Intrinsic tumor biomarkers revealed by novel double-differential spectroscopic analysis of near-infrared spectra

Shwayta Kukreti,[a,b,c] **Albert Cerussi,**[a] **Bruce Tromberg,**[a,d] **and Enrico Gratton**[a,b,c,d,*]

[a]University of California, Irvine, Beckman Laser Institute, Irvine, California 92612
[b]University of California, Irvine, Laboratory for Fluorescence Dynamics, Irvine, California 92612
[c]University of Illinois, Urbana-Champaign, Biophysics and Computational Biology Program, Urbana-Champaign, Illinois 61801
[d]University of California, Irvine, Department of Biomedical Engineering, Irvine, California 92612
E-mail: egratton22@yahoo.com

Abstract. We develop a double-differential spectroscopic analysis method for broadband near-infrared (NIR, 650 to 1000 nm) absorption spectra. Application of this method to spectra of tumor-containing breast tissue reveals specific cancer biomarkers. In this method, patient-specific variations in molecular composition are removed by using the normal tissue as an internal control. The effects of concentration differences of the four major tissue absorbers (oxyhemoglobin, deoxyhemoglobin, water, and bulk lipid) between the tumor and normal tissue are accounted for to reveal small spectral components unique to cancer. From a pilot study of 15 cancer patients, we find these spectral components to be characterized by specific NIR absorption bands. Based on the spectral regions of absorption at about 760, 930, and 980 nm, we identify these biomarkers with changes in state or addition of lipid and/or water. To quantify spectral variation in the absorption bands, we construct the specific tumor component (STC) index. The STC index identifies regions of the breast with tumors. © *2007 Society of Photo-Optical Instrumentation Engineers.* [DOI: 10.1117/1.2709701]

The basis for near-infrared (NIR) spectroscopy and imaging in breast cancer has been that tumor alterations of tissue vascularization/angiogenesis and oxygen consumption can be measured through hemoglobin concentration and oxygenation state, respectively. Several investigators have developed diffuse optical imaging (DOI) and diffuse optical spectroscopy (DOS) instruments at discrete wavelengths in the 650- to 850-nm range to detect and characterize breast tumors due to the absorption of both oxy- and deoxyhemoglobin. NIR tissue absorption spectra are typically fitted with hemoglobin extinction spectra obtained *in vitro* to quantify tissue hemoglobin concentrations.

Increased levels of hemoglobin are often found in malignant breast tissues.[1] However, hemoglobin is not a specific marker for tumors. Several groups have increased the wavelength range and number of absorbing components to which the absorption data are fitted to include water and bulk lipid.[1–3] These four components form the basis spectra for breast tissues. However, like hemoglobin, water and bulk lipid are present in both diseased and normal tissues, and are not specific to cancer.

In our experience, fitting broadband absorption spectra to the known basis spectra for breast chromophores produces small mismatch residuals. There are several possible sources of these mismatches. One possibility is that the four component basis spectra, which have been obtained *in vitro*, do not correctly quantify absorption *in vivo*. Published hemoglobin extinction spectra have small but clear differences between their shapes.[4] There is debate whether or not the spectrum of water found in tissues is the same as that of bulk water. Ideally a patient-specific basis spectra set should be used for analysis.
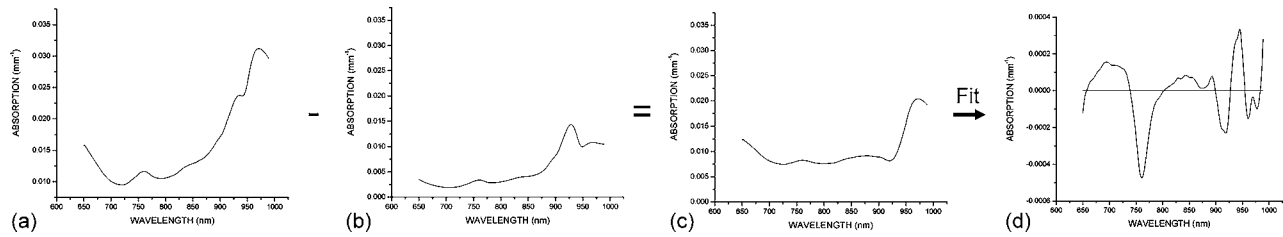
Another possibility for the mismatches is that the current basis spectra do not fully account for all NIR tissue absorption. For example, tumors alter protein expression, modify lipid and water states, and generate hemoglobin breakdown products.[5] Multivariate spectral analysis methods have addressed this shortcoming of currently used basis spectra, but finding the meaning behind cancer-related coordinates can be difficult.

