The University Of Sheffield.

Spectroscopic analysis of Breast Cancer: Study of Tissue microarrays (TMA)

By:

Daniela Lazaro Pacheco

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Faculty of Engineering
Department of Material Science and Engineering,

July 2019

ABSTRACT

Breast cancer is the most common cancer among women worldwide. The current gold standard for breast cancer screening is triple assessment using clinical examination, histological assessment and imaging (a combination of X-ray mammography and ultrasound). However, this approach limits the understanding of the subtypes and progression of the disease in terms of biochemical changes. Vibrational spectroscopy has demonstrated its potential to provide diagnostic information. Raman and FTIR spectroscopy can facilitate the prediction of the biochemical progression for different diseases in a rapid non-destructive manner. Raman and FTIR spectroscopy was used to characterise and differentiate normal breast and breast cancer as well as different subtypes and grades of breast cancer. Tissue microarrays (TMAs) are valued in cancer research due to the minimal tissue requirements and excellent correlation with whole tissue sections. TMAs containing breast cancer and normal breast samples were analysed with Raman and FTIR spectroscopy. The cancerous samples include different grades and subtypes and represent similar tribal origins in Nigeria. A total of 578 samples were analysed with Raman spectroscopy, while a total of 273 samples were analysed with FTIR. Biochemical changes were detected for luminal, HER2+, and triple negative breast cancer for grade I, II and III. Spectral data revealed key differences in the concentration of biochemical compounds, such as, lipids, nucleic acids and proteins. Principal component analysis (PCA) was able to identify differences that lead to accurate and reliable characterisation of normal breast tissue, and different subtypes and grades of breast cancer. Linear discriminant analysis (LDA) confirmed 86% sensitivity and 89% specificity for the Raman spectroscopy results and 92% sensitivity and 86% specificity for the FTIR results. The identified differences represent a stepping stone towards a complete understanding behind the nature of different subtypes and cancer progression. In summary, vibrational spectroscopy showed a good potential in characterising and identifying new spectral markers. These findings provide an insight into different subtypes of breast cancer and the chemical pathway to the cancer progression associated with different grades.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my supervisors Prof. Ihtesham ur Rehman and Dr Abeer Shaaban for the continuous support of my PhD study and related research, for their patience, motivation, and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. I could not have imagined having better supervisors and mentors.

This PhD program and thesis have been made possible thanks to the financial support of the student grants 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.

Appreciation is due to my fellow labmate and best friend Armine Garcia, for the stimulating discussions, the sleepless nights, and for all the fun we have had in the last four years. Also, I thank my friends in the research group and the Kroto Research Institute.

I want to thank my family for their unconditional love and support. Thank you for putting up with me and always encouraging me to continue. Being away from you is the hardest thing I have had to do, thank you for believing in me. Mamá, gracias por ser mi ejemplo a seguir. Papá, gracias por enseñarme a luchar por lo que quiero. Agradecimientos especiales para mis segundos padres, mis tios. Eduardo, espero disfrutes los chistes. Tom, you are my rock. Thank you for your love and patience, I couldn't have done this without you. This thesis is dedicated to them.

ABBREVIATIONS

AI	Artificial intelligence
AR	Androgen receptor
ATR	Attenuated total reflectance
BC	Breast cancer
BL1	Basal-like 1
CA	Cancerous area
CBCS	Carolina breast cancer study
CK14	Cytokeratin 14
CK17	Cytokeratin 17
CK17	Cytokeratin 17
CK5	Cytokeratin 5
CK6	Cytokeratin 6
COST	Consider One Separate variable at a Time
DCIS	Ductal carcinoma in situ
ECM	Extracellular matrix
EGFR	Epidermal growth factor receptor
ER	Oestrogen receptor
FDA	Food and Drug Administration
FIR	Far infrared
FISH	Fluorescence in situ hybridization
FTIR	Fourier transform infrared
GE	Gene expression
HER2	Human epidermal growth factor
HR	Hormonal receptors
IDC	Invasive ductal carcinoma
IHC	Immunochemistry methods
ILC	Infiltrating lobular carcinoma
IMS	Industrial methylated spirits
IR	Infrared
LAR	Luminal androgen receptor
LCIS	Lobular carcinoma in situ
LDA	Linear discriminant analysis
LN	Lobular neoplasia
M	Mesenchymal
MIR	Mid infrared
MRI	Magnetic resonance imaging
MSL	Mesenchymal stem–like
NIR	Near infrared
NST	No special type
PARP	Poly(adenosine diphosphate–ribose) polymerase
PC	Principal component
PCA	Principal component analysis

PR	Progesterone receptor
RS	Raman spectroscopy
SEER	Surveillance Epidemiology and End Results
TDLU	Terminal duct lobular unit
TMA	Tissue microarray
TN	Triple negative
TNBC	Triple negative breast cancer
WHO	World health organization
WTS	Whole tissue sections

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

CHAPTER 1. INTRODUCTION

The incidence of breast cancer (BC) has risen significantly in the past years. Nowadays BC is the second most frequent cancer among women.[1] Evidence suggests that early detection and diagnosis are among the most important factors for a favourable prognosis. Once a tumour is detected and it is suspected to be cancerous, biopsy is the typical procedure to follow. Currently histopathological assessment of biopsy samples remains the gold standard for diagnosis. However, several drawbacks such as subjectivity of the technique, time delay, tissue trauma, patient stress and the need of further invasive procedures still remain.

As part of the diagnosis, the receptor status is assessed. Oestrogen receptor (ER) and Human epidermal growth factor receptor 2 (HER2) are compulsory to assess. Progesterone receptor (PR) is optional, but it also forms part of this status. These 3 receptors are used to target the cancer cells in addition to chemotherapy. Triple negative breast cancer (TNBC) is a subtype of BC in which hormonal and HER2 receptors are not expressed and therefore its current treatment is limited to chemotherapy. Black and Hispanic young women present a higher incidence of TNBC in comparison with other ethnicities and age groups. In addition, this subtype is aggressive, and it is related to a poor prognosis. It is clear that a better understanding of breast cancer, its heterogeneity and molecular drivers is needed in order to develop effective targeted treatments and therapies.

Vibrational spectroscopy has been used to study different tissues and the progression of several diseases including cancer. The use of vibrational spectroscopy techniques such as Raman and FTIR spectroscopy could be a significant asset in the field of oncology research. The use of vibrational techniques could allow identifying chemical changes in the breast tissue related to the presence, development, and progression of breast cancer and therefore facilitating our understanding of the disease. The patient status includes clinical data and pathological parameters such as laterality, tumour size, type, grade and stage of BC. The spectral data managing through a combination of statistical analysis and chemometrics could allow the inclusion of patient status and as consequence a better characterisation of BC.

AIMS AND OBJECTIVES

1. Raman spectroscopic analysis of human breast cancer and normal breast biopsies on tissue microarrays (TMA)
2. Analysis of spectroscopic results with multivariate analysis employing Principal Component Analysis (PCA) and cluster analysis and Linear Discriminant Analysis (LDA) to distinguish specific cancer related changes.
3. Fourier-transform infrared (FTIR) spectroscopic analysis of human breast cancer and normal breast biopsies on tissue microarrays (TMA)
4. Chemometric analysis using statistical analysis: Principal Components Analysis (PCA), Linear Discriminant Analysis (LDA) and Cluster analysis to distinguish specific cancer related changes.

Literature review

CHAPTER 2. LITERATURE REVIEW

2.1 CANCER OVERVIEW

Cancer is a disorder of cells that ends up in the formation of a tumour, which is considered a mass of cells with abnormal proliferation. The lack of capacity to follow normal differentiation and growth processes distinguishes a tumour cell from a normal cell. In the cell cycle, there are 3 well-known checkpoints to diverge out or programme to death these abnormal cells. Unfortunately when cancers arise these checkpoint are compromised due to disruptions or key proteins alterations in the cell cycle.[2]

According to the Globocan2018 statistics performed by the International Agency for Research on Cancer in charge of the World Health Organization, 18.1 million cancer cases were predicted to be diagnosed worldwide in 2018. In this year 9.6 million deceases were predicted on the same report.[1] These numbers reveal that this disease is a leading cause of death among different populations, affecting people of a wide range of ages in developed and developing countries. According to the World Cancer report 2014, 22 million of cancer cases will exist within the next two decades.

Early detection has been the measure taken around the globe to decrease cancer patient´s mortality by giving treatment at early stages. Some of the risk factors for cancer have been well identified, but some others remain unproven making it difficult to stablish successful prevention programs.

2.2 BREAST CANCER

When cancer cells form a tumour in the breast epithelial tissue this cancer is known as breast carcinoma.[3] Carcinoma is a term that could be used for any epithelial cancer. However, if the cancer origin is a gland tissue the term given is adenocarcinoma.[2] Breast cancer is considered a heterogeneous malignancy, not only because of its morphological and physical characteristics, but also at the molecular level. All these variations affect the progression and prognosis of the cancer, in addition to determine the treatment options suitable for its particular characteristics. [4], [5]

According to the Global Cancer Statistics 2018 report, breast cancer is the second most frequent cancer worldwide.[1] As consequence of its tremendous heterogeneity the understanding, study and research of this disease is vital to improve and develop efficient and specific treatments.[6] The understanding of the specifics and general aspects of BC is essential to aboard this public health problem. Identifying all the risk factors, facilitating easier and faster detection techniques, and offering cancer specific therapies remain priorities in health research.[7]

2.3 BREAST STRUCTURE

The breast is a combination of several tissues located on the top of the pectoralis major forming protuberances at the chest level. Enlargement of the female breast occurs during puberty. The mammary gland undergoes several changes before, during and after pregnancy

and nursing. The breast gland development occurs during pregnancy, with the milk production purpose followed by its atrophy when the nursing process is ceased.[8] The breast undergoes several hormonal changes through the puberty, pregnancy, lactation, and after menopause.[9] Owing to these regenerative and remodelling processes, several diseases can arise in the breast tissue.[10]

As it can be seen in Figure 1, the breast is formed of the nipple, lobes and their duct system, adipose tissue, suspensory ligaments and connective tissue stroma. The lobes contain the duct system and can be found in a radial form around the nipple. The duct system is formed by collecting ducts, lactiferous sinus, segmental duct, sub-segmental ducts, terminal ducts and lobules. The nipple has around 8-24 collecting ducts represented by orifices, which represent a lobe each. In the nipple smooth muscle fibres surround the lactiferous ducts with the purpose of erecting the nipple for suckling. Due to its circular arrangement contracting and emptying the lactiferous sinuses is possible.[10] External to the ducts and lobules there is connective tissue known as stroma. The stroma contains collagen, fibroblast, plasma cells, fat cells, blood vessels, lymphocytes, foam cells, and sometimes pigment-laden histiocytes and multinucleated stromal giant cells. The stroma occupies an intra- and inter- lobular position. This structure is a combination of adipose and fibrous tissue and it is the responsible for the bulky shape in women breast. The adipose component of the stoma it is usually located in the interlobar stroma. The stromal fibrous tissue supports the gland with it mechanical needs. [11]

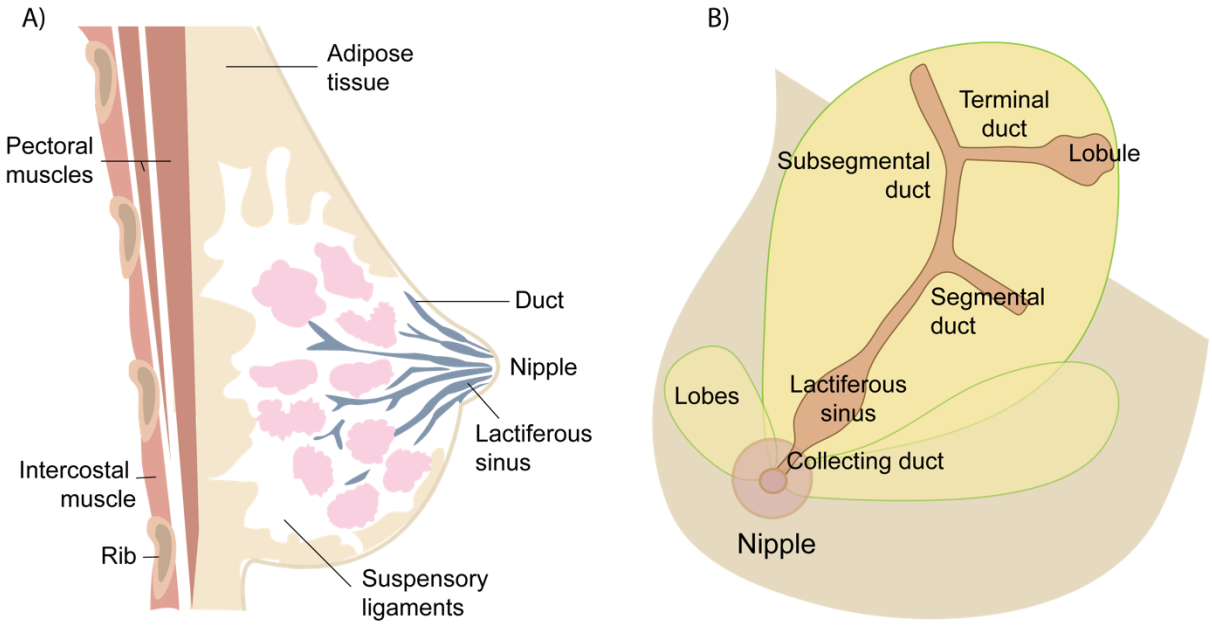

Figure 1. (A) Structure of the adult female breast. (B) Duct system within a lobe.

The function of the collecting ducts resides in draining the mammary lobe ducts to the nipple. The accumulation of milk during lactation is in charge of the lactiferous sinus. The lactiferous sinus can be found right after the collecting duct in the duct system. The segmental duct divides into sub-segmental ducts that end in terminal ducts. The terminal duct is in charge of

draining the lobule. Each lobe can host around 20-40 lobules. These lobules are arranged in a grapelike cluster form. Each lobule is formed by 10-100 acini. Modified from [8], [10], [12]

The terminal duct lobular unit (TDLU) consist on a lobule and its corresponding terminal duct. The TDLU is where most diseases processes arise due to the changes suffered by the epithelial cells and lobules under hormonal stimulation.[10]

The lymphatic channels of the breast drain in their majority to the axillary mammary nodes. There are 3 levels of nodes. The importance of the breast relation with the lymphatic channels resides in the possibility of cancer metastases to the lymphatic system.[8]

2.4 Epidemiology

In 2018 was determined that breast cancer had a 24% incidence among all cases of women´s cancer, converting it in the most common cancer among women. It has also been reported that nearly 2.1 million cases will be diagnosed in 2018. [1]

With a total of 630,000 deaths in 2018, breast carcinomas had the first mortality rank among all types of cancers in females. The incidence and mortality in different regions of the world can be seen in Figure 2, Figure 3, and Figure 4.

