

Discrimination between healthy and tumor tissues on formalin-fixed paraffin-embedded breast cancer samples using IR imaging

Audrey Bénard^{a,*}, Christine Desmedt^b, Virginie Durbecq^b, Ghizlane Rouas^b, Denis Larsimont^c, Christos Sotiriou^b and Erik Goormaghtigh^a

^a *Laboratory for the Structure and Function of Biological Membranes, Center for Structural Biology and Bioinformatics, Université Libre de Bruxelles, Brussels, Belgium*

^b *Functional Genomics and Translational Research Unit, Department of Medical Oncology, J. Bordet Institute, Brussels, Belgium*

^c *Pathology Department, J. Bordet Institute, Brussels, Belgium*

Abstract. This work presents a pilot study to illustrate the potential of Fourier transform infrared (FT-IR) imaging in breast cancer research. Using this technique, we have acquired infrared (IR) microspectroscopic images from healthy and cancerous breast tissue section from one patient. First of all, a Student *t*-test was applied, showing DNA/RNA spectral region (1400–1000 cm^{-1}) as the most discriminant for the differentiation between healthy and tumor samples. Afterwards, a supervised pattern recognition method, Partial Least Squares (PLS) was used to develop an automated classifier to discriminate the two classes of data. Infrared spectra of independent IR measurements were used to test the classifier. The class identity was correlated with information obtained by histopathologic gold standard. The results showed that more than 95% of the training and validation spectra were correctly identified. We demonstrate that combination between IR microspectroscopic imaging and multivariate data analysis can be used as a complement to present diagnostic tools for breast cancer.

Keywords: Infrared microspectroscopy, infrared imaging, breast cancer, cancer diagnosis

1. Introduction

Breast cancer is the most frequently diagnosed cancer in women in Western countries. This cancer represents approximately 30% of all cancers diagnosed and 16% of all cancer deaths [3]. The prognosis and diagnostic of breast cancer has always been based on clinical variables such as histologic type and grade, lymph node involvement, tumor size and status of hormonal receptors. After surgery or biopsy, a specimen is prepared using histology procedures for viewing under a microscope and the diagnosis of tumors relies on the visual inspection of stained tissue sections by a trained pathologist. The visualisation of the structure and distribution of cellular components in tissue sections using light microscopy has become the histopathologic gold standard procedure.

*Corresponding author: Audrey Bénard, Laboratory for the Structure and Function of Biological Membranes, Center for Structural Biology and Bioinformatics, Campus Plaine, Université Libre de Bruxelles, Bld du Triomphe Acces 2, CP 206/02, B1050 Brussels, Belgium. Tel.: +32 2 650 53 65; Fax: +32 2 650 53 82; E-mail: Audrey.Benard@ulb.ac.be.

Because of its potential to probe tissues and cells at the molecular level without requirement for extrinsic contrast agents, infrared spectroscopy could become an attractive tool in clinical and diagnostic analysis to complement the existing methods. Hundreds of biological applications of this technology have been published since it was demonstrated, in the eighties, that FT-IR spectra of bacteria provide a unique fingerprint that allows the identification of bacteria species [4,9,14]. In cancer research, it was demonstrated that different tumor cell lines can be identified by statistics analysis of their FT-IR spectra [10,15,16]. Furthermore, tumor cell lines with different biological behaviour can be separated thanks to this technique. For instance, *in vivo* aggressiveness and *in vitro* migration of glioma cell lines were successfully predicted from their IR spectra while no molecular biology technique was available [5]. As IR spectroscopy is based on the absorption of infrared light by vibrational transitions in covalent bonds, characteristic spectral features are correlated with biological properties of the sample. So, useful diagnostic information can be extracted from infrared spectra in case of different pathologies [7,12].

The recent availability of IR-sensitive multi-channel array of detectors allows to obtain spatially resolved chemical and structural information of tissue sample. IR imaging presents several advantages. (i) This technology is rapid as IR data can be collected and interpreted within minutes by statistical analysis; (ii) neither staining of the samples nor chemical reagent additions are necessary as the intrinsic molecular vibrations probe the chemical composition and structural properties of the sample; (iii) this technology is non-destructive. Because of the subtle alterations in the morphological and biochemical composition of the tissue associated with cancer transformation, collection of high-quality spectra and high spatial resolution is necessary for the diagnostic of cancer disease.

Here, we applied mid-IR microspectroscopic imaging to breast cancer tissue and demonstrate that the differentiation between normal and malignant tissue can be obtained by statistical analysis of their IR features. Because the distinction between tumor and normal tissue is crucial in surgery to maximally remove tumor cells and minimally affect normal breast tissue, we evaluate in the present work the potential of IR microspectroscopy for diagnostic of breast cancer disease. Importantly, we show that the identification of cell type can be achieved on formalin-fixed paraffin-embedded tissues (FFPE) which gives access to a large tumor tissue bank and allows direct comparison with histopathologic gold standard procedures.

