sampling accuracy. The idea is that this optical sensor will provide non-destructive diagnosis of tissue pathology, which the physicians can use to determine the optimal tissue site for biopsy. My research has been dedicated to develop the prototype optical sensor, and to test the feasibility of *in vivo* optical spectroscopy of breast tissues during a percutaneous core needle biopsy and its potential for distinguishing between malignant and benign breast tissues.

Body

In the first stage of the research (reported in annual report 2006), we have developed a novel fiber optic probe, which is compatible with a 9-gauge core biopsy needle and used in clinical breast biopsy procedures for *in vivo* fluorescence spectroscopy of breast tissues. The clinical trials and sample collection are ongoing and by the end of July 2007, we have performed optical spectroscopy on 81 patients, from which a total of 132 tissue samples have been collected for analysis.

The tissue spectra (both fluorescence and diffuse reflectance) contain a wealth of information about the physiological, morphological and biochemical information about the tissue properties. The diffuse reflectance spectra reflect the absorption and scattering properties of the tissue, and the fluorescence spectra contain the biochemical composition of the tissue [1]. Some of these properties are the “fingerprints” of breast cancer in the tissue spectra that can be used for the diagnosis of this disease. One of the key problems in tissue spectroscopy and its diagnostic application is to explore this diagnostically useful information by systematically analyzing the measured tissue spectra. This has also been a specific goal of my research.

In the past year, the primary focus of my work has been: (1) to develop the methods for extracting the diagnostically useful information from the measured tissue spectra, (2) to explore the spectroscopic contrast between malignant and non-malignant breast tissues, and (3) to diagnose breast malignancy based on the extracted spectroscopic contrast. This work has been carried out using the tissue spectra measured *ex vivo* from patients undergoing a breast surgery. Two approaches have been explored for the analysis of tissue spectra, one is empirical spectral analysis that has been commonly used in previous studies, and the other is a model based approach which is novel and recently developed by our group [2, 3].

In the empirical spectral analysis, I have employed a Partial Least Squares (PLS) analysis method to extract a set of principal components (PCs) that can be used to represent the tissue spectra with dramatically reduced dimension yet with minimal information loss [4]. The extracted PCs characterized the difference in the spectral line shapes between malignant and non-malignant breast tissues. Results from the study showed that for the sample set investigated, the

extracted PCs could provide a sensitivity of up to 87% and a specificity of up to 89% for discriminating breast malignancy.

In the model based approach, I have employed a Monte Carlo based inverse model of diffuse reflectance and a Monte Carlo model of fluorescence, both of which were developed by our group [2, 3], to extract the intrinsic absorption, scattering and fluorescence properties of breast tissues. These intrinsic tissue properties include β -carotene concentration, total hemoglobin concentration, hemoglobin saturation, mean reduced scattering coefficients, and fluorescence contributions of collagen, NADH and a third component we attributed to retinol. The relative fluorescence contributions of individual fluorophores, as well as β -carotene concentration, hemoglobin saturation and mean reduced scattering coefficient displayed statistically significant difference between malignant and non-malignant breast tissues. Classification based on these tissue properties yielded sensitivity and specificity of up to 89% for discriminating breast malignancy [5].

Both approaches extracted diagnostically useful features from the fluorescence and diffuse reflectance spectra and yielded comparable classification accuracy, which suggests that both approaches are equally effective for the discrimination of breast malignancy. A correlation study also showed that the features extracted using both approaches are significantly correlated, suggesting that both approaches may probe the same spectroscopic contrast in the tissue that discriminate between malignant and non-malignant breast tissues albeit in different ways. Each approach has its advantages and disadvantages though. The model based approach extracts quantitative information and is computationally intensive, however it allows for a comprehensive

understanding of the biochemical, physiological and morphological changes associated with the disease. The empirical approach can only provide qualitative information about the spectral features, however it is fast and computational effective, which has important implication to the clinical applications such as providing real-time feedback to guide the breast biopsy, where processing speed is critical.

The work described above demonstrated that optical spectroscopy could be used to effectively discriminate malignant from non-malignant breast tissues, which provided solid foundation to move the research forward. One other specific goal of my research is to demonstrate the potential of this technique for *in vivo* diagnosis of breast cancer during the core needle breast biopsy. During the first two years of this research, we have established the feasibility of performing optical spectroscopy during the clinical biopsy procedure, and a total of 132 tissue samples have been collected for analysis. The next step of my research is to apply the approaches that we have developed in the past year to analyze the spectra data and demonstrate the potential of diagnosing breast malignancy using the *in vivo* tissue spectra. Given the advantage of computational effectiveness of the empirical approach, I will use this method for the analysis of tissue spectra measured *in vivo* from patients undergoing a core needle biopsy. This work is expected to demonstrate the potential and effectiveness of *in vivo* diagnosis of breast cancer.

Key Research Accomplishment

- Continued the clinical trials and collected *in vivo* tissue spectra from a total of 81 patients

- ❑ Developed empirical approach for the analysis of tissue spectra to extract the diagnostically useful spectral features
- ❑ Explored the use of Monte Carlo model based approach for the extraction of intrinsic optical properties, including absorption, scattering and fluorescence properties of breast tissues.
- ❑ Explored the contrast in spectral features and intrinsic tissue properties between malignant and non-malignant breast tissues
- ❑ Tested the diagnostic potentials of using the extracted spectral features or intrinsic tissue properties for discriminating breast malignancy
- ❑ Investigated the relationship between the spectral features and the intrinsic tissue properties to understand the underlying physiological, structural and biochemical basis of the diagnostically useful spectral features

Reportable Outcomes

Conference presentations

- ❑ Changfang Zhu, Gregory M. Palmer, Tara M. Breslin, Josephine Harter, Nirmala Ramanujam, Diagnosis of Breast Cancer using Fluorescence Spectroscopy: a Physical Model Based Approach, Engineering Conference International, Naples, FL, June 10 – 14, 2007
- ❑ Changfang Zhu, Elizabeth S. Burnside, Gale Sisney, Josephine Harter, Nirmala Ramanujam, Fluorescence Spectroscopic Characterization of Breast Tissue Pathology during Core Needle Breast Biopsy, Engineering Conference International, Naples, FL, June 10 – 14, 2007

Journal publications

- Changfang Zhu, Gregory M. Palmer, Tara M. Breslin, Josephine Harter, and Nirmala Ramanujam, "Diagnosis of Breast Cancer using Fluorescence and Diffuse Reflectance Spectroscopy: a Monte Carlo Model Based Approach", submitted to Journal of Biomedical Optics, 2007.
- Changfang Zhu, Tara M. Breslin, Josephine Harter, and Nirmala Ramanujam, "Model Based and Empirical Spectral Analysis for the Diagnosis of Breast Cancer", in preparation.

Conclusions

Optical spectroscopy has the potential to be used as a diagnostic tool for breast cancer. There are statistically significant differences in the fluorescence, absorption and scattering properties between malignant and non-malignant breast tissues, which are also reflected in the different spectral features observed in the fluorescence and diffuse reflectance spectra between normal and diseased breast tissues. For the sample set investigated in our study, both tissue properties and spectral features can provide sensitivity and specificity of up to 89% for discriminating breast malignancy. These outcomes provided the foundation for using the optical sensor based on tissue fluorescence and diffuse reflectance spectroscopy as an adjunct diagnostic tool, which has the potential to provide guidance for core needle breast biopsy.

References

- [1] R. Richards-Kortum and E. Sevick-Muraca, "Quantitative optical spectroscopy for tissue diagnosis," *Annu Rev Phys Chem* 47(555-606 (1996)

- [2] 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)
- [3] G. M. Palmer and N. Ramanujam, "Monte Carlo based Model to Extract Intrinsic Fluorescence from Turbid Media: Theory and Phantom Validation," *J Biomed Opt* submitted (2007)
- [4] C. Zhu, G. M. Palmer, T. M. Breslin, J. Harter, and N. Ramanujam, "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, 714-724 (2006)
- [5] C. Zhu, G. M. Palmer, T. M. Breslin, J. Harter, and N. Ramanujam, "Diagnosis of Breast Cancer using Fluorescence and Diffuse Reflectance Spectroscopy: a Monte Carlo Model Based Approach," *J Biomed Opt* submitted (2007)

Appendices – Manuscript Submitted for Journal Publication

Diagnosis of Breast Cancer using Fluorescence and Diffuse Reflectance

Spectroscopy: a Monte Carlo Model Based Approach

Changfang Zhu¹

¹Department of Electrical and Computer Engineering

University of Wisconsin – Madison, WI 53705

Email: czhu@cae.wisc.edu

Gregory M. Palmer²

²Department of Radiation Oncology

Duke University, Durham, NC 27710

Tara M. Breslin³

³Division of Surgical Oncology

University of Michigan, MI 48109

Josephine Harter⁴

⁴Department of Pathology

University of Wisconsin – Madison, WI 53705

***Nirmala Ramanujam⁵**

⁵Department of Biomedical Engineering

Duke University, Durham, NC 27708

For correspondence:

