

ORE Open Research Exeter

TITLE

Deciphering the structural and chemical composition of breast cancer using FTIR spectroscopy

AUTHORS

Lazaro-Pacheco, D; Shaaban, A; Baldwin, G; et al.

JOURNAL

Applied Spectroscopy Reviews

DEPOSITED IN ORE

16 December 2020

This version available at

http://hdl.handle.net/10871/124134

COPYRIGHT AND REUSE

Open Research Exeter makes this work available in accordance with publisher policies.

A NOTE ON VERSIONS

The version presented here may differ from the published version. If citing, you are advised to consult the published version for pagination, volume/issue and date of publication

Deciphering the structural and chemical composition of breast cancer using FTIR spectroscopy

Daniela Lazaro-Pacheco^a, Abeer Shaaban^b, Gouri Balden^b, Nicholas Akinwale Titiloye^c, Shazza Rehman^d and Ihtesham ur Rehman^{e*}

^a Department of Engineering, University of Exeter, Exeter, UK; ^b Department of Histopathology, University Hospitals Birmingham NHS Foundation Trust, Queen Elizabeth Hospital Birmingham, and University of Birmingham. Birmingham, UK; ^c Department of Pathology, School of Medicine and Dentistry, Kwame Nkrumah University of Science & Technology. Kumasi, Ghana; ^d Department of Medical Oncology, Airedale NHS Foundation Trust, Airedale General Hospital, Steeton, West Yorkshire, UK; ^e Engineering Department, Faculty of Science and Technology, Lancaster University.

*+44 (0) 1524 594038 i.u.rehman@lancaster.ac.uk

IUR, SR, and AS were involved in the conception and design.

AS and NAT were involved in sample collection and histopathological analysis.

GB was involved in the TMAs design and preparation.

DLP was involved in spectral data collection, analysis and interpretation of the data; and the drafting of the paper

All authors were involved in revision and approval of the paper contents.

Deciphering the structural and chemical composition of breast cancer using FTIR spectroscopy

A critical factor to favor a good prognosis in breast cancer (BC) patients is early detection. Offering prompt diagnosis, treatment, and delaying or stopping the progression of the disease are key aspects of breast cancer management. The use of FTIR spectroscopy on ex vivo breast samples can elucidate important biochemical information associated with the presence and progression of breast cancer. In this study, tissue microarrays of breast cancer biopsy samples (n = 378) and normal breast (n=134) were analyzed using FTIR spectroscopy, principal component analysis (PCA) for feature extraction and validation employing linear discriminant analysis (LDA). The differentiation between normal breast and breast cancer was successfully achieved with a sensitivity of 92% and specificity of 86%. Lipids and proteins spectral bands were identified as the main differentiators between normal and breast cancer. FTIR results also highlighted that chemical structural changes formed an important part of breast cancer evolution. Regardless of the heterogeneity of breast cancer, the use of multivariate analysis and FTIR spectroscopy offer a suitable and reliable tool for BC monitoring and diagnosis. The integration of spectroscopic techniques such as FTIR in cancer diagnosis and monitoring could add useful information enabling the diagnosis and management of breast cancer patients

Keywords: Breast cancer; FTIR spectroscopy; infrared spectroscopy; multivariate analysis.

Word count: 3961

1. Introduction

Breast cancer is the most common type of cancer among women accounting for 30% of all female cancers ^[1]. Early disease detection and diagnosis are associated with lower mortality and a more favorable prognosis ^[2]. Once a tumor is suspected, imaging, biopsy and histopathological analysis are the typical procedures to follow. Currently, histopathological assessment of biopsy samples remains the gold standard for diagnosis. However, health care systems in developed and developing countries are experiencing a shortage of specialist

pathologists to carry the preoperative diagnosis and post-operative assessments of breast cancer specimens ^[3–5]. For example, in the UK the Royal Colleague of Pathologist recently reported, that only 3% of the NHS histopathology departments are sufficiently staffed to meet clinical demand ^[6]. Furthermore, in the UK a quarter of trained pathologist are 55 or over which can exacerbate the current shortage once they retire ^[6]. New technologies have been slowly integrated to assist with pathological assessment and to achieve a better characterization and understanding of the different cancer cases. A highly promising technology for the analysis of breast sections is Fourier-transform infrared spectroscopy better known as FTIR spectroscopy. FTIR spectroscopy allows identifying the chemical bond and functional groups within the tissue, facilitating its chemical characterization ^[7]. By identifying the chemical changes at the tissue level, correlation with different diseases, including breast cancer, can be identified ^[8].

Breast cancer is a highly heterogeneous disease including different histological types, grades and molecular profiles ^[9]. In-depth characterization of breast tumors can facilitate translational knowledge to assist with breast cancer diagnosis and management. The chemical structure characterization of breast cancer cases can help with response prediction to some therapies ^[10,11].

