

Original article:

**ELUCIDATING THE CHEMICAL AND STRUCTURAL  
COMPOSITION OF BREAST CANCER USING  
RAMAN MICRO-SPECTROSCOPY**

Daniela Lazaro-Pacheco<sup>1</sup>, Abeer M. Shaaban<sup>2</sup>, Nicholas Akinwale Titiloye<sup>3</sup>,  
Shazza Rehman<sup>4</sup>, Ihtesham ur Rehman<sup>5\*</sup>

<sup>1</sup> Department of Engineering, University of Exeter, Exeter, UK

<sup>2</sup> Department of Cellular Pathology, Queen Elizabeth Hospital Birmingham, and University of Birmingham, Birmingham, UK

<sup>3</sup> Department of Pathology, School of Medicine and Dentistry, Kwame Nkrumah University of Science & Technology, Kumasi, Ghana

<sup>4</sup> Department of Medical Oncology, Airedale NHS Foundation Trust, Airedale General Hospital, Steeton, West Yorkshire, UK

<sup>5</sup> Engineering Department, Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YW, U.K.

\* **Corresponding author:** Ihtesham ur Rehman, Engineering Department, Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YW, U.K.;  
Tel. +44 (0) 1524 594038; E-mail: [i.u.rehman@lancaster.ac.uk](mailto:i.u.rehman@lancaster.ac.uk)

<http://dx.doi.org/10.17179/excli2021-3962>

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>).

**ABSTRACT**

The current gold standard for breast cancer (BC) diagnosis is the histopathological assessment of biopsy samples. However, this approach limits the understanding of the disease in terms of biochemical changes. Raman spectroscopy has demonstrated its potential to provide diagnostic information and facilitate the prediction of the biochemical progression for different diseases in a rapid non-destructive manner. Raman micro-spectroscopy was used to characterize and differentiate breast cancer and normal breast samples. In this study, tissue microarrays of breast cancer biopsy samples (n=499) and normal breast (n=79) were analyzed using Raman micro-spectroscopy, and principal component analysis (PCA) was used for feature extraction. Linear discriminant analysis (LDA) was used for feature validation. Normal breast and breast cancer were successfully differentiated with a sensitivity of 90 % and specificity of 78 %. Dominance of lipids, specifically fatty acids, was identified in the normal tissue whereas proteins dominated the malignant spectra. Higher intensities of carotenoids,  $\beta$ -carotenoids, and cholesterol were identified in the normal breast while ceramide related peaks were mostly visible in the BC spectra. The biochemical characterization achieved with Raman micro-spectroscopy showed that this technique is a powerful and reliable tool for the monitoring and diagnosis of BC, regardless of the cohort heterogeneity. Raman spectroscopy also provided a powerful insight into the biochemical changes associated with the BC progression and evolution.

**Keywords:** Breast cancer, Raman spectroscopy, multivariate analysis

**INTRODUCTION**

Breast cancer (BC) is the second most frequent cancer worldwide and the commonest

cancer in women, according to the Global Cancer Statistics 2018 report (Bray et al., 2018). With 630,000 estimated deaths in

2018, breast carcinomas had the first mortality rank among all types of cancers in females. The number of cases identified in 2040 is predicted to increase by a million, while mortality has been estimated to go up by 400,000 in comparison with the cases foreseen in 2020, according to the World Health Organization statistics (IARC, 2020). Along with the increase in BC incidence, the shortage of trained pathologists has been estimated to worsen in the coming years. In the UK, a quarter of trained pathologists is less than ten years from retirement age. Moreover, only 3 % of the UK histopathological departments do currently have enough staff to meet clinical demand (The Royal College of Pathologists, 2018).

An up-and-coming technology for the identification of BC and the assessment of cancerous sections is Raman spectroscopy. This vibrational spectroscopic technique focuses on the inelastic scattering of photons produced at relaxation after being excited with near-IR, visible or near-UV energy (Talari et al., 2015a). Raman spectroscopy allows identifying the chemical bond and functional groups within the tissue, facilitating its chemical characterization (Rehman, 2013). The identification of chemical changes at the tissue level can be correlated with the presence of disease even before morphological changes take place (Brozek-Pluska et al., 2018). Furthermore, Raman spectroscopy allows the label-free analysis of mounted tissue, and with the coupling of confocal microscopy and mechanical controlled stages, it can facilitate high-throughput analysis of samples (Lazaro-Pacheco et al., 2020a). The combination of other sampling techniques such as tissue microarrays (TMA) can reduce the amount of tissue needed for assessment while allowing a cost-effective, high throughput analysis (Camp et al., 2000). The combination of both technologies, Raman spectroscopy and TMAs, could not only help to alleviate the pressure in histopathological department assisting with the identification of BC but also help to understand the underlying biochemis-

try associated with the progression and evolution of the disease as well as early detection (Keller et al., 2006).

The use of Raman spectroscopy has allowed the identification of biological components within a variety of tissue in addition to their quantities, quality and structure (Talari et al., 2019). The classification of different breast cancer and normal breast cell lines representative of different receptor status has been successfully classified with a 100 % sensitivity and 91 % specificity (Talari et al., 2015b). The classifications achieved when combining Raman spectroscopy and multivariate analysis have enabled differentiating between normal and malignant tissue using basic differences in the concentration of biochemical compounds within the cells. Validating the technique in normal mammary tissue and patient tumor samples would therefore be an important step towards understanding the chemical changes associated with mammary tumorigenesis.

In this study, African breast cancer samples assembled in TMAs were analyzed using Raman micro-spectroscopy. Principal component analysis (PCA) was used to identify the biochemical changes and spectral biomarkers associated with the presence of cancer, and linear discriminant analysis (LDA) was used to validate the findings.

## MATERIALS AND METHODS

### *Sample information and preparation*

The study investigated cancer samples and normal breast tissue arranged in seven different TMAs. The cancerous samples corresponded to breast samples from Nigerian patients, and the normal breast tissue was obtained from surgical breast reductions. The cohort in use was arranged into TMAs under the Ethical approval REC No. 06/Q1206/180. The tissue cores were 5  $\mu\text{m}$  thick and 0.6 mm in diameter. To help with the orientation, different human and bovine tissues were added to the TMAs. The TMAs were prepared as reported by Pinder et al., (2013) and Titiloye et al. (2016).

Six TMAs containing breast cancer samples (n=378) and one TMA composed of normal breast (n=134) were received in duplicates. Continuous cuts of the same breast section were placed on each duplicate slide. From each duplicate set, one slide was used for Haematoxylin and Eosin (H&E) staining and the other for spectroscopic analysis after dewaxing. H&E staining and assessment was done to confirm the areas of interest such as cancerous and normal epithelial tissue. Mian's dewaxing protocol (Mian et al., 2014) was modified by reducing the time in xylene to 10 minutes, considering the sample thickness.

### ***Spectra acquisition***

Some loss of TMA cores and/or folding of sections, a well recognized issue associated with TMA sectioning and preparation, was observed. The eligibility criteria for Raman analysis required the availability of both corresponding H&E and its dewaxed tissue core and in the cancerous TMA's, identifiable cancerous areas by H&E staining.

Twenty Raman spectra were collected from the epithelial tissue of each sample, either cancerous or healthy. The Raman spectral data was collected using a Raman microscope DXR™ (Thermo Scientific, Waltham, MA, USA) using a 50X long work distance objective, a 532 nm diode laser with a 10 mW laser power and a 50 μm pinhole within 3500-400 cm<sup>-1</sup> spectral range.