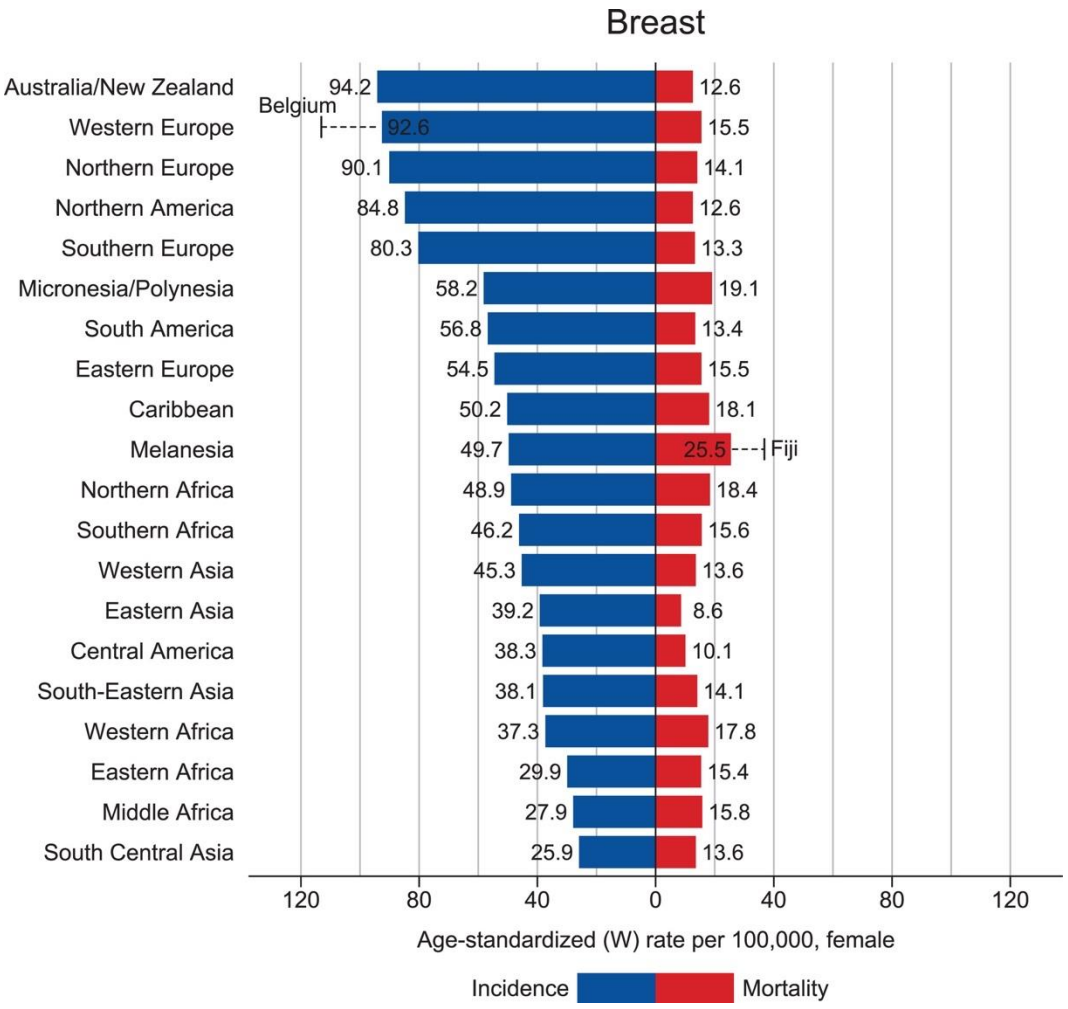

Figure 2. Incidence and mortality rates for breast cancer. [1]

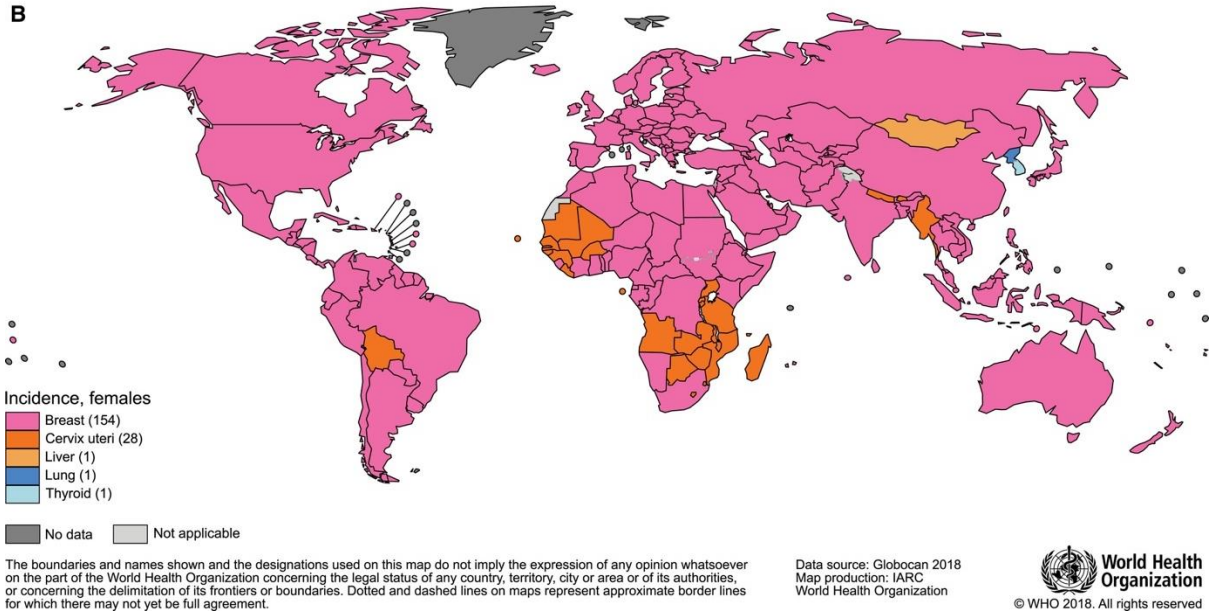

Figure 3. Global representation of the most common type of cancer incidence in 2018 by country among females. The numbers of countries represented in each ranking group are included in parenthesis the legend. [1]

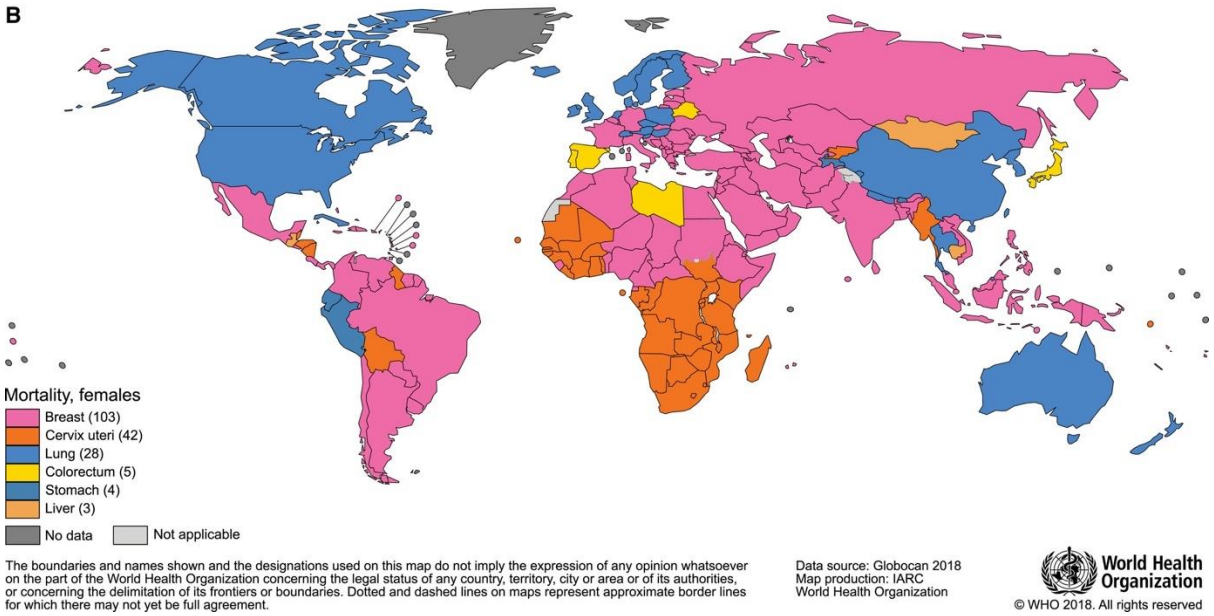

Figure 4. Global representation of the most common type of cancer mortality in 2018 by country among females. The numbers of countries represented in each ranking group are included in parenthesis the legend. [1]

The lower rates of breast cancer incidence correspond to younger ages (<50), increasing as the age does. The incidence of breast cancer rates also differs by ethnicity and race. BC incidence for white women is higher than for black and South Asian women. The incidence for women younger than 40 years is higher for black population than for white background women. However, for women older than 40 this tendency is reversed. The mortality for black women suffering from breast cancer is higher in comparison with white patients.[13]–[16]. According with the Cancer incidence and survival by major ethnic group, England, 2002 - 2006

report, in England young Asian women (15-64 years) have a lower survival rate at 3 years in comparison with white ethnicity women. This same report showed that Black women between 15-64 years presented a reduced survival at 1 and 3 years compared to white women. The survival comparison of black and white women older than 64 years showed no significant differences.[13] The same tendency has been demonstrated by different studies and programs such as the population-based Surveillance, Epidemiology, and End Results (SEER) Program, and the Disease Control and Prevention's National Program of Cancer Registries (NPCR) in USA among many others around the globe. [16], [17]

In accordance with reports from patients in UK and USA, young black women are more likely to be diagnosed with tumours associated with the worst prognosis, such as triple negative breast cancer. In addition to the influence of the ethnicity, race, tumour type, and cancer stage; the biologic variations in the tumour, socioeconomic factors, time of detection, access to healthcare and treatment services play a vital role in breast cancer survival. [14], [17] The rates in mortality in countries deprived of early detection programs or standardized guidelines to handle and report breast samples is significantly higher than in developed countries with rigid protocols and programmes. An example of this is Nigeria among other African countries, where education and training in pathologic reporting needs to be promoted.[18]

2.5 Aetiology

Several breast cancer risk factors have been recognised including, age at menarche and natural menopause, birth rate, pregnancy age, lactation histories, the menstrual cycle patterns, ovarian activity, the use of hormone therapies as contraceptives and menopausal hormone use, obesity, diet patterns, genetics, tobacco use and alcohol intake.[10], [19], [20]

Different studies have suggested that the BC aetiologies may be as heterogeneous as the disease. For example reproductive factors such as early menarche, nulliparity and delayed childbearing represent risk factors for hormone receptor-positive tumours, but not for hormone receptor–negative cancers.[21]

2.6 Future predictions

The International Agency for Research on Cancer has predicted an important increase of BC rates in the following years. The expected number of breast cancer cases and mortality for the next two decades are displayed in Figure 5. Based on these figures a 46% increase of BC cases will be occur by 2040 in comparison with 2018, while an increment of 58% in mortality is expected.[22]

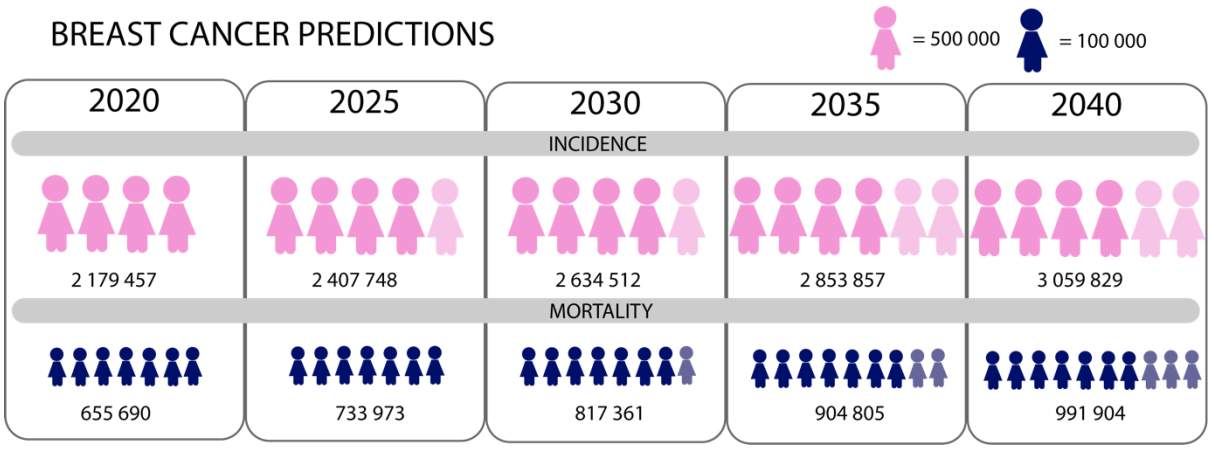

Figure 5. Breast cancer incidence and mortality prediction worldwide for females of all ages. Based on the Globocan IARC online analysis prediction tool.