***Nimmi Ramanujam**

Dept. Biomedical Engineering

Duke University

136 Hudson Hall

Box 90281, Campus

Durham, NC 27708-0281

Email: nimmi@duke.edu

Phone:(919) 660-5307

Fax:(919) 684-4488

Abstract

We explored the use of Monte Carlo model based approaches for the analysis of fluorescence and diffuse reflectance spectra measured *ex vivo* from breast tissues, to extract the absorption, scattering and fluorescence properties of malignant and non-malignant tissues and to diagnose breast cancer based on these intrinsic tissue properties. Absorption and scattering properties were derived from diffuse reflectance spectra using a Monte Carlo based inverse model of diffuse reflectance, which include β -carotene concentration, total hemoglobin concentration, hemoglobin saturation and mean reduced scattering coefficient. A Monte Carlo model of fluorescence developed by our group and described in the companion manuscript was employed to retrieve the intrinsic fluorescence spectra of breast tissues. The intrinsic fluorescence spectra were decomposed into several contributing components, which we attribute to endogenous fluorophores present in breast tissues including collagen, NADH and retinol/Vitamin A. The model based approaches remove the dependency of fluorescence and diffuse reflectance measurements on instrument and probe geometry. These approaches were applied independently to three sets of tissue spectra measured with two different instruments and probe systems, and the tissue properties extracted from two sets of spectra were pooled together for discrimination analysis. The relative fluorescence contributions of individual fluorophores, as well as β -carotene concentration, hemoglobin saturation and mean reduced scattering coefficient displayed statistically significant difference between malignant and adipose breast tissues. The hemoglobin saturation and the reduced scattering coefficient displayed statistically significant differences between malignant and fibrous/benign breast tissues. A linear SVM classification using: 1)

fluorescence properties alone, 2) absorption and scattering properties alone, and 3) the combination of all tissue properties achieved comparable classification accuracies of 81-84% in sensitivity and 75-89% in specificity for discriminating malignant from non-malignant breast samples, suggesting each set of tissue properties are diagnostically useful for the discrimination of breast malignancy.

Key words: fluorescence, diffuse reflectance, spectroscopy, intrinsic fluorescence, optical property, Monte Carlo, breast cancer

1. Introduction

Fluorescence and diffuse reflectance spectroscopy in the ultraviolet-visible (UV-VIS) spectrum, are increasingly being investigated for breast cancer diagnosis (1-15). This technology is cost-effective, and can be deployed through fiber-optic probes to quickly, non-destructively and quantitatively detect the structural, physiological and biochemical changes that are associated with the progression of disease. Several studies have demonstrated that there are significant differences in the fluorescence (1-3, 5, 7, 9, 11) and diffuse reflectance (4, 13-17) spectra of normal, benign and malignant breast tissues, which could be used for the diagnosis of breast malignancy.

In most previous studies, fluorescence (1-3, 5, 7, 9, 11) and/or diffuse reflectance spectra (1, 4, 5, 8, 13, 15) of breast tissues were analyzed using empirical methods, in which the spectral intensities and/or line shapes were examined to extract spectral patterns (such as intensities, ratio of intensities, or principal components) that consistently discriminate between malignant and non-malignant tissues. For example, in one of the first reported studies, Alfano et al (11) used the ratio of fluorescence intensities at 340 and 440 nm, excited at 300 nm, to distinguish malignant from normal breast tissues. While several studies (1-4, 7, 11, 13) examined the spectral intensities, there are other studies where spectral line shapes were explored for breast cancer diagnosis. For example, Palmer et al. (5) employed Principal Component Analysis (PCA) to extract a set of principal components that dramatically reduced the spectral data dimension with minimal information loss, and showed that a subset of principal components obtained from diffuse reflectance spectra and fluorescence spectra at excitation wavelengths of

300, 400, 420 and 460 nm displayed the statistically most significant differences between malignant and non-malignant breast tissues. The collection of studies above examined spectral intensities or line shapes of turbid tissue fluorescence and/or diffuse reflectance spectra. Although empirical analyses can reveal important spectroscopic features for tissue characterization and disease diagnosis, they do not relate the measured spectra directly to the physically meaningful information that are contained in the spectra.

The diffuse reflectance spectrum reflects the absorption and scattering properties of the tissue. The absorption coefficient is directly related to the concentration of physiologically relevant absorbers in the tissue such as oxygenated and deoxygenated hemoglobin. The scattering coefficient reflects the size and density of scattering centers in tissue, such as cells and nuclei. The fluorescence spectrum contains biochemical information about a number of molecules that are relevant to the structural properties of tissues or respiratory metabolism of cells. However, there are distortions in the fluorescence spectra due to the interplay of tissue absorption and scattering, which makes it difficult to interpret the biochemically relevant information contained in the fluorescence spectra. Extracting and quantitatively assessing the intrinsic and physically meaningful information from the fluorescence and diffuse reflectance measurements will provide insight into the biochemical, physiological and morphological changes that take place during the progression of disease.

Several studies have been reported in which the absorption and scattering properties were quantified from diffuse reflectance spectra to characterize breast tissues.

Ghosh et al. (16) used a model based on the standard diffusion approximation to extract the absorption and scattering coefficients from spatially resolved diffuse reflectance measurements of human breast tissue samples. They showed that malignant tissues had increased absorption and scattering relative to non-malignant breast tissues. However, they did not report the classification accuracy for diagnosing breast cancer based on the parameters extracted from the model. More recently our group, in two independent studies (17, 18) applied a Monte Carlo based inverse model for the extraction of absorption and scattering properties from the diffuse reflectance spectra of malignant and non-malignant breast tissues. Both studies showed that malignant tissues have decreased hemoglobin saturation and increased mean reduced scattering coefficient compared to non-malignant tissues. Classification based on machine learning using these absorption and scattering properties as inputs achieved an unbiased sensitivity and specificity of 82% and 92%, respectively in the study by Palmer et al (18), and 86% and 80%, respectively in the study by Zhu et al (17), for discriminating between malignant and non-malignant breast tissues.

For fluorescence, a number of studies (19-25) have been carried out to disentangle the effects of absorption and scattering from the measured fluorescence spectrum to recover the intrinsic fluorescence of the tissue. The intrinsic fluorescence is due only to fluorophores present in the medium, from which fluorophore concentrations can be extracted using simple analytical models. A majority of these studies used analytical methods to derive mathematical expressions for the intrinsic fluorescence, from which the intrinsic fluorescence spectrum can be calculated (20-22, 24, 25). These methods are

limited in that they are valid for a limited range of absorption and scattering, not flexible in their applicability to various probe geometries, or require extensive empirical calibration. Biswal et al (23) proposed a method that was based on simultaneous measurements of polarized fluorescence and polarized elastic scattering spectra, and turbidity free fluorescence was obtained by normalizing the polarized fluorescence spectrum by the polarized elastic scattering spectrum. However to our knowledge, there have been no studies that have looked at the diagnostic value of the intrinsic fluorescence contrast between malignant and non-malignant human breast tissues. Our group (19) has developed a fluorescence model based on scalable Monte Carlo simulations, which is capable of extracting the intrinsic fluorescence spectra of a turbid medium, and is applicable for a wide range of optical properties and any illumination-collection geometry. The details of the model and its validation in synthetic phantoms are described in the preceding companion manuscript. This paper uses the model to recover the intrinsic fluorescence of human breast tissues and to extract the relative fluorescence contribution of individual fluorophores present in breast tissue.

The specific goals of this study were to: 1) apply Monte Carlo model based approaches developed by our group for the analysis of both fluorescence and diffuse reflectance spectra to extract the intrinsic fluorescence, absorption and scattering properties of breast tissues and 2) test the diagnostic value of these intrinsic tissue properties for discriminating malignant from non-malignant breast tissues. This work is distinct from previous publications by our group (5, 17, 18) in that first, it reports on the application of a new fluorescence model to recover the intrinsic fluorescence and then

extract the fluorophore contributions to breast tissue spectra, second it compares the merits of the extracted fluorophore contributions vs. the tissue absorption and scattering properties for diagnosing breast cancer and finally, it demonstrates the consistency of our model based approach for the analysis of diagnostically useful information contained in fluorescence and diffuse reflectance spectra measured with two different instruments and probe geometries.

2. Methods

2.1. Sample collection

Three sets of breast tissue samples were obtained from patients undergoing either breast cancer surgery (lumpectomy or mastectomy) or breast reduction surgery in two independent breast studies; both were approved by the Institutional Review Board at the University of Wisconsin - Madison. The two studies followed similar protocols for tissue handling and spectral measurements. In both studies, fluorescence and diffuse reflectance spectra were measured from freshly excised breast tissue samples immediately after excision and accompanying histological diagnoses were obtained for each specimen from microscopic evaluation of hematoxylin and eosin (H&E) sections. For tissue specimens obtained from cancer surgery, optical measurements were made at a cancerous site and a remote normal adipose tissue site, and for tissue specimens obtained from breast reduction surgery, optical measurements were attempted at a fibro glandular site and an adipose site. The tissue measurement and handling protocol for the first breast study has been described in detail elsewhere (1, 17), and that for the second breast study can also be found in an earlier publication (5). Based on the histological diagnosis, the breast tissue

specimens were broadly categorized as malignant, benign, fibrous and adipose. When malignancy presented at the site of measurement, the sample diagnosis was designated as being malignant. If a benign or normal sample exhibited heterogeneous tissue composition, the sample histology was designated to be the predominant tissue type immediately under the tissue surface where the optic probe was placed. In addition, the percent adipose, fibro-connective and percent glandular tissue content was quantified in all breast reduction surgery samples.