Infrared spectroscopy is the one of the most widely used spectroscopy technique for chemical structural characterization of natural and synthetic materials. In the last 15 years, its potential to analyze biological molecules has been recognized. Infrared spectroscopy is a vibrational spectroscopy technique and is routinely employed by researchers when it comes to solving a problem involving having to discover the molecular structure, molecular behavior and/or the identification of unknown natural and synthetic materials. When IR beam hits the sample, molecules vibrate to

absorption and these molecular vibrations produce a dipole moment change in the molecules. This change is detected, the energy absorption is recorded and spectral data processed ^[12]. The energy absorption can only happen when the molecular vibrations and the frequencies of the IR radiation match. Due to the extensive details obtained in the spectra with vibrational spectroscopy, biochemical changes can be detected in tissues. Additionally, infrared spectroscopy allows the label-free analysis of mounted tissue, and with the coupling of microscopy and mechanical controlled stages, it can facilitate high-throughput analysis of samples ^[13]. The study of tissue with FTIR spectroscopy enables the biochemical characterization of the breast tissue meeting the clinical requirements for diagnosis in terms of objectivity, accuracy, and repeatability while preserving the tissue for further histopathological analysis. The combination with other technologies such as tissue microarrays (TMA) can reduce the amount of tissue needed for analysis while allowing a cost-effective examination ^[14]. Extensive research in breast oncology has been performed employing infrared spectroscopy. Ductal carcinoma in situ (DCIS) and invasive ductal carcinoma sections have been characterized, confirming differences in band position and intensity, which suggested biochemical and compositional changes among the cancerous and normal tissues. Higher grades had an overall increased intensity in comparison with lowgrade samples. This study also validated the technique efficacy to support diagnosis despite the sample's $age^{[15]}$.

The use of infrared spectroscopy facilitated the identification of changes in the protein structure, which have been suggested to offer an indication on the progression from healthy to cancerous epithelial tissue ^[16,17]. The differences in the content of different biomolecules (lipids, proteins, nucleic acids, etc) have been explored to differentiate

benign and cancerous lesions in a variety of tissues ^[18]. The chemical characterization could provide further information about cell biology processes such as cell cycle, differentiation, transcriptional nature, and apoptosis. For example, protein patterns according to breast pathology ^[19], cancer cell growth and cycle phases ^[20], cancerinduced modifications in collagen ^[21], global effects of miRNAs in MCF7 breast cancer cells ^[22], and fibroblast behavior under cancer stimuli ^[23] have been explored using this type of spectroscopy.

The progression of cancer encompasses a complex assortment of mechanisms where inflammation, immune response, hormones, angiogenesis, and structural changes play essential roles. These mechanisms will result in chemical changes at tissue level. The identification of these changes could confirm the presence and progression of BC, offering valuable predictive information and facilitating diagnosis. The use of infrared spectroscopy for the assessment of breast sections could provide a reliable and straightforward diagnosis method. This approach could work as a first line of action and prioritization in busy and understaffed histopathological departments. In this study TMA containing African breast cancer samples and normal breast sections were analyzed using Fourier Transform Infrared (FTIR) micro-spectroscopy. The obtained spectra were analyzed using principal component analysis (PCA) to identify spectral biomarkers associated with the presence of cancer, followed by linear discriminant analysis (LDA) for validation.

2. Materials and methods

2.1 Sample information and preparation

The study cohort comprised breast cancer samples and normal breast sections assembled in tissue microarrays (TMAs). The cancerous samples correspond to breast samples from Nigerian patients. The normal mammary tissue samples are derived from reduction mammoplasty specimens. The cohort in use was obtained under the Ethical approval REC No. 06/ Q1206/180. The TMA cores were 5µm thick and 0.6 mm in diameter and arranged in duplicates. Different human and bovine tissues were used to help with the TMA orientation. The TMAs were prepared as previously described by Titloye et al ^[24] in 2016.

Six TMAs composed of breast cancer samples (n =378) and one TMA containing normal breast from mammary reductions (n =134) were received in duplicates placed in CaF₂ and MirrIR (Kevley Technologies, USA) slides. Duplicate slides included continuous cuts from the same breast section. One slide of each set was stained by Haemotoxylin and Eosin. The stained slide was scanned and used for the identification of cancerous and normal epithelia and breast tissue components. The second slide was dewaxed using Mian's ^[25] modified dewaxing protocol. The modification in the protocol consisted on reducing the exposition time in xylene due to the reduced thickness of the breast samples to only 10 minutes followed by alcohol baths at different concentrations for 5 minutes.

2.2 Spectra acquisition

Spectroscopic analysis was performed on samples with an existing H&E stained core duplicate which presented either cancerous or healthy epithelial tissue. Twenty spectra were obtained from the epithelial tissue of each sample. The spectral collection was performed using an iN10 MX Nicolet Infrared Imaging Microscope using the reflection mode with a

gold plate background and a KBr beamsplitter. Spectra were acquired by accumulating 16 number of scans, at 8.0 cm⁻¹ resolution by using a liquid nitrogen-cooled MCT-A detector in the 4000-675 cm⁻¹ spectral range. All the spectral data was corrected using the Kramers-Kronig correction by OMNICTM PictaTM (Thermo Scientific, USA) software.