2.7 Breast tumours classification

Tumours are masses formed by cells with abnormal characteristics such as atypical proliferation. These could be classified as benign or malignant. Benign tumours are formed by abnormal cells, but they are not identified as cancerous. On the contrary, the malignant tumours are formed by cancer cells. If these cells are contained or limited to the ducts without invading the basement membranes the tumour is known as in situ. The most common cases are ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS). When the cancer cells invade the basement membranes and adjacent fat/stroma this is regarded as invasive carcinoma. The commonest examples are infiltrating ductal carcinoma (IDC) and infiltrating lobular carcinoma (ILC). Invasive tumours have the possibility of invading surrounding tissue, for epithelial tumours the invasion through the blood vessels and lymphatic system bring high chances of carrying malignant cancer cells to other tissues causing secondary tumours. This process is better known as metastasis.[2]

Reliable and consistent tumour detection techniques have been used for tumour detection. Breast screening is carried out using a triple assessment including imaging, X-ray mammography and ultrasound, clinical examination and histopathological assessment.[23] Tumour detection can be done rapidly and effectively using imaging. However, determining the nature of the tumour and tumour assessment requires the use of invasive labour-intensive and rigid protocols. If a detected lesion presents with characteristic that suggest cancer, a sample must be taken. The tumour sample is obtained using a thin needle to retrieve the tissue. Needle core biopsies have some limitations such as insufficient material retrieval, sampling errors, and risk associated with the procedures.

Biopsy schedule, long waiting for results, and missampling are reasons for uncertainty, anxiety and emotional and physical stress in patients.[24] Unfortunately, real-time diagnosis cannot be performed with the current techniques. As consequence intraoperative margin evaluation during tumour removal represents an extra challenge for surgeons. Until this day, there is only one Food and Drug Administration (FDA) approved device for intraoperative margin assessment. This device is named MarginProbe and works based on differences in the electromagnetic wave reflection among normal and cancerous tissue. The most important

limitation lies in the fact that although this device provided a 75% sensitivity its 46% specificity is still very low.[25] Histological examination remains the gold standard for confirming the diagnosis and for margin assessment.

Breast tumours may be classified based on their histological characteristics according to the WHO classification of tumours of the breast.[26] In order to achieve a better understanding of the invasive and non-invasive tumour nature a simple schematic representation of breast cancer progression is shown in Figure 6. In the UK the Royal College of Pathologists reporting guidelines are used to identify the tumour type and perform the classification.

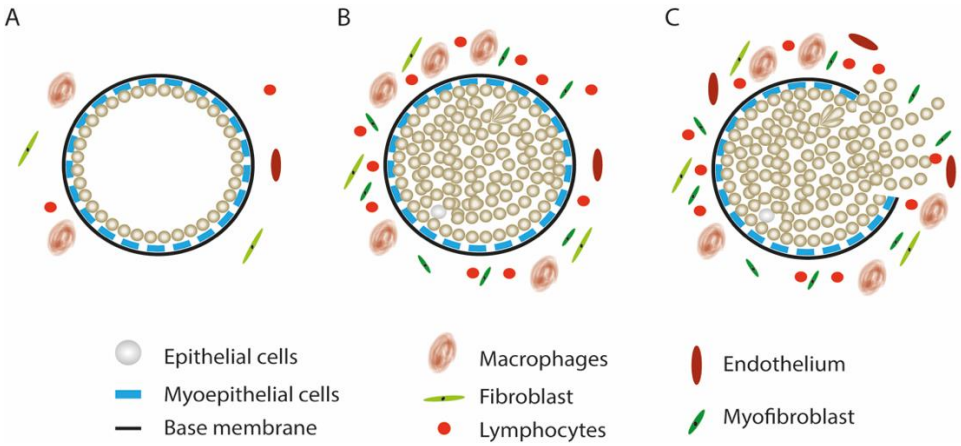

Figure 6. Breast cancer progression. (A) Normal breast tissue. (B) Ductal carcinoma in situ, proliferation of abnormal (epithelial and stromal) cells. (C) Invasive carcinoma, base membrane and myoepithelial cells partial degradation allowing cancerous cell invasion into ECM.

Diagnosis and assessment of tumours is doing by microscopic evaluation. These tasks have as an objective to determine the existence and type of cancer and provide oncologists with enough information to select the treatment needed by that specific subject. The type of treatment is determined based on the histoprognostic index, the histological parameters and the molecular profile. To achieve this, tumour staging, and grading are required. The staging evaluates the spread of the tumour. According to internationally accepted staging protocols clinical characteristics such as the size of the neoplasm and near lymph nodes status must be considered. The stage is assigned based on the scores obtained by the TNM system. This approach includes 3 different aspects to assess: T for Primary Tumour, N for Regional Lymph Nodes and M for Distant Metastasis. A simplified representation of the TNM system based on the Royal College of Pathologists reporting guidelines is presented in Table 1. [27] Based on this, there are four stages of breast cancer, with the stage 1 being the earliest state of malignancy and stage 4 being metastatic. [26] The possible stages and how are they calculated based on the TNM systems are represented in Table 2.

Table 1. TNM classification of tumours of the breast based on the Royal College of Pathologists reporting guidelines. [27]

Primary tumour (T)	
T0	No evidence of primary tumour
Tis	Carcinoma in situ /DCIS/LCIS

T1	Tumour ≤20 mm in greatest dimension
T2	Tumour >20 mm but ≤50 mm in greatest dimension
T3	Tumour >50 mm in greatest dimension
T4	Tumour of any size with direct extension to the chest wall and/or to the skin (ulceration or skin nodules) [NB. Invasion of the dermis alone does not qualify as T4]
Nodes (pN)	
pN0	No regional lymph node metastasis identified histologically
pN1	Micrometastases OR
	Metastases in 1–3 axillary lymph nodes AND/OR
	Metastases in internal mammary nodes with metastases detected by sentinel lymph node biopsy but not clinically detected
pN2	Metastases in 4–9 axillary lymph nodes OR
	Metastases in clinically detected internal mammary lymph nodes in the absence of axillary lymph node metastases.
pN3	Metastases in ≥10 axillary lymph nodes OR
	Metastases in infraclavicular (level III axillary) lymph nodes OR
	Metastases in clinically detected ipsilateral internal mammary lymph nodes in the presence of one or more positive level I, II axillary lymph nodes
	OR
	Metastases in >3 axillary lymph nodes and in internal mammary lymph nodes with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected OR
	Metastases in ipsilateral supraclavicular lymph nodes.
Distant metastases (M)	
M0	No clinical or radiological evidence of distant metastases.
M1	Distant detectable metastases as determined by classic clinical and radiographic means and/or histologically proven >0.2 mm

Table 2. Breast tumour stage based on the Royal College of Pathologists reporting guidelines. [27]

Stage	T	N	M
0	Tis	N0	M0
IA	T1	N0	M0
IB	T0	N1mi	M0
	T1	N1mi	M0
IIA	T0	N1	M0
	T1	N1	M0
	T2	N0	M0
IIB	T2	N1	M0
	T3	N0	M0
IIIA	T0	N2	M0

	T1	N2	M0
	T2	N2	M0
	T3	N1	M0
	T3	N2	M0
IIIB	T4	N0	M0
	T4	N1	M0
	T4	N2	M0
IIIC	Any T	N3	M0
IV	Any T	Any N	M1

Tumour grading evaluates how similar/different the cancer cells appear in comparison with normal cells. The tumour grade is established using histological and cytological criteria. The cancer grade reflects the tumour morphology and its proliferative capacity. [2], [28] In the UK the Nottingham criteria is used to assess the tumour grade. The Nottingham criteria is based on the scoring of gland (acinus) formation, nuclear atypia/ pleomorphism, and mitosis count from 1 to 3 depending on a specific qualitative evaluation. When the scores from acinus formation, nuclear atypia, and mitosis count are added the total score must be in the range of 3-9. Depending on the score the grade is assigned as indicated in Table 3.

Table 3. Tumour grading according to the Nottingham criteria and the Royal College of Pathologists reporting guidelines. [27]

Score	**Grade**	
3, 4, 5	I	Cancer cells resemble normal cells. Slow growth
6, 7	II	Cancer cells look clearly different to normal cells. Higher proliferation
8, 9	III	Abnormal look. Spread more aggressively, Higher proliferation.

Breast tumours are also assessed to identify the receptor status. The biomarker receptor status identification is performed to score the oestrogen receptor (ER), and the human epidermal growth factor (HER2) protein overexpression. The progesterone receptor (PR) can also be assessed, but it is not compulsory. According to the receptor status, tumours can have different clinical behaviour, response to endocrine treatment an anti her2 therapy and different metastatic probability. Owing to this fact the identification of biomarker receptor status is an imperative factor to consider in clinical decision-making.[29]

A classification based on immunopathologic characteristics assessing the biomarker receptor status is an important task to provide an accurate diagnostic. Scoring of PR, ER and HER2 is performed through immunochemistry methods (IHC) and Fluorescence In Situ Hybridization test (FISH).[29] Several research groups have contributed to the tumour subtype classification including markers such as Ki-67 and PCNA (related to cell proliferation), cytokeratins5/6/14/17 and epidermal growth factor receptors. Considering contributions

from different literature Table 4 shows the breast cancer subtypes by molecular profiling and IHC. [29]–[32]

Table 4. Subtype classification based on IHC/FISH.

Subtype	Features
Luminal A	(ER+ /PR+ & Ki-67<14%)
Luminal B	(ER+ /PR+ & Ki-67≥14%)
Luminal/HER2	(ER+/PR+/HER2+)
HER2 enriched	(ER-/PR-/HER2+)
Basal-like	(ER-/PR-/HER2- & EGFR + and/or CK5/6 +)
Triple-negative (TN)	(ER-/PR-/HER2- & EGFR- & CK5/6 -)

In the UK the histopathological assessment is very active and functional due to an external quality assurance programme. The implementation of the programme has supported the delivery of high-quality care to breast cancer patients. However, the use of strict protocols is inconsistent around the world with variations in the results being reported. These variations are caused by differences among fixation times, methods, staining times and antigen retrieval.[33] Constant measures are taken to improve the reporting of breast samples. For example, exploring new technologies to complement the current histopathological assessment gold standard.

Molecular classification has been achieved in high resolution by using microarray-based gene expression profiling. When unsupervised analysis has been performed for different types of BC, the ER status has proven to be the most critical discriminator. However, clear clusters have been obtained identifying several receptor status subtypes.[29], [31], [34], [35] Based on the identification of specific gene sequence on different BC subtypes, disease progression (information related to proliferation rates, high/low grade tendency, metastatic spread patterns and prognosis) can be better understood through the molecular changes occurring and underlying the biology of BC.[29], [30] Although gene signatures confirmed that the receptor status classification offers a good reference to suggest treatment, it has also confirmed the heterogeneity of the disease and it has helped expand the subtype classification with other three subgroups: claudin-low, molecular apocrine and interferon. The molecular subtype features are display in Table 5. This table has been modified from Alizart et al. [29]

Table 5. Features of molecular profiled breast cancer subtypes. [29]

Hormonal status	Molecular Subtype	HER2 by IHC/FISH	Ki-67 IHC	Additional features
ER-positive	Luminal A	HER2-	Low	Luminal cytokeratin +; E-cadherin +/-
	Luminal B	HER2-/+	High	Luminal cytokeratin +; TP53 mutations

ER-negative	HER2	HER2+	High	TP53 mutations
	Basal-like	HER2-	High	Basal cytokeratin +; TP53 mutations; DNA repair loss; EGFR +/- ; KIT +/-
	Molecular apocrine	HER2-	High	Androgen receptor þ
	Claudin-low	HER2-	High	Cancer stem cell-like; EMT-like; E-cadherin low
	Interferon-related	HER2-	High	STAT 1

Different studies involving different cohort sizes had been performed reporting a similar classification. However, consensus and concordance about the classification is non-existing. Although several studies with big cohorts had been carried out involving gene profiling, the molecular subtyping is not standardized. The variability showed by molecular subtyping reflects that further evidence is required before implementing them in routine use. In addition, the time and costs of those technique and tumour sampling will represent a limitation to its application into clinic.[29], [34]. Recently, molecular tests have been approved by the National institute for Health and Care Excellence (NICE) for the routine pathway of breast cancer management in selected patients including EndoPredict (EPclin score), Oncotype DX Breast Recurrence Score, Prosigna, MammaPrint and IHC4+C. [36]

Taken together, these classifications suggest that there is an association between the hormonal receptor status, treatment and prognosis. Positive hormonal receptors are used to target the cancer cell and treat the disease with endocrine or aromatase inhibitors therapy following surgery. It has been shown that the use of adjuvant systemic therapy has a positive effect on reducing breast cancer mortality and recurrence in selected patients.[37], [38] Therapy for negative hormonal receptors differs from positive hormone cases in a number of important ways. Triple negative breast cancer (TNBC) has no receptors to be targeted therefore therapy is limited to chemotherapy, radiation and surgery.[39] The limited therapy options, the aggressiveness of an only cytotoxic approach treatment and the TN phenotype itself have made the study of TNBC increasingly important in the oncology area.[40]

2.8 Triple negative breast cancer /phenotype (TNBC)

This type of breast cancer is characterized by tumours with negative expression of progesterone receptor (PR), oestrogen receptor (ER) and that does not overexpress Human epidermal growth factor receptor 2 (HER2), hence its name. These triple negative (TN) tumours represent the 10-20% of breast tumours. TN Tumours are usually larger and with higher grades in comparison with those corresponding to other BC subtypes. [41], [42] TNBCs are resistant to hormonal therapies maintaining chemotherapy as the only course of action for its treatment.[43] This molecular subtype is considered biologically aggressive and its heterogeneity hinders the molecular target identification in order to develop treatments. [42]

2.8.1 Classification

TNBC can be subdivided in two types, the basal and non-basal type. To be considerate as triple negative the basal type must have a coral basal phenotype which consist in the triple negative profile and the expression of one or more of the basal cytokeratins such as cytokeratin 5 (CK5), cytokeratin 6 (CK6), cytokeratin 14 (CK14), cytokeratin 17 (CK17) and epidermal growth factor receptor (EGFR), . On the contrary the non-basal TNBC does not express any of these receptors and maintains the ER-/PR-/HER2- phenotype. Basal-like TNBC represents 70-90% of TN phenotype, while 75% of basal-like tumours are TN. It is highly relevant to indicate that basal-like tumours can also express positive hormonal receptors (HR) or HER2, however these would not be considered as TN tumours. Some features of TNBC are displayed in Table 6.[41], [43]–[45] The identification of these TN phenotypes is performed using IHC with HER2 FISH analysis. Some evidence from molecular studies reflects that TNBC encompasses other molecular subtypes such as basal-like, Claudin-low, interferon-rich and normal-breast like.[43], [44]

Table 6. Features and characteristics associated to TNBC phenotype.