In the first study (17), fluorescence and diffuse reflectance spectroscopy was performed on 18 breast reduction patients and 45 cancer patients. According to the type of surgery, tissue samples were divided into two sets: the first set (Set #1) contained only normal samples from healthy women undergoing breast reduction surgery, and the second (Set #2) contained both cancerous and normal samples from cancer patients. A total of 64 normal samples were collected from the 18 breast reduction patients however 3 samples had to be excluded due to saturation in the spectral measurements, resulting in a total of 61 samples including 22 fibro-glandular and 39 adipose samples in Set #1. A total of 97 breast samples were collected from the 45 cancer patients, among which 14 samples had to be excluded from further analysis, due to either saturation of the measured spectra or due to the visible presence of lymphazurin on the measurement site. Thus a total of 83 tissue samples, including 37 malignant, 2 benign, 6 fibrous and 38 adipose samples were included in Set #2 for further analysis.

In the second study (5), a total of 56 tissue samples were obtained from 32 cancer or breast reduction patients, however 15 samples had to be excluded because they were measured with a demo instrument and the calibration information needed for the model analysis was not available. Thus a total of 41 samples obtained in the second breast study, including 17 malignant, 4 benign, 6 fibrous and 14 adipose breast tissues, were included in Set #3.

Set #1 was used to carry out a correlation analysis between the tissue properties extracted from the model based approach and histological parameters including percent adipose, fibro-connective and glandular content in normal breast tissues from healthy women. Set #2 and Set #3 were used in the discrimination analysis for the purpose of breast cancer diagnosis. Combining sets #2 and #3 yielded a sample size of 54 malignant, 6 benign, 12 fibrous and 52 adipose tissues. Table 1 displays the histological break down of the breast tissue samples in this study.

(Table 1)

2.2 Fluorescence and diffuse reflectance measurements

In the two breast studies, different spectrometers and probe geometries were used for spectral measurements, both of which have been described in our earlier publications (1, 5, 17). In the first breast study (corresponds to Set #1 and Set #2), fluorescence and diffuse reflectance spectra were measured with a multi-wavelength optical spectrometer and the fiber optic probe used consists of a central illumination core and a surrounding collection ring (1, 17). In the second breast study (corresponds to Set #3), tissue spectral

measurements were made using a Skinskan spectrometer (JY Horiba, Edison, NJ) and a fiber optic probe that consists of a central collection core and a surrounding illumination ring (5). The measured fluorescence and diffuse reflectance spectra were calibrated in order to correct for the (1) background spectrum, (2) wavelength dependence, and (3) throughput of the system. Detailed description of instrument settings and calibrations for the fluorescence and diffuse reflectance measurements can be found in earlier publications (1, 5, 17). In both breast studies, fluorescence measurements covered excitation wavelengths of 300 – 460 nm in 20 nm increments, and emission wavelengths from 20 nm beyond each excitation wavelength up to 600 nm in an increment of 5 nm. Diffuse reflectance measurements covered the wavelength range of 350 – 600 nm. Therefore, spectral analyses were carried out on the three sets of tissue spectra over these common wavelength ranges.

2.3 Extraction of tissue optical properties

A Monte Carlo inverse model of diffuse reflectance was used to extract the tissue optical properties. Details about the Monte Carlo based inverse model of diffuse reflectance can be found in references (17, 18, 26). The accuracy of the inverse model has been verified using experimental tissue phantom studies. The results indicated that for phantoms with a wide range of absorption coefficients ($0 \sim 20 \text{ cm}^{-1}$) and reduced scattering coefficients ($7 \sim 33 \text{ cm}^{-1}$), optical properties could be extracted within an average error of 3% for phantoms containing hemoglobin and within 12% for phantoms containing Nigrosin (26). The diffuse reflectance spectra were fit to the Monte Carlo model in the wavelength range of 350 – 600 nm. The primary intrinsic absorbers in the

model over this wavelength range were assumed to be oxygenated and deoxygenated hemoglobin and beta-carotene. Lymphazurin (Cat# 00592358, Tyco Healthcare, Mansfield, MA), a dye used to locate the sentinel lymph node during surgery, was included as an extrinsic absorber since it was found to be present in some of the tissue samples. A Gaussian function with a mean of 515 nm and a standard deviation of 13 nm was included as an additional absorbing component to account for a deviation in the fit between 500-530 nm, which was found in a large number of tissue samples. In addition, a baseline tissue absorption coefficient, describing protein absorption in the absence of blood (27) was added to the absorption coefficient calculated from the concentrations of the chromophores mentioned above. This baseline absorption is given by:

$$m_a(l) = a * \left(0.244 + 85.3 \exp\left(\frac{-(l - 154)}{66.2}\right) \right) \quad [cm^{-1}] \quad (\text{Eq. 1})$$

The concentrations of each absorber and the magnitude of the baseline absorption described in equation 1 were the free parameters related to absorption in the fit, which were extracted from the inversion process.

It should be pointed out that for the first two data sets (Set #1 and Set #2), NADH was also included in the model as an absorber for the calculation of absorption coefficient, while this was not done for the Set #3. This discrepancy was due to the difference in model selection by individual researchers, and this was the only difference in the model parameters for fitting the diffuse reflectance spectra of Set #1-2 and Set #3. However the absorption due to NADH alone was found to account for only 0 ~ 3% in average of the total absorption across the wavelength range of 350 - 600 nm, with a mean

of 0.4% and a standard deviation of 2%. This suggested that NADH did not contribute significantly to the absorption in this wavelength range, and the absorption and scattering properties obtained from the three sets of tissue spectra are still comparable. The absorption properties yielded from the model and used in further data analysis then include β -carotene concentration, total hemoglobin concentration and hemoglobin saturation, and the latter two were calculated directly from the concentrations of oxygenated and deoxygenated hemoglobin.

For scattering, the refractive indices of the scatterers and the surrounding medium were fixed parameters and were assumed to be 1.4 and 1.36 (28), respectively. The scatterer size and density were the free parameters in the model related to scattering, and the scatterer size was constrained to be between 0.35 and 1.5 μm in diameter (29). The wavelength dependent scattering coefficient and anisotropy factor were calculated from scatterer size, density and the refractive index mismatch between the scatterer and surrounding medium using the Mie theory for spherical particles. The mean reduced scattering coefficient was then calculated from the wavelength dependent scattering coefficient and anisotropy factor, and used to describe the bulk scattering properties of tissue.

It should be noted that the tissue spectra of Set #2 and Set #3 were measured with different spectrometers and probes, and the spectral intensities were recorded on different scales. In order to compare directly the results obtained from the two data sets, it is important to remove the system dependency. The Monte Carlo inverse model of diffuse

reflectance takes into account the specific probe geometry and incorporates a phantom calibration to account for the difference in system throughput thus allowing for a direct comparison of the absorption and scattering properties extracted from the two data sets (26).

2.4 Extraction of intrinsic fluorescence spectra

A Monte Carlo based model of fluorescence has been developed for the extraction of intrinsic fluorescence spectra of turbid tissues and is described in the preceding companion manuscript (19). This model is capable of removing the distorting effects of absorption and scattering and is valid for a wide range of optical properties. Using a phantom study, it was found that this model could recover the intrinsic fluorescence of the tissue phantoms with reasonably good accuracy and also retrieve the fluorophore concentration within an RMS error of 11% as described in the companion manuscript (19). The Monte Carlo model simulates photon propagation (including both incident photons and fluorescence re-emission photons) within a turbid medium, given a set of known absorption and scattering coefficients and illumination-collection geometry. In our approach, the set of absorption and scattering coefficients of tissue were derived from the concomitantly measured diffuse reflectance spectrum using the Monte Carlo based inverse model of diffuse reflectance, which was described in the previous section. Next, these optical properties were used in the Monte Carlo simulation of fluorescence to account for the effects of absorption and scattering on the fluorescence emission.

In this study, fluorescence spectra within excitation wavelength range of 340 – 460 nm and emission wavelength range of 350 – 600 nm were processed with the model to retrieve the intrinsic fluorescence, since diffuse reflectance was measured within 350 – 600 nm and absorption and scattering properties can be determined only for this wavelength range. The absorption and scattering coefficients at 340 nm were estimated using the chromophore concentrations, scatterer size and scatterer density extracted from the diffuse reflectance spectra within 350 – 600 nm. Inputs into the fluorescence model include the absorption and scattering coefficients at both the excitation and emission wavelengths. The model then used a scaling technique reported by Graaff et al. (30) to simulate the propagation of photons in a turbid medium with given absorption and scattering properties. A correction curve is generated from the output of the Monte Carlo simulation, which accounts for the effects of absorption and scattering in the medium at both excitation and emission wavelengths, as well as the collection efficiency defined by the specific probe geometry used. This correction curve corresponds to the denominator in equation (8) in the companion manuscript (19). The measured tissue fluorescence spectra were then divided point-by-point by the correction curve to retrieve the intrinsic fluorescence spectra independent of the absorption and scattering properties of tissue and probe geometry. The derived intrinsic fluorescence spectra is a function (product) of the effective quantum yield (corresponding to the left part in equation 8 in the companion manuscript (19) and a scaling factor S that accounts for the difference in magnitude between the Monte Carlo simulations (which are on an absolute scale) and the measured result (which is instrument dependent).