### ***Pre-processing and analysis***

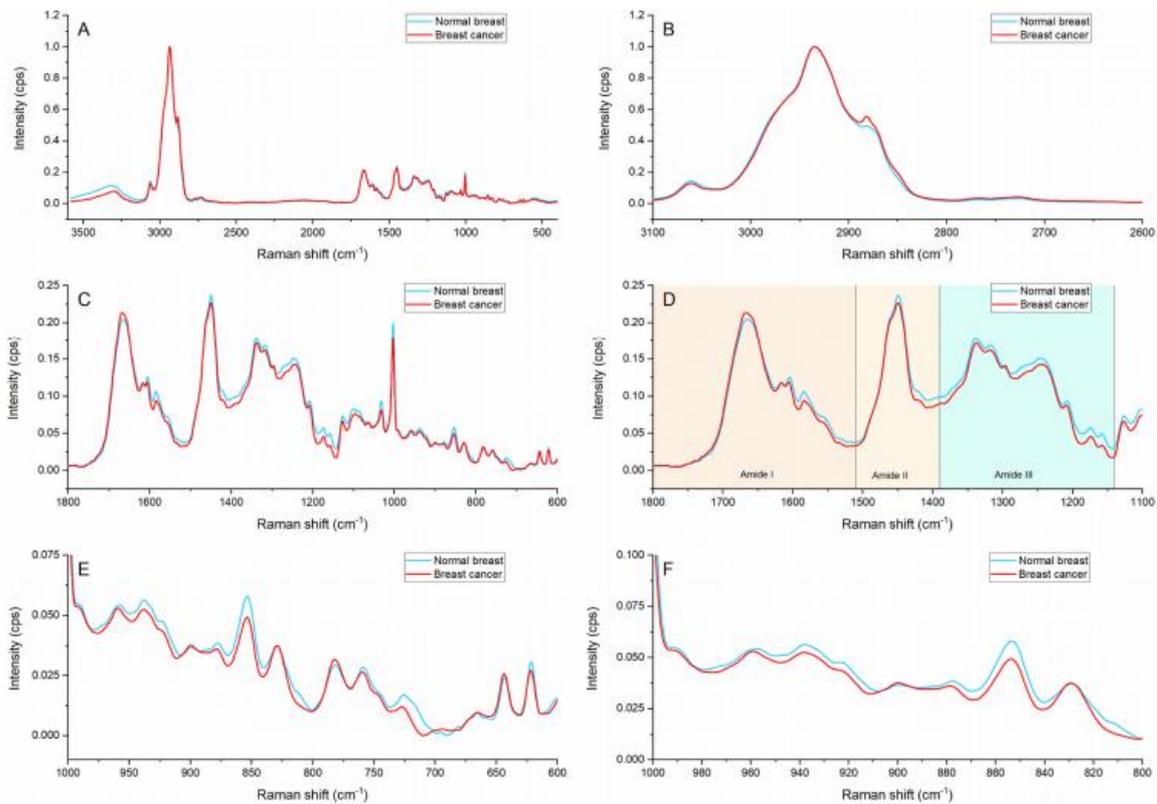
Fluorescence can be present in biological samples, therefore, all the spectral data was fluorescence corrected using a six polynomial algorithm, by employing Omnic™ (Thermo Scientific, Waltham, MA, USA) software. Twenty spectra per sample were averaged and used for multivariate analysis. The Unscrambler X 10.2™ software (Camo software, Oslo, Norway) was used for statistical analysis and pre-processing. The spectral data was pre-processed using baseline, Savitzky-Golay smoothing and standard normal variate

(SNV) corrections. Principal component analysis (PCA) was performed on the following spectral regions: fingerprint 1800-500 cm<sup>-1</sup>, amides 1800-1140 cm<sup>-1</sup>, amino acids and nucleic acids 980-600 cm<sup>-1</sup>, lipids 3100-2680 cm<sup>-1</sup>, and hydroxyproline /proline 960-810 cm<sup>-1</sup>. PCA was followed by linear discriminant analysis (LDA) for validation. LDA was performed leaving out five samples from each group at each pass until the total number of spectra of each type were predicted. The sensitivity and specificity of the model were calculated using the correctly classified false positive, false negative and true negative.

## **RESULTS AND DISCUSSION**

Based on the inclusion criteria, a total of 578 samples were analyzed with Raman microspectroscopy comprising 499 breast cancer tumors and 79 normal breast sections. The cancerous TMAs were received with the following anonymised information: histological tumor type, tumor grade, estrogen receptor (ER), progesterone receptor (PR) and HER2 status, and patient sex and age at diagnosis. Based on the molecular profile, the invasive carcinomas were classified into luminal, her2 positive or triple negative (TNBC). The clinicopathological features of the cancerous and normal cohort are presented in Table 1. Ductal NST (no-special-type) carcinoma was the dominant histological subtype accounting for 87.3 % of the cancerous cases. Ductal carcinoma in situ (DCIS) was also included in the analysis.

The average spectra of the cancerous epithelia and the normal epithelia within the breast samples are presented in Figure 1. Breast cancer and normal breast spectral profiles showed subtle differences in intensity and peak position. The spectral profile of breast cancer showed higher intensity on the lipid area (Figure 1B) and the peaks located at ≈1667 and ≈1295 cm<sup>-1</sup>. Noticeable changes in peak ratios and peak shapes can be observed in the amide region (Figure 1D). Furthermore, subtle peak shifts are present on the amino and amino acid region (Figure 1E).



**Figure 1:** Raman spectra of cancerous breast and normal breast. (A) General region (3600-400  $\text{cm}^{-1}$ ), (B) Lipid region (3,100-2,680  $\text{cm}^{-1}$ ), (C) Fingerprint region (1,800-500  $\text{cm}^{-1}$ ), (D) Amides region (1,800-1,140  $\text{cm}^{-1}$ ) indicating: Amide I (1,800-1,510  $\text{cm}^{-1}$ ), Amide II (1,510-1,390  $\text{cm}^{-1}$ ) and Amide III (1,390-1,140  $\text{cm}^{-1}$ ) regions. (E) Amino acid and nucleic acids regions (980-600  $\text{cm}^{-1}$ ), and (F) Hydroxyproline and proline region (810-960  $\text{cm}^{-1}$ ).

**Table 1:** Clinicopathological features of cancerous and normal tissue cohort. Values expressed as n (%), unless otherwise indicated

Mean age, years	48.62 (SD 12.41)
Mean tumor size	63.33 mm (SD 1.24)
Total samples	578
Cancerous samples	499 (86.33 %)
Normal breast	79 (13.66 %)
Histological grade	395
I	31 (7.84 %)
II	153 (38.73 %)
III	211(53.41 %)
Molecular subtype	395
Luminal	121 (30.63 %)
HER2+	21 (5.31 %)
TNBC	253 (64.05 %)

The peak located at  $\approx 1295 \text{ cm}^{-1}$  is only visible in the breast cancer spectral profile. This peak is representative of ceramides which form part of the structural components of sphingolipids and are considered vital biomolecules in the carcinogenesis process. Sphingolipids are membrane lipids and work as signaling molecules that are involved in cell growth and differentiation, apoptosis, senescence and proliferation (Hartmann et al., 2012). Sphingolipids regulate various cellular processes linked to cancer development, progression, metastasis and therapy resistance, identification. Ceramide levels are statistically significantly increased (12-fold higher) in breast cancer tissue in comparison with benign tumors and normal breast (Schiffmann et al., 2009). The identification of ceramides via Raman spectroscopy at tissue can therefore be

suggested as a robust biomarker for breast cancer.

The  $I_{1667}/I_{1449}$  ratio compares the protein to lipid content being represented by the C=C stretch in amide I and the CH<sub>2</sub> scissoring in lipids. The calculated ratios showed that the normal breast had a lower protein to lipid proportion in comparison with the breast cancer tissue (0.86 vs 0.93).