Basal-like TNBC
Core basal phenotype (TN and expression of any CK5/6, CK14, CK17 and EGFR) Arise from luminal progenitor compartment, or from normal basal cells. Abnormal chromosome X inactivation Node-negative likely
TNBC (ER-/PR-/HER2-)
Incidence risk increased by societal and genetic factors Cell surface profile similar to the breast cancer stem cells CD44+CD24- and ALDH1A1 expression. A subset may represent a BRCA1 germline mutation tumours Relationship between the survival probability and primary tumour size. Tumour presents as high grade, increased mitotic count, scarce stromal content, central necrosis, lymphocytic response visible in stroma. Patient characteristics: Young age at diagnosis, mutation carriers Tumour characteristics: Usually ductal or mixed histology and high grade Treatment/prognosis; Poorer prognosis, chemotherapy sensitive, no targeted therapy and high risk of early relapse

Gene expression (GE) profiling reveals the heterogeneity of the TNBC subtype. Six TNBC subtypes with different ontologies and unique profiles were identified by Lehman et al when 587 TNBC cases were profiled and cluster analysis was performed.[42] This classification is shown in Table 7.

Table 7. TNBC subtypes displaying unique gene expression (GE) and ontologies. Distinct gene ontologies are associated with each TNBC subtype.

TNBC Subtype		**Top gene ontologies for TNBC subtypes**
Basal-like	BL 1	Heavily enriched in cell cycle and cell division components and pathways Expression of genes associated with proliferation. Elevated DNA damage response pathways Increased proliferation and cell-cycle checkpoint loss in consistency with elevated expression of the DNA damage response genes

	BL 2	Growth factor signalling and receptor gene expression as well as glycolysis/gluconeogenesis. This subtype has features suggestive of basal/myoepithelial origin.
Immunomodulatory (IM)		Enriched immune cell processes/signalling. Core immune signal transduction pathways. Immune cell-surface antigens, cytokine signalling, complement cascade, chemokine receptors and ligands, and antigen presentation.
Mesenchymal (M)		Heavily enriched in components and pathways involved in cell motility, ECM receptor interaction, and cell differentiation pathways. M and MSL share features with metaplastic breast cancer, characterized by mesenchymal/sarcomatoid or squamous features and its chemoresistance.
Mesenchymal stem–like (MSL)		Enrichment of genes for cell motility, differentiation, and growth pathways. Unique to the MSL express growth factor signalling pathways. Genes involved in angiogenesis. Low levels of proliferation genes in comparison with M, accompanied by enrichment in the expression of genes associated with stem cells, numerous HOX genes, and mesenchymal stem cell–specific markers.
Luminal androgen receptor (LAR)		Characterized by androgen receptor (AR) signalling. Although this subtype is ER negative, gene ontologies richly express hormonally regulated pathways including steroid synthesis, porphyrin metabolism, and androgen/oestrogen metabolism.

2.8.2. EPIDEMIOLOGY

TN tumours including basal-like TN account for 10-20% of BC cases. TNBC has a higher incidence in young black, African, Indian and Hispanic women in comparison with other ethnic groups of young age. [42], [44], [46]–[49] The Carolina Breast Cancer Study (CBCS) reflect that the group at higher risk to present basal-like TNBC are premenopausal African–American women. In this group the tumour incidence comprises 27-47% of all the cases. On the contrary the group of postmenopausal non-African–American women is at a lower risk of basal-like TNBC, with only 14% of these tumours among all cases.[43] When age is considered as predictor of hormonal receptor status, women younger than 40 years are predicted as candidates for TN status.[50]

Among African American female population, TN basal-like cancer patients have an earlier relapse and overall survival in comparison with other subtypes. This tendency is increased if the patient is in a pre-menopausal stage.[14] The higher predominance of TNBC has been demonstrated in studies involving a dominant black population in African countries with BC incidence in young population. For example, Titloye et al reported that >80% of a total of 835 tumours belonged to patients younger than 50 years old. This study determined that the predominant subtype among Nigerian breast samples was the high-grade TN phenotype, including basal-like TNBC. In accordance with studies performed with other African population

samples, the TN phenotype proves to have an important incidence among young black women.[50]–[53]

The risk factor for TNBC have not been identified clearly due to the heterogeneity of the disease. However, Parity, full-term pregnancy at young age, suppressing lactation drug use, and interrupted breastfeeding after multiple births proved to be more influential as risk factors for basal-like breast cancer. While long breastfeeding on multiple pregnancies has been reported to reduce risk of basal-like breast cancer, but not luminal A. [14], [43], [54]

2.8.3. TNBC PROGNOSIS AND TREATMENT

Basal-like TNBC has been associated to poor prognosis by molecular characterisation and translational studies. This subtype showed high recurrence scores and expressed an activated wound response signature. This signature is related to a shortened overall survival and distant metastasis-free survival .[43], [55] TNBC is associated with a higher risk of central nervous system and viscera metastasis (mainly brain and lungs, 30% and 40%), in comparison with bone metastasis (40%) which is more common in non-TNBC. Additionally, lower survival rates during the 5 years after diagnosis are associated to TNBC in comparison with other BC subtypes, while the distant relapse possibility decrease after it.[47]

The lack of hormone receptors and HER2 targeted agents preclude the use of hormonal therapy as adjuvant therapy. Thereby adjuvant chemotherapy and neoadjuvant chemotherapy approaches are the treatment options offered to women with TNBC. It has been demonstrated that some TN tumours are highly sensitive to chemotherapy. As result they present excellent outcomes once there is no residual disease after the neoadjuvant chemotherapy treatment. The presence of residual disease after chemotherapy administration is associated with an unfavourable outcome. [44]

Recent research has suggested the association between TNBC and the BRCA1 gene mutation.[56], [57]The similitudes and relationship found with BRCA1 mutated cases arise the therapy possibilities. Patients with this germline mutation present a deficient break repair of double-stranded DNA. The use of platinum salts, antiangiogenic agents and DNA repair strategies such as Poly(adenosine diphosphate–ribose) polymerase (PARP) enzyme inhibitors for TNBC with BRCA1 dysfunctional pathways have demonstrate beneficial outcomes.[43], [44], [58] In order to overcome the TNBC drug resistance and find friendlier adjuvant therapies, the understanding of its heterogeneity is vital.

2.9 BREAST CANCER HETEROGENEITY

Breast cancer is a highly heterogeneous disease. There can be several variations among and within BC cases. Intratumour heterogeneity can occur not only at different parts of the tumour itself, but also as an evolution over time.[59] Breast tumours can present with different histological types, subtypes, grades, and at different stages. This diversity of factors has an impact on the progression risk and the resistance to treatment. [5]

The relevance of the histological characterisation can be considered limited if we take into consideration that more than 70% of BC cases are classified as invasive ductal carcinoma of no special type (NST) and present different clinical behaviour.[59] As mentioned in the

previous section, the determination of the tumour phenotype may be a better indicator for treatment selection.

Wider and deeper characterisation of BC tumours, using techniques such as molecular profiling and whole-genome sequencing, can provide translational knowledge to assist with BC management. BC characterisation offers advantages such as predicting responses to some therapies and further research on targeted therapies, but this is not a straight forward task. [5], [32], [33]

Molecular profiling of breast cancer samples has demonstrated that there are variations in the genetic composition of tumours. Further genetic characterisation can provide more information on BC growth rate, signalling pathways, cellular composition and proliferation rates.[31], [32] There are suitable treatments to destroy the cancer cells based on their molecular subtype. However, there is not consensus on the classification of the different molecular subtypes.

Cancer progression is a representation of treatment failure and is defined by the growing and spread of cancer after treatment.[60] Disease progression indicates phenotypic changes. These changes involve modification at the morphological, molecular and functional level. Several processes such as angiogenesis, inflammation, immune and hormonal responses take place during the cancer progression.[61], [62] All these processes can be tracked and research through the biochemical changes caused at cellular and tissue level.[63]–[65] Hereafter, if those biochemical changes can be identified a better understanding of the aggressiveness and metastatic potential of BC at different time-points can be achieved.[66]

The development and use of personalised and tailored therapies remain a must in cancer management. The use of "one-fits-all" approaches is impossible due to the high heterogeneity of the disease. The implications of the tumour heterogeneity at intra- and inter-tumour level are displayed in Table 8 which was adapted from Zardavas et al. [59]

Table 8. Clinical implications of tumour heterogeneity in breast cancer.

Type of heterogeneity	**Clinical implications**	**Potential solution**
Intertumour	Need for patient stratification	High-throughput molecular profiling technique Molecular classifiers
	Need for therapy selection/clinical development of targeted agents	Innovative trial designs: Master protocols Basket trials Adaptive trial design N-of-1 studies
Intratumour	Need to define the phenotype of the recurrent disease	Metastatic biopsy
	Molecular evolution of the disease	Repeated tumour biopsies Geographically separated biopsies Liquid biopsies
	Identification of driver events	Next-generation sequencing Bioinformatic tools and algorithms

		Systems biology Animal models/functional validation
	Identification of predictive biomarkers	Deep sequencing Single-cell sequencing
	Emergence of treatment resistance	Combination of targeted agents Exploiting passenger events

A Biomarker is defined as an objective, usually quantifiable characteristic of biological processes such as diseases. These markers work as reliable predictors of absence or presence of diseases. [67] Biomarkers are sought after to facilitate diagnosis, decide treatment, obtain prognostic factors, and predict cancer progression.[68], [69] Additionally, a finer classification will benefit those groups whose prognosis cannot be forecasted by conventional parameters.

The characterisation of the different subtypes can lead to a better understanding of the pathways responsible for the outcome, development, spread patterns, and metastases. The diagnoses, treatment, and development of drugs and adjuvant regimens will be facilitated and simplified. In addition, the discovery of new markers can help to develop new detection techniques and support potential drug discovery.

2.10 Vibrational spectroscopy

Vibrational spectroscopy comprises techniques where samples are exposed to electromagnetic radiation causing the stable vibrational states of the matter to change to an excited state. In order for this to occur the quantum radiation (energy applied) must be higher or equal to the excitation energy.[70] The chemical nature and structure of the analysed specimen, in addition to the applied energy is strongly related to the excited states produced in the compound.[71] Therefore vibrational spectroscopy allows identifying the chemical composition of a wide range of materials and tissues.

Owing to the high detail in the spectra obtained with vibrational spectroscopy, biochemical changes can be detected in tissues. These changes can be related to diseases and can be easily detected with these characterisation techniques. Vibrational spectroscopy has demonstrated potential to provide diagnostic information in addition to facilitate the prediction of biochemical progression on different diseases in a rapid non-destructive manner.[70], [72]

Mid-infrared FTIR and Raman spectroscopy (RS) are the most popular techniques in this area and will be revised briefly in this section.

2.10.1 Raman spectroscopy (RS)

When electromagnetic energy in the near-IR, visible or near-UV range is applied to a sample 2 types of scattering are expected: elastic or Rayleigh and inelastic or Raman. As its name suggest the elastic scattering maintains the same frequency as the incident photon. On the contrary a loss or gain in energy in comparison with the incident photon occurs in the Raman scattering.[70] RS bases its fundament on inelastic scattering and polarizability changes. The polarizability change is produced by vibrations when the molecules are irradiated.[73]

When light interacts with the electron cloud and bonds of the analysed specimen an excitation of the vibrational states occurs. Once the molecule relaxes it emits a photon. The photon's frequency is shifted in comparison with the excitation wavelength owing to the change in state. If the scattered photon ends at a less energetic state it is known as Stokes shift or red-shifted and has a lower frequency. If the photon finish at a more energetic state during the scattering the shift it is known as anti-Stokes or blue-shift and has a higher frequency. A representation of this it is shown Figure 7. The energy needed to promote the change in the vibrational state is the difference between the incident and scattered photon. The difference in frequency between the incident and the Raman-scattered lights is known as Raman shift. This shift is specific for each individual molecule. Owing to this specificity Raman spectroscopy allows the characterization of the chemical composition through the identification of chemical bonds and functional groups. [70]

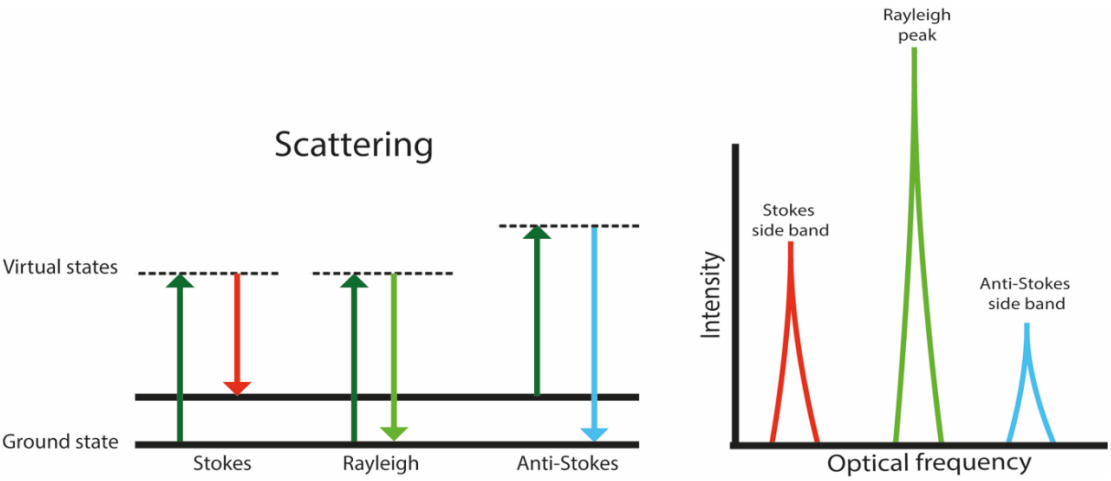

Figure 7. Types of light scattering in terms of energy; maintained (Raleigh), loss due vibration (Stokes), or gained by vibration (anti-Stokes), and intensity versus frequency.