2.3 Pre-processing and analysis

Once the twenty spectra for each sample were collected, they were average and the spectral data analyzed using a multivariate analysis. The Unscrambler X 10.2[™] software (Camo software, Oslo, Norway) was used for statistical analysis. The spectral data was pre-processed using baseline corrections, Savitzky-Golay smoothing and standard normal variate (SNV). Unsupervised (PCA) and supervised multivariate methods (LDA) were used in data analysis. PCA focus on the following spectral regions: lipids 3500-2550 cm⁻¹, proteins 1700-1500 cm⁻¹, nucleic acids 1250-1000 cm⁻¹, and 1450-600 cm⁻¹. LDA was used for validation of the PCA model, leaving out five samples from each group at each pass until the total number of spectra of each type were predicted. The predictions were used to define correctly classified false positive, false negative, true negative and true positive and to calculate the sensitivity and specificity of the model.

3. Results

A total of 282 samples were available for analysis with FTIR micro-spectroscopy, 245 breast cancer tumors and 37 normal breast sections. The following information was collected on invasive breast carcinomas: histological tumor type, grade, ER and HER2 status, and patient sex and age at diagnosis (Table 1). The commonest histological subtype was Ductal NST (no special type) carcinoma accounting for 87.3% of the malignant cases. The mean patient's age at diagnosis was 48.62 ± 12.41 years, while the mean tumor size was $63.33 \text{ mm} \pm 1.24$.

The normalized average spectra of the cancerous epithelia and the normal epithelia presented bands characteristic of carbohydrates, lipids, nucleic acids and proteins as highlighted in Figure 1A. Overall, cancerous (CA) bands were shifted towards lower wavenumbers in comparison with the normal breast (NB) bands. Additionally, the breast cancer spectrum presented higher overall intensity, representing biochemical changes associated with the disease. The increased intensity in C-H bands served as an indicator of an increase of cellular components such as lipids, nucleic acids and proteins.

A clear shift and increased intensity in the 1700-1500 cm⁻¹ amide area was identified for the breast cancer samples. Furthermore, the band representing the asymmetric stretching of CH₂ of acyl chains in lipids was downshifted (2922 to 2918 cm⁻¹) on the cancerous area. Spectral peaks present at \approx 1334, \approx 1204, and \approx 1033 cm⁻¹, which are associated to the collagen moieties were considerably more noticeable in the normal breast samples.

The cancerous spectra presented a higher wave number shift from 1160 cm⁻¹ to 1171 cm⁻¹ reported as an increase of the cancerous cells in malignant tissue ^[26]. This band is representative of C–O stretching of the parenchymal cells phosphorylated proteins. However, this band has also been assigned to the asymmetric stretching of ester bonds in cholesteryl esters ^[22].

The ratio of the methyl stretching band to the methylene band (A_{2958}/A_{2853}) was smaller for the CA tissue in comparison with the NB. This reduction could suggest DNA hypomethylation and an overall malignant cellular content increase ^[26]. The higher content of collagen relative to the total content of protein in NB, (A_{1657}/A_{1204} and A_{1657}/A_{1278}), suggesting protein fibers reduction in CA tissue ^[26].

The deconvoluted IR spectra for the amide area shown in Figure 2 presents the downshift of the CA α -helix and β -sheet protein structure deconvoluted peaks (1659 and 1632 cm⁻¹) when compared with the NB bands (1666 and 1635 cm⁻¹). Likewise, the higher intensities of the β -sheet CA peaks were responsible for the broader amide I peak. A greater contribution of α -helix conformation in the NB it is suggested by the A₁₆₅₇/A₁₆₃₅ ratio when compared with the CA tissue. The deconvoluted FTIR spectrum showed a band located at 1593 cm⁻¹ which was only identified in the normal tissue. This band is associated with the vibrations of the C-C stretch of the phenyl ring, which can be attributed to collagen ^[7]. The lower contributions of collagen relative to protein in NB was confirmed by the A₁₆₅₇/A₁₂₀₄ and A₁₆₅₇/A₁₂₇₈ ratios. CA presented a shift towards a higher wavenumber (1160 cm⁻¹ vs 1171 cm⁻¹) of the C-OH stretching band caused by the increased non-hydrogen bonded C-OH groups in the CA tissue ^[16].

The cancerous A_{1640}/A_{1460} (Amide I/Lipids) and A_{1550}/A_{1460} (Amide II/Lipids) ratios were larger than the ones for NB. These ratios indicate a lower content of lipids relative to the protein contribution for the CA tissue, whereas the opposite trend was factual for the NB sections.

3.1 Multivariate analysis

The PCA model created by using the entire spectral range (4000-650 cm⁻¹) was validated using Linear discriminant analysis (LDA). The spectral differences were used to differentiate between malignant and normal samples. These differences were identified using the score plots and loadings for each principal component. False positives, false negative, true positives and true negatives were obtained and used to calculate the accuracy of the spectroscopic model. A sensitivity of 92% and specificity of 86% was confirmed.