2. Materials and methods

2.1. Sample preparation

Human breast tissue was obtained from the FFPE Tumor Bank of the Institute J. Bordet in Brussels. The breast sample corresponds to one surgical piece of one patient. A 3 μm tissue section was cut from paraffin-embedded tissue block mounted on a glass slide and stained with Hematoxylin and Eosin (H&E) for histological assessment. Regions of interest (healthy and tumor tissue) were identified using the stained section by a trained pathologist. A second 3 μm tissue section was mounted on a barium fluoride (BaF_2) disk for IR imaging analysis. This section was subsequently deparaffinized, rehydrated and dried but not stained.

2.2. Spectroscopic data acquisition and processing

The IR data were collected using a Hyperion 3000 IR imaging system (Bruker Optics, Ettlingen, Germany), equipped with a 64×64 Mercury Cadmium Telluride (MCT) Focal Plane Array (FPA)

detector. The data were collected in transmission mode from sample regions of $170 \times 170 \mu\text{m}^2$. Every individual element of the array detector covers an area of $2.7 \times 2.7 \mu\text{m}^2$. The spectral resolution was set to 8 cm^{-1} . One IR image results in 4096 spectra, each one being the average of 256 scans (ca 5 min recording). All the spectra were preprocessed as follows. The water vapour contribution was subtracted as described previously [6] with $1956\text{--}1935 \text{ cm}^{-1}$ as reference peak. Spectra were retained for further analysis when, on the Amide I and II region, the absorbances were superior to -0.2 and inferior to 1.4 absorbance units and when the Signal-to-Noise ratio was better than 500:1. This ratio is calculated using $1750\text{--}1480 \text{ cm}^{-1}$ spectral range for signal and $2000\text{--}1900 \text{ cm}^{-1}$ region for noise determination. Finally the spectra were normalized for equal area between 1481 and 1725 cm^{-1} (Amide I peak) and converted into second derivative spectra.

A Student t -test is used in this work to determine whether two population means are equal or not. This supervised statistical test requires a normal distribution of the IR data which was checked with success for the two populations by a Kolmogorov–Smirnov test. The significance level (α level) was set to 0.1% .

Discriminant Partial Least Squares (PLS) is a supervised statistical analysis which requires a priori identification of groups of data contained within the training samples; for example, in our case, the identification of IR spectra belonging to healthy tissue and those representing tumor tissue. This technique will group the IR data into classes predefined by the operator and construct a discriminant model which will be tested on a validation data set (IR images not included in the training data set).

3. Results and discussion

Four IR images were recorded in the healthy region of the tissue slice and four images in the tumor (see Section 2.2). To gain more insight into the biochemical differences between the two breast tissue regions of interest, a Student t -test was performed. After computation of the mean spectra for each condition, significant differences ($\alpha = 0.1\%$) were observed mainly in the DNA/RNA spectral region ($1400\text{--}1000 \text{ cm}^{-1}$) (Fig. 1).

On a second step, a discriminant PLS classifier was built based on the histopathological information of the IR measurements. Four IR images per condition were used for the training data set. For the discrimination between the healthy and tumor samples, the PLS model was constructed on the most significant spectral region determined by the Student t -test ($1400\text{--}1000 \text{ cm}^{-1}$). After pretreatments and filters (see Section 2.2), 100 spectra per IR image were randomly extracted and used as the training set i.e. 400 spectra for each condition. Table 1 reports the results of the PLS classification for the training set of spectra. For the validation data set, two additional IR images per condition were recorded from the breast tissue sample. The results of the assignments for the validation data set are shown in Table 1, after elimination of spectra with low intensity, saturation and low Signal-to-Noise ratio as explained in Section 2. Between 3400 and 4000 spectra were retained per IR image and used for the statistical analysis. Table 1 reports that more than 95% of the spectra were classified correctly by the PLS discriminant model built on the DNA/RNA spectral region ($1400\text{--}1000 \text{ cm}^{-1}$).