Additionally, Raman spectroscopy can provide information about the quality, quantity of certain species and some structural properties such as orientation and symmetry.[74]

Raman spectroscopy has several advantages for use in biological analysis. RS is considered less invasive and the sample preparation is minimum or not required, avoiding the use of reagents. In vivo and ex vivo analysis is possible when special accessories such as external probes are used. In addition, the depth and spatial high resolution offered by this spectroscopy technique provide accurate qualitative and quantitative results.[74]

The analysis of small volumes of sample has been facilitated through the merge of Raman spectroscopy and confocal microscopy. A diagram and brief functioning explanation are shown in Figure 8.

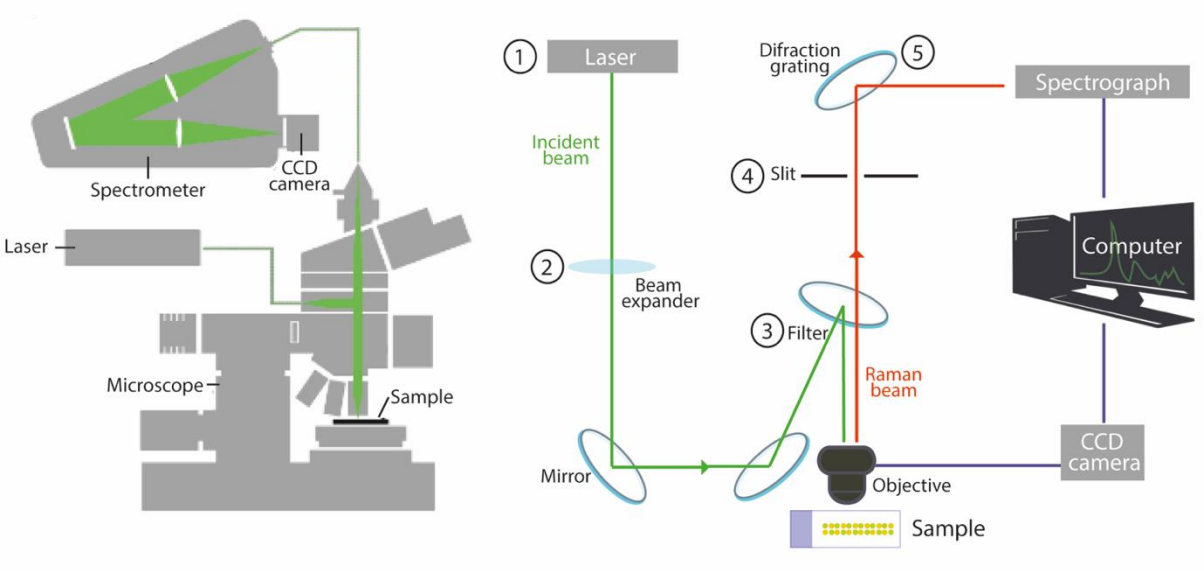

Figure 8. Schematic representation of confocal Raman microscope. A representation of the incident beam is coming from the laser (1) and being expanded (2). After this the laser goes through mirrors and optical arrangements to pass through a filter (3) before arriving to the objective and interacting with the sample. The inelastic Raman scattering is collected though the objective and passed through the filter. Once there is sent through a slit to the grating (5) before arriving to the spectrograph. Once in the spectrograph the beam is imaged.

2.10.2 FTIR SPECTROSCOPY

Nowadays infrared spectroscopy is the most widely spread spectroscopy technique. Its popularity is attributed to its outstanding capacity to provide qualitative and quantitative information regarding chemical analysis. FTIR spectroscopy offers extensive sample techniques possibilities, rapid analysis, and its fast running speed.[71]

There are 3 regions of infrared (IR), the near infrared (NIR), the mid infrared (MIR) and the far infrared (FIR). The wavelengths and wavenumbers covered in each region are showed in Figure 9.

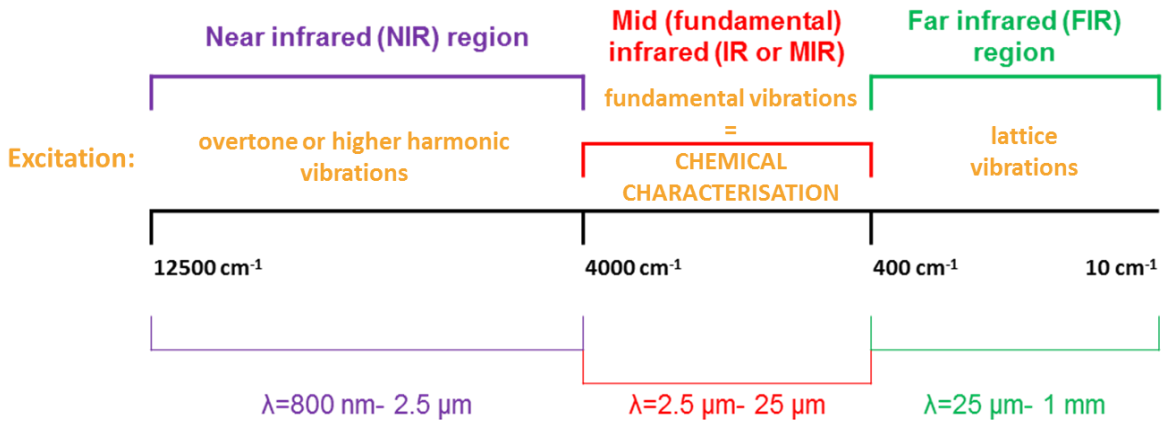

Figure 9. Infrared regions and excitations caused by them.

When a sample is exposed to infrared radiation excitation in the vibrational modes occurs. The molecular vibration causes a change in the dipole moment of the molecules. When this change occurs the absorption of energy is obtained and processed. This absorption is dependant of the coincidence between the molecular vibrations and the frequencies of the IR radiation.[73]

As this technique is dependent on the dipole moment changes, it is imperative that dipole conditions exist in the sample to be IR active.[70]

Fourier transform infrared (FTIR) spectroscopy is a type of IR spectroscopy in which a mathematical process involving Fourier Transform is used for data processing. FTIR spectroscopy generates an interferogram as result of the measure of the radiation absorbed with and without the sample. The infrared source produces the radiation, passing through the interferometer to interact with the sample and then being detected. The detector identifies the amount of energy absorbed, then this is amplified, filtered and processed using Fourier Transform.[75] The infrared spectrum is a plot of radiant absorbance or transmittance versus frequency (cm^{-1}). Figure 10 shows the components of the FTIR Instrumentation.

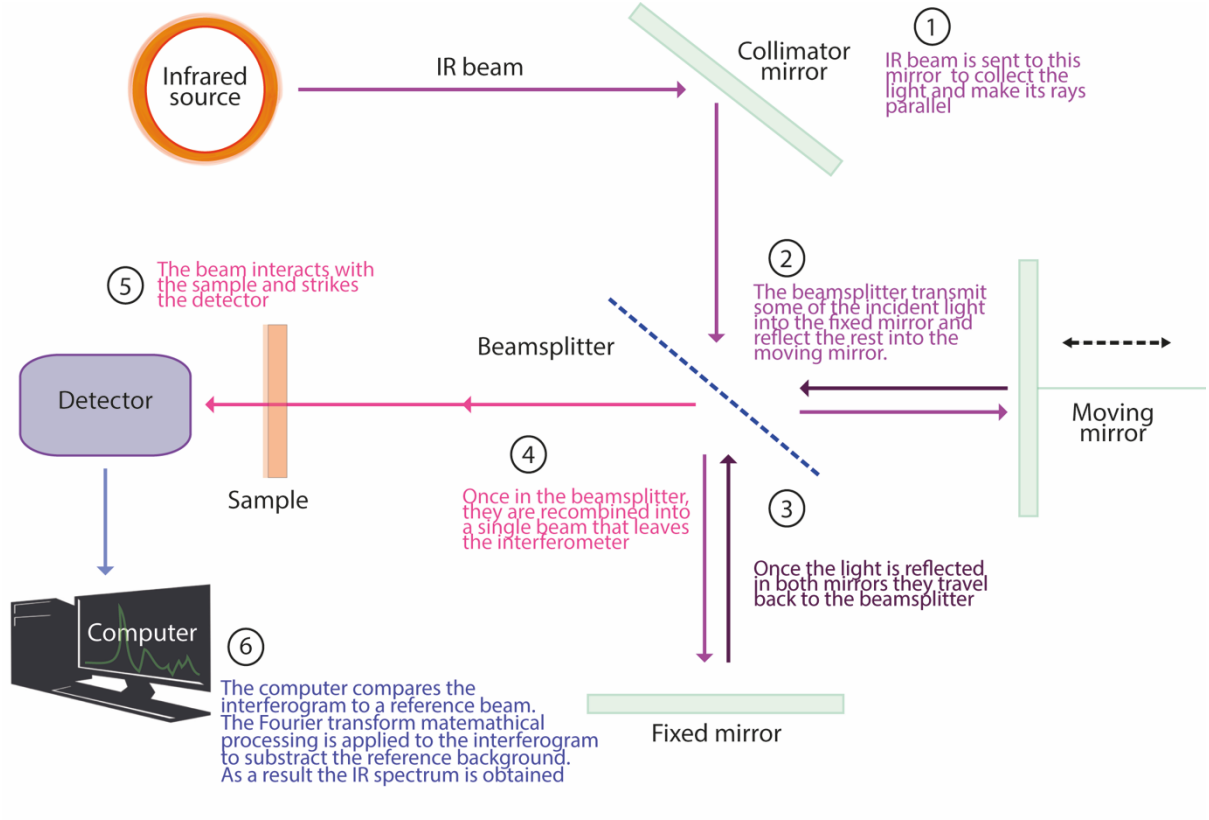

Figure 10. Schematic representation of FTIR instrument.

2.10.3 FTIR and Raman spectroscopy comparison

Raman and FTIR are complementary techniques. As both arise from different phenomena, they provide different information. Both techniques have demonstrated to be suitable to analyse samples in different states, and to facilitate quantitative analysis and sample

fingerprint elucidation. A comparison of both techniques it is shown in Table 9 which has been adapted from Larkin et al, 2011 and Rehman et al, 2013. [70], [73]

Table 9. FTIR and Raman spectroscopy comparison.

Features	Raman	FTIR
Physical effect	Scattering Vibrational state relaxation/ Polarisation change.	Absorption Radiation frequency/molecular dipole moment change.
Sample preparation	Minimum or not required	Required in majority of cases
Problems	Fluorescence can be an issue in this technique.	Samples without remaining water are needed owing to its strong signal
Materials	Wide range of synthetic and natural materials in different states (liquid, gas, solids)	Natural and synthetic dry materials, excluding non-aqueous samples.
Resolution	Varies depending on laser, usually around 0.05-8 µm	Varies with beam splitters, usually around 1-20 µm
Frequency range	4000-50 cm^{-1}	Mid-IR: 4000-400 cm^{-1}
Best vibrations	Best at symmetric vibrations of non-polar groups. Strong for covalent bonds	Best at asymmetric vibrations of polar groups. Strong for ionic bonds

A brief description of some techniques and accessories related to these techniques can be found in Table 10.[76]

Table 10. FTIR and Raman spectroscopy most popular approaches.

Sampling approach	**Feature**
Fourier transform infrared spectroscopy in the mid-IR	
Transmission and reflection based	Sample is analysed onto a ZnSe or Si sample carrier.
Attenuated total reflectance (ATR)	Sample is in contact with an ATR crystal. It provides accurate information about the chemical composition of the sample (in contact with the crystal).
Imaging	The equipment is attached to a microscope system. It allows mapping analysis of areas, points and lines.
Raman spectroscopy	

Normal/ Imaging	Laser power must be selected according to needs and sample. Visible lasers can present problems related to fluorescence. Imaging allows mapping and chemogram obtaining.
Resonance enhanced	Enhancement of signal is based on the resonance. Excitation of this type is in resonance with the electronic transition. Achieve 10^4 signal enhancement over normal Raman scattering
Surface enhanced	Requires using metal surfaces/colloidal solutions/electrodes. Enhancement is achieved by electromagnetic enhancement effect and chemical enhancement (which is a charge transfer mechanism). In addition to enhance the signal by 10^3-10^6, it has a fluorescence quenching effect.

2.11 Use of vibrational spectroscopy for BC analysis

The use of vibrational spectroscopy allows the characterisation at biological level content. Biological characterisation provides further information about cell biology processes such as cell cycle, differentiation, transcriptional nature, and apoptosis.