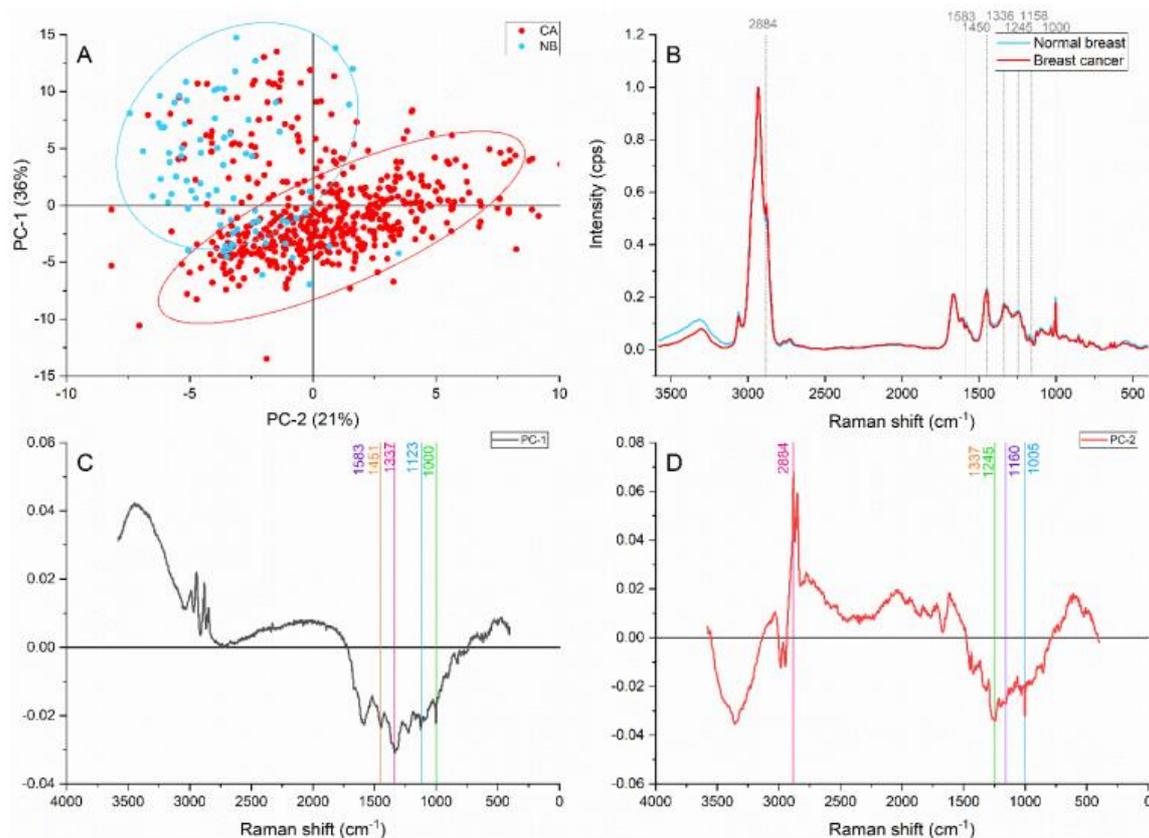
### Multivariate analysis

Comparison of breast cancer and normal breast tissue in different spectral regions were performed using PCA. Different scores and loadings are produced as output of this analysis. Both are complementary to each other to interpret the data. The scores show the dispersion of the samples based on the variance

along the analyzed area. The loadings are a representation of the contribution of the Raman bands. In the loading for a particular principal component (PC), it can be assumed that the higher the intensity, the higher the contribution of that specific peak for the classification. The PCA model created showed a sensitivity of 90 % and a specificity of 78 %.

### General region (4000-400 cm<sup>-1</sup>)

When PCA was applied to the general spectral region, a clear separation and grouping were achieved with the PC-1 and PC-2 accounting for 57 % of the variance (Figure 2). The peaks located at 2884, 1583, 1450, 1336, 1245, 1159, 1122, and 1004 cm<sup>-1</sup> were identified as responsible for the grouping and separation.



**Figure 2:** Principal component analysis on the general region (4,000-400 cm<sup>-1</sup>). (A) Score plot using PC-1 and PC-2 accounting for 57 % of the variance-ellipses were drawn subjectively based on visual trends, (B) Average Raman spectral profile for breast cancer and normal breast. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-1 loading, (D) PC-2 loading

The peak located at  $\approx 2884\text{ cm}^{-1}$  attributed to the  $\text{CH}_2$  and  $\text{CH}$  stretching of lipids and proteins (Talari et al., 2015a) contributed more in the cancer area than in the normal breast. Tryptophan and the  $\text{C}=\text{C}$  bending mode of phenylalanine (Lau et al., 2003) identified by the  $\approx 1583\text{ cm}^{-1}$  band also showed greater intensities on the cancer samples. The content differences represented by these two peaks might be caused by the accelerated and active cell component production characteristic of cancer development and progression as well as the structural components in highly dense samples. Furthermore, the cancerous band located at  $\approx 1450\text{ cm}^{-1}$  presented higher intensities than the normal counterpart. This band has been reported as a  $\text{CH}_2$  bending characteristic of malignant tissues (Talari et al., 2015a).

The  $\approx 1336\text{ cm}^{-1}$  peak has been assigned to different components, such as the amide III, and polynucleotide chains, and  $\text{CH}_2$  wagging vibrations from glycine backbone and proline side chain in collagen (Lazaro-Pacheco et al., 2020a). Breast cancer samples presented higher intensities in this and which suggests higher content of proteins, nucleic acids, and collagen in the malignant samples. The amide III contribution in breast cancer was confirmed by the  $\approx 1245\text{ cm}^{-1}$  peak, which was identified by the PC-2.

The normal breast samples presented higher intensities than the average on the  $1159\text{ cm}^{-1}$  peak associated with the  $\text{C}=\text{C}$  stretch mode characteristic of carotenoids and  $\beta$ -carotenoids. Besides, the band located at  $1122\text{ cm}^{-1}$  representing  $\text{C}-\text{C}$  stretch on carotene and breast lipids also presented a similar behavior. Carotenoids offer oxidative damage protection, upregulate connexion junctional communication and prevent the lipid oxidation. Moreover, normal breast tissue is known to possess a higher quantity of carotenoids than cancerous tissue, acting as a fatty acids reservoir (Abramczyk et al., 2012). The decreased intensity of carotenoids contributions suggested a lower content of these components in the malignant tissue. Brożek-Płuska

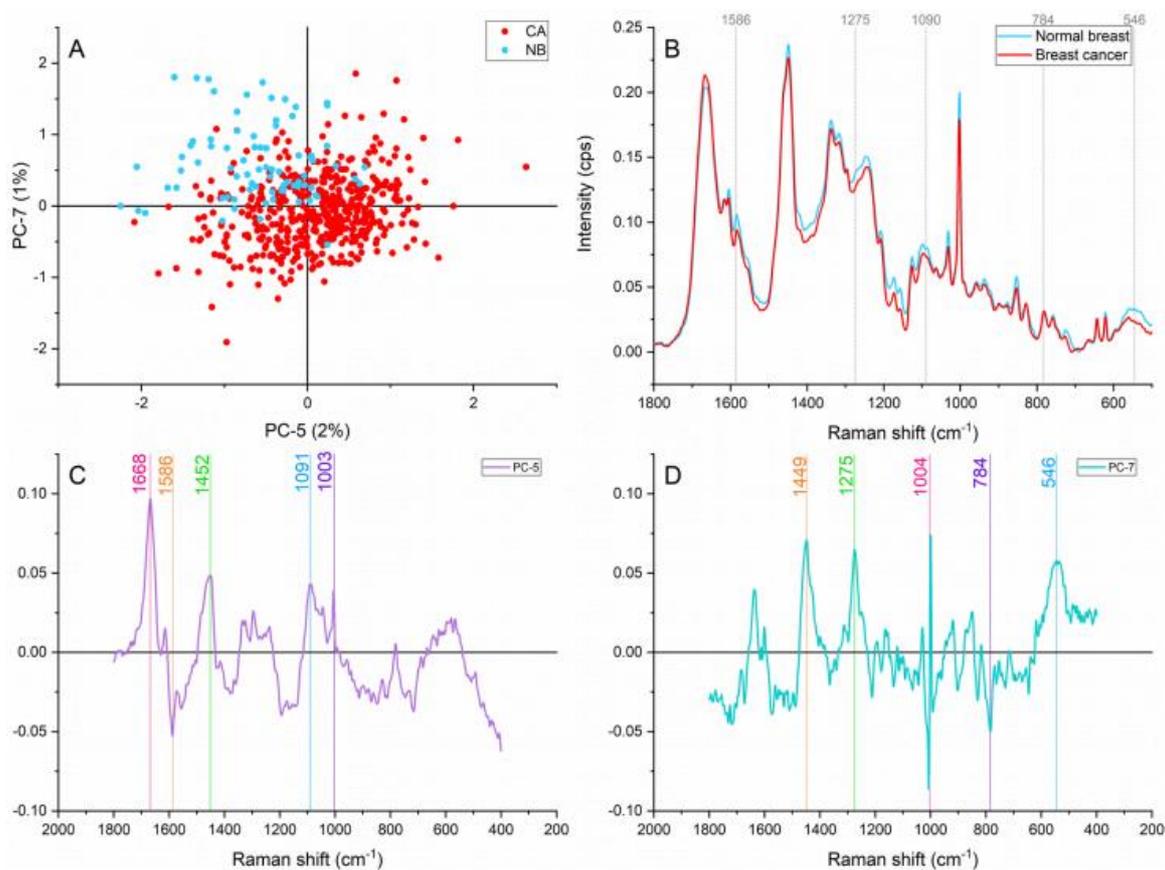
et al. have suggested that the oxidation of aging-related pigments and the peroxidation of lipids are responsible for the decline of carotenoids in cancerous breast (Brożek-Płuska et al., 2008).