The Monte Carlo model of fluorescence also accounts for the specific probe geometry in the modeling process, however the model includes a scaling factor S (19), which accounts for the difference in magnitude between the Monte Carlo simulations and the measured result. This factor can be determined by comparing the measured and simulated fluorescence of a reference phantom with known optical properties, and needs to be known in order to compare, on an absolute scale, the intrinsic fluorescence measured with the two different instruments and probe geometries. In this study, the intrinsic fluorescence spectra of each tissue sample were normalized to the integrated intensity of the EEM, thus the difference in magnitude between the two sets of tissue spectra (sets #2 and 3) are removed, and a direct comparison of the relative (fractional) fluorescence contributions from individual fluorophores is possible.

2.5 Extraction of relative fluorescence contribution from individual fluorophores

The intrinsic fluorescence spectrum can be viewed as a linear combination of the fluorescence emission by one or more contributing fluorescent components (fluorophores). A Multivariate Curve Resolution (MCR) method (31) was used to decompose the intrinsic tissue fluorescence into several spectral profiles of the contributing components. MCR refers to a group of techniques which intend the recovery of response profiles, e.g. spectra, of the components in an unresolved mixture when no or limited prior information is available about the nature and composition of these mixtures. In this study, three components were assumed to contribute to the tissue fluorescence within excitation wavelength range of 340 – 460 nm and emission wavelength range of 350 – 600 nm, as there are three visible fluorescence peaks present in tissue EEMs in

these spectral ranges. A MCR toolbox available from the Chemometrics Group at the University of Barcelona (31) was used to extract the relative fluorescence contribution of each component, as well as their intrinsic spectral line shape.

Prior to MCR analysis, the intrinsic fluorescence spectra of each tissue sample were normalized to the integrated intensity over the entire EEM, that is, the fluorescence intensities over all excitation-emission wavelength pairs sum up to unity. This normalization removed the patient-to-patient variation and the difference in the magnitude of fluorescence spectra measured with different spectrometers and probe geometries. MCR analysis was then carried out on the normalized fluorescence EEM, and the intrinsic tissue fluorescence can be represented as: $F_{\text{tissue}}(\lambda_{\text{ex}}, \lambda_{\text{em}}) = a_1 F_1(\lambda_{\text{ex}}, \lambda_{\text{em}}) + a_2 F_2(\lambda_{\text{ex}}, \lambda_{\text{em}}) + a_3 F_3(\lambda_{\text{ex}}, \lambda_{\text{em}}) + e(\lambda_{\text{ex}}, \lambda_{\text{em}})$, where a_i represents the contribution and $F_i(\lambda_{\text{ex}}, \lambda_{\text{em}})$ represents the intensity of fluorescence from the i th component at excitation wavelength of λ_{ex} , and emission wavelength of λ_{em} , and $e(\lambda_{\text{ex}}, \lambda_{\text{em}})$ is the residual component. In the MCR process, the basis fluorescence profile of each contributing component was normalized to have a peak of unity. Since all spectra were normalized to the integrated intensity within an EEM, a_i then represents the fractional fluorescence contribution from an individual component to that sample. It is worth noting that the value of a_i is related to the quantum yield and absorption coefficient (a function of fluorophore concentration and extinction coefficient) of the individual fluorescing component at the excitation wavelength. The quantities a_i are the fluorescence properties that we expect to obtain from the fluorescence model analysis.

2.6 Correlation between extracted tissue properties and histological tissue composition

The extracted optical properties and fluorescence contributions were evaluated for their correlation with the histological tissue composition in the normal tissues obtained from breast reduction surgery (Set #1). The tissue composition within the sensing volume of the optical measurements was recorded as %adipose, %fibro-connective, and %glandular. Spearman correlations were used to determine the correlation coefficients and p-values for the relationship between the extracted tissue properties and the histological tissue composition.

2.7 Statistical analysis and classification using fluorescence properties

The tissue optical properties and fluorophore contributions extracted from the tissue Set #2 and Set #3 were pooled together for the purpose of discriminating malignant from non-malignant breast tissues. A Wilcoxon rank-sum test was performed to identify which extracted features from the diffuse reflectance and fluorescence spectra show statistically significant differences between malignant and non-malignant breast tissues. The optical properties and/or fluorophore contributions that displayed statistically significant differences were input to a linear Support Vector Machine (SVM) classifier to test the diagnostic accuracy of using these tissue properties for discriminating malignant from non-malignant breast tissues. Classification was carried out on (1) absorption and scattering properties only; (2) fluorophore contributions only; and (3) combination of fluorophore contribution with absorption and scattering properties. For each case, two cross validation schemes, i.e. holdout validation and leave-one-out cross validation, were employed to perform an unbiased evaluation of the classification accuracy. In the holdout

validation, the entire data set was randomly divided into training and testing sets, with each set containing half of the breast samples of each tissue type (i.e., 50% of the total malignant, 50% of the total fibrous/benign, and 50% of the total adipose tissue samples). Such a random partition was repeated 20 times, and the average classification accuracy was evaluated. In the leave-one-out cross validation, a single sample was used as the testing data and the remaining samples were used as the training data. This was repeated such that each sample was used once as the test data.

3. Results

Figure 1 shows the average absorption coefficient (a) and average reduced scattering coefficient (b) as a function of wavelength, for malignant ($n = 37$), fibrous/benign ($n = 8$) and adipose ($n = 38$) tissues in sample Set #2. For all tissue types, significant hemoglobin absorption was observed at approximately 420, 530 and 575 nm. Adipose tissues displayed an additional absorption band in the wavelength range of 440 – 500 nm. This absorption band is primarily attributed to beta-carotene, which present primarily in fat or adipose tissues. The scattering coefficients of all tissue types decreased monotonically with increasing wavelength. Malignant tissues had increased scattering coefficient relative to that of both fibrous/benign and adipose tissues. Scattering coefficients of fibrous/benign tissues were also higher as compared with that of adipose tissues. The average absorption and reduced scattering coefficients of the three tissue types in sample Set #3 displayed similar features but are not shown here.

(Figure 1)

Figure 2 shows the representative raw (a) and intrinsic (b) EEMs of a malignant tissue sample, each normalized to the integrated intensity of the EEM and plotted in log scale. In comparing the raw and intrinsic EEMs, the most obvious change is the disappearance of the fluorescence peak at the wavelength pair around (440, 510) nm in the intrinsic EEM. In addition, the dip shown in the raw EEM at around 420 nm, which is due to hemoglobin absorption, is also corrected for in the intrinsic EEM.

(Figure 2)

Figure 3 displays the average intrinsic fluorescence spectra obtained at excitation wavelength of (a) 340 nm and (b) 360 nm, for malignant ($n = 37$), fibrous/benign ($n = 8$) and adipose ($n = 38$) tissues in sample Set #2. At both excitation wavelengths, the average fluorescence intensities of malignant and fibrous/benign tissues were both higher than that of adipose tissues ($p < 0.05$). The average fluorescence intensity of malignant tissues was also higher relative to that of fibrous/benign tissues however the difference is not statistically significant. Malignant and fibrous/benign tissue samples share similar spectral profiles and displayed strong fluorescence at the emission wavelengths around 395 nm at 340 nm excitation and around 450 nm at 360 nm excitation. In adipose tissue samples, there was a weak presence of fluorescence at wavelengths around 395 nm, but strong fluorescence at the wavelengths between 450 – 500 nm. The humps at wavelengths around 420 nm in the spectra of malignant and fibrous/benign tissue samples are not fluorescence peaks but rather artifacts that may be introduced from the over-correction of the Soret band of hemoglobin absorption. The average intrinsic fluorescence spectra of

the three tissue types in sample Set #3 (not shown here) displayed similar spectral features but the intensities were on a different scale.

(Figure 3)

A three-component MCR analysis was carried out on the intrinsic fluorescence EEM to extract the relative contribution of the constituent fluorescing component, as well as their intrinsic spectral line shape. Figure 4 (a) – (c) shows the EEMs of the three fluorescing components (denoted as F_1 , F_2 , and F_3) obtained from the MCR analysis on intrinsic EEMs of sample Set #2. Each intrinsic tissue EEM is a linear combination of the component EEMs. These three components accounted for a vast majority of variance in the tissue EEMs, as the residual was only about 0.2% of the tissue EEM. The first component F_1 (Figure 4a) displayed excitation-emission maximum at a wavelength pair of (340, 395) nm. The second component F_2 (Figure 4b) displayed strong fluorescence at the emission wavelengths ranging from 450 to 500 nm over an excitation range of 360 – 420 nm. The third component F_3 (Figure 4c) had peak fluorescence at an emission wavelength of 480 nm at 340 – 360 nm excitation. This component was found to be prevalent in adipose tissues.