Breast cancer models had been studied with Raman spectroscopy. A vast number of studies have focused on describing the spectral fingerprint of breast tissue. Lipids and proteins have been identified as the main differentiation between normal and cancerous breast tissue. A major presence of certain lipids such as fatty acids like triolein and oleic acid characterise the normal breast while a dominant presence of protein has been reported in cancerous tissue. The protein plays an important role not only in cancerous tumours but also in other types of lesions such as fibrosis and benign neoplasms. [74], [77], [86]–[90], [78]–[85]

Grading of breast cancer samples has been attempted successfully using RS. Rehman et al reported changes in important biomolecules, such as, nucleic acids, proteins and fatty acids. Changes in intensities and peak shifts were observed helping to determine different grades. High nuclear grade (HNG) DCIS was characterised by a higher content of lipids represented by fatty acids and acylglyceride peaks, while low nuclear grade (LNG) DCIS samples were protein-rich [89]. These results support the presence of this lipids as energy reservoirs. [91]–[94]. In addition, the lipid contain has proved to have an effect on the membrane fluidity and therefore on the preparation of the ECM. When the ECM was analysed with Raman spectroscopy the surrounding connective tissue presented peaks suggesting an overproduction of proteins. Paidi et al reported the preparation of the ECM based on changes associated with collagen and proteoglycan components. [95], [96]

RS allowed disclosing biochemical variations between cancerous and normal cell lines by using Chemometric analysis. Talari et al analysed two cancerous cell lines (ER+/PR+ and TN) and one normal cell line. This study characterised the chemical differences among cell lines using a Raman confocal microscope system. Cell lines differentiation was achieved with statistical analysis. Disease status related to lipid metabolism was portrayed based on the different lipid vibrations intensities. An increase in the fatty acids in the cancerous cells suggested cancer

progression. Content increment was represented by the corresponding peak intensity increasing. In addition, this intensity increment was a sign of rapid cell proliferation due to low lipid degradation (caused by lipid peroxidation). Amide groups are structural-functional groups of proteins and nucleic acids; therefore, the amide content can provide a good insight into the protein and nucleic acids levels. High levels of amides represent a hallmark of cancer cells. Principal component analysis of the finger print region suggested a large variation among both cancer cell lines. The variation was consistent with the heterogeneity of the disease. Therefore, narrower spectral regions were evaluated to identify components that allowed differentiation. As result of this, the concentration of proteins and amino acids was suggested to be the main difference between malignant and normal cells in agreement with the lack of significant difference in the lipid concentration between the triple negative subtype and the normal cell line. These results were used to stablish a predictive model using linear discriminant analysis which yielded 100% sensitivity and 91% specificity. Talari results confirmed that RS has the capacity not only of differentiate among cancer and normal cells, but also among BC subtypes.[74]

The analysis of biopsies using Raman spectroscopy has been exploited due to the enhanced resonance of phospholipids and carotene derivatives caused by the use of visible lasers. The contribution of these components decreases when carcinoma samples are analysed allowing its identification.[78] Haka et al. used Raman spectroscopy to distinguish among benign and cancerous breast lesions. This research group used a set of ex vivo samples from 58 subjects. A total of 130 spectra were taken with an optical fibre Raman probe system. The diagnosis was based on results obtained with the Shafer-Peltier et al spectroscopic model of breast tissue.[97] This model uses basis spectra of morphological structures expected in the breast tissue to compare with the spectra obtained from the breast tissue sample. The basis spectra include cell nucleus, collagen, epithelial cell cytoplasm, fat, β-carotene, calcium hydroxyapatite, calcium oxalate dihydrate, cholesterol-like lipid deposits, and water. Based on the relation between signal intensity and chemical concentration fit coefficients were obtained for each sample. The basis spectra have a value equal to 1 representing the 100%, while the sample will have a value >1 for the same component. This fit coefficient will represent the contribution of that specific substance or structure in the sample. The obtained coefficients were used to stablish the lesion chemical makeup. By using maximum likehood estimation the sample probability of representing a benign or malignant lesion was determined. The fit coefficients of this study for different lesions allowed identifying some chemical changes related to different breast pathologies. The results of this research indicate that normal breast is mainly composed of fat while an increase on the collagen content was found in all abnormal breast samples. The collagen increment is consistent with the proliferation of fibroblast and collagen overproduction caused during scaring in breast pathologies. In infiltrating carcinoma, the collagen increment is a consequence of the invasion of stromal tissue by cancerous cells which promote the proliferation of fibroblast. In cancerous and benign tumours, the fit coefficient corresponding to cell nucleus and epithelial cell cytoplasm increased, in concordance with the neoplasm trademark of nuclei enlargement and abnormal cell proliferation. The comparison between the nuclear-to-cytoplasm ratios reflected that the ratio corresponding to the carcinoma was significantly higher differentiating cancer from benign neoplasm. This study suggests that precancerous lesions can be diagnosed

by considering the intermediate nuclear-to-cytoplasm ratio. However, this statement might not be strictly true as low grade cancers will also present a low ratio. Another significant component that allowed the nature differentiation among tumours was the fat. As the nature of the tumour become malignant the infiltration of cancerous cells occurs in between fat cells. On the contrary, the benign tumour pushes the fat tissue aside to expand the lesion. By cross-validating the results of this study with the pathologist diagnosis, the use of this Raman spectroscopic model had 86% accuracy, a specificity of 96% and a sensitivity of 94%. These results indicated the potential of this technique to be incorporate to clinical diagnosis after further optimization.[98] A few years later the same group used a new set of samples to validate the model. In this occasion a specificity of 93% and a sensitivity of 84% were achieved, therefore the potential of Raman spectroscopy for diagnoses was confirmed.[86]

Haka research has suggested that applying the principles of their proposed diagnosis model the diagnosis of BC can be performed clinically by incorporating optical fibres into biopsy needles or even to develop a transdermal technology. This suggestion is based on the fact that Raman spectroscopy can provide information to an immediate diagnosis of cancer. It is important to mention that the weakness of the signal in comparison with other techniques such as fluorescence spectroscopy and the use of safe laser powers must be overcome by a careful and appropriate system design.[86], [98]

When fit coefficients are used in spectroscopic models, several factors must be taken into consideration. For example, the affectation of chemotherapy drugs and re-excision surgery, that leaded to false negatives in Haka et al validation research. The use of fresh or frozen tissue has a minimum change in the density-scattering relation. However, it needs to be optimized. [86]

Raman spectroscopy has demonstrated to be a powerful technique to characterise the chemical progression of BC. Oncogenic cellular senesce and proliferation has been characterised by identifying changes in the configuration of unsaturated lipid isomers (-cis and –Trans respectively).[99] As senescence/proliferation is a key process for cancer development this information can improve and support the diagnosis process.

Intensive research in breast oncology has been perform employing infrared spectroscopy. For example, different breast pathologies such as hyperplasia, fibroadenoma and carcinoma were analysed by Eckel et al using FTIR spectroscopy. The amide region was identified as differentiator for these three diseases and normal breast. Shifts to lower frequencies in the 3000-3600 cm^{-1} region was characteristic of the malignant samples. This fact has been backed up by other groups.[100]–[102] In addition, in the cancerous tissue the α-helix amide I frequency increased while the β-sheet for the same amide band increased.[102] The FTIR study of the secondary structure and conformational changes of proteins might offer indication on the progression from normal to cancerous epithelial tissue.[103], [104] The fibroadenoma was differentiated by the significantly higher absorbance ratio for the bands at 3300 and 3075 cm^{-1}. The A_{1680}/A_{1657} absorbance ratio allowed to establish a trend based on intensities where carcinoma, fibroadenoma, hyperplasia and normal breast were decreasing in intensity as listed. The content of collagen determined using the A_{1657}/A_{1204} and A_{1657}/A_{1278} absorbance ratios increased in intensity as listed in normal, hyperplasia, carcinoma and

fibroadenoma.[102]

Ling et al found a decrease in the intensity of the symmetric and antisymmetric vibrations of the $-CH_2$, $(-CH2)_n$, and C=O vibration in the cancerous samples in comparison with normal breast. Moreover, a shift of the P=O or P-O-C bands towards lower wavenumbers was identified in the cancerous sample.[100]

Several useful ratios had been used to differentiate benign and cancerous lesions. Some examples are the following presented a significant difference ($P < 0.05$): for proteins and its structure $A_{1640/}$ A_{1550} and A_{1160}/A_{1120}; for lipid and protein content A_{1640}/A_{1460} and A_{1550}/A_{1460}; for lipid structureA_{1460}/A_{1400}; and finally for nucleic acids content A_{1310}/A_{1240}.[101]The methylation ratio determined by A_{2958}/A_{2853} has also been used as a carcinoma biomarker. This ratio increases when the cancerous epithelia are analysed. On the contrary, a decrease on the A_{2958}/A_{2853} ratio is found when the stromal component of cancerous samples is analysed and compared with the connective tissue of normal breast samples. Hypomethylation can act as an indicator of cancerous processes. This methylation behaviour can be explained by hypomethylation of DNA, an increase of malignant epithelial cells and the decrease of protein fibres in cancerous cases.[105] Also, the area under the band at 2925 cm^{-1} has been reported as a biomarker with a sensitivity and specificity of 100 % when normal and cancerous samples are being compared.[104]

Analysis of BC with FTIR had been carried out with accessories such as Attenuated total reflectance (ATR). Rehman et al characterised DCIS and IDC processed breast samples with different grades. This study confirms the considerable contribution of DNA, lipids, proteins and fatty acids into the FTIR spectrum. Spectra obtained in this study revealed difference in positioning and intensity of the peaks suggesting biochemical and compositional changes among the cancerous and normal tissues. Different spectral ranges able to identify the grade were clearly identified in this research. For example, 1,250 to 960 cm^{-1} which represent changes in collagen, glycogen, ribose, phosphodiester and DNA. The intensity at different peaks such as 1030 cm^{-1} also serves as an indicative of grade. Higher grades had an overall increased intensity in comparison with low grade samples. In addition to this, the appearance of new peaks such as at 1055 cm^{-1} as the grade increase represent spectral evidence of identification among different types and grades by using FTIR spectroscopy. An interest highlight of this study is the suggestion of the validation of the technique and its efficacy to support diagnosis disregarding samples age.[106]

The BC microenvironment has also been studied with FTIR. Kumar et al focused their analysis on the extra cellular matrix (ECM) owing to its importance as supporter for growth, progression, invasion and metastasis of the tumour. In this study section of BC, samples were analysed by FTIR imaging considering the cancer area, and the collagen close to- and far -from the cancer tumour. As major findings three bands were identified as markers, the band at 1653 cm^{-1} with an increased intensity in the epithelial cell in comparison with its content in both ECM-areas and two bands at 1640 and 1630 cm^{-1} which ratio suggests a progression in the collagen as it goes away from the cancerous area. These results propose the use of these bands to identify cancer-induced changes in the collagen/stroma component.[107]

Devi et al analysed frozen /dried samples of different breast pathologies including BC with FTIR. The following bands were used as markers as they differentiated among benign and malignant tissue, 538, 853, 935, 1005, and 1080 cm^{-1}. In agreement with the literature a higher content of collagen was found in cancerous samples. Furthermore, once more differentiation of benign and malignant lesion was confirmed.[102], [108], [109]

Vibrational spectroscopy has been used extensively to characterise and study different process related to BC, a brief summary of relevant findings in this field involving BC and vibrational spectroscopy can be found in Table 11.

Table 11. Summary of Breast cancer findings using Raman and FTIR spectroscopy.

Findings	**Technique**	**Reference**
Malignant tissue spectra have a lack of proteins peaks	FTIR-Raman	[110]
Shifts related to malignancy development	Near-IR Raman	[78]
Lipid vs protein dominance in normal and abnormal tissue	Near-IR Raman	[80]
Senescent/proliferating state of BC cells by configuration in lipids	Micro-Raman	[99]
Calcification identification	Near-IR Raman	[111]
Cancer dominated by arachidonic acid metabolism products vs monounsaturated oleic acid and derivatives from normal tissue	Raman imaging	[112]
Feasibility of Raman spectroscopy for in vivo intraoperative margin assessment	Optical fibre Raman probe	[113]
Suggest relevant peak intensity ratios of serum spectra for breast cancer diagnosis	Raman imaging	[114]
Non-invasively detection and identification of calcifications at different deepness	Spatially offset Raman spectroscopy	[115]
Malignant samples have excess of lipids whereas benign present excess of proteins	Near-IR Raman	[90]
Biochemical assessment of heterogeneous lymph node features	Raman imaging	[116]

Higher DNA duplication activities in tumorigenic cell nuclei in comparison to normal cells.	Confocal Raman microspectroscopy	[117]
Proliferation biochemical changes in mammalian cells	Optical fibre Raman probe/ FTIR	[118], [119]
Spectroscopic detection of early neoplastic changes in animal model	Confocal Raman microspectroscopy	[84]
Spectral range to identify differences among different DCIS and IDC grades	ATR FTIR/ Micro-Raman	[89], [106]
Protein bands patterns according to breast pathology	FTIR	[102]
Breast cell cancer growth/cell cycle phases	ATR FTIR	[120]
Cancer-induced modifications in collagen	FTIR imaging	[107]
Benign vs cancerous characteristic peaks	FTIR	[108]
Global effects of miRNAs in MCF7 breast cancer cells	ATR FTIR	[121]
Fibroblast behaviour under cancer stimuli	FTIR Imaging	[122]
BC samples analysis and formalin-fixation and Paraffin-embedding effect on spectra	FTIR Imaging	[123]
Spectral differentiation among different BC stages	Fibre optic evanescent wave Fourier transform IR (FEW-FTIR)	[124]
Ensembles for different types and stages of breast cancer	Mid-FTIR microspectroscopy	[125]
Sentinel lymph node (SLN) cancer metastases analysis	FTIR	[126]

Specific absorption peaks for breast cancer tissues	Synchrotron radiation based FTIR (SR–FTIR) absorption spectroscopy	[127]
Spectral characterization of whole tissue sections including cancer and its microenvironment	FTIR Imaging	[109]
ECM spectral study in 3D BC models	FTIR Imaging	[128]

2.12 Chemometrics/ Multivariate data analysis

The term Chemometrics is defined by the application of mathematical and statistical analysis to chemical data to extract and obtain useful information efficiently. Usually the large number of data requires the use of computationally intensive methods.[129] Due to its analytical nature several components in one single sample can be detected by using spectroscopic techniques. When chemical data is analysed there are several variables to be considered. Approaches such as Consider One Separate variable at a Time (COST) are not an efficient way to process the data to obtain accurate information. The data managing in a multivariate form will always be preferred.[130]

Analytical characterisation techniques such as spectroscopy measure several variables, and as a consequence yields multivariate data. Once this data is processed using different statistic methods different approaches can be achieved. For example discrimination or classification of the obtained data in order to obtain differences among samples or finding similarities to help its classification.[131], [132]

Data reduction is the aim of using computational multivariate analysis in spectroscopic characterisation. This aim is intended due to the great number of values obtained, such as thousands of intensity values.[131]

2.12.1 Unsupervised algorithms (PCA & CA)

These types of algorithms are used when the process of data exploration is done without prior expectations. The result of this process is the formation of clusters and gradients providing information on similitudes and differences.[132] Principal component analysis (PCA) and Cluster analysis (CA) are examples of this type of algorithms as they do not know the groups prior to the analysis.[131]

PCA is a multivariate analysis in which a set of related variables are reduced into a smaller set of uncorrelated variables.[133] This algorithm is used when correlation exist between the variables, allowing to reduce the amount of data. As its name indicates the PCA aim is to find principal components. This is achieved by obtaining a principal component (PC) value as function of the given variables linear combination and coefficients used to avoid correlation between the new variables. PC1 shows the greater variation among the data set, followed by PC2, PC3, etc. On the contrary, as the PC score increases the variation decreases. The PCs result from the covariance matrix, which is a matrix obtained from the measures of the joint

variations. Nevertheless, the correlation matrix can also be used for this purpose, having the zero mean variable standardization and unit variance as consequence. The need for standardization occurs when different scales are used or when there is a big difference between the variance and the domination of a PC will be compromised. The standardization will ensure that equal weight would be carried by all variables allowing their comparison. Once the PCA data processing has been done, the number of groups identified can be used to explain the chemical structure difference among the analysed samples. [131]

PCA not always provide a good separation between groups; in those cases, other methods such as cluster analysis can be used to search for specific groups. Cluster analysis divides the specimens into groups of the same class. This grouping is made based on the closeness in the variable space based on the Euclidean distance, which is known as the distance between two points in an n-dimensional space. The standardizing of the data can also be used for cluster analysis, and has the same effect than in PCA, the use of a common scale without one variable domination.[131]

The clustering process can be represented in a dendrogram illustrating the stages of clustering, plotting the observations against the distance or the similarity. The similarity is considered a function of the Euclidian distance and the maximum distance.[131]

2.12.2 Supervised algorithms (LDA)

These algorithms are used when a previous classification is known. Linear discriminant analysis (LDA) is an example of supervised algorithms and is considered a dimension reduction method. To distinguish among the different classes a unique label or output value is assigned to each of them. After this step, known as training stage, the data is better known as labelled observations. This algorithm has as goals to discriminate the data by using classification rules (classifiers) that will intend to separate the data to its maximum capacity, and to classify new unlabelled data using previous processes.[134], [135] Due to class prediction is the aim of this technique and certain rules are defined prior analysis these algorithms are known as supervised and are used for validation of chemometric models.