Furthermore, the phenylalanine peak found at  $1000\text{-}1004\text{ cm}^{-1}$  (Talari et al., 2015a) presented greater contributions in cancerous samples. This peak could be associated with stromal components and with structural features of the epithelial cells. Collagen fibre deposition, rearrangement of fibres and overproduction occurring when tumors develop and grow have been reported (Conklin et al., 2011), and it could be expected in our samples. However, as the epithelial tissue was targeted on this study, the assignment is representative of higher cell density as a result of abnormal proliferation.

#### *Fingerprint region (1800-500 $\text{cm}^{-1}$ )*

Narrowing down the analysis to the fingerprint region revealed  $1275$ ,  $1090$ ,  $784$ , and  $546\text{ cm}^{-1}$  as new peaks responsible for the separation and grouping of samples as shown in Figure 3. Representing nucleic acids bands, the peaks located at  $\approx 1090$  and  $784\text{ cm}^{-1}$  are assigned to the symmetric phosphate stretching vibrations and ring breathing modes in DNA and RNA bases (Rehman, 2013; Yu et al., 2006). These two bands showed higher intensities in the cancer samples signifying higher cell density and genetic material overproduction. This overproducing behaviour is characteristic of early and late stages of carcinogenesis. Besides, the phosphate assignment can also serve as an indicator of higher metabolic activity (Talari et al., 2015b).

The band found at  $\approx 1275\text{ cm}^{-1}$  assigned to amide III contributions is a protein characteristic vibration but is also considers a collagen specific assignment (Talari et al., 2015a). Higher contributions of this collagen band were identified in the normal breast in comparison with the cancerous breast. Amide III represents  $\alpha$  helix conformations (Chen and Lord, 1974). The lower intensity in  $\alpha$  helix contributions in the cancerous breast suggested conformational changes in the cancerous tissue.



**Figure 3:** Principal component analysis on the fingerprint region (1,800-500  $\text{cm}^{-1}$ ). (A) Score plot using PC-5 and PC-7 accounting for 3 % of the variance, (B) Average Raman spectral profile for breast cancer and normal breast. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-5 loading, (D) PC-7 loading

This finding is in agreement with our FTIR results (Lazaro-Pacheco et al., 2020b).

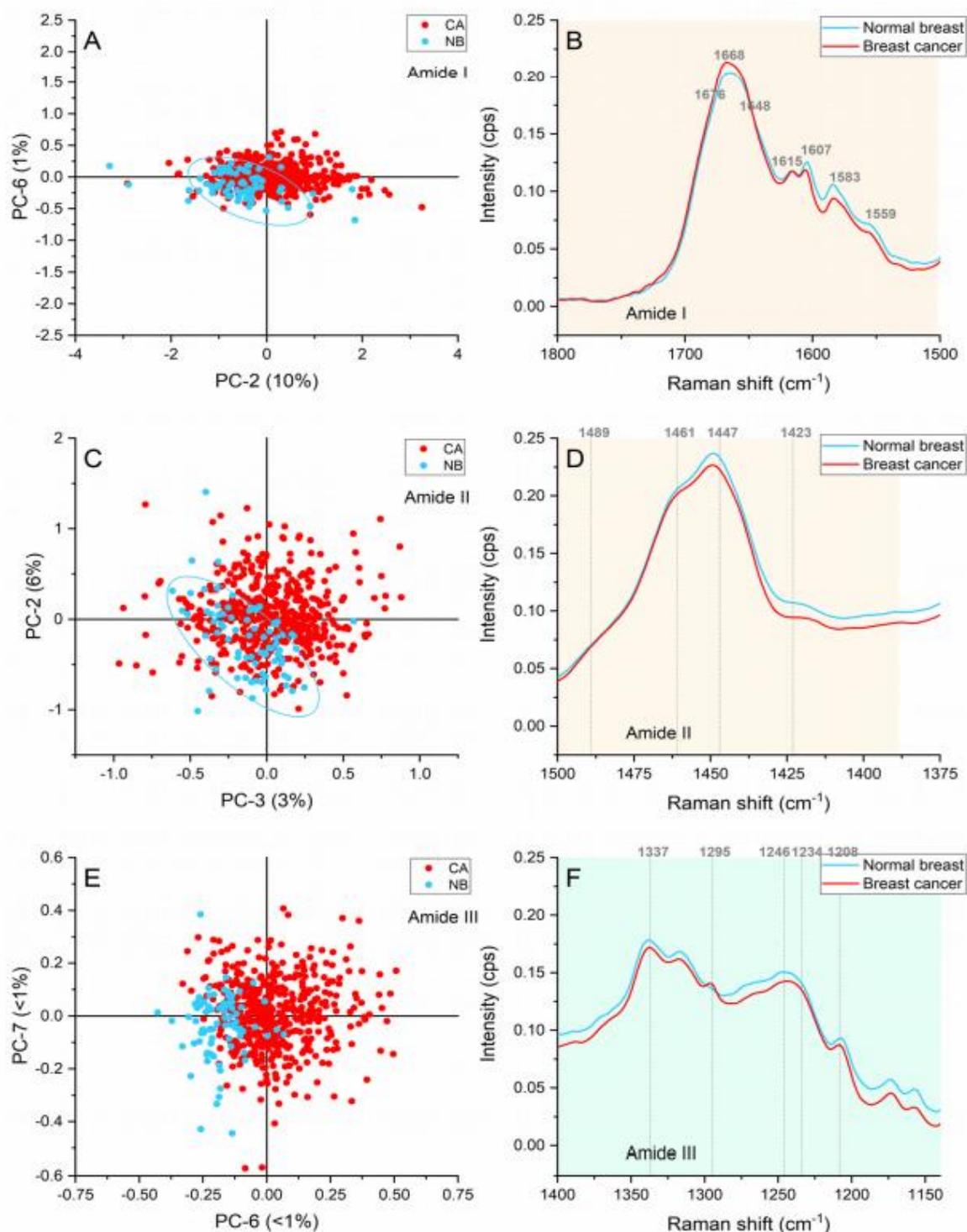
A cholesterol related peak located at  $\approx 546 \text{ cm}^{-1}$  presented greater contributions in the normal breast. Starting stable cholesterol content is expected in normal breast (Kneipp et al., 2003). However, oxidized and modified cholesterol has been reported as a sign of histological progression from normal epithelium to hyperplasia and could explain the low contribution of this component in the malignant tissue (Wrensch et al., 1989).

#### Amide region (1800-1140 $\text{cm}^{-1}$ )

The bands found to contribute towards the grouping of samples when the amide region was analyzed are presented in Figure 4.