3.1.1 Higher-wavenumber region $(3,500-2,550 \text{ cm}^{-1})$

Principal component analysis was used to identify changes and variations in the higherwavenumber, the amide and the fingerprint region. The higher-wave number region analysis shown in Figure 3, identified the bands at 3264, 3173, 2930, 2903, 2861, 2842 cm⁻¹ accountable for 82% of the variance and the grouping of CA and NB samples.

The spectral peak at \approx 3264 cm⁻¹ was attributed to the N–H stretching of amide A in proteins and nucleic acids ^[19]. CA areas seemed to present higher vibrational activity in comparison with NB reflecting greater cell density. N-H stretching bands with cis configuration have been assigned to the \approx 3173 cm⁻¹ band. NB had a lower contribution of the 3173 cm⁻¹ band. However, CA samples equally distributed along PC-4 indicating a dynamic behavior among the cancer samples in our cohort. The peaks positioned at \approx 2930 and \approx 2842 cm⁻¹ correspond to the antisymmetric and symmetric CH₂ stretching of hydrocarbon chains ^[7]. These assignments are characteristic of lipids with long chains such as ceramides, glycolipids and phospholipids ^[27].

The ≈ 2903 and ≈ 2861 cm⁻¹ bands represented the CH₃ symmetric stretch, and aliphatic C-H stretching in fatty acids and lipids and presented higher contributions in CA tissue confirming modification on the lipid metabolism.

3.1.2 Amide I and amide II region $(1,700-1,500 \text{ cm}^{-1})$

Good separation and grouping were achieved when analyzing the amide region, as presented in Figure 4. In other spectral regions, these peaks are the product of the convolution of different molecules' contributions. The peaks at 1668, 1646, 1612, 1568, 1552, and 1528 cm⁻ 1 were identified as responsible for the separation and grouping of both sample types. The $\approx 1668 \text{ cm}^{-1}$ band corresponded to the normal breast shifted peak assigned to α helix found in proteins. The normal breast spectra had more contributions towards this specific band. Distribution of CA samples across the negative PC-3 axis suggests a dynamic behavior concerning the conformational changes in proteins at different stages of carcinogenesis. CA samples showed higher intensities on the ≈ 1646 band, mainly associated with β structures and unordered random coils ^[19].

The $\approx 1612 \text{ cm}^{-1}$ band representative of adenine vibration in DNA ^[7] presented higher contributions on the CA samples. The band presented at $\approx 1552 \text{ cm}^{-1}$ is characteristic of the C-N stretching and the N-H bending of the amide II ^[28]. The normal breast samples presented higher contributions on this band, which can be attributed to collagen components in the stromal area. Inclusion of collagen in the spectroscopic analysis due to the low cell density (in comparison with CA) in the normal epithelium is suspected.

The band at $\approx 1528 \text{ cm}^{-1}$ (presented in the deconvoluted spectra) representative of the C=C and C=N specific for guanine and adenine nucleic bases stretching ^[29], showed higher contributions in half of the cancerous samples. This suggested elevated cell density, but also active proliferation characterized by increased cell components.

3.1.3 Fingerprint region $(1,450-600 \text{ cm}^{-1})$

Clear separation and grouping were achieved when the fingerprint was analyzed as it can be seen in Figure 5. The bands responsible for the grouping and separation of samples were present at 1427, 1385, 1366, 1332, 1236, 1194, 1051, and 882 cm⁻¹. The peak at \approx 1427 cm⁻¹ assigned to fatty acids CH₂ vibrations and lipids ^[7] presented higher intensities on the CA samples in comparison with NB contributions. Similarly, the stretching of C-O, deformation of C-H and N-H can be confirmed by the spectral peaks at ≈ 1385 and ≈ 1366 cm⁻¹ peaks ^[7] and are mainly associated with proteins and nucleic acids presented greater intensities in the CA samples, which is related to both, the increased cell contents and higher cell density ^[15]. Important differences in the ≈ 1332 cm⁻¹ and ≈ 1194 cm⁻¹peaks were presented on the NB samples, which showed greater intensities. These bands are considered collagenspecific peaks and their increased intensity is attributed to higher contents of collagen ^[30].

The band found at $\approx 1236 \text{ cm}^{-1}$ has been assigned to the phosphodiester asymmetric stretching in nucleic acids, but is also representative of amide III and CH₂ wagging vibrations of collagen ^[26]. Nonetheless, half of the cancer samples presented increased contributions towards that band. On the other hand, the $\approx 1236 \text{ cm}^{-1}$ band is highly representative of nucleic acids content and showed good inclusion of CA samples caused by the cancerous higher cell density and cell component hallmark of high-grades, malignancy and aggressive behavior ^[31]. This representation of increased nuclear acid content was confirmed with the $\approx 1051 \text{ cm}^{-1}$ greater contributions on the CA samples. This peak is assigned to the C-O-C stretching of DNA and RNA and the combination of the C-O stretching and C-O bending ^[7].

4. Discussion

In this study, FTIR spectroscopy has proven to be an outstanding technique for breast tissue analysis. The use of FTIR spectroscopy on ex vivo TMA breast samples provided promising results and elucidated important biochemical information.