The optimization of LDA allows to reduce the distance between different classes and to reduce it between components in the same class.[135] Through this method and some variations the classification of data corresponding to microarrays, and several spectroscopic techniques analysing different tissues has been done successfully.[136]–[139]

2.13 Tissue Microarrays (TMA)

Tissue microarrays (TMAs) have been integrated in cancer research due to their many advantages and effective use of limited tissue. TMAS are blocks that contain many paraffin-embedded tissue core samples. Owing to the large number of samples contained in the TMA, a rapid, efficient and cost-effective analysis can be performed.[140] This methodology allows to analyse hundreds to thousands cases in single sets.

TMAs can be arranged by clinical information association and have demonstrated to be a valid technique to confirm clinic-pathologic correlations in comparison with whole tissue sections.

This kind of arrangement allows analysing samples of tumour biopsies at different stages, and has been also used to validate the expression of protein biomarkers.[141]

TMA analysis of different carcinomas has been done in the last years validating this technique for cancer diagnosis and research.[142]–[144] The advantages and limitations of this technique are shown in Table 12.

Table 12. Tissue microarrays (TMA) advantages and limitations for tissue analysis.

Advantages	Disadvantages
Small cores are representative of whole tissue sections (WTS).[140], [141]	Issues related to the size, number, spacing and layout of cores.[140]
Time and cost effective. Allows resource-efficient use.[140], [145]	Trained and experienced histotechnologists are required to prepare and analyse the samples in order to include relevant lesions.[140]
Efficient and reliable platform for sub-classifying breast cancers into relevant subsets.[141]	Some tissues (breast, colon, and skin) are resistant to being punched, which require extra careful preparation.[146]
Can reproduce clinico-pathologic correlations shown by standard sections and controls.[141], [145]	There is not standardization on preparation methods (eg. storing and transferring temperatures). [146]
High throughput profiling.[140]	Duplicate cores are needed to achieve valid results representative of WTS.[143]
Multiple analyses of large number of samples.[140], [145]	Other tissues for orientation purposes are recommended.[144]
Improvement of reproducibility, speed and reliability of interpretation as consequence of having all the samples in one slide.[145]	
Comparison of adjacent sections by staining of consecutive slide.[145]	
Virtually all tissues are suitable for TMA.[144]	

Materials and methods

CHAPTER 3. MATERIALS AND METHODS

This study was performed using tissue microarrays (TMAs) containing a large number of breast cancer samples and normal breast tissue. The cancerous cohort corresponds to breast samples from Nigerian patients. The samples in use were obtained under the Ethical approval REC No. 06/ Q1206/180. An example of the TMA aspect is displayed in Figure 11.

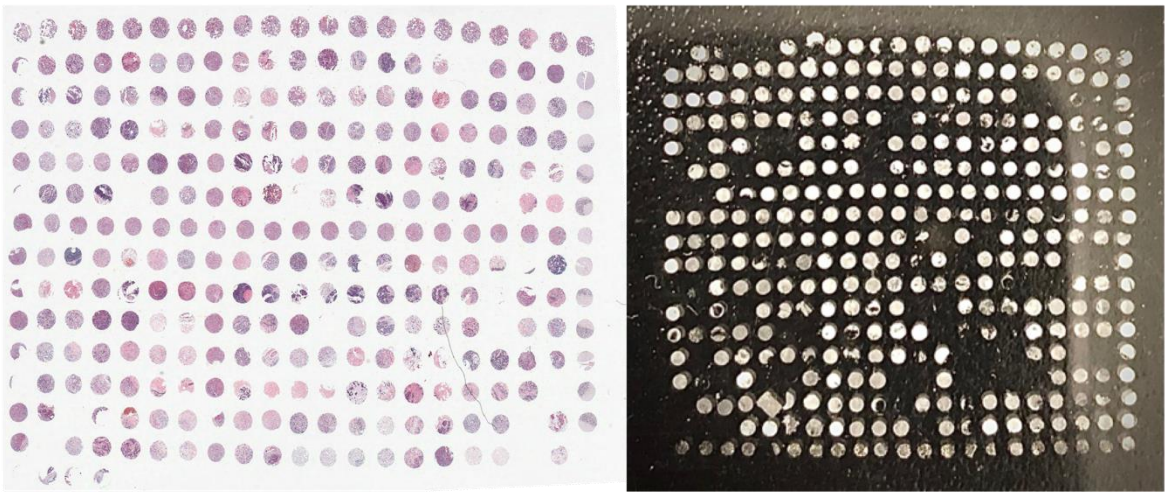

Figure 11. TMA appearance, H&E stained TMA (left) and dewaxed TMA for spectral analysis (right).

The TMAs include an orderly arrangement of duplicate circular cores. These samples were removed from hundreds of different breast tumours, normal breast tissue. Other tissues were used as controls and for orientation. After selecting and removing the punched samples from donor blocks, they were transferred into one empty recipient paraffin blocks as described before.[50] Blocks were then sectioned on glass, calcium fluoride slides and MirrIR slides. The TMAs were provided by Dr Abeer Shaaban, Consultant pathologist at University Hospitals Birmingham NHS Foundation Trust.

3.1 TMA ASSESSMENT

Six TMAs containing breast cancer samples and one TMA containing normal breast sections were received in duplicates. Theses samples were received with the following information: histological tumour type, grade, ER and HER2 status, and patient sex and age at diagnosis.

The TMAs nomenclature and intended number of samples and orientation cores in each slide are shown in Table 13. All the TMAs in this study were received in glass slides. TMA´s A2, B2, C2 and normal were additionally received in a second substrate. The additional TMA NGB A2 and normal were placed in calcium fluoride slides. Calcium fluoride is transparent in FTIR and therefore this set was used for this analytical purpose. On the other hand, the additional NGB B2 and C2 TMA´s were placed in MirrIR slides. A MirrIR slide (Kevley Technologies, Chesterland, Ohio, USA) is a glass slide coated with a thin Ag/SnO2 layer working as a reflective substrate which is inert and transparent to visible light. When working on infrared reflection

mode this substrate reflects almost completely making it perfect for tissue analysis with FTIR spectroscopy.

Table 13. Tissue microarrays conforming the study cohort. Shaded TMA´s were received in two different substrates and were included in both spectroscopic analyses.

TMA		Substrate	Spectroscopic analysis	Total of Cores	Samples of Interest	Orientation Cores
Breast cancer	NGB A2	Glass	Raman	378	252	126
		CaF_2	FTIR			
	NGB B2	Glass	Raman	378	252	126
		MirrIR	FTIR			
	NGB C2	Glass	Raman	378	252	126
		MirrIR	FTIR			
	NGB D2	Glass	Raman	315	188	127
	NGB F2	Glass	Raman	273	154	119
	NGB G2	Glass	Raman	336	202	134
Control	Normal breast	Glass	Raman	376	268	108
		CaF_2	FTIR			

The samples cores have 5µm thickness and 0.6 mm diameter. The samples were taken from different tumours from different patients. Different human and bovine tissues were used to help with the orientation. An orientation grid example can be found in Figure 12. This grid in combination with images of the duplicated slides were used to ensure the proper orientation and identification of samples.

In all the slides, duplicates of each tumour were placed next to each other. Duplicated slides representing the same sections were used for this study; one slide was used for the spectroscopic analysis after dewaxing and the other one for H&E staining. The staining was done to identify the cancerous and normal epithelia. Figure 13 shows a schematic representation of the use of the duplicate sets.

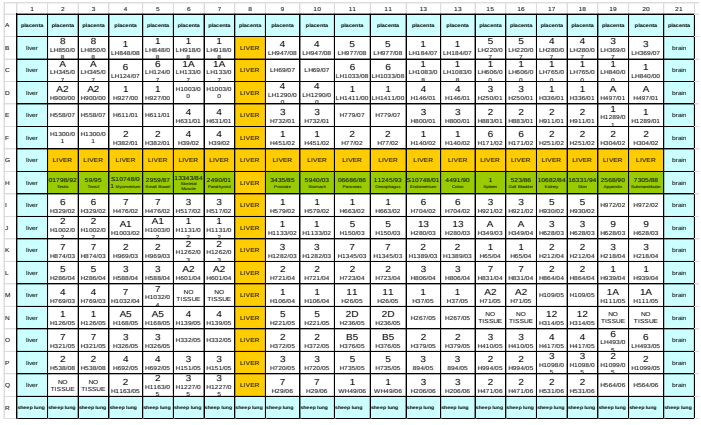

Figure 12. Orientation grid example. Highlighted cells indicate orientation cores and white cells represent breast samples.

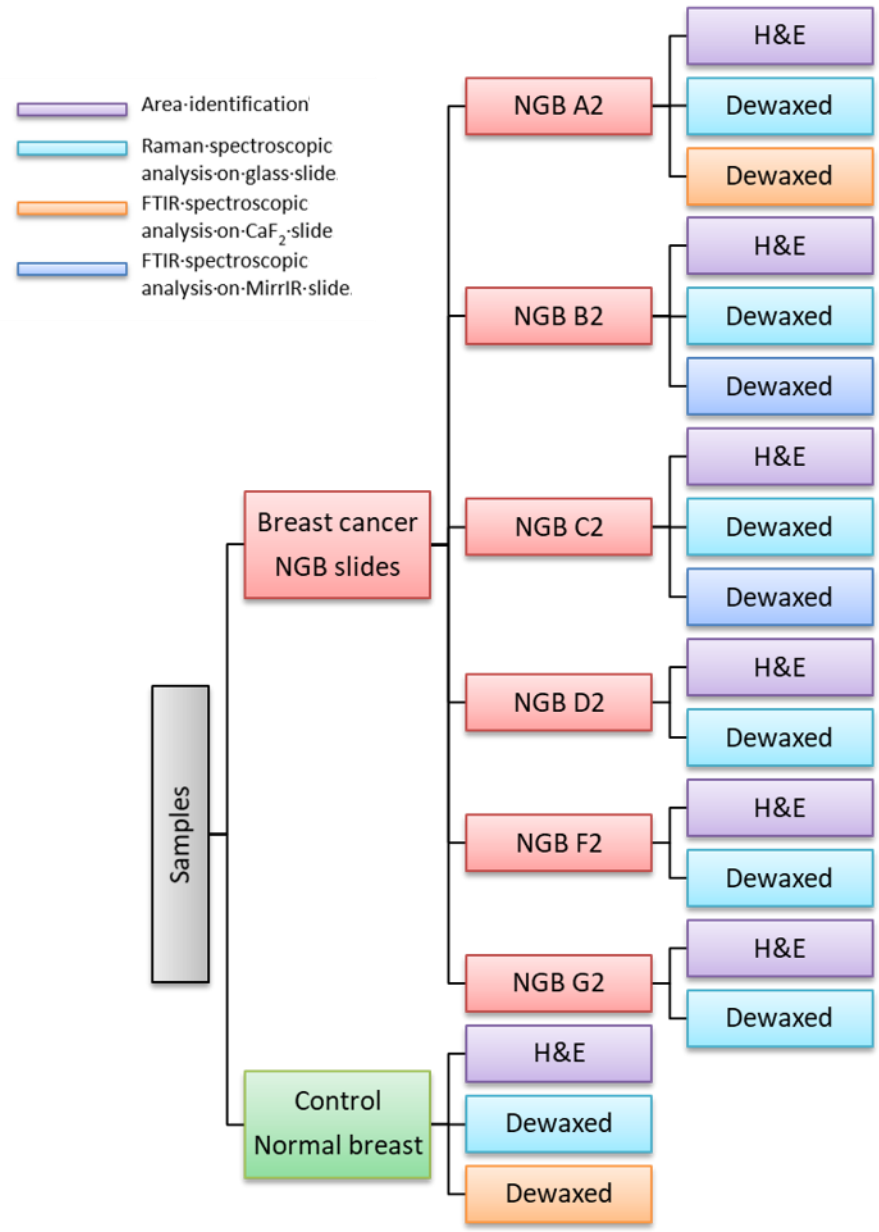

Figure 13. Use of duplicate sets for spectroscopic analysis and area identification.

3.2 Samples information

Samples from tumours and normal sections were used for spectroscopic analysis. A total of 578 samples were analysed with RS (499 breast cancer tumours and 79 normal breast sections) while a total of 273 samples (245 breast cancer tumours and 28 normal breast samples) were analysed with FTIR. Seven TMA's were analysed with RS, while four TMA's were analysed with FTIR. The difference in samples analysed is due to tissue availability in FTIR transparent substrates.

.

Ductal NST (no-special-type) carcinoma was the dominant histological subtype representing 87.3% of the cancerous cases. The mean patient´s age at diagnosis was 48.62 ± 12.41 years. The clinicopathological features of the Nigerian patients and the normal breast samples are presented in Table 14.

Table 14. Clinicopathological features of the studied cancerous and normal tissue. Values expressed as n (%), unless otherwise indicated.