#### Amide I region (1,800- 1,510 $\text{cm}^{-1}$ )

The first peak located at  $\approx 1667 \text{ cm}^{-1}$  is caused by the carbonyl (C=O) stretching vibrations of the amide I, but it is also representative of the  $\alpha$ -helix structures of proteins (Rehman et al., 2007). The spectral profiles showed an upshift from  $1663 \text{ cm}^{-1}$  in the normal breast to  $1667 \text{ cm}^{-1}$  in the cancerous area as shown in Figure 4B. Manoharan has previously reported this shift towards higher wavenumbers in addition to a widening of the peak (Manoharan et al., 1998). In our case, a minimal widening can be appreciated in the breast cancer spectral profile, but a higher intensity has been confirmed in the score plot. The higher intensities in the malignant samples suggest a greater contribution of proteins with dominant  $\alpha$ -helix structure.



**Figure 4:** Principal component analysis of the amide region. (A) Score plot of the amide I region (1,800-1,510 cm<sup>-1</sup>), (B) Average Raman spectral profile for breast cancer and normal breast with peaks responsible for the separation in the amide I region, (C) Score plot of the amide II region (1,510-1,390 cm<sup>-1</sup>), (D) Average Raman spectral profile for breast cancer and normal breast with peaks responsible for the separation in the amide II region, (E) Score plot of the amide III region (1,390-1,140 cm<sup>-1</sup>), (F) Average Raman spectra profile for breast cancer and normal breast with peaks responsible for the separation in the amide III region. Ellipses were drawn subjectively based on visual trends.

The peak located at  $\approx 1648\text{ cm}^{-1}$  related to random coil vibrations presented higher intensities in comparison with breast cancer samples. In addition, the  $\approx 1676\text{ cm}^{-1}$  peak assigned to amide I  $\beta$ -sheet conformations showed higher intensities in the cancerous spectra. These conformational changes are caused by ECM adaptation in hypoxia (Talari et al., 2017) and have been differentiators of normal and cancerous cells in *in vitro* studies (Talari et al., 2015b).

The peaks located at  $\approx 1615\text{ cm}^{-1}$  and  $1605\text{ cm}^{-1}$  representing tyrosine phenylalanine, tryptophan, and C=C vibrations in proteins (Rehman, 2013) had higher intensities in the cancerous area of the spectra. These responses serve as indications of overproduction of cellular components leading to abnormal proliferation.

#### *Amide II region (1,510-1,390 $\text{cm}^{-1}$ )*

The bands located at  $\approx 1461$  and  $1489\text{ cm}^{-1}$  presented greater intensities in the cancerous samples. These bands represent deoxyribose vibrations and DNA vibrations. The increased intensities in the cancerous samples suggest an accelerated genetic material production leading to cell division and aberrant behavior. Similarly, variable intensities in the CH deformation-representative band located at  $\approx 1448$ , suggested changes linked to protein and lipid content.

Normal breast samples presented higher intensities in the  $1423\text{ cm}^{-1}$  band associated with the N-H in plane deformation in proteins (Talari et al., 2015a). This freedom for deformation might be caused by the vibrational modes found in strict intensely packed structures in normal tissue in comparison with the much-relaxed  $\beta$ -structures in malignant tissue (Rygula et al., 2013).

#### *Amide III region (1,390-1,140 $\text{cm}^{-1}$ )*

Ceramides ( $\approx 1295\text{ cm}^{-1}$ ), amide III from a concerted ring mode ( $\approx 1234\text{ cm}^{-1}$ ), and tryptophan, phenylalanine, and amide III vibrations characteristic of proteins ( $\approx 1208\text{ cm}^{-1}$ )

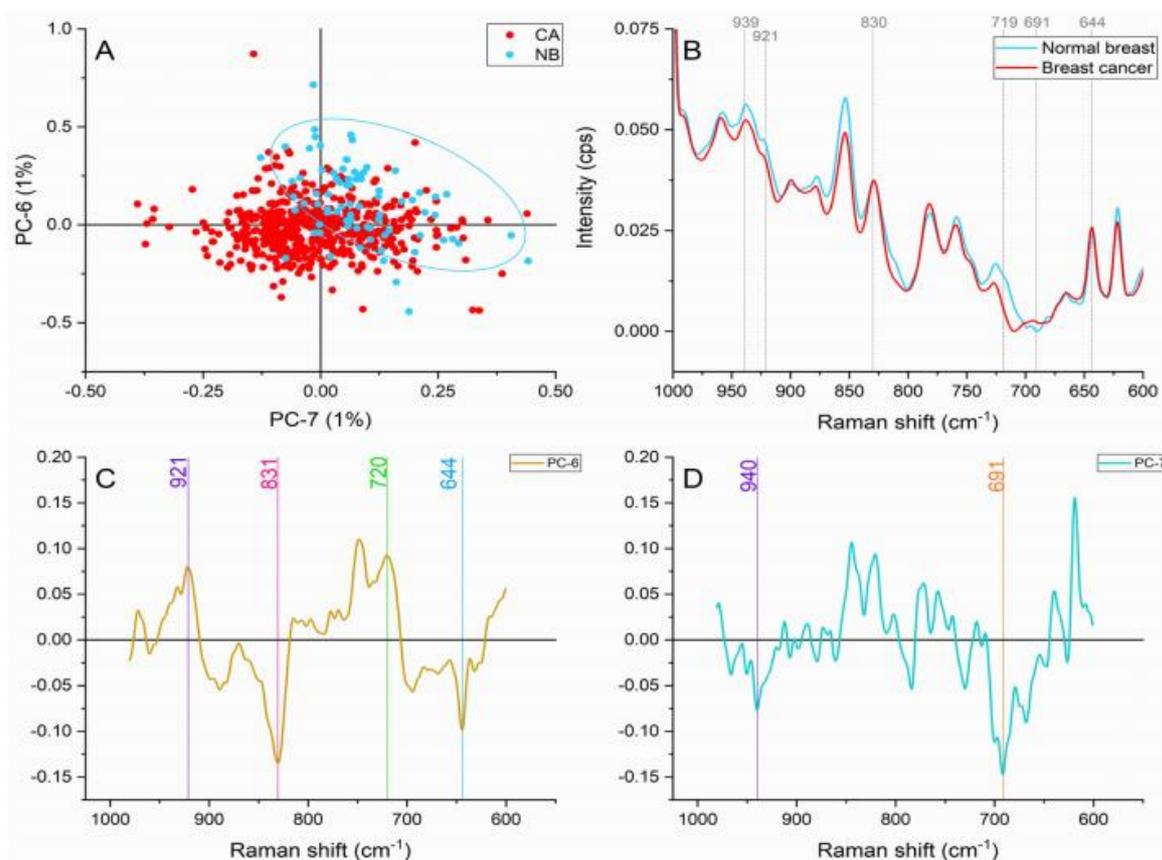
were identified as responsible of the grouping of the cancerous and normal breast. These peaks presented larger intensities in the cancerous areas, which suggest higher metabolic activity and cell density in the malignant tissue.

Nucleic acids variation represented by the  $\approx 1291\text{ cm}^{-1}$  peak linked to cytosine vibrations (Rehman, 2013) were confirmed for the cancerous areas. The production of genetic material and cellular content is expected as part of the aberrant proliferation in malignancies.

The bands found at  $\approx 1243$  and  $\approx 1204\text{ cm}^{-1}$  are considered collagen assignments. The  $\text{CH}_2$  wagging of the glycine backbone and proline side changes contribute considerably to the total intensity of this peak (Talari et al., 2015a). Normal breast presented higher contributions of these bands suggesting the inclusion of collagen in the Raman data collection. This inclusion has been considered due to the lower cell density in the normal epithelia. Nevertheless, some cancerous samples presented considerable contributions to these two peaks. These contributions in the cancer tissue could represent the extracellular matrix preparation to facilitate further invasion and hypoxia survival (Kumar et al., 2013; Talari et al., 2017). Among the adaptation changes, the overproduction of collagen is present by fibroblast stimulation (Provenzano et al., 2008).