In this article, we have applied a Student t -test, supervised statistical test to gain more insight into the biochemical differences between diseased and healthy tissue which is mainly composed of collagen and fibroblasts. Yet most of these differences were located in the DNA/RNA spectral region. Several studies have previously reported spectral differences for cells at different stages of the eukaryotic cell division cycle [1,2,13]. These differences are caused by the morphological and biochemical changes occurring within the cell division cycle. It was suggested that cells DNA is detectable mainly during the replication

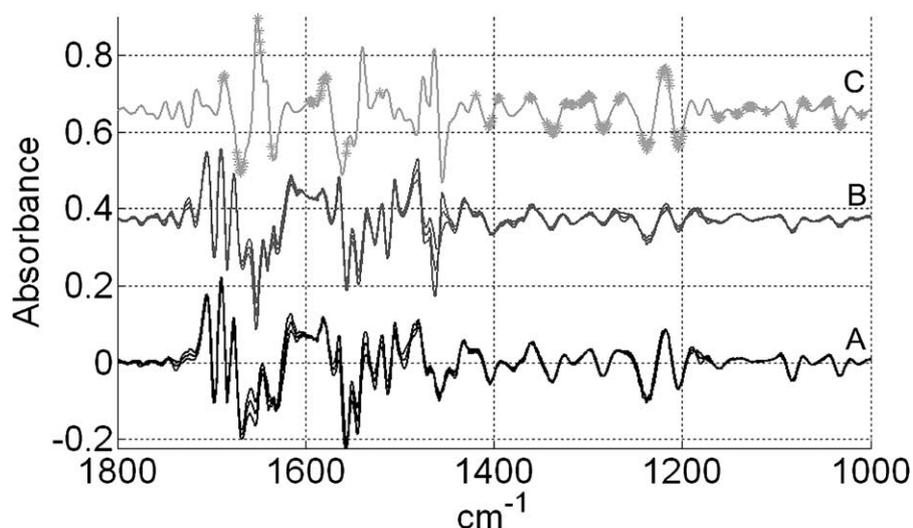


Fig. 1. Spectral features comparison by a Student *t*-test between second derivative spectra recorded in healthy and tumor regions of the breast tissue slice. Mean spectrum corresponding to healthy (A) and tumor class (B), obtained after averaging the spectra of 4 IR images per condition. The standard deviations for the mean spectra are plotted in dashed lines. (C) The difference spectrum between the two conditions was multiplied by a factor 2. The Student *t*-test was computed at every wavenumber on the 1800–1000 cm^{-1} spectral region with a significance level of 0.1%. Each marked wavenumber (gray star) represents a statistically significant difference between the two means.

Table 1
Identification accuracy of the training and validation data sets

Attribution by a trained pathologist	Number of spectra included in the training/validation data set		Attribution by the PLS model % of correctly classified spectra
Healthy tissue	Training set	400	100
	Validation set #1	3782	99.9
	Validation set #2	3447	95
Tumor	Training set	400	100
	Validation set #1	3982	100
	Validation set #2	3742	99.9

Notes: Proportion of correct identification (in %) in the training and validation data sets by PLS discriminant analysis applied to healthy and tumor breast tissues. For the training data set, 400 spectra per condition (healthy and diseased samples) were used for the construction of the PLS model. For the validation set, two additional IR images per condition were recorded.

stage. Otherwise, the nucleic acid contribution of the spectrum is mostly due to cytoplasmic RNA, DNA being too tightly packed in the nucleus to allow the infrared radiation to cross through [1]. Meanwhile, Mourant et al. demonstrated that the ratio RNA/lipid could be used to determine the proliferative status of mammalian cells [13]. A diagnostic advantage of the FT-IR spectroscopy in cancer research is that the development of a cancer DNA phenotype can be observed prior to the tumor formation, notably for prostate carcinomas [11]. As tumorigenesis in humans is a multistep process, cells acquire several capabilities leading to the progressive conversion of normal human cells into cancer cells, notably the limitless potential of growth [8]. As cancer cells are in high proliferative stage and DNA is more detectable in the S-phase of growth, it is not surprising that the DNA/RNA spectral region shows the most

discriminative differences between healthy and tumor samples. Anyway, a PLS model based to this spectral region shows good classification of the validation data set.

In conclusion, the present study demonstrates that healthy and tumor regions in breast tissue section can be differentiated on the basis of their IR spectra. There are only small differences between mean spectra of the two conditions, but thanks to the high Signal-to-Noise ratio obtained, the discriminant PLS model was able to recognize characteristic spectral features which allow the construction of an automated classifier able to separate healthy from tumor tissue. Importantly, we show here that IR imaging can be applied on formalin-fixed paraffin-embedded tissues paving the way to retrospective studies. We also demonstrate that IR microspectroscopic imaging is a powerful tool for biodiagnostics respecting the tissue microheterogeneity and requiring only a few minutes for IR data acquisition. In turn, this technology has a great potential for the detection of isolated cells present in the healthy tissue like metastatic cells.

Acknowledgements

This research has been supported by a grant from Interuniversity Attraction Poles (IAP) P6/19 (Belgium) and the National Fund for Scientific Research (FRFC 2.4533.10 and 2.4588.06). Audrey Bénard is Research Fellow for the Fund for Research in the Industry and Agriculture (Communauté Française de Belgique). Erik Goormaghtigh is Research Director with the National Fund for Scientific Research (Belgium). The work of Christine Desmedt, Ghizlane Rouas and Christos Sotiriou is supported by “La Région Bruxelloise”, “les Amis de l’Institut Jules Bordet” and National Fund for Scientific Research, respectively.

