TITLE PAGE

Citation Format:

P. Taroni, A. Pifferi, R. Cubeddu, F. Ieva, A. M. Paganoni, F. Abbate, and E. Cassano, "Optical Quantification of Collagen and Breast Cancer: Lesion Classification and Risk Estimate," in Biophotonics Congress: Biomedical Optics Congress 2018 (Microscopy/Translational/Brain/OTS), OSA Technical Digest (Optical Society of America, 2018), paper OTh4D.1.

Copyright notice:

DOI abstract link:

https://doi.org/10.1364/OTS.2018.OTh4D.1

Optical Quantification of Collagen and Breast Cancer: Lesion Classification and Risk Estimate

P. Taroni, A. Pifferi, R. Cubeddu
Politecnico di Milano − Dipartimento di Fisica, Milano, Italy
F. Ieva, A.M. Paganoni
Politecnico di Milano − Dipartimento di Matematica, Milano, Italy
F. Abbate, E. Cassano
European Institute of Oncology – Divisione di Radiologia Senologica, Milano, Italy
Author e-mail address: paola.taroni@polimi.it

Abstract: Collagen content quantified through 7-wavelength (635-1060 nm) time domain diffuse optical mammography in 200 women proved key to discriminate malignant from benign breast lesions, to measure breast density, and to estimate breast cancer risk.
OCIS codes: (170.6510) Spectroscopy, tissue diagnostics; (170.3830) Mammography; (170.6920) Time-resolved imaging; (170.5280) Photon migration.

1. Introduction

Breast cancer is commonly associated to drastic alterations of the architecture and composition of the extracellular matrix (ECM), and collagen is one of the major ECM constituents [1]. The possibility to quantify collagen non-invasively may thus be of great interest both to diagnose breast cancer and to estimate the risk of developing it.

After a preliminary characterization of collagen absorption properties in the red and near-infrared (NIR) [2], we started including collagen among tissue constituents quantified from *in vivo* data. In particular, collagen was quantified for 200 subjects (healthy and with malignant and benign breast lesions) in a clinical study aiming at the optical characterization of breast lesions, and at the optical assessment of breast density, *i.e.* the fibro-glandular fraction of breast tissue, which is recognized as a strong independent risk factor for developing breast cancer [3].

2. Experimental set-up and procedures

Instrument set-up
The instrument is designed to collect projection images in compressed breast geometry. Time-resolved transmittance measurements were performed at seven wavelengths (635, 685, 785, 905, 930, 975, 1060 nm) using picosecond pulsed diode lasers and two PC boards for time-correlated single photon counting. The compressed breast was raster-scanned continuously, moving an illumination fiber and a collecting bundle in tandem and recording data every millimeter. Optical images were routinely acquired from both breasts in cranio-caudal and medio-lateral oblique (45°) views. Details on the instrument and its performance can be found in [4].

Patient study
The Institutional Review Board at the European Institute of Oncology approved the clinical study. Written informed consent was obtained from all participants. The study had twofold aim: i) the non-invasive assessment of breast density by optical means, and ii) the optical characterization of malignant and benign lesions.

218 subjects enrolled, but only 200 met the inclusion criteria for the estimate of breast density, and 84 for the characterization of breast lesions (45 malignant lesions and 39 benign lesions).

Data analysis
Absorption and reduced scattering coefficients at each wavelength were estimated by fitting the experimental data to an analytical solution of the diffusion approximation (with the extrapolated boundary condition) for an infinite homogeneous slab.

Information on tissue composition and structure was obtained directly from time-resolved transmittance curves measured at 7 wavelengths. The Beer law was used to relate the absorption properties to the concentrations of the main tissue constituents, while the scattering properties were modeled through the “power law” obtained as an approximation to Mie theory. A spectrally constrained global fitting procedure was applied, where free parameters were the concentrations of oxy- and deoxy-hemoglobin, water, lipids, and collagen, together with the scattering amplitude *a* and power *b*.

For the estimate of breast density, tissue composition and scattering parameters were averaged over each image, excluding regions close to the boundary of the compressed breast, which cannot correctly be described by the

infinite slab model. Then the average breast composition and scattering parameters of each subject were obtained averaging the results for the four views (cranio-caudal and oblique views for both breasts).

The statistical significance of the difference between groups of women with different breast density based on tissue composition was assessed using the Mann-Whitney (with $p < 0.05$). To identify women at high risk for their dense breast tissue, the best regression logistic model based on a tissue composition and scattering parameters was chosen via a stepwise variable selection performed minimizing the AIC (Akaike Information Criterion) [5].