#### *Aminoacid and nucleic acid region (980-600 $\text{cm}^{-1}$ )*

The band at  $\approx 921\text{ cm}^{-1}$  caused by the C-C stretch of proline ring, and the peak found at  $\approx 939\text{ cm}^{-1}$  represent the C-C stretch of the skeletal collagen backbone (proline and hydroxyproline) presented important contributions in both tissue types as shown in Figure 5. The collagen contributions can be explaining with the stromal inclusion in the normal breast samples and the overproduction and hyperactive activity of carcinoma-associated fibroblasts (CAFs) in the malignant tissue (Orimo et al., 2005).



**Figure 5:** Principal component analysis on amino acids and nucleic acids region (980-600  $\text{cm}^{-1}$ ). (A) Score plot using PC-6 and PC-7 accounting for 2 % of the variance-ellipses were drawn subjectively based on visual trends, (B) Average Raman spectral profile for breast cancer and normal breast. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-6 loading, (D) PC-7 loading

Breast cancer samples presented higher intensities in comparison with normal breast spectra on the peaks located at  $\approx 830 \text{ cm}^{-1}$  produced by  $\text{PO}_2^-$  stretching of nucleic acids,  $\approx 644 \text{ cm}^{-1}$  representing C-C twisting mode of tyrosine and phenylalanine characteristic of proteins, and  $\approx 691 \text{ cm}^{-1}$  caused by cholesterol ester vibrations. The increased intensities in the malignant tissue are caused by a higher metabolic activity as reported in previous works (Rehman et al., 2007; Talari et al., 2015b).

Normal breast spectra presented greater intensities in the  $\approx 719 \text{ cm}^{-1}$  peak assigned to the C-N vibration in the membrane phospholipid head (Talari et al., 2015a). A number of authors have proposed a decrease in phospholipids in malignant breast tissue (Glunde et al., 2004; Kast et al., 2008; Ruiz-Cabello and

Cohen, 1992; Ting et al., 1996). As a cause of this reduction, the synthesis and breakdown of choline phospholipids caused by specific genetic and phospholipids alterations affecting the membrane fluidity have been proposed.

#### Lipid region (3100-2680 $\text{cm}^{-1}$ )

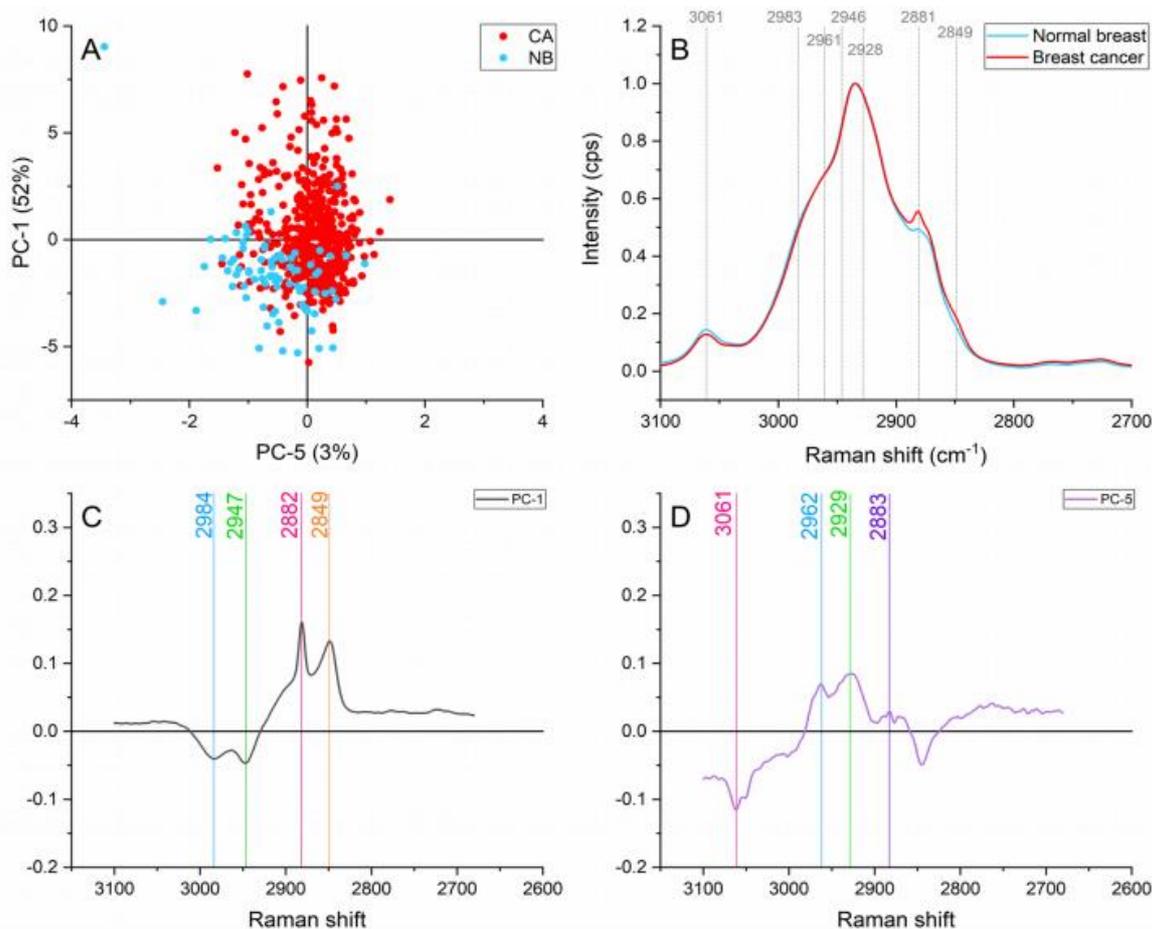
The normal breast spectra presented higher contributions of aromatic CH stretching modes ( $\approx 3061 \text{ cm}^{-1}$ ) and  $\text{CH}_2$  asymmetric stretching of lipids and fatty acids ( $\approx 2946 \text{ cm}^{-1}$ ) as presented in Figure 6. The dominance of lipids in normal tissue has been previously reported when comparing normal and cancerous tissue (Frank et al., 1994; Marro et al., 2014; Talari et al., 2015b). The  $\approx 2961 \text{ cm}^{-1}$  band caused by Out-of-plane chain end anti-symmetric  $\text{CH}_3$  stretch band in lipids and fatty

acids vibrations are responsible for this band (Talari et al., 2015a) also showed differences in the normal breast and cancer spectra.