References

- [1] S. Boydston-White, T. Gopen, S. Houser, J. Bargonetti and M. Diem, Infrared spectroscopy of human tissue. V. Spectroscopic studies of myeloid leukemia cells at different phases of the cell cycle, *Biospectroscopy* **5** (1999), 219–227.
- [2] S. Boydston-White, M. Romeo, T. Chernenko, A. Regina, M. Miljkovic and M. Diem, Cell-cycle-dependent variations in FT-IR micro-spectra of single proliferating Hela cells: Principal component and artificial neural network analysis, *Biochim. Biophys. Acta* **1758** (2006), 908–914.
- [3] J.P. Brettes, C. Mathelin, B. Gairard and J.P. Bellocq, *Cancer du sein*, Masson, Issy-les-Moulineaux, France, 2007.
- [4] L.P. Choo-Smith, K. Maquelin, T. Van Vreeswijk, H.A. Bruining, G.J. Puppels, N.A.G. Thi, C. Kirschner, D. Naumann, D. Ami, A.M. Villa, F. Orsini, S.M. Doglia, H. Lamfarraj, G.D. Sockalingum, M. Manfait, P. Allouch and H.P. Endtz, Investigating microbial (micro)colony heterogeneity by vibrational spectroscopy, *Appl. Environ. Microbiol.* **67** (2001), 1461–1469.
- [5] A. Gaigneaux, C. Decaestecker, I. Camby, T. Mijatovic, R. Kiss, J.M. Ruysschaert and E. Goormaghtigh, The infrared spectrum of human glioma cells is related to their *in vitro* and *in vivo* behaviour, *Exp. Cell Res.* **297** (2004), 294–301.
- [6] E. Goormaghtigh, V. Cabiaux and J.M. Ruysschaert, Determination of soluble and membrane protein structure by Fourier transform infrared spectroscopy. II. Experimental aspects, side chain structure and H/D exchange, *Subcell. Biochem.* **23** (1994), 363–403.
- [7] D.M. Haaland, D.T.J. Howland and E.V. Thomas, Multivariate classification of the infrared spectra of cell and tissue samples, *Appl. Spectrosc.* **51** (1997), 340–345.
- [8] D.M. Hanahan and R.A. Weinberg, The hallmarks of cancer, *Cell* **100** (2000), 57–70.
- [9] C. Kirschner, K. Maquelin, P. Pina, N.A.G. Thi, L.P. Choo-Smith, G.D. Sockalingum, C. Sandt, D. Ami, F. Orsini, S.M. Doglia, P. Allouch, M. Manfait, G.J. Puppels and D. Naumann, Classification and identification of enterococci: a comparative phenotypic, genotypic and vibrational spectroscopic study, *J. Clin. Microbiol.* **39** (2001), 1763–1770.
- [10] D.C. Malins, N.L. Polissar and S.J. Gunselman, Models of DNA structure achieve almost perfect discrimination between normal prostate, benign prostatic hyperplasia (BPH), and adenocarcinoma and have a high potential for predicting BPH and prostate cancer, *Proc. Natl. Acad. Sci. USA* **94** (1997), 259–264.

- [11] D.C. Malins, K.M. Anderson, N.K. Gilman, V.M. Green, E.A. Barker and K.E. Hellström, Development of a cancer DNA phenotype prior to tumor formation, *Proc. Natl. Acad. Sci. USA* **101**(29) (2004), 10721–10725.
- [12] L.M. McIntosh, M. Jackson, H.H. Mantsch, M.F. Stranc, D. Pilavdzic and A.N. Crowson, Infrared spectra of basal cell carcinomas are distinct from non-tumor-bearing skin components, *J. Invest. Dermatol.* **112** (1999), 951–956.
- [13] J.R. Mourant, Y.R. Yamada, S. Carpenter, L.R. Dominique and J.P. Freyer, FT-IR spectroscopy demonstrates biochemical differences in mammalian cell cultures at different growth stages, *Biophys. J.* **85** (2003), 1938–1947.
- [14] D. Naumann, D. Helm and H. Labischinski, Microbiological characterizations by FT-IR spectroscopy, *Nature* **351** (1991), 81–82.
- [15] B. Rigas and P.T.T. Wong, Human colon adenocarcinoma cell lines displays infrared spectroscopic features of malignant colon tissues, *Cancer Res.* **52** (1992), 84–88.
- [16] B. Rigas, S. Morgello, I.S. Goldman and P.T.T. Wong, Human colorectal cancers display abnormal Fourier-transform infrared spectra, *Proc. Natl. Acad. Sci. USA* **87** (1990), 8140–8144.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