To account for all possible compositional differences between tumor and normal tissues, we hypothesize that a unique NIR tumor component spectrum exists, which we refer to as the specific tumor component (STC) spectrum. The STC spectrum would be highly specific to the tumor-containing tissue, and should not be found in normal tissues. However, since mismatches between absorption spectra and the basis spectra fits exist for both tumor and normal tissues, we need a method that can account for patient-specific shifts in basis spectra, as well as individual biochemical variations in normal tissue. To detect a potential STC spectrum, we have developed a double-differential method. We begin with the difference spectrum between tumor and normal tissue, fit this spectrum to the four component basis spectra, then focus on the residuals to this fit.

It is imperative that we obtain a broadband tissue absorption spectrum that is corrected for the effects of multiple scattering. We used a DOS instrument to measure broadband NIR absorption and scattering spectra from 650 to 1000 nm.[1] Using a handheld probe absorption spectra were obtained at several spatial locations over both tumor-containing and normal tissue for each subject, with the normal region serving as a patient-specific control. Fifteen subjects were studied, and all lesions were confirmed by standard pathology. Data were acquired in compliance with an institutionally approved human subjects research protocol [University of California, Irvine (UCI) 02-2306 and 95-563].

Here we present the algorithm for the double-differential spectroscopic method for computing the residual due to components that are unaccounted for by the basis spectra. Tissue

*Tel: 1-949-824-2674; E-mail: egratton22@yahoo.com

Fig. 1 Algorithm for the double-differential method: (a) Scatter-corrected absorption spectrum at a single spatial location over tumor-containing breast tissue. (b) Representation of average normal absorption spectra. (c) Difference spectrum obtained by subtracting (b) from (a). (d) STC spectrum obtained from subtracting the fit of the difference spectrum using the four basis components from the difference spectrum (c).

absorption and scattering spectra were calculated using custom software in MATLAB™.[1] The resulting absorption spectra were analyzed by in-house Elantest software to calculate the STC spectra.

We begin by assuming that the absorption spectrum at a particular location can be expressed as a linear combination of the basis spectra plus the unaccounted STC:

$$A_T(\lambda, x, y) = \sum_i a_i(x, y) \times S_i^*(\lambda) + \text{STC}(\lambda, x, y), \quad (1)$$

where A_T is the absorption spectrum of tumor-containing tissue as a function of position on the breast given by x and y coordinates, and wavelength λ. Following, a_i is the fractional contribution to the overall absorption. S_i^* contains the four component basis spectra. The * in this notation indicates that the spectra are patient specific (i.e., not known). In other words, we account for individual molecular differences of hemoglobin, water, and lipid by defining basis spectra for each subject. The goal of our analysis is to recover the STC.

In Fig. 1 we show a schematic of the algorithm. We begin by using a representative A_T (650 to 990 nm) at a single spatial location over the tumor [Fig. 1(a)]. This algorithm should also be performed for normal tissue. We then calculate the average absorption spectrum of the normal breast $A_N(\lambda)$. This is obtained by averaging the spectra taken at the various positions on the unaffected normal breast tissue. By taking an average we acquire a baseline spectrum of "normal" tissue for the patient [Fig. 1(b)]:

$$A_N(\lambda) = \sum_i c_i \times S_i^*(\lambda), \quad (2)$$

where c_i is the fractional contribution of the patient-specific basis spectra to the overall absorption in normal tissue.