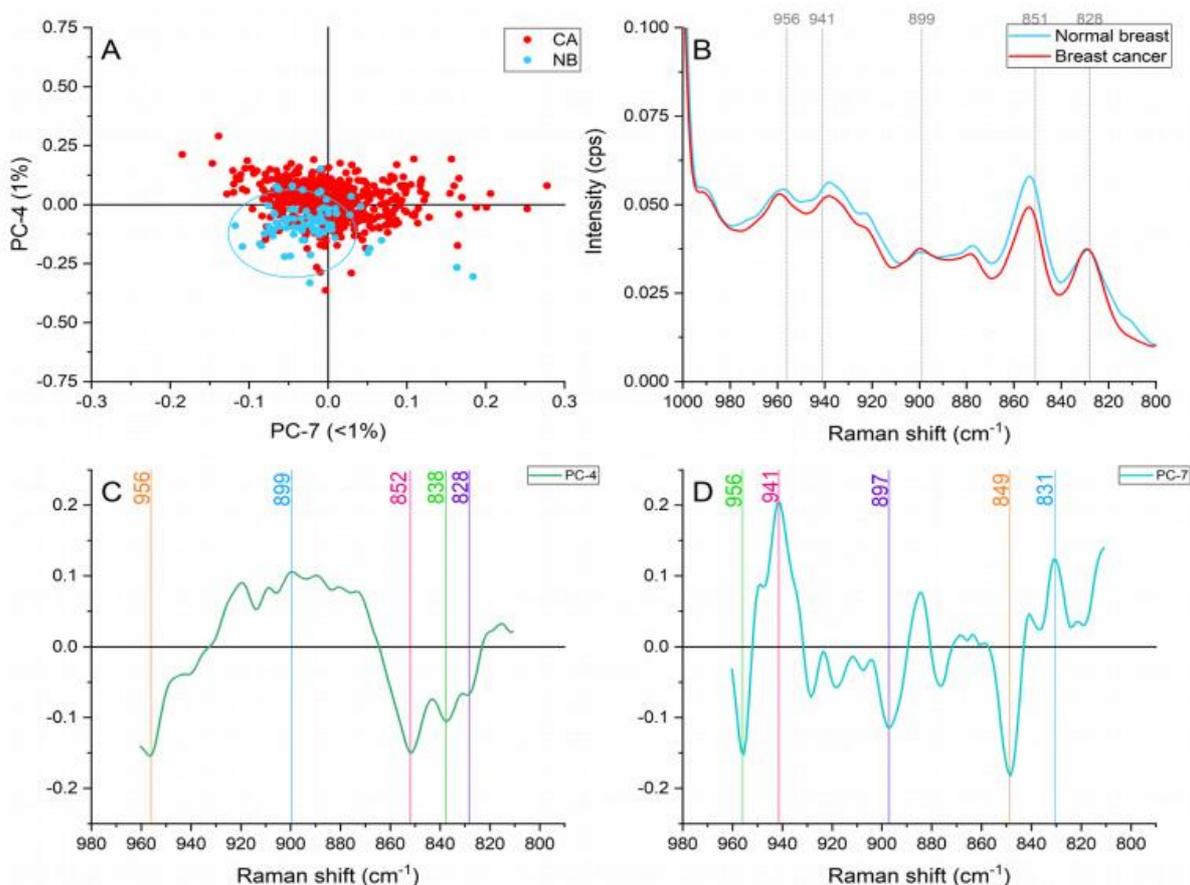
Protein vibrations such as  $\approx 2928$ ,  $2881$ , and  $2849\text{ cm}^{-1}$  produced by the symmetric  $\text{CH}_3$  stretch primarily due to protein,  $\text{CH}_2$  asymmetric stretch of lipids and proteins, and symmetric stretch of C-H respectively presented greater contributions on the cancerous spectra. These protein contributions can be attributed to; uncontrolled and abnormal cell proliferation, cell division, and migration in the malignant tissue (Abramczyk and Brozek-Pluska, 2013; Allred et al., 1993).

#### Hydroxyproline/proline ( $960\text{-}810\text{ cm}^{-1}$ )

Normal breast samples showed increased intensities in the bands located at  $\approx 956$  and  $851\text{ cm}^{-1}$  in Figure 7. The former band is representative of carotenoids (Talari et al., 2015a). Oleic acid and carotenoids have been identified to dominate normal breast tissue (Zhang et al., 1997) and have demonstrated to play key roles in cellular processes that prevent the development of pathologies in the breast tissue (Abramczyk et al., 2012). The  $\approx 851\text{ cm}^{-1}$  peak is produced by proline, hydroxyproline and glycogen vibration characteristics of collagen, and which offer structural integrity and biochemical properties to the connective tissue in this case stoma (Martinez et al., 2019).



**Figure 6:** Principal component analysis on the lipid region ( $3,100\text{-}2,680\text{ cm}^{-1}$ ). (A) Score plot using PC-1 and PC-5 accounting for 55 % of the variance, (B) Average Raman spectral profile for breast cancer and normal breast. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-1 loading, (D) PC-5 loading



**Figure 7:** Principal component analysis on the hydroxyproline and proline region (960-810 cm<sup>-1</sup>). **(A)** Score plot using PC-4 and PC-7 accounting for 2 % of the variance-ellipses were drawn subjectively based on visual trends, **(B)** Average Raman spectral profile for breast cancer and normal breast. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in **C&D**, **(C)** PC-4 loading, **(D)** PC-7 loading

Within this region, the cancer samples presented greater contributions to the spectral peaks found at  $\approx 941$ , 899, and 828 cm<sup>-1</sup>. The stretching of C-C skeletal backbone of collagen assigned to the  $\approx 941$  cm<sup>-1</sup> band suggests an adaptation in the ECM on the cancerous samples (Kumar et al., 2013). Structural rearrangements in collagen, resulting in the exposure of proline and hydroxyproline amino acids have shown to affect the intensities of this band (Martinez et al., 2019). The  $\approx 828$  cm<sup>-1</sup> band representing tyrosine vibrations and stretching of PO<sub>2</sub><sup>-</sup> in nucleic acids in combination with the adenine peak placed at  $\approx 899$  cm<sup>-1</sup> act as an evidence of higher metabolic activity and cell density in the cancerous samples.

## CONCLUSIONS

In this study, Raman spectroscopy proved to be a powerful technique for the analysis of *ex vivo* breast tissue. The combination of spectroscopy and a heterogeneous group of TMA breast samples resulted in the successful differentiation of normal breast and breast cancer with a sensitivity of 90 % and specificity of 78 %. DCIS and invasion were both included for analysis, impacting on the heterogeneity of the cohort.

Dominance of lipids, specifically fatty acids, in the normal tissue is also evident from the results, whereas, protein dominated the malignant breast. The normal breast tissue presented greater contributions of carotenoids,  $\beta$ -carotenoids, and cholesterol which

has been reduced in breast cancer cases likely due to lipid peroxidation. Higher protein and nucleic acid contributions were identified in the cancerous samples. Furthermore, greater amide I  $\beta$ -sheet conformations were present in the malignant samples. Results obtained by Raman spectroscopy in this study are comparable those obtained using FTIR on the same cohort of samples reported previously (Lazaro-Pacheco et al., 2020b), in which structural changes from  $\alpha$ -helix in the normal tissue to  $\beta$ -structures in breast cancer were identified. Ceramide related peaks characterized the breast cancer spectra. The ceramide band could serve as a breast cancer biomarker due to its participation as a structural component of sphingolipids, which are linked to various cellular processes such as cancer development, progression, metastasis and resistance to therapy.

This model allowed to identify significant biochemical differences associated with the presence of breast cancer. The application of this technology for the study and assessment of breast tissue could facilitate screening of a large number of samples and thus prioritize likely cancerous cases for full examination by the pathologists and alleviate the work within pathology departments. In addition, further work into the atypical and precancerous breast lesions will help identify the earliest chemical signs of onset of malignancy enabling timely prophylactic/interventional strategies. Furthermore, this technology could also be used to assess the early effects of response to neoadjuvant therapies at the tissue level; an area of clinical need in breast cancer management.

### **Acknowledgments**

DLP doctoral studies were supported by 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. AMS is supported by the Birmingham CRUK Centre (C17422/A25154). We thank Gouri Baldin for the technical assistance in cutting sample blocks.