The differentiation between normal breast and breast cancer was successfully achieved with a sensitivity of 92% and specificity of 86%. This model could help as a

cost-effective triage system, alleviating the workload and reducing the pressure in histopathological departments while resulting in a faster and effective first point of action for breast cancer diagnostics. In this study, DCIS and invasion were not studied separately, which could represent a limitation. The inclusion of both case types accounts for some of the heterogeneity presented in our cohort. Furthermore, studying atypical lesions could help to improve the predictive models as well as understand the chemical signature of these changes. More broadly, research is also needed to determine the biochemical response at the tissue level as a response to neoadjuvant therapy.

FTIR spectroscopy not only provides a biochemical profile of the breast tissue at a particular time, but it could also help to track the changes over time based on the chemical alterations a tissue level. The progression of cancer involves a wide variety of mechanisms where all the breast components are involved. For example, angiogenesis, inflammation, immune and endocrine response play an important role and will cause modifications at the chemical level. Identifying the chemical pathway to the progression of the disease could help the clinician to make decisions, select the treatment course and to give the patient a better chance against the disease. The FTIR results showed a lower content of lipids relative to the protein content for the cancer tissue. In contrast, the Amide I/Lipids and Amide II/Lipids ratios confirmed the opposite trend for the normal breast. Additionally, lipid transformations in breast cancer were identified by changes associated with these molecular components. Shifts of the CH₂ of acyl chains have been associated with an increment of trans- lipids conformations. A higher-order in lipids is usually represented by an important number of trans-lipids, which reduces the flexibility of the acyl chain. This

modifications in flexibility are important as they can affect the cell shape ^[32]. Consequently, the cell membrane fluidity can be modified, helping and supporting malignant activities as well as signaling pathways associated with carcinogenesis ^[22,32]. Furthermore, the cancerous peak increased intensity at 1171 cm⁻¹ could suggest changes on the cell membrane fluidity as well as higher cholesterol requirement reported on malignant tissue ^[22]. Besides, a higher contribution of acyl chains in the cancerous area suggested affectations on the cell membrane, which could affect the cell adhesion molecules, and the interactions between these molecules and other lipid components in the membrane.

Changes in the CH₂ stretching in long-chain lipids could suggest the presence of active biomolecules like ceramides. Ceramides play an essential role in carcinogenesis by acting as precursors for regulators of cell growth, proliferation and motility ^[33]. Moreover, the conformational changes in lipids will impact the membrane fluidity, promoting the signaling and stimuli to facilitate proliferation and cell adhesion ^[34–36].

Breast cancer is characterized by the stimulated production of novo lipids. Lipids involvement in the membrane synthesis saturation has a straight impact on cell growth, proliferation and oxidative stress resistance. Likewise, the triacylglycerides storage in lipid droplets has been suggested to work as fuel source after reoxygenation during intermittent hypoxia. Whereas, fatty acids promote redox balance supporting a high-glycolytic rate in CA tumors. In addition, lipids form the structural basis of paracrine hormones and growth factors which stimulate tumor growth, neovascularization, invasion, and metastatic spread ^[36]. Higher CA contributions towards lipid components identified in the higher wavenumber and fingerprint support the idea of active lipid metabolism and transformations.

The overall downshift of cancerous peaks has been reported by different research groups, agreeing with the presented results ^[15,19,26]. The shifting and higher intensity in the area associated with nucleic acids and amide stretchings (1700-1500 cm⁻¹) could suggest changes in the conformational structure of proteins in the cancerous tissue ^[15].

Changes in intensity on the representative of conformational proteins structures in addition to the spread of CA samples on both axes hints for a transition from α -helix into β -sheet structures as cancer progresses ^[19]. The varying contributions of the conformational peaks reflect changes in the number of hydrogen intermolecular bonds in the secondary structural conformation of proteins ^[19]. Our results suggest that β -sheet structural proteins are more present in cancer tissue and that structural changes form an essential part of the carcinogenesis process. Tumor microenvironment studies had reported changes from α -helix towards β -sheet conformations as they move closer to the tumor ^[21]. Furthermore, the identification of the 1593 cm⁻¹ deconvoluted band highlights the changes in the hydrogen intermolecular bonds proposed during the carcinogenesis. Besides, the collagen-rich areas could also contribute to this band. The phosphorylation of the cellular proteins could have been reflected by the increase of non-hydrogen bonded C-OH groups in the CA tissue. Larger phosphate groups replace the –OH groups leaving behind unordered structures as a result of this process, which is distinctive of the carcinogenesis process ^[16].

Higher contributions of protein and nucleic acid bands were identified in the cancerous samples in comparison with the normal breast samples on different spectral

regions. DNA methylation has been suggested as responsible for the reduction in the methylation ratio from normal to malignant tissue in the FTIR analysis. Higher intensities presented by CA samples on peaks representing nucleic acid content could be translated as a higher content of those biomolecules, which is in agreement with the increased cellular content expected in abnormal proliferation. Furthermore, a higher cell density is characteristic of higher grades and can explain the spreading of the cancerous samples in the different regions.