For the optical characterization of breast lesions, they were described as localized absorption perturbations, embedded in an otherwise homogeneous diffusive medium, and a perturbation model based on the high-order calculation of the pathlength of photons inside the lesion was applied, as described in detail in [6]. The inhomogeneity (*i.e.*, the lesion) was assumed as spherical (as only the maximum diameter was known) and located halfway between the injection and detection points (as no information on the depth was available from the projection images). Once the absorption difference between the lesion and the average tissue in the same breast had been calculated, the Beer law was used to estimate the difference in composition between lesion and average breast tissue.

To achieve lesion discrimination, we exploited both optically derived tissue composition (Hb and HbO_2, lipid, water, and collagen) or absorption properties and information known from the patient's demographics and anamnesis (age, body mass index (BMI), parity, family history of breast cancer, use of oral contraceptives (OC), and use of preventive Tamoxifen (TAM). Classification analysis was performed implementing the most common version of the AdaBoost procedure with decision trees, the so called Discrete AdaBoost [7].

3. Results and discussion

Mammographic density reflects breast tissue composition. More specifically, it is determined by the amount of connective and epithelial tissue (opaque to x-rays) as compared to adipose tissue (translucent). Thus, the sensitivity of optical techniques to tissue composition can at least in principle be exploited to assess non-invasively breast density and the related risk of developing breast cancer.

Our results show that increasing mammographic density (*i.e.*, increasing BI-RADS category) corresponds to increasing amounts of water, collagen, and total hemoglobin content (*tHb*). Also scattering amplitude and power increase, denoting higher density and equivalent size of the scattering centers. Only the amount of lipids decreases, and the blood oxygenation level (SO_2) does not change significantly [8]. All these observations are in agreement with adipose tissue being replaced by fibro-glandular tissue in denser breasts.

Identifying early in their lifetime women that are at high risk for their high breast density would allow designing for them dedicated screening paths, and also interventions to reduce breast density and the related risk. Thus, we looked for the best regression logistic model to identify women in BI-RADS category 4 (highest risk), based on their tissue composition and scattering parameters. The following model was obtained [5]:

$$logit(p_i) = \alpha_0 + \alpha_1 * [Collagen]_i + \alpha_2 * a_i + \alpha_3 * b_i \quad (1)$$

Table 1: Parameters of the regression logistic model

Coefficient	α_0 (Intercept)	α_1 (Collagen)	α_2 (a)	α_3 (b)
Estimate	-10.4609	0.0157	0.2278	6.2876
Std Error	1.9685	0.0069	0.1141	1.218
z-value	-5.314	2.289	1.997	5.162
p-value	1.07*10-7	0.0221	0.0458	2.44*10-7

The model depends only on collagen content and scattering parameters (which in turn depend on collagen content and possibly also structure and organization in tissue). Thus, collagen proved to be a key tissue constituent to distinguish women at high risk for their high breast density. The model reaches a specificity of 94% and sensitivity of 69%. The latter looks rather low, but it should be taken into account that the aim is not diagnosing cancer, but rather recognizing women who deserve specific screening/diagnostic attention, and – for economic sustainability – only a small group could reasonably be considered.

The non-invasive optical assessment of collagen offers an effective alternative way to measure breast density, but it may turn out to be even more important if collagen proves to be also a risk factor independent of mammographic density. To investigate this possibility, we compared information on breast tissue in healthy women ($N = 53$) and in breast cancer patients ($N = 56$). The average collagen content does not differ significantly between cancer patients and healthy women. A higher collagen content is observed in cancer patients after normalization to correct for age dependence, as age is the strongest risk factor for breast cancer. Still, the significance is not very high ($p = 0.0366$). Similar results are obtained also for mammographic density, and that seems to suggest that both

collagen content and mammographic density are sensitive to the same risk factor. However, women at high risk for their mammographic density are essentially those in BI-RADS category 4, namely 10-15% of the entire population. If we rank women for decreasing collagen content (or mammographic density) and consider the top 15% of them, the two top subgroups (collagen and density) contain a much higher percentage of women with cancer than the entire population of the study: 81% *vs* 51%, confirming that both collagen and density measure risk. It is interesting to observe (Fig. 1) that different "high-risk" women are identified based on high collagen content or high density, suggesting that the two figures may explore different risk factors.