(Figure 4)

Table 2 shows the correlations between the extracted scattering, absorption and fluorescence (marked with *) properties and the constituent tissue composition, which were evaluated on breast tissues in sample Set #1 (Note +: positive correlation; -: negative correlation; Ø: no significant correlation). Correlations were considered

significant for $p < 0.01$. The mean reduced scattering coefficient (μ_s') and the fluorescence contribution from F_1 and F_2 were negatively correlated with %adipose, while positively correlated with %fibro-connective and %glandular tissue content. The β -carotene concentration and the fluorescence contribution of F_3 both had significant positive correlations with %adipose content, while they showed negative correlations with %fibro-connective and %glandular tissue content.

(Table 2)

Table 3 shows the pair wise cross-correlation between extracted scattering, absorption (including baseline absorption component) and fluorescence (marked with *) properties of the combined sample sets (Note +: positive correlation; -: negative correlation; $\emptyset$: no significant correlation). Optical properties obtained from the diffuse reflectance did not show statistically significant correlations with each other. For fluorescence properties, the fluorescence contribution of F_3 was negatively correlated with that of F_1 ($r = -0.67$, $p < 1e-16$) and F_2 ($r = -0.70$, $p < 1e-16$). In addition, the mean reduced scattering coefficient was positively correlated with the fluorescence contribution of F_2 ($r = 0.23$, $p < 0.01$), while negatively correlated with the fluorescence contribution of F_3 ($r = -0.26$, $p < 0.005$). β -carotene concentration had a strong positive correlation with the fluorescence contribution of F_3 ($r = 0.67$, $p < 1e-16$), and negative correlations with that of F_1 ($r = -0.56$, $p < 1e-10$) and F_2 ($r = -0.38$, $p < 0.00005$). There was also a positive correlation between the total hemoglobin concentration and F_2 ($r = 0.23$, $p < 0.01$), and a negative correlation between hemoglobin saturation and F_1 ($r = -0.30$, $p < 0.001$). The baseline absorption component did not show statistically significant

correlation with other extracted properties except for the fluorescence contribution of F_3 ($r = -0.23$, $p < 0.005$).

(Table 3)

Figure 5 shows the bar graphs of (a) β -carotene concentration, (b) mean reduced scattering coefficient (mean μ_s'), (c) hemoglobin saturation ($H_bO_2\%$) and (d) relative fluorescence contribution of individual fluorophores, as a function of tissue type. Data shown here were obtained from the combined sample sets (Set #2 and Set #3). β -carotene concentration (Figure 5a) was statistically higher in adipose breast tissues compare to that in malignant ($p < 1e-10$) and fibrous/benign breast tissues ($p < 0.001$). The mean reduced scattering coefficients μ_s' (Figure 5b) of malignant tissues were higher than that of fibrous/benign ($p < 0.05$) and adipose breast tissues ($p < 1e-9$). Hemoglobin saturation (Figure 5c) was higher in both fibrous/benign and adipose breast tissues in comparison to that in malignant tissues ($p < 0.0001$). In Figure 5d, it was shown that relative fluorescence contribution of F_1 and F_2 were higher in malignant and fibrous/benign tissues compare to that in adipose tissues ($p < 0.0001$), while the inverse trend was observed for F_3 ($p < 1e-7$).

(Figure 5)

Table 4 lists the results from Wilcoxon rank-sum test on extracted absorption, scattering and fluorescence properties for malignant vs. non-malignant and malignant vs. fibrous/benign breast tissues (properties are listed in the order of decreased p-value for differentiating between malignant and non-malignant tissues, and fluorescence properties

are marked with *). The difference in data distribution is considered statistically significant for $p < 0.05$. All extracted properties, except for total hemoglobin concentration and the baseline absorption component, displayed statistically significant differences between malignant and non-malignant breast tissues ($p < 0.0001$). However, only hemoglobin saturation and the mean reduced scattering coefficient displayed statistically significant difference between malignant and fibrous/benign breast tissues. This suggests that the fluorescence contrast shown between malignant and non-malignant breast tissues may be primarily associated with the differences between malignant and adipose tissues.

(Table 4)

Table 5 shows the results from the (a) holdout validation and (b) leave-one-out cross validation of a linear SVM classification on the combined data sets for discriminating malignant from non-malignant breast tissue samples using: (1) absorption and scattering properties only, (2) fluorescence properties only, and (3) combination of fluorescence, absorption and scattering properties that show statistically significant differences between malignant and non-malignant breast tissues. The classification rate (C Rate), sensitivity and specificity were averaged over the repeated training-and-testing trials. In the holdout validation (Table 5a), classification using absorption and scattering properties alone provided an average sensitivity and specificity of 82.6% and 90.1% respectively in training, and 81.7% and 88.6% respectively in testing for discriminating malignant from non-malignant breast tissues. The classification using fluorescence properties alone yielded slightly higher sensitivity but relatively lower specificity (average

SE and SP of 83.4% and 78.7% respectively in training, and 82.6% and 75.7% respectively in testing) compared to that achieved using absorption and scattering properties alone. It was also shown that combining fluorescence properties with absorption and scattering properties for classification slightly improved the diagnostic accuracy, and the average sensitivity and specificity were improved to 87.6% and 90.6% respectively in training, and 83.9% and 88.6% in testing, for discriminating malignant from non-malignant breast samples. In the leave-one-out cross validation (Table 5b), the classification accuracy for the three cases were comparable to those yielded from the holdout validation. To summarize, the classification using absorption and scattering properties alone and that using combined fluorescence, absorption and scattering properties yielded similar classification accuracy, and both outperformed the classification using fluorescence properties alone.

(Table 5)

Table 6 lists the number of tissue samples of each tissue type that were frequently (repeatedly) misclassified in the SVM classification using the three sets of inputs: (1) absorption and scattering properties only, (2) fluorescence properties only, and (3) combination of fluorescence, absorption and scattering properties. Most of the malignant samples that are misclassified are invasive ductal cancers. However, 25% ~ 40% of misclassified malignant samples were carcinoma *in situ* or lobular cancer. These subtypes of malignancy only accounted for a very small portion of the malignant samples and may be underrepresented. When fluorescence properties alone were used for

classification, most of the frequently misclassified non-malignant samples were benign or fibrous tissues.

(Table 6)

4. Discussion

In this study we explored the use of Monte Carlo model based approaches to extract a set of tissue properties, including intrinsic fluorophore contributions and absorption and scattering properties, which include β -carotene concentration, total hemoglobin concentration, hemoglobin saturation and mean reduced scattering coefficient from fluorescence and diffuse reflectance measurements of malignant and non-malignant breast tissues. For normal samples obtained from breast reduction patients, the mean reduced scattering coefficient decreased with the %adipose tissue content, while it increased with the %fibro-connective and %glandular tissue content. This agrees with findings from previously published studies. For example, Peters et al (14) estimated the reduced scattering coefficients at 540 nm for normal fibrous and normal adipose breast tissues and showed that fibrous tissues had higher reduced scattering coefficients as compared to adipose tissues. Cerussi et al (32) and Durduran et al (33) also showed in their studies that the scattering coefficients decreased with increasing BMI. β -carotene concentration was positively correlated with the %adipose tissue content, while negatively correlated with %fibro-connective and %glandular tissue content. This agrees with the fact that the majority of body reserves of β -carotene are thought to be in adipose tissues (34). In comparing the malignant, fibrous/benign and adipose breast tissues, β -carotene had an increased concentration in adipose breast tissues compared to that of

malignant ($p < 1e-10$) and fibrous/benign breast tissues ($p < 0.001$). This is expected since β -carotene is primarily stored in fatty tissues (34). The hemoglobin saturation showed a significant decrease in malignant tissues relative to that in non-malignant (including both fibrous/benign and adipose) tissues ($p < 0.0001$), which is likely due to the oxygen extraction of rapidly proliferating tumor cells. This property agreed with the findings from our previous studies (17, 18), and has also been reported in several previous studies using other techniques such as pO_2 measurements (35, 36) and NIR diffuse optical spectroscopy (32, 37, 38). Malignant breast tissues had an increased mean reduced scattering coefficient compared to that of non-malignant breast tissues, which is also consistent with the findings from earlier studies published by Peters et al (14) and Ghosh et al (16).

Three fluorescence components were extracted from the MCR analysis of the intrinsic fluorescence spectra. The first component has peak fluorescence at an excitation-emission pair of (340, 395) nm, which coincides with the typical excitation-emission maximum of collagen (39). The EEM of this component is similar to that reported for collagen type 1 (40). Therefore this fluorescence component is likely attributed to collagen. Collagen is the structural protein present in tissue extracellular matrix. Certain collagens, especially Type 1 collagen, is a major constituent of the dermis and fibrous stroma of breast (41). It was shown that in normal breast tissues the fluorescence contribution of collagen decreased with increasing adipose tissue content, while it increased with increasing fibro-connective and glandular tissue content (as seen in Table 2). This agrees with the fact that collagen resides primarily in the dense fibrous stroma of

the breast (41). Results of our study also showed that collagen has statistically higher contribution to the fluorescence of malignant and fibrous/benign breast tissues, as compared with that of normal adipose tissues ($p < 1e-6$). Similar observations were also reported in previous studies using other optical spectroscopy techniques for breast tissue characterization (42, 43). Haka et al (42) used Raman spectroscopy for the diagnosis of breast cancer, and found that there was increased contribution from collagen in benign and cancerous breast tissues, relative to that in normal breast tissues. Taroni et al (43) measured absorption of collagen in healthy breast tissues using time-resolved transmittance spectroscopy, and showed that less fatty tissues are characterized by higher collagen concentration.