The difference spectrum D [Fig. 1(c)], at each spatial location is obtained by subtracting the A_N [Fig. 1(b)] from the A_T [Fig. 1(a)]:

$$D(\lambda, x, y) = \sum_i \Delta_i(x, y) \times S_i^*(\lambda) + \text{STC}(\lambda, x, y). \quad (3)$$

Here, the symbol $\Delta_i(x, y) = (a_i - c_i)$ indicates differences between the basis component concentrations at different spatial locations. In calculating STC, the spectra are referenced to a patient-specific baseline of normal tissue. In other words, we correct the tumor spectra for any intrasubject variation to reveal only differences between tumor and normal tissue. Solving Eq. (3) for the STC spectrum leaves:
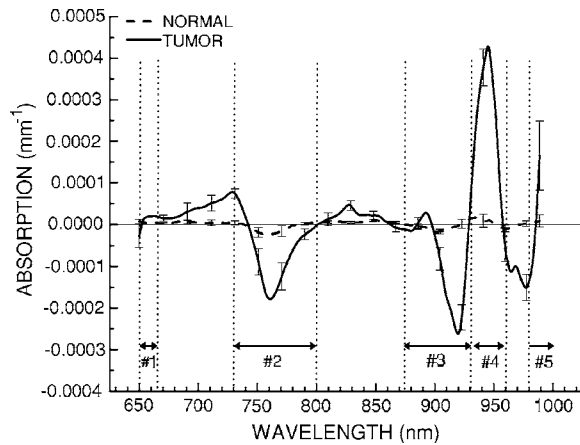
$$\text{STC}(\lambda, x, y) = D(\lambda, x, y) - \sum_i \Delta_i(x, y) \times S_i^*(\lambda). \quad (4)$$

In Eq. (4), the last term is the fit of D to the standard basis spectra. By fitting the difference spectrum D as opposed to A_T, the variation due to differences between the S and S^* basis spectra are minimized; thus the STC component can be recovered with relatively high precision. If we had fitted Eq. (1) instead of Eq. (3) using the standard basis spectra, the coefficients of the fit a_i would have been large and the differences between S^* and S (the patient-specific basis spectra and the standard basis spectra) would overwhelm the subtle differences due to the STC component. Thus the single differential residual mainly reflects intersubject differences, thereby masking the subtle differences between tumor and normal tissue for a given patient. In the double-differential method, we subtract out the common residual for tumor and normal from fitting to the standard basis spectra; thus the selection of the basis spectra has little impact in revealing spectral differences due only to tumors.

For the normal tissue, the difference spectrum (at each location) has been completely fitted by using the four components, resulting in residual spectra which are essentially a flat, featureless line with values close to zero. This means that the spectral differences between the normal and reference spectrum should be accounted for by the "natural" compositional differences, thus no new component should be needed.

Two crucial internal controls must be performed for each patient. If we calculate the residual from the double-differential method for different locations of the unaffected breast, we expect the residual spectra to be that of normal tissue. If we repeat this second differential analysis for the breast with tumor, we expect to observe a spatially dependent residual. The outcome for the normal surrounding tissue should be the same as that for the normal tissue spectra of the unaffected side. However, at the tumor regions we expect a residual different from normal tissue. We performed a phantom experiment to verify that the double-differential method can recover an unknown spectrum that is not part of the basis set. Unaccounted for components were recovered using the double-differential method.

In Fig. 2, we show a comparison of the average of 15 tumor spectra along with the average of normal spectra obtained from the contralateral normal side. Patients ranged in age from 32 to 57 years, nine pre- and six postmenopausal, with pathologically confirmed diagnosis of invasive ductal carcinoma, with one patient diagnosed with adenocarcinoma with lobular features.

Fig. 2 Average STC spectra (double-differential residual) for 15 patients at tumor locations exhibiting the largest variation in comparison to surrounding normal tissue on tumor-containing breast, as well as normal tissue at the equivalent position on normal contralateral breast. Standard error of mean bars calculated every 20 nm indicate STC spectral variation in population. Regions are defined in the text.