The FTIR spectral profile of the breast cancer samples suggested a lower content of collagen relative to protein content. This technique indicated that the normal breast samples were dominated by the presence of collagen, while a higher content of proteins constituting the malignant epithelia was present in the cancerous samples. Collagen-specific vibrations presented higher intensities in the normal breast samples. reinforcing the idea of the stromal spectral acquisition when the normal breast was analyzed. As mention before, the inclusion of collagen into the NB analysis occurred due to low cell density in epithelia, and therefore NB samples are expected to contribute towards collagen related bands such as this amide III peak arising from the C–N stretching and N–H deformation ^[37].

5. Conclusions

A combination of supervised and unsupervised algorithms in the form of principal component and linear discriminant analysis in combination with FTIR spectroscopy proved to be a powerful tool to support the diagnosis of breast cancer on ex vivo samples. Despite the wide heterogeneity presented in the studied cohort, FTIR spectroscopy allowed for the identification of spectral biomarkers on the malignant breast tissue. The biochemical differences in lipids, amides, nucleic acids and collagen were responsible for the

classification of cancerous and normal breast. Structural protein changes identified with this technique proved to be important differentiators for malignant and normal breast. The use of TMAs, along with FTIR microspectroscopy allows for a high throughput analysis of biological tissue, resulting in the rapid and accurate spectral analysis and classification of breast samples. The integration of these approaches on the triage of patients suspected with breast cancer could result in a fast, cost-effective diagnosis method.

6. Acknowledgements

DLP received the financial support from the Mexican Consejo Nacional de Ciencia y Tecnología (CONACYT; CVU 610378), the Secretaría de Educación Pública (Beca complemento 2017-2018 y 2018-2019) and the Mexican Government for the completion of her studies. AB is supported by the Birmingham CR-UK Centre.

7. References

[1] Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2020. CA. Cancer J. Clin. 2020,

70(1), 7-30. DOI: 10.3322/caac.21590.

[2] Caplan, L. Delay in breast cancer: implications for stage at diagnosis and survival. *Front. public Heal.* **2014**, *2*, 87. DOI: 10.3389/fpubh.2014.00087.

[3] Egevad, L.; Delahunt, B.; Samaratunga, H.; Leite, K.R.M.; Efremov, G.; Furusato, B.;

Han, M.; Jufe, L.; Tsuzuki, T.; Wang, Z. The International Society of Urological Pathology

Education web—a web-based system for training and testing of pathologists. Virchows Arch.

2019, 474(5), 577–584. DOI: 10.1007/s00428-019-02540-w.

[4] Adesina, A.; Chumba, D.; Nelson, A.M.; Orem, J.; Roberts, D.J.; Wabinga, H.; Wilson, M.; Rebbeck, T.R. Improvement of pathology in sub-Saharan Africa. *The Lancet Oncology*. **2013**, *14*(4), e152-e157. DOI: 10.1016/S1470-2045(12)70598-3.

[5] Metter, D.M.; Colgan, T.J.; Leung, S.T.; Timmons, C.F.; Park, J.Y. Trends in the US and Canadian pathologist workforces from 2007 to 2017. *JAMA Netw. open* **2019**, *2*(5), e194337–e194337. DOI: 10.1001/jamanetworkopen.2019.4337.

[6] The Royal College of Pathologists. College report finds UK wide histopathology staff shortages. https://www.rcpath.org/discover-pathology/news/college-report-finds-severe-staff-shortages-across-services-vital-to-cancer-diagnosis.html (accessed Apr. 2020).

[7] Movasaghi, Z.; Rehman, S.; Rehman, I.U. Fourier transform infrared (FTIR) spectroscopy of biological tissues. *Appl. Spectrosc. Rev.* **2008**, *43*(2), 134–179. DOI:

10.1080/05704920701829043.

[8] Kendall, C.; Isabelle, M.; Bazant-Hegemark, F.; Hutchings, J.; Orr, L.; Babrah, J.; Baker,

R.; Stone, N. Vibrational spectroscopy: a clinical tool for cancer diagnostics. Analyst 2009,

134(6), 1029–1045. DOI: 10.1039/B822130H.

[9] Polyak, K. Heterogeneity in breast cancer. J. Clin. Invest. 2011, 121(10), 3786. DOI: 10.1172/JCI60534.

[10] Brenton, J.D.; Carey, L.A.; Ahmed, A.A.; Caldas, C. Molecular classification and molecular forecasting of breast cancer: Ready for clinical application? *J. Clin. Oncol.* 2005, 23(29), 7350–7360. DOI: 10.1200/JCO.2005.03.3845.

[11] Pusztai, L.; Mazouni, C.; Anderson, K.; Wu, Y.; Symmans, W.F. Molecular classification of breast cancer: Limitations and potential. *Oncologist* 2006, *11*(8), 868–877.
DOI: 10.1634/theoncologist.11-8-868.

[12] Larkin, P.J. Infrared and Raman Spectroscopy: Principles and Spectral Interpretation.Infrared Raman Spectrosc. *Princ. Spectr. Interpret.* 2011, 1–228.