The second component displayed strong fluorescence at the emission wavelengths ranging in 450 – 500 nm at excitation wavelengths of 360 – 420 nm. The fluorescence feature is spectrally very similar to that of NADH (39, 40, 44) thus the primary fluorophore that contributes to this fluorescent component is likely NADH. Like collagen, the fluorescence contribution from NADH was also statistically higher in malignant and fibrous/benign breast tissues than that in adipose breast tissues. Since NADH is one of the important coenzymes for a large number of metabolic activities in cells, the relative concentration of this fluorophore varies according to the oxidative metabolic status of the cells. The higher NADH fluorescence in malignant tissues may result from the increased metabolic activity and oxygen depletion in cancer cells, and thus accumulation of NADH. Uppal et al (45) carried out an enzymatic measurement of NADH concentration in malignant and normal breast tissues, and found that NADH is significantly higher in

malignant tissues as compared with normal breast tissues. In the normal breast, NADH may also increase with the increased metabolism associated with the duct proliferation and secretory activities of the breast lobules, while adipose cells have an indolent metabolism (41). Therefore fluorescence of adipose tissues would be expected to have lower contribution from NADH than that of fibrous/benign breast tissues. This explains the negative correlation observed between the fluorescence contribution of this component and adipose tissue content, as well as the positive correlation observed between this fluorescence component and fibro-connective and glandular tissue content in healthy breast tissues.

The third component displayed an excitation-emission maximum at (360, 480) nm. This fluorescence peak or shoulder was observed in the intrinsic fluorescence spectra of a majority of adipose breast tissues. However, to our knowledge there is very limited work discussing the source of this fluorescent component. Possible sources for this fluorescence feature can be FAD or retinol/Vitamin A, both reported to have fluorescence emission maxima at 520 nm. We measured the fluorescence EEMs of commercially available FAD (F-6625, Sigma-Aldrich Co.) and retinol acetate (R-7882, Sigma-Aldrich Co.). By comparing the extracted EEM of the third component with those of FAD and retinol, we found noticeable similarity between the third component EEM and that of retinol (Figure 6). Also the relative fluorescence contribution of the third component was highly correlated with the percent adipose tissues content in the breast tissues ($r = 0.83$, $p < 1e-18$, evaluated on sample Set #1, and $r = 0.76$, $p < 1e-16$, evaluated on samples Set #2, not evaluated on sample Set#3 since the percent tissue composition was not available

for this sample set), as well as the β -carotene concentration ($r = 0.67$, $p < 1e-16$, evaluated on combined data sets), which was derived from the Monte Carlo inverse model based analysis of concomitantly measured diffuse reflectance. β -carotene is a precursor of Vitamin A, and is primarily stored in adipose tissues (46). Based on these observations and facts, we have tentatively attributed this component to retinol, or Vitamin A. A systematic investigation of this fluorescence component may be required in the future in order to fully identify the source and also explore the potential confounding effects of FAD or other factors that may be present in the tissue.

(Figure 6)

Results from our study suggest that there may be a statistically significant increase in fluorescence contribution from retinol in normal adipose breast tissues compared to that in other tissue types, indicating the prevalence of retinol content in normal adipose tissues. Since retinol is fat-soluble, it is primarily present in adipose tissues, and most women carry stores of retinol or Vitamin A in their fat cells. Lunetta et al (34) have estimated and compared the retinol concentrations in normal and cancerous breast tissues in breast cancer patients, and found that the retinol concentration in normal breast tissues was slightly higher (on average) than that in malignant breast tissues, however the differences were not statistically significant.

The negative correlation observed between the fluorescence contribution of collagen/NADH and the β -carotene concentration may be attributed to fact that collagen/NADH is found predominantly in fibro glandular tissue while β -carotene is

found in fatty tissues. The mean reduced scattering coefficient has a positive correlation with the fluorescence contribution of NADH, while it has a negative correlation with the retinol fluorescence. This may be due to decreasing scattering (14, 33, 47), decreased NADH fluorescence (41) and increased retinol fluorescence in adipose tissues. The physiological basis for the negative correlation observed between fluorescence contribution of collagen and hemoglobin saturation is not clear. However, collagen was found to be prevalent in malignant tissues, which has lower hemoglobin saturation than normal tissues. This may contribute to the correlation observed between the two variables.

In this study we extracted the contribution of absorbers, scatterers and fluorophores in breast tissue and correlated them with histological changes in the normal breast as well as histopathology of the breast. It was not possible to directly compare the extracted tissue properties to tissue constituent concentrations, because this is a study involving clinical samples, on which disruptive biochemical assays cannot be performed. However, we are currently doing a study to compare the extracted tissue properties to immunohistochemical assays (for example, hypoxia, micro vessel density, collagen distribution, etc) to demonstrate that the optical technique is non-destructively able to quantify biological and morphological information from breast tissues. This will provide a means to directly validate the quantitative physiological aspects of this technique.

The intrinsic fluorescence spectra of each tissue sample were normalized to the integrated intensity over the entire EEM prior to MCR analysis. This normalization

removed the inter-sample variations in the spectral intensity so that the extracted fluorescence contribution only reflected the fractional contributions of individual fluorophores to the fluorescence of each tissue sample. MCR analysis was also carried out on un-normalized intrinsic fluorescence spectra from set #2 only, and it was observed that the relationships of the extracted fluorophore contribution with different tissue types were the same as those observed with the normalized spectra.

The relative fluorescence contribution of all three fluorophores displayed statistically significant differences between malignant and non-malignant breast tissues, these differences however, may be associated with the differences between malignant and adipose tissue only, as a large portion (52 out of 70) of non-malignant breast samples are adipose tissues. The fluorescence spectra of malignant and fibrous/benign tissue samples shared similar spectral line shapes, and the Wilcoxon rank sum test did not show statistically significant difference between malignant and fibrous/benign breast tissues in the fluorescence contributions of individual fluorophores. In the holdout validation, the classification based on fluorophore contributions alone achieved an average sensitivity and specificity of 83.4% and 78.7% respectively in training, and 82.6% and 75.7% respectively in testing for discriminating malignant from non-malignant breast tissues. The separation boundary yielded from this classification may discriminate primarily between malignant and normal adipose tissues and a retrospective look at the misclassified samples indicate that most of the misclassified non-malignant samples are fibrous or benign tissues.

Classification using absorption and scattering properties that displayed statistically significant difference between malignant and non-malignant breast tissues provided a slightly decreased sensitivity (82.6% in training and 81.7% in testing using holdout validation) but higher specificity (90.1% in training and 88.6% in testing using holdout validation) for discriminating breast malignancy, as compared to that achieved using fluorescence properties alone. The improvement to the specificity may primarily be attributed to the inclusion of hemoglobin saturation and mean reduced scattering coefficient, especially the former, which displayed statistically significant differences between malignant and fibrous/benign breast tissues.

The comparable classification performance using either fluorescence properties, or absorption and scattering properties alone, or both suggest that fluorescence and diffuse reflectance spectra both yield diagnostically useful information for the discrimination of breast malignancy. However, the model based approach does allow for a full exploration of the physiologically relevant information contained in the fluorescence and diffuse reflectance spectra, thus providing a comprehensive understanding of the biochemical, physiological and morphological changes that take place in the tissue, which has implications in applications such as monitoring tumor response to therapy where both metabolism and hemoglobin oxygenation, for example are important.

Besides providing physiologically meaningful information about the tissue composition and pathology, another advantage of the model based approach for the

analysis of tissue spectra is that it can significantly eliminate the instrument dependency, therefore results obtained using different instruments and probes are directly comparable. In our study, we have combined the fluorescence, absorption and scattering properties extracted independently from two sets of tissue spectra, which were measured with two different instruments and fiber optic probes. The extracted properties were comparable and a statistical t-test showed that there was no statistically significant difference between the two data sets. This allowed for an increased sample size by combining the two data sets.

In this study, the sensitivity and specificity for discriminating breast malignancy were evaluated using two validation methods, i.e., holdout validation and leave-one-out cross validation, both of which provide an unbiased evaluation of the classification accuracy and the robustness of the classification algorithm. The classification accuracies were consistent across the repeated trials in which training samples were randomly selected from the entire sample set (combined Set #2 and Set #3). The classification accuracy achieved using holdout validation and leave-one-out cross validation were also comparable.

In conclusion, we presented Monte Carlo model based approaches for the analysis of fluorescence and diffuse reflectance spectra of breast tissues, which enable the extraction of intrinsic fluorescence, absorption and scattering properties that provide the biochemical and morphological information about the tissue. The diagnostically useful absorption/scattering properties and fluorescence properties can be used alone or in

combination for the discrimination of breast malignancy. It was also demonstrated that the model based approach has the advantage of eliminating the dependency of fluorescence and diffuse reflectance measurements on the instrument and probe geometry, which makes it a more generalized approach for the analysis of tissue spectra.

Acknowledgement

This work is supported by the NIH grant 1R01CA100559-01 and the Department of Defense pre-doctoral traineeship DAMD17-02-1-0628 and W81XWH-05-1-0380.