### **REFERENCES**

- Abramczyk H, Brozek-Pluska B. Raman imaging in biochemical and biomedical applications. Diagnosis and treatment of breast cancer. *Chem Rev.* 2013;113:5766–81.
- Abramczyk H, Brozek-Pluska B, Surmacki J, Jablonska-Gajewicz J, Kordek R. Raman ‘optical biopsy’ of human breast cancer. *Prog Biophys Mol Biol.* 2012;108:74–81.
- Allred DC, Clark GM, Elledge R, Fuqua SAW, Brown RW, Chamness GC, et al. Association of p53 protein expression with tumor cell proliferation rate and clinical outcome in node-negative breast cancer. *J Natl Cancer Inst.* 1993;85:200–6.
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68:394–424.
- Brożek-Pluska B, Placek I, Kurczewski K, Morawiec Z, Tazbir M, Abramczyk H. Breast cancer diagnostics by Raman spectroscopy. *J Mol Liq.* 2008;141:145–8.
- Brozek-Pluska B, Kopec M, Surmacki J, Abramczyk H. Histochemical analysis of human breast tissue samples by IR and Raman spectroscopies. *Protocols discussion. Infrared Phys Technol.* 2018;93:247–54.
- Camp RL, Charette LA, Rimm DL. Validation of tissue microarray technology in breast carcinoma. *Lab Invest.* 2000;80:1943–9.
- Chen MC, Lord RC. Laser-excited Raman spectroscopy of biomolecules. VI. Some polypeptides as conformational models. *J Am Chem Soc.* 1974;96:4750–2.
- Conklin MW, Eickhoff JC, Riching KM, Pehlke CA, Eliceiri KW, Provenzano PP, et al. Aligned collagen is a prognostic signature for survival in human breast carcinoma. *Am J Pathol.* 2011;178:1221–32.
- Frank CJ, Redd DCB, Gansler TS, McCreery RL. Characterization of human breast biopsy specimens with near-IR Raman spectroscopy. *Anal Chem.* 1994;66:319–26.
- Glunde K, Jie C, Bhujwalla ZM. Molecular causes of the aberrant choline phospholipid metabolism in breast cancer. *Cancer Res.* 2004;64:4270–6.
- Hartmann D, Lucks J, Fuchs S, Schiffmann S, Schreiber Y, Ferreirós N, et al. Long chain ceramides and very long chain ceramides have opposite effects on human breast and colon cancer cell growth. *Int J Biochem Cell Biol.* 2012;44:620–8.

- IARC, International Agency for Research on Cancer W. Cancer tomorrow - global cancer observatory. 2020. <https://gco.iarc.fr/>. Accessed 22 March 2021.
- Kast RE, Serhatkulu GK, Cao A, Pandya AK, Dai H, Thakur JS, et al. Raman spectroscopy can differentiate malignant tumors from normal breast tissue and detect early neoplastic changes in a mouse model. *Biopolymers*. 2008;89:235–41.
- Keller MD, Kanter EM, Mahadevan-Jansen A. Raman spectroscopy for cancer diagnosis. *Spectroscopy*. 2006;21(11):33–41.
- Kneipp J, Schut TB, Kliffen M, Menke-Pluijmers M, Puppels G. Characterization of breast duct epithelia: a Raman spectroscopic study. *Vib Spectrosc*. 2003;32(1):67–74.
- Kumar S, Desmedt C, Larsimont D, Sotiriou C, Goormaghtigh E. Change in the microenvironment of breast cancer studied by FTIR imaging. *Analyst*. 2013;138:4058–65.
- Lau DP, Huang Z, Lui H, Man CS, Berean K, Morrison MD, et al. Raman spectroscopy for optical diagnosis in normal and cancerous tissue of the nasopharynx: Preliminary findings. *Lasers Surg Med*. 2003;32:210–4.
- Lazaro-Pacheco D, Shaaban AM, Rehman S, Rehman I. Raman spectroscopy of breast cancer. *Appl Spectrosc Rev*. 2020a;55:439-75.
- Lazaro-Pacheco D, Shaaban A, Baldwin G, Titiloye NA, Rehman S, Rehman I ur. Deciphering the structural and chemical composition of breast cancer using FTIR spectroscopy. *Appl Spectrosc Rev*. 2020b; epub ahead of print.
- Manoharan R, Shafer K, Perelman L, Wu J, Chen K, Deinum G, et al. Raman spectroscopy and fluorescence photon migration for breast cancer diagnosis and imaging. *Photochem Photobiol*. 1998;67(1):15–22.
- Marro M, Nieva C, Sanz-Pamplona R, Sierra A. Molecular monitoring of epithelial-to-mesenchymal transition in breast cancer cells by means of Raman spectroscopy. *Biochim Biophys Acta*. 2014;1843:1785–95.
- Martinez MG, Bullock AJ, MacNeil S, Rehman IU. Characterisation of structural changes in collagen with Raman spectroscopy. *Appl Spectrosc Rev*. 2019;54:509-42.
- Mian SA, Colley HE, Thornhill MH, 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;49:614–7.
- Orimo A, Gupta PB, Sgroi DC, Arenzana-Seisdedos F, Delaunay T, Naeem R, et al. Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 Secretion. *Cell*. 2005;121:335–48.
- Pinder SE, Brown JP, Gillett C, Purdie CA, Speirs V, Thompson AM, et al. The manufacture and assessment of tissue microarrays: suggestions and criteria for analysis, with breast cancer as an example. *J Clin Pathol*. 2013;66(3):169.
- Provenzano PP, Inman DR, Eliceiri KW, Knittel JG, Yan L, Rueden CT, et al. Collagen density promotes mammary tumor initiation and progression. *BMC Med*. 2008;6(1):11.
- Rehman I ur. *Vibrational spectroscopy for tissue analysis*. Boca Raton, FL: CRC Press, 2013.
- Rehman S, Movasaghi Z, Tucker AT, Joel SP, Darr JA, Ruban A V, et al. Raman spectroscopic analysis of breast cancer tissues: identifying differences between normal, invasive ductal carcinoma and ductal carcinoma in situ of the breast tissue. *J Raman Spectrosc*. 2007;38:1345–51.
- The Royal College of Pathologists. College report finds UK wide histopathology staff shortages. 2018. <https://www.rcpath.org/discover-pathology/news/college-report-finds-severe-staff-shortages-across-services-vital-to-cancer-diagnosis.html>. Accessed 14 April 2020.
- Ruiz-Cabello J, Cohen JS. Phospholipid metabolites as indicators of cancer cell function. *NMR Biomed*. 1992;5:226–33.
- Rygula A, Majzner K, Marzec KM, Kaczor A, Pilarczyk M, Baranska M. Raman spectroscopy of proteins: a review. *J Raman Spectrosc*. 2013;44:1061–76.
- Schiffmann S, Sandner J, Birod K, Wobst I, Angioni C, Ruckhäberle E, et al. Ceramide synthases and ceramide levels are increased in breast cancer tissue. *Carcinogenesis*. 2009;30:745–52.
- Talari A, Movasaghi Z, Rehman S, ur Rehman I, Talari ACS, Movasaghi Z, et al. Raman spectroscopy of biological tissues. *Appl Spectrosc Rev*. 2015a;50:46–111.
- Talari ACS, Evans CA, Holen I, Coleman RE, Rehman IU. Raman spectroscopic analysis differentiates between breast cancer cell lines. *J Raman Spectrosc*. 2015b;46:421–7.
- Talari ACS, Raza A, Rehman S, Rehman IU. Analyzing normal proliferating, hypoxic and necrotic regions of T-47D human breast cancer spheroids using Raman spectroscopy. *Appl Spectrosc Rev*. 2017;52:909–24.

Talari ACS, Rehman S, Rehman IU. Advancing cancer diagnostics with artificial intelligence and spectroscopy: identifying chemical changes associated with breast cancer. *Expert Rev Mol Diagn.* 2019;19:929–40.

Ting YL, Sherr D, Degani H. Variations in energy and phospholipid metabolism in normal and cancer human mammary epithelial cells. *Anticancer Res.* 1996;16:1381–8.

Titloye NA, Foster A, Omoniyi-Esan GO, Komolafe AO, Daramola AO, Adeoye OA, et al. Histological features and tissue microarray taxonomy of nigerian breast cancer reveal predominance of the high-grade triple-negative phenotype. *Pathobiology.* 2016;83:24–32.

Wensch MR, Petrakis NL, Gruenke LD, Miike R, Ernster VL, King EB, et al. Breast fluid cholesterol and cholesterol beta-epoxide concentrations in women with benign breast disease. *Cancer Res.* 1989;49:2168–74.

Yu C, Gestl E, Eckert K, Allara D, Irudayaraj J. Characterization of human breast epithelial cells by confocal Raman microspectroscopy. *Cancer Detect Prev.* 2006;30:515–22.

Zhang S, Tang G, Russell RM, Mayzel KA, Stampfer MJ, Willett WC, et al. Measurement of retinoids and carotenoids in breast adipose tissue and a comparison of concentrations in breast cancer cases and control subjects. *Am J Clin Nutr.* 1997;66:626–32.