[13] Rehman, I. ur. *Vibrational spectroscopy for tissue analysis*. Boca Raton : CRC Press:Boca Raton, 2013.

[14] Camp, R.L.; Charette, L.A.; Rimm, D.L. Validation of tissue microarray technology in breast carcinoma. *Lab. Investig.* 2000, *80*(12), 1943–1949. DOI: 10.1038/labinvest.3780204.
[15] Rehman, S.; Movasaghi, Z.; Darr, J.A.; Rehman, I.U. Fourier transform infrared spectroscopic analysis of breast cancer tissues; identifying differences between normal breast, invasive ductal carcinoma, and ductal carcinoma in situ of the breast. *Appl. Spectrosc. Rev.* 2010, *45*(5), 355–368. DOI: 10.1080/05704928.2010.483674.

[16] Ci, Y.; Gao, T.; Dong, J.; Kan, X.; Guo, Z. FTIR assessment of the secondary structure of proteins in human breast benign and malignant tissues. *Chinese Sci. Bull.* 1999, 44(24), 2215–2221. DOI: 10.1007/BF02885924.

[17] Elshemey, W.M.; Ismail, A.M.; Elbialy, N.S. Molecular-Level Characterization of Normal, Benign, and Malignant Breast Tissues Using FTIR Spectroscopy. *J. Med. Biol. Eng.* **2016**, *36*(3), 369–378. DOI: 10.1007/s40846-016-0133-0.

[18] Zhou, S.; Xu, Z.; Ling, X.-F.; Li, Q.-B.; Xu, Y.-Z.; Zhang, L.; Zhao, H.-M.; Wang, L.-

X.; Hou, K.-Y.; Zhou, X.-S.; Wu, J.-G. FTIR spectroscopic characterization of freshly removed breast cancer tissues. *Zhonghua Zhong Liu Za Zhi* **2006**, *28*(7), 512–4.

[19] Eckel, R.; Huo, H.; Guan, H.-W.; Hu, X.; Che, X.; Huang, W.-D. Characteristic infrared spectroscopic patterns in the protein bands of human breast cancer tissue. *Vib. Spectrosc.*

2001, *27*(2), 165–173. DOI: 10.1016/S0924-2031(01)00134-5.

[20] Yang, W.; Xiao, X.; Tan, J.; Cai, Q. In situ evaluation of breast cancer cell growth with 3D ATR-FTIR spectroscopy. *Vib. Spectrosc.* **2009**, *49*(1), 64–67. DOI:

10.1016/j.vibspec.2008.04.016.

[21] Kumar, S.; Desmedt, C.; Larsimont, D.; Sotiriou, C.; Goormaghtigh, E. Change in the microenvironment of breast cancer studied by FTIR imaging. *Analyst* 2013, *138*(14), 4058–4065. DOI: 10.1039/C3AN00241A.

[22] Ozek, N.S.; Tuna, S.; Erson-Bensan, A.E.; Severcan, F. Characterization of microRNA125b expression in MCF7 breast cancer cells by ATR-FTIR spectroscopy. *Analyst* 2010, *135*(12), 3094–3102. DOI: 10.1039/C0AN00543F.

[23] Kumar, S.; Goormaghtigh, E.; Shabi, T.S. A FTIR imaging characterization of fibroblasts stimulated by various breast cancer cell lines. *PLoS One* **2014**, *9*(11). DOI:

10.1371/journal.pone.0111137.

[24] Titloye, N.A.; Foster, A.; Omoniyi-Esan, G.O.; Komolafe, A.O.; Daramola, A.O.;

Adeoye, O.A.; Adisa, A.O.; Manoharan, A.; Pathak, D.; D''Cruz, M.N. Histological Features

and Tissue Microarray Taxonomy of Nigerian Breast Cancer Reveal Predominance of the

High-Grade Triple-Negative Phenotype. *Pathobiology* **2016**, *83*(1), 24–32. DOI:

10.1159/000441949.

[25] Mian, S.A.; Colley, H.E.; Thornhill, M.H.; Rehman, I. u. Development of a dewaxing protocol for tissue-engineered models of the oral mucosa used for Raman spectroscopic analysis. *Appl. Spectrosc. Rev.* 2014, *49*(8), 614–617. DOI: 10.1080/05704928.2014.882348
[26] Gao, T.; Feng, J.; Ci, Y. Human breast carcinomal tissues display distinctive FTIR spectra: implication for the histological characterization of carcinomas. *Anal. Cell. Pathol.* 1999, 18(2), 87-93.

[27] Garidel, P. Mid-FTIR-Microspectroscopy of stratum corneum single cells and stratum corneum tissue. *Phys. Chem. Chem. Phys.* **2002**, *4*(22), 5671–5677. DOI:

10.1039/B207478H.

[28] Chadefaux, C.; Hô, A.-S. Le; Bellot-Gurlet, L.; Reiche, I. Curve-fitting micro-ATR-FTIR studies of the amide I and II bands of type I collagen in archaeological bone materials. *E-Preservation Sci.* 2009, 6, 129–137. HAL: hal-01676897

[29] Talari, A.C.S.; Evans, C.A.; Holen, I.; Coleman, R.E.; Rehman, I.U. Raman

spectroscopic analysis differentiates between breast cancer cell lines. J. *Raman Spectrosc.* **2015**, *46*(5), 421-427. DOI: 10.1002/jrs.4676.