From inspection of the average STC spectra, we find systematic differences in roughly five spectral regions: region 1: 650 to 665 nm; region 2: 730 to 800 nm; region 3: 875 to 930 nm; region 4: 930 to 960 nm; and region 5: 980 to 990 nm (Fig. 1). In an effort to quantify the amount of STC spectrum, or the amount of "tumor biomarker," we calculate the local residual variance for each spectral region, defined by:

$$L_k(x,y) = \sum_{i,k} STC_i(\lambda_i,x,y)^2 \Big/ N_k. \qquad (5)$$

The local variance L_k is a function of position on the breast given by x and y coordinates. The index k indicates a given spectral region and N_k indicates the total number of wavelengths in the spectral region. $STC_i(\lambda_i,x,y)$ is the value of the STC spectra at a given wavelength. The STC index is the sum of all local variance L_k. Across all patients, the STC index displays the maximum value over regions of tumor-containing tissue, as opposed to normal tissue surrounding the lesion or normal contralateral breast tissue. Note that the STC index is our first step toward quantifying the unique spectrum and using it to identify lesions. Improved methods of spectral content analysis must be further developed.

With regards to tissue heterogeneity, we make two observations. First, we used reference spectra from different spatial locations and found that tissue heterogeneity does not affect the recovered STC spectra. Second, we believe that the amount of the STC component could depend on the particular geometry of the lesion. However, the overall spectra shape of the STC should be less affected.

The investigation of the biochemical/physical origin of the STC spectrum will be the subject of a future manuscript. Changes in region 1 (650 to 665 nm) may be due to other types of hemoglobins or breakdown products in tumors. In region 2 (730 to 800 nm), we observe a distinct negative peak in the STC spectra. As per the definition of STC, a negative peak indicates absence of a component or a broadening of the NIR band. Spectral changes in this region may be suggestive of different lipid composition in the tumor. With regards to region 3 (875 to 930 nm) and region 4 (930 to 960 nm), we see a negative peak immediately followed by a positive peak. This may be indicative of a spectral shift toward longer wavelengths. This is also the spectral region in which lipids have characteristic absorption spectra due to absorption of C-H bonds at 920 to 930 nm. Thus if there are differences in lipid composition in the tumor-containing tissue with respect to the normal tissue, then this spectral region is likely to be affected. In region 5 (980–990 nm) we observe spectral changes that may be attributed to differences in the O-H overtone region. Two possible candidates for these changes are the O-H of water and lipid oxidation products.

In conclusion, we present an algorithm for a novel analysis method of NIR absorption spectra. We show an application of method to breast tissue absorption spectra. Through this method we retrieve the spectral differences between normal and diseased tissue for a patient by fitting the difference spectra using the four basis component spectra and then analyzing the residuals of this fit. This differential approach requires regions of normal and tumor breast tissue within a patient. With this method, we show the presence of spatially localized intrinsic specific spectroscopic biomarkers of breast tumors.

References

1. A. Cerussi, N. Shah, D. Hsiang, A. Durkin, J. Butler, and B. J. Tromberg, "In vivo absorption, scattering and physiologic properties of 58 malignant breast tumors determined by broadband diffuse optical spectroscopy," *J. Biomed. Opt.* **11**(4), 044005 (2006).
2. B. W. Pogue, S. Jiang, H. Dehghani, C. Kogel, S. Soho, S. Srinivasan, X. Song, T. D. Tosteson, S. P. Poplack, and K. Paulsen, "Characterization of hemoglobin, water, and NIR scattering in breast tissue: analysis of intersubject variability and menstrual cycle changes," *J. Biomed. Opt.* **9**(3), 541–52 (2004).
3. P. Taroni, G. Danesini, A. Torricelli, A. Pifferi, L. Spinelli, and R. Cubeddu, "Clinical trial of time-resolved scanning optical mammography at 4 wavelengths between 683 and 975 nm," *J. Biomed. Opt.* **9**(3), 464–73 (2004).
4. J. G. Kim, M. N. Xia, and H. L. Liu, "Extinction coefficients of hemoglobin for near-infrared spectroscopy of tissue," *IEEE Eng. Med. Biol. Mag.* **24**(2), 118–121 (2005).
5. R. Katz-Brull, P. T. Lavin, and R. E. Lenkinski, "Clinical utility of proton magnetic resonance spectroscopy in characterizing breast lesions," *J. Natl. Cancer Inst.* **94**(16), 1197–1203 (2002).