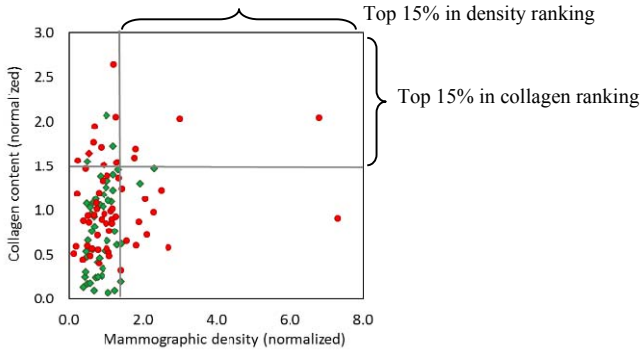

Fig. 1. Age-normalized collagen content *vs* mammographic density in cancer (red) and healthy (green) groups. From [9].

For the discrimination between malignant and benign breast lesions, the multivariate comparison of the means of tissue constituent concentrations was not effective. Thus, we tested a machine-learning algorithm, the Discrete AdaBoost procedure, operating on tissue composition and information available from the patient's anamnesis [10]. Sensitivity of 88% and specificity of 79% were obtained. For lesion classification, tissue composition proved to be more important than other parameters, and collagen was again the most important tissue constituent (Table 2).

Table 2: Average variable importance rank for lesion discrimination with the Discrete AdaBoost procedure.

Collagen	**Lipid**	**BMI**	**HbO_2**	**Water**	**Hb**	**Age**	**Parity**	**OC**	**Familiarity**	**TAM**
0.09	0.0885	0.0871	0.0865	0.0851	0.0836	0.0739	0.0639	0.0531	0.0400	0.0274

In conclusion, diffuse optical spectroscopy is a powerful tool to investigate tissue non-invasively, and the optical assessment of collagen certainly deserves further investigation for breast cancer diagnosis and risk estimate.

4. References

1. C. Luparello, "Aspects of Collagen Changes in Breast Cancer," J. Carcinog. Mutagen. **S13**, (2013).
2. 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**, 14021 (2007).
3. V. A. McCormack and I. dos Santos Silva, "Breast Density and Parenchymal Patterns as Markers of Breast Cancer Risk: A Meta-analysis," Cancer Epidemiol. Biomarkers Prev. **15**, 1159–1169 (2006).
4. P. Taroni, A. Pifferi, E. Salvagnini, L. Spinelli, A. Torricelli, and R. Cubeddu, "Seven-wavelength time-resolved optical mammography extending beyond 1000 nm for breast collagen quantification.," Opt. Express **17**, 15932–46 (2009).
5. P. Taroni, G. Quarto, A. Pifferi, F. Ieva, A. M. Paganoni, F. Abbate, N. Balestreri, S. Menna, E. Cassano, and R. Cubeddu, "Optical identification of subjects at high risk for developing breast cancer.," J. Biomed. Opt. **18**, 60507 (2013).
6. G. Quarto, L. Spinelli, A. Pifferi, A. Torricelli, R. Cubeddu, F. Abbate, N. Balestreri, S. Menna, E. Cassano, and P. Taroni, "Estimate of tissue composition in malignant and benign breast lesions by time-domain optical mammography.," Biomed. Opt. Express **5**, 3684–98 (2014).
7. J. Friedman, T. Hastie, and R. Tibshirani, "Additive Logistic Regression: A Statistical View of Boosting on JSTOR," Ann. Stat. **28**, 337–374 (2000).
8. P. Taroni, G. Quarto, A. Pifferi, F. Abbate, N. Balestreri, S. Menna, E. Cassano, and R. Cubeddu, "Breast tissue composition and its dependence on demographic risk factors for breast cancer: non-invasive assessment by time domain diffuse optical spectroscopy.," PLoS One **10**, e0128941 (2015).
9. P. Taroni, A. M. Paganoni, F. Ieva, F. Abbate, E. Cassano, R. Cubeddu, and A. Pifferi, "Is collagen an independent risk factor for breast cancer?," in *Biomedical Optics 2016, OSA Technical Digest (Online)* (2016), p. TTh4B.3.
10. P. Taroni, A. M. Paganoni, F. Ieva, A. Pifferi, G. Quarto, F. Abbate, E. Cassano, and R. Cubeddu, "Non-invasive optical estimate of tissue composition to differentiate malignant from benign breast lesions: A pilot study[1] P. Taroni, A. M. Paganoni, F. Ieva, A. Pifferi, G. Quarto, F. Abbate, E. Cassano, and R. Cubeddu, "Non-invasive optical estimate of," Sci. Rep. **7**, (2017).