[30] Camacho, N.P.; West, P.; Torzilli, P.A.; Mendelsohn, R. FTIR microscopic imaging of collagen and proteoglycan in bovine cartilage. *Biopolymers* 2001, 62(1), 1–8. DOI: 10.1002/1097-0282(2001)62:1<1::AID-BIP10>3.0.CO;2-O

[31] Dovbeshko, G.I.; Chegel, V.I.; Gridina, N.Y.; Repnytska, O.P.; Shirshov, Y.M.;

Tryndiak, V.P.; Todor, I.M.; Solyanik, G.I. Surface enhanced IR absorption of nucleic acids

from tumor cells: FTIR reflectance study. *Biopolymers* 2002, 67(6), 470–486. DOI:

10.1002/bip.10165

[32] Spector, A.A.; Yorek, M.A. Membrane lipid composition and cellular function. *J. Lipid Res.* **1985**, *26*(9), 1015–35.

[33] Derenne, A.; Vandersleyen, O.; Goormaghtigh, E. Lipid quantification method using

FTIR spectroscopy applied on cancer cell extracts. Biochim. Biophys. Acta - Mol. Cell Biol.

Lipids **2014**, *1841*(8), 1200–1209. DOI: 10.1016/j.bbalip.2013.10.010.

[34] Clarke, R.; Berg, H.W. van den; Murphy, R.F. Reduction of the membrane fluidity of human breast cancer cells by tamoxifen and 17β-estradiol. JNCI J. Natl. *Cancer Inst.* **1990**, 82(21), 1702–1705. DOI: 10.1093/jnci/82.21.1702.

[35] Hilvo, M.; Denkert, C.; Lehtinen, L.; Muller, B.; Brockmoller, S.; Seppanen-Laakso, T.;

Budczies, J.; Bucher, E.; Yetukuri, L.; Castillo, S.; Berg, E.; Nygren, H.; Sysi-Aho, M.;

Griffin, J.L.; Fiehn, O.; Loibl, S.; Richter-Ehrenstein, C.; Radke, C.; Hyotylainen, T.;

Kallioniemi, O.; Iljin, K.; Oresic, M. Novel Theranostic Opportunities Offered by

Characterization of Altered Membrane Lipid Metabolism in Breast Cancer Progression.

Cancer Res. 2011, 71(9), 3236–3245. DOI: 10.1158/0008-5472.CAN-10-3894.

[36] Santos, C.R.; Schulze, A. Lipid metabolism in cancer. FEBS J. 2012, 279(15), 2610-

2623. DOI: 10.1111/j.1742-4658.2012.08644.x.

[37] Chang, M.C.; Tanaka, J. FT-IR study for hydroxyapatite/collagen nanocomposite cross-

linked by glutaraldehyde. Biomaterials 2002, 23(24), 4811–4818. DOI: 10.1016/S0142-

9612(02)00232-6.

Table 1. Clinicopathological features of breast samples. Values expressed as n (%), unless otherwise indicated.

Cancerous samples	245 (86.88%)
Normal breast	37 (13.12%)
Histological grade	205
Ι	22 (10.73%)
Π	69 (33.66%)
III	114 (55.61%)
Molecular subtype	179
Luminal	64 (35.75%)
HER2+	15 (8.38%)
TNBC	100 (55.86%)

Figure 1. FTIR characterisation of breast cancer and normal breast. (A) General region (4,000-675 cm⁻¹), (B) High wavenumber region (3,500- 2,550 cm⁻¹), (C) Amide region (1,700-1,500 cm⁻¹), and (D) Fingerprint region (1,450-600 cm⁻¹).



Figure 2. Stacked deconvoluted peaks of FTIR spectra in the amide area.



Figure 3. Principal component analysis of the higher wavenumber region (3,500 -2,550 cm⁻¹). (A) Score plot using PC₋₁ and PC-4 accounting for 82% of the variance, (B) Average spectral profile of the higher wavenumber region for breast cancer and normal breast. Dotted lines represent the peaks presented on the loadings in C&D, (C) PC₋₁ Loading, (D) PC-4 loading.



Figure 4. Principal component analysis of the amides region (1,700-1,500 cm⁻¹). (A) Score plot using PC-2 and PC-3 accounting for 17% of the variance, (B) Average spectral profile of the amides region for breast cancer and normal breast. Dotted lines represent the peaks presented on the loadings in C&D, (C) PC-3 Loading, (D) PC-2 loading.



Figure 5. Principal component analysis of the fingerprint region (1,450-600 cm⁻¹). (A) Score plot using PC-4 and PC-5 accounting for <2% of the variance, (B) Average spectral profile of the amides region for breast cancer and normal breast. Dotted lines represent the peaks presented on the loadings in C&D, (C) PC-4 Loading, (D) PC-5 loading