References

- 1.C. Zhu, G. M. Palmer, T. M. Breslin, F. Xu and N. Ramanujam, "The Use of a Multi-separation Fiber Optic Probe for the Optical Diagnosis of Breast Cancer," *J Biomed Opt* 10(2), 024032 (2005)
- 2.Y. Yang, E. J. Celmer, M. Zurawska Szczepaniak and R. R. Alfano, "Excitation spectrum of malignant and benign breast tissues: a potential optical biopsy approach," *Lasers in the Life Sciences* 7(4), 249-265 (1997)
- 3.Y. Yang, A. Katz, E. J. Celmer, M. Zurawska-Szczepaniak and R. R. Alfano, "Fundamental differences of excitation spectrum between malignant and benign breast tissues," *Photochem Photobiol* 66(4), 518-522 (1997)
- 4.Y. Yang, A. Katz, E. J. Celmer, M. Zurawska Szczepaniak and R. R. Alfano, "Optical spectroscopy of benign and malignant breast tissues," *Lasers in the Life Sciences* 7(2), 115-127 (1996)

- 5.G. M. Palmer, C. Zhu, T. M. Breslin, F. Xu, K. W. Gilchrist and N. Ramanujam, "Comparison of multiexcitation fluorescence and diffuse reflectance spectroscopy for the diagnosis of breast cancer (March 2003)," *IEEE Trans Biomed Eng* 50(11), 1233-1242 (2003)
- 6.G. M. Palmer and N. Ramanujam, "Diagnosis of Breast Cancer Using Optical Spectroscopy," *Medical Laser Application* 18(233-248 (2003)
- 7.P. K. Gupta, S. K. Majumder and A. Uppal, "Breast cancer diagnosis using N2 laser excited autofluorescence spectroscopy," *Lasers Surg Med* 21(5), 417-422 (1997)
- 8.C.-H. Liu, B. B. Das, W. L. Sha Glassman, G. C. Tang, K. M. Yoo, H. R. Zhu, D. L. Akins, S. S. Lubicz, J. Cleary, R. Prudente, E. J. Celmer, A. Caron and R. R. Alfano, "Raman, fluorescence, and time-resolved light scattering as optical diagnostic techniques to separate diseased and normal biomedical media," *J Photochem Photobiol, B: Biol.* 16(2), 187-209 (1992)
- 9.S. K. Majumder, P. K. Gupta, B. Jain and A. Uppal, "UV excited autofluorescence spectroscopy of human breast tissues for discriminating cancerous tissue from benign tumor and normal tissue," *Lasers in the Life Sciences* 8(4), 249-264 (1999)
- 10.I. J. Bigio and J. R. Mourant, "Ultraviolet and visible spectroscopies for tissue diagnostics: fluorescence spectroscopy and elastic-scattering spectroscopy," *Phys Med Biol* 42(5), 803-814 (1997)
- 11.R. R. Alfano, B. B. Das, J. Cleary, R. Prudente and E. J. Celmer, "Light sheds light on cancer--distinguishing malignant tumors from benign tissues and tumors," *Bull N Y Acad Med* 67(2), 143-150 (1991)

- 12.S. V. Pushkarev, S. A. Naumov, S. M. Vovk, V. A. Volovodenko and V. V. Udut, "Application of laser fluorescence spectroscopy and diffuse reflection spectroscopy in diagnosing the states of mammary gland tissue," *Optoelectronics,-Instrumentation-and-Data-Processing* 2(71-76 (1999)
- 13.I. J. Bigio, S. G. Bown, G. Briggs, C. Kelley, S. Lakhani, D. Pickard, P. M. Ripley, I. G. Rose and C. Saunders, "Diagnosis of breast cancer using elastic-scattering spectroscopy: preliminary clinical results," *J Biomed Opt* 5(2), 221-228 (2000)
- 14.V. G. Peters, D. R. Wyman, M. S. Patterson and G. L. Frank, "Optical properties of normal and diseased human breast tissues in the visible and near infrared," *Phys Med Biol* 35(9), 1317-1334 (1990)
- 15.Y. Yang, E. J. Celmer, J. A. Koutcher and R. R. Alfano, "UV reflectance spectroscopy probes DNA and protein changes in human breast tissues," *J Clin Laser Med Surg* 19(1), 35-39 (2001)
- 16.N. Ghosh, S. K. Mohanty, S. K. Majumder and P. K. Gupta, "Measurement of optical transport properties of normal and malignant human breast tissue," *Applied Optics* 40(1), 176-184 (2001)
- 17.C. Zhu, G. M. Palmer, T. M. Breslin, J. Harter and N. Ramanujam, "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)
- 18.G. M. Palmer, C. Zhu, T. M. Breslin, F. Xu, K. W. Gilchrist and N. Ramanujam, "Monte Carlo-based inverse model for calculating tissue optical properties. Part II: Application to breast cancer diagnosis," *Appl Opt* 45(5), 1072-1078 (2006)

- 19.G. M. Palmer and N. Ramanujam, "Monte Carlo based Model to Extract Intrinsic Fluorescence from Turbid Media: Theory and Phantom Validation," *J Biomed Opt* submitted((2007)
- 20.Q. Zhang, M. G. Muller, J. Wu and M. S. Feld, "Turbidity-free fluorescence spectroscopy of biological tissue," *Opt Lett* 25(19), 1451-1453 (2000)
- 21.N. N. Zhadin and R. R. Alfano, "Correction of the internal absorption effect in fluorescence emission and excitation spectra from absorbing and highly scattering media: theory and experiment," *J Biomed Opt* 3(2), 171-186 (1998)
- 22.J. Wu, M. S. Feld and R. Rava, "Analytical model for extracting intrinsic fluorescence in turbid media," *Appl Opt* 32(19), 3585-3595 (1993)
- 23.N. C. Biswal, S. Gupta, N. Ghosh and A. Pradhan, "Recovery of turbidity free fluorescence from measured fluorescence: an experimental approach," *Optics Express* 11(24), 3320-3331 (2003)
- 24.C. M. Gardner, S. L. Jacques and A. J. Welch, "Fluorescence spectroscopy of tissue: recovery of intrinsic fluorescence from measured fluorescence," *Appl Opt* 35(10), 1780-1792 (1996)
- 25.M. G. Muller, I. Georgakoudi, Q. Zhang, J. Wu and M. S. Feld, "Intrinsic fluorescence spectroscopy in turbid media: disentangling effects of scattering and absorption," *Applied Optics* 40(25), 4633-4646 (2001)
- 26.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)
27. R. Huang and S. Jacques, "Skin Optics Summary," *J Biomed Opt* (1998)

- 28.F. P. Bolin, L. E. Preuss, R. C. Taylor and R. J. Ference, "Refractive index of some mammalian tissues using a fiber optic cladding method," *Applied Optics* 28(12), 2297-2303 (1989)
- 29.J. R. Mourant, J. P. Freyer, A. H. Hielscher, A. A. Eick, D. Shen and T. M. Johnson, "Mechanisms of light scattering from biological cells relevant to noninvasive optical-tissue diagnostics," *Applied Optics* 37(16), 3586-3593 (1998)
- 30.R. Graaff, M. H. Keolink, F. F. M. de Mul, W. G. Zijlstra, A. C. M. Dassel and J. G. Aarnoudse, "Condensed Monte Carlo simulations for the description of light transport," *Appl Opt* 32(4), 426-434 (1993)
- 31.R. Tauler, "Multivariate Curve Resolution, MCR-ALS Command Line Toolbox," (2006)
- 32.A. E. Cerussi, A. J. Berger, F. Bevilacqua, N. Shah, D. Jakubowski, J. Butler, R. F. Holcombe and B. J. Tromberg, "Sources of absorption and scattering contrast for near-infrared optical mammography," *Acad Radiol* 8(3), 211-218 (2001)
- 33.T. Durduran, R. Choe, J. P. Culver, L. Zubkov, M. J. Holboke, J. Giammarco, B. Chance and A. G. Yodh, "Bulk optical properties of healthy female breast tissue," *Phys Med Biol* 47(16), 2847-2861 (2002)
- 34.J. M. Lunetta, R. A. Zulim, S. R. Dueker, Y. Lin, V. Flaig, P. D. Schneider, B. M. Wolfe and A. J. Clifford, "Method for the simultaneous determination of retinol and beta-carotene concentrations in human tissues and plasma," *Anal Biochem* 304(1), 100-109 (2002)

- 35.W. Mueller-Klieser, K. H. Schlenger, S. Walenta, M. Gross, U. Karbach, M. Hoeckel and P. Vaupel, "Pathophysiological approaches to identifying tumor hypoxia in patients," *Radiother Oncol* 20 Suppl 1(21-28 (1991)
- 36.P. Vaupel, K. Schlenger, C. Knoop and M. Hockel, "Oxygenation of human tumors: evaluation of tissue oxygen distribution in breast cancers by computerized O₂ tension measurements," *Cancer Res* 51(12), 3316-3322 (1991)
- 37.B. J. Tromberg, N. Shah, R. Lanning, A. Cerussi, J. Espinoza, T. Pham, L. Svaasand and J. Butler, "Non-invasive in vivo characterization of breast tumors using photon migration spectroscopy," *Neoplasia* 2(1-2), 26-40 (2000)
- 38.D. Grosenick, K. T. Moesta, H. Wabnitz, J. Mucke, C. Stroszczynski, R. Macdonald, P. M. Schlag and H. Rinneberg, "Time-domain optical mammography: initial clinical results on detection and characterization of breast tumors," *Appl Opt* 42(16), 3170-3186 (2003)
- 39.N. Ramanujam, "Fluorescence Spectroscopy In Vivo," in Encyclopedia of Analytical Chemistry R. Meyers, Ed., pp. 20-56, John Wiley and Sons, Ltd. (2000).
- 40.I. Georgakoudi, B. C. Jacobson, M. G. Muller, E. E. Sheets, K. Badizadegan, D. L. Carr-Locke, C. P. Crum, C. W. Boone, R. R. Dasari, J. Van Dam and M. S. Feld, "NAD(P)H and collagen as in vivo quantitative fluorescent biomarkers of epithelial precancerous changes," *Cancer Res* 62(3), 682-687 (2002)
- 41.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)

- 42.A. S. Haka, K. E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R. R. Dasari and M. S. Feld, "Diagnosing breast cancer by using Raman spectroscopy," *Proc Natl Acad Sci U S A* 102(35), 12371-12376 (2005)
- 43.P. Taroni, D. Comelli, A. Pifferi, A. Torricelli and R. Cubeddu, "Absorption of collagen: effects on the estimate of breast composition and related diagnostic implications," *J Biomed Opt* 12(1), 014021 (2007)
- 44.T. G. Scott, R. D. Spencer, N. J. Leonard and G. Weber, "Synthetic spectroscopic models related to coenzymes and base pairs. V. Emission properties of NADH. Studies of fluorescence lifetimes and quantum efficiencies of NADH, AcPyADH, [reduced acetylpyridineadenine dinucleotide] and simplified synthetic models," *J. Am. Chem. Soc.* 92(3), 687-695 (1970)
- 45.A. Uppal and P. K. Gupta, "Measurement of NADH concentration in normal and malignant human tissues from breast and oral cavity," *Biotechnol Appl Biochem* 37(Pt 1), 45-50 (2003)
- 46."Medline Plus Medical Encyclopedia online."
<http://www.nlm.nih.gov/medlineplus/ency/article/002400.htm>
- 47.S. Srinivasan, B. W. Pogue, S. Jiang, H. Dehghani, C. Kogel, S. Soho, J. J. Gibson, T. D. Tosteson, S. P. Poplack and K. D. Paulsen, "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)

Tables

Table 1 Histological break down of the breast tissue samples examined in this study.

Type and Sub-type	Set #1	Set #2	Set #3	Set#2+Set#3
Total Malignant		37	17	54
Invasive ductal carcinoma (IDC)		29	15	44
Invasive lobular carcinoma (ILC)		2	2	4
Ductal carcinoma in situ (DCIS)		2	0	2
Lobular carcinoma in situ (LCIS)		2	0	2
Infiltrating tubulolobular		2	0	2
Total Non-malignant		46	24	70
Benign		2	4	6
Fibro glandular	22	6	6	12
Adipose	39	38	14	52
Total samples	61	83	41	124

Table 2 Correlations between the extracted scattering, absorption and fluorescence (marked with *) properties and the histological tissue composition, evaluated on breast tissues in sample Set #1 (Note +: positive correlation; -: negative correlation; Ø: no significant correlation). Correlations were considered significant for $p < 0.01$.

	% adipose	% fibro-connective	% glandular
Mean reduced scattering coefficient μ_s' (cm^{-1})	-0.61 (-)	0.63 (+)	0.39 (+)
β -carotene concentration (μM)	0.66 (+)	-0.61 (-)	-0.64 (-)
Total hemoglobin concentration (μM)	Ø	Ø	Ø
Hemoglobin saturation ($\text{HbO}_2\%$)	Ø	Ø	Ø
F_1 contribution *	-0.42 (-)	0.37 (+)	0.44 (+)
F_2 contribution *	-0.80 (-)	0.81 (+)	0.58 (+)
F_3 contribution *	0.83 (+)	-0.82 (-)	-0.65 (-)
Baseline absorption	Ø	Ø	Ø

Table 3 Pair wise cross-correlation between extracted scattering, absorption (including baseline absorption component) and fluorescence (marked with *) properties [Note: +: positive correlation; - negative correlation; Ø: no significant correlation]

	μ_s'	β -carotene	Total Hb	HbO ₂ %	F ₁ *	F ₂ *	F ₃ *	Baseline
μ_s'								
β -carotene	Ø							
Total Hb	Ø	Ø						
HbO ₂ %	Ø	Ø	Ø					
F ₁ *	Ø	− 0.56	0	− 0.30				
F ₂ *	+ 0.23	− 0.38	+ 0.23	Ø	0			
F ₃ *	− 0.26	+ 0.67	Ø	Ø	− 0.67	− 0.70		
Baseline	Ø	Ø	Ø	Ø	Ø	Ø	− 0.23	

Table 4 Results from Wilcoxon rank-sum test on extracted absorption, scattering and fluorescence properties for malignant vs. non-malignant and malignant vs. fibrous/benign breast tissues (properties are listed in the order of decreasing p-value for differentiating between malignant and non-malignant tissues, and fluorescence properties are marked with *).

Extracted Tissue Properties	Malignant vs. non-malignant	Malignant vs. fibrous/benign
F ₃ contribution *	p < 1e-8	Not significant
β -carotene concentration (μM)	p < 1e-7	Not significant
F ₁ contribution *	p < 1e-7	Not significant
Hemoglobin saturation	p < 1e-7	p < 1e-4
Mean reduced scattering coefficient μ_s' (cm ⁻¹)	p < 1e-7	p < 0.05
F ₂ contribution *	p < 1e-4	Not significant
Total hemoglobin concentration (μM)	Not significant	Not significant
Baseline absorption	Not significant	Not significant

Table 5 Results from the (a) holdout validation and (b) leave-one-out cross validation of a linear SVM classification on the combined data sets for discriminating malignant from non-malignant breast tissue samples using (1) absorption and scattering properties only, (2) fluorescence properties only, and (3) combination of fluorescence, absorption and scattering properties that show statistically significant differences between malignant and

non-malignant breast tissues. The classification rate (C Rate), sensitivity and specificity were averaged over the repeated training-and-testing trials (CV: cross validation).

(a) Holdout validation

	(1) Abs. & scatt. properties only		(2) Fluorescence properties only		(3) All significant properties	
	Training	Testing	Training	Testing	Training	Testing
C Rate	86.8±3.1%	85.6±3.2%	80.1±2.5%	78.1±2.5%	89.3±3.0%	86.5±2.5%
Sensitivity	82.6±6.0%	81.7±5.2%	83.4±8.3%	82.6±7.5%	87.6±4.8%	83.9±5.8%
Specificity	90.1±2.7%	88.6±5.5%	78.7±4.9%	75.7±4.9%	90.6±3.6%	88.6±4.7%

(b) Leave-one-out cross validation

	(1) Abs. & scatt. properties only		(2) Fluorescence properties only		(3) All significant properties	
	Training	CV	Training	CV	Training	CV
C Rate	86.4±0.5%	85.5%	80.4±0.6%	77.4%	86.4±0.6%	85.5%
Sensitivity	83.4±0.5%	83.3%	85.2±2.3%	81.5%	83.3±0.8%	81.5%
Specificity	88.7±0.6%	87.1%	76.7±2.1%	74.3%	88.8±0.7%	88.6%

Table 6 Number of tissue samples of each tissue type that were frequently misclassified in the SVM classification using the three sets of inputs: (1) absorption and scattering properties only, (2) fluorescence properties only, and (3) combination of fluorescence, absorption and scattering properties.

Type and Sub-type	(1) Abs. & scatt. properties only	(2) Fluorescence properties only	(3) All significant properties
Malignant			
IDC	6	6	5
ILC	2	0	2
DCIS	0	1	1
LCIS	1	1	1
Non-malignant			
Benign	3	4	2
Fibro glandular	2	8	2
Adipose	3	3	2

Figure Captions

Figure 1 Average absorption coefficient (a) and average reduced scattering coefficient (b) as a function of wavelength, for malignant (n = 37), fibrous/benign (n = 8) and adipose (n = 38) tissues in sample Set #2.

Figure 2 Raw (a) and intrinsic (b) fluorescence EEMs of a representative malignant tissue sample, each normalized to the integrated intensity of the EEM and plotted in log scales.

Figure 3 Average intrinsic fluorescence spectra obtained at excitation wavelength of (a) 340 nm and (b) 360 nm, for malignant (n = 37), fibrous/benign (n = 8) and adipose (n = 38) tissues in sample Set #2.

Figure 4 EEMs of fluorescing components (a) F_1 , (b) F_2 and (c) F_3 obtained from a three-component MCR analysis on intrinsic EEMs of sample Set #2.

Figure 5 Bar graphs of (a) β -carotene concentration, (b) mean reduced scattering coefficient (mean μ_s'), (c) hemoglobin saturation ($H_bO_2\%$), (d) relative fluorophore contribution of individual fluorophores, as a function of tissue type.

Figure 6 Fluorescence EEM of retinol acetate (R-7882, Sigma-Aldrich Co.)