Mean age, years	48.62 (SD 12.41)		
Mean tumour size	63.33 mm (SD 1.24)		
Raman (n=578)		**FTIR (n=273)**	
Cancerous samples	499 (86.33%)	**Cancerous samples**	245 (89.74%)
Normal breast	79 (13.66%)	**Normal breast**	28 (10.25%)
Histological grade	395	**Histological grade**	205
I	31 (7.84%)	I	22 (10.73%)
II	153 (38.73%)	II	69 (33.66%)
III	211(53.41%)	III	114 (55.61%)
Molecular subtype	395	**Molecular subtype**	179
Luminal	121 (30.63%)	Luminal	64 (35.75%)
HER2+	21 (5.31%)	HER2+	15 (8.38%)
TNBC	253 (64.05%)	TNBC	100 (55.86%)

3.3 Dewaxing and H&E staining of paraffin sections

One of the TMA in each duplicate set was used for spectroscopic analysis and the second one was used for area identification. The slide intended for spectroscopic analysis was used after dewaxing. In order to achieve free-wax samples, the modified Mian et al dewaxing protocol for Raman analysis of tissue was used.[147] The modification in the protocol involved reducing the exposition time in xylene based on the thickness of the breast samples. In addition, the rehydration of samples was performed using the organic phase, industrial methylated spirit (IMS), first and moving on to the aqueous solutions in order to wash all the xylene. The chemicals and exposition times used for the dewaxing process is show in Table 15.

Table 15. Dewaxing protocol for spectroscopic analysis set.

Time	Chemical
10 min	Xylene
5 min	100% IMS
5 min	70% IMS
5 min	50% IMS

The second slide of the set was used to identify the cancerous and normal epithelia and components of the breast tissue. To allow this identification, Haematoxylin and Eosin (H&E) staining was performed after wax removal with Xylene for 5 min, and hydration of the tissue sections in 100% (1 min), 70% IMS (30 s) and distillate water (1 min). The orientation slides were stained in Harris haematoxylin solution for 1.5 minutes and washed in running tap water for 4 minutes. After this, the slides were counterstained in eosin solution for 5 minutes and dunked in tap water twice. The staining process was followed by dehydration through 70% and 100% IMS (30s each). As a final step, the slides were cleared in Xylene for 10 min and mounted with xylene based mounting medium. Furthermore, the staining helped with the identification of the orientation cores and general spatial awareness within the TMA.

3.4 Adequacy of cores for inclusion

When the TMAs are prepared, the loss of cores and the folding of samples during sectioning can be observed. These issues can be related to the tissue type, and the preparation of the sample´s variables such as thickness or drying. The core retention can be affected by poor process and by the tissue nature. Considering all this, it is important to define the terms used on this study.

- **Core loss:** include all the missing samples on the TMAs and cores with less than 10% of the original core size remaining. Owing to the orientation template and the grid positioning of the cores, it is easily identifiable which cores are missing.
- **Incomplete core:** There are considered incomplete cores those whom maintain less than 50% of the original core size, but more than 10%.
- **Complete cores:** cores with sample quantity greater than or equal to 50% of the original core.

The eligibility criteria for Raman and FTIR analysis required cores to have:

- An H&E stained core and its dewaxed duplicate.
- In the NGB TMA´s, identifiable cancerous areas by using the H&E stained slide.
- A complete or incomplete core.

As it was previously indicated, the samples are presented as duplicates; since core loss can exist only for one of these duplicates the spectroscopic analysis was performed for all the samples that meet the inclusion criteria.

3.5 Area identification

Once one slide of each duplicate set was stained using the H&E protocol previously reported the TMAs were scanned. The comparison of the H&E and the light sample was performed to identify the regions where the spectra needed to be taken. In the cancerous samples, the cancerous epithelium was the area of interest while in the normal breast the normal epithelium was the one analysed. In some cancerous samples, infiltration of cancerous cell into stroma was present. However, the analysis of these cancerous cells was not performed as it lies outside the scope of this research.

A tutorial on morphology of normal mammary tissue, in situ and invasive cancer including its various types was received to ensure the adequate differentiation between normal and

cancerous tissues. This tutorial was conveyed by Dr Abeer Shaaban, Consultant pathologist in the Histopathology University Hospitals Birmingham NHS Foundation Trust department at Queen Elizabeth Hospital Birmingham. A subsequent joint analysis of TMAs including difficult cases revealed excellent concordance.

The identification of the cancerous areas and the normal tissue was performed for all the samples where they existed both, the H&E stained sample core and its dewaxed analogue sample.

3.6 Vibrational spectroscopic characterisation

3.6.1 Raman spectroscopy data acquisition

Raman spectroscopy was used only in TMA´s placed in glass as indicated in Table 13 and Figure 13. Raman spectra were collected using a Thermo Scientific high performance DXR confocal Raman microscope. Once the cancerous epithelial cells in the NGB TMA´s and the normal epithelium in the normal TMA were identified with the help of the H&E images, 20 random points within each area were analysed with the conditions displayed in Table 16 as it is illustrated in Figure 14. Each random point spectrum took 150 seconds, and the 20 sample points were completed in 50 minutes. The sample points represented in Figure 14 were obtained with 50X long work distance objective and a 3D spatial resolution of 1 µm (x, y) and 2 µm depth resolution. Each core shown in Figure 14 took 10s to image

Table 16. Raman spectroscopy analysis conditions.

Raman conditions	
Laser wavelength	532 nm diode laser
Alignment and calibration	Polystyrene standard
Objective	50x long work distance (LWD)
Laser power	10 mW
Aperture	50 µm pinhole
Exposure time	10 s
Number of exposures	15
Number of background exposures	512
Corrections	Fluorescence polynomial 6

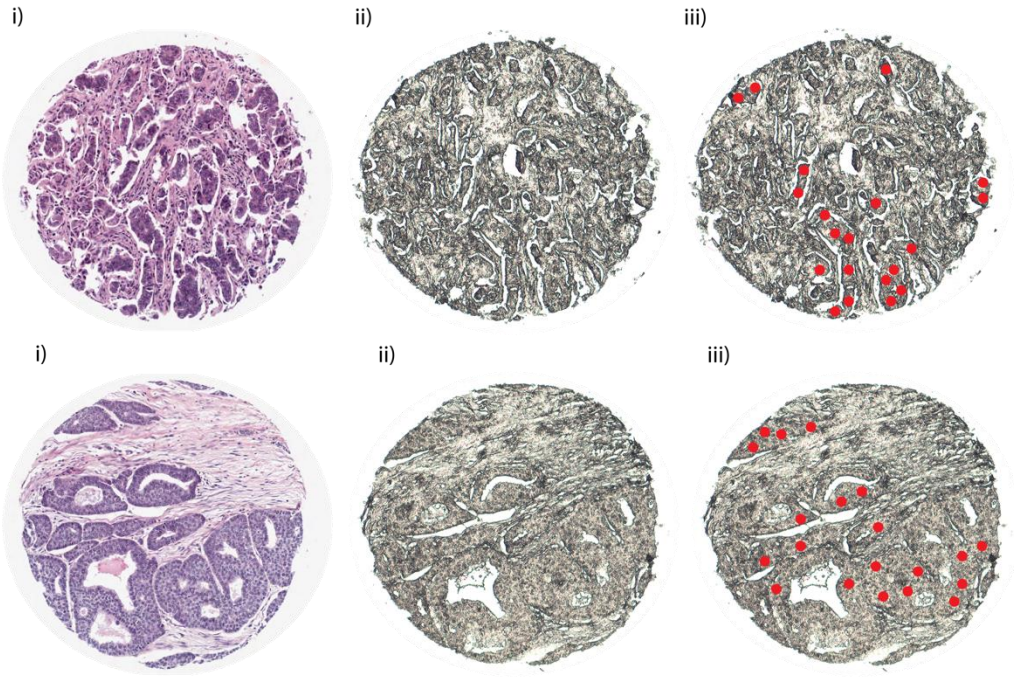

Figure 14. Example of area identification and random points selection for spectral analysis of an invasive ductal carcinoma. (i)H&E for area identification, (ii) Light image of core for spectroscopic analysis (iii) Random points selected for spectral acquisition in the cancerous area (red).

Considering the thickness of the samples, glass was sometimes detected in the spectra. The software Omnic for dispersive Raman (Thermo Scientific, USA) was used to subtract the glass contribution and to obtain averages of each sample, duplicate or single cores depending on availability.

3.6.2 FTIR SPECTROSCOPY DATA ACQUISITION

FTIR spectral acquisition was collected only in the CaF_2 and MirrIR slides as indicated in Table 13 and Figure 13. Spectral collection was performed using an iN10 MX Nicolet Infrared Imaging Microscope using the reflection mode. Twenty spectra were collected randomly from the cancerous epithelia in the NGB samples and from the normal epithelia in the normal TMA. The conditions used for spectra collections are shown in Table 17 and the point selection is illustrated in Figure 14. Using the UltraFast Mapping mode, a 1.2 × 1.2 mm area took 4.5 minutes to scan, with a spatial resolution 25 µm and a spectral resolution of 8 cm^{-1}.

Table 17. FTIR spectroscopy analysis conditions.

FTIR conditions	
Mode	Reflection
Background standard	Gold plate
Beamsplitter	KBr
Number of scans	16
Spectral range	675-4000 cm^{-1}
Resolution	8.0 cm^{-1}
Detector	Liquid nitrogen cooled MCT-A

Corrections	Kramers-Kronig

3.6.3 DATA ANALYSIS

The statistical data analysis was performed using Unscrambler X 10.2software (Camo software, Oslo, Norway). Unsupervised (PCA & Cluster analysis) and supervised multivariate approaches (LDA) were used in data analysis. The data was pre-processed using baseline, Savitzky-Golay smoothing and standard normal variate (SNV) corrections.

Principal component analysis (PCA) was performed on the spectral regions showed in Table 18 for Raman spectroscopy and in Table 19 for FTIR spectroscopy. The dots presented throughout the score-plots represent the average spectrum of the spatial points collected for a specific sample.

Table 18. Spectral regions used for Raman spectroscopic analysis.

Spectral region	Spectral range (cm^{-1})
General	3500-400
Fingerprint	1800-500
Amides	1800-1140
Aminoacids and nucleic acids	980-600
Lipids	3100-2680
Hydroxyproline and proline	960-810

Table 19. Spectral regions used for FTIR spectroscopic analysis.

Spectral region	Spectral range (cm^{-1})
Lipids	3500-2550
Proteins	1700-1500
Nucleic acids	1250-1000
Fingerprint	1450-600

Linear discrimination analysis (LDA) was used to validate the chemometric model over the full spectral range. Five samples from each group were left out at each pass until the total number of spectra of each type were predicted.

Sensitivity of this model was calculated using correctly classified false positive, false negative and true negative. Formulas used for calculation of sensitivity and specificity are:

$$\text{Sensitivity} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}$$

$$Specificity = \frac{True\ negative}{False\ positive + True\ negative}$$

Results and discussion

CHAPTER 4. FTIR ANALYSIS OF NORMAL BREAST AND BREAST CANCER

The results presented in this section include the FTIR spectral analysis of 3 tissue microarrays containing African breast tumours from similar tribal origins in Nigeria including different grades and subtypes and one TMA containing normal breast sections.

Twenty spectra were obtained from the epithelial tissue of each sample. A total of 245 cancerous tumours (359 cancerous sections) and 28 normal breast (37 normal breast sections) were analysed with FTIR spectroscopy. The average spectra of the cancerous epithelia named cancerous area (CA) and the normal epithelia which will be referred to as normal breast (NB) are presented in Figure 15.

When compared, the spectral profile of both tissue types presented characteristic bands associated with the presence of carbohydrates, lipids, nucleic acids and proteins as shown in Figure 15. The variance presented in the spectroscopic data has been plotted and shown in Figure 16. The spectral bands and their assignments are listed in Table 20 with the assignments taken from Movasaghi et al and Rehman et al [148], [149]. Overall, the bands found in the cancerous area spectrum are shifted towards lower wavenumbers in comparison with the normal breast bands. This downshift has been reported in cancerous samples by different research groups. [102], [105], [106]

Furthermore, differences in the intensity of peaks were identified between normal and cancerous tissue as it can be seen in Figure 17. The cancerous area spectra had a higher overall intensity which suggested biochemical changes associated with the disease. The increased intensity in C-H bands can be considered as an indicator of an increase of cellular components such as lipids, nucleic acids and proteins.

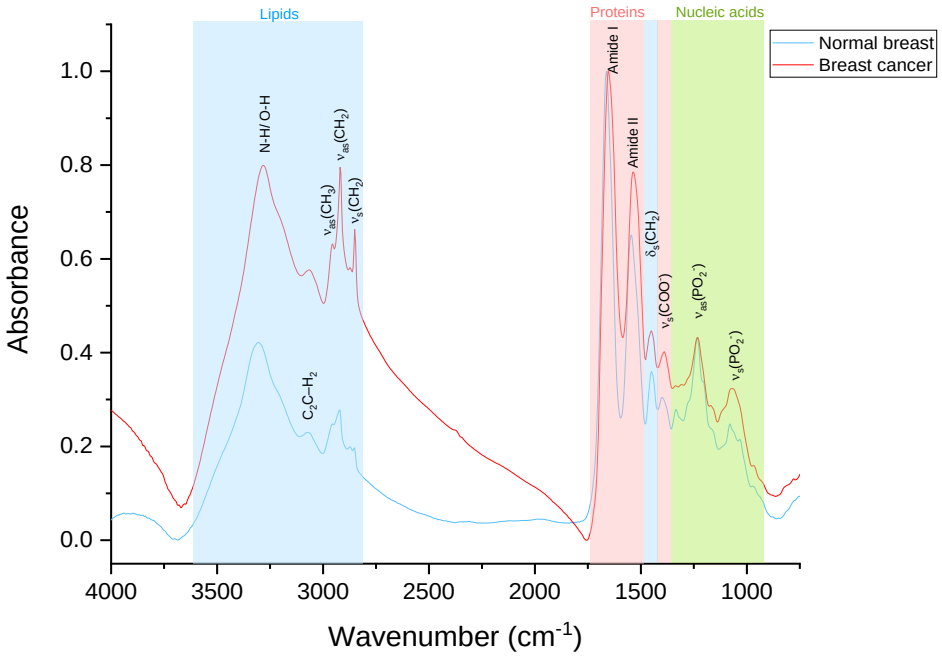

Figure 15. FTIR average spectra of cancerous area (CA) and normal breast (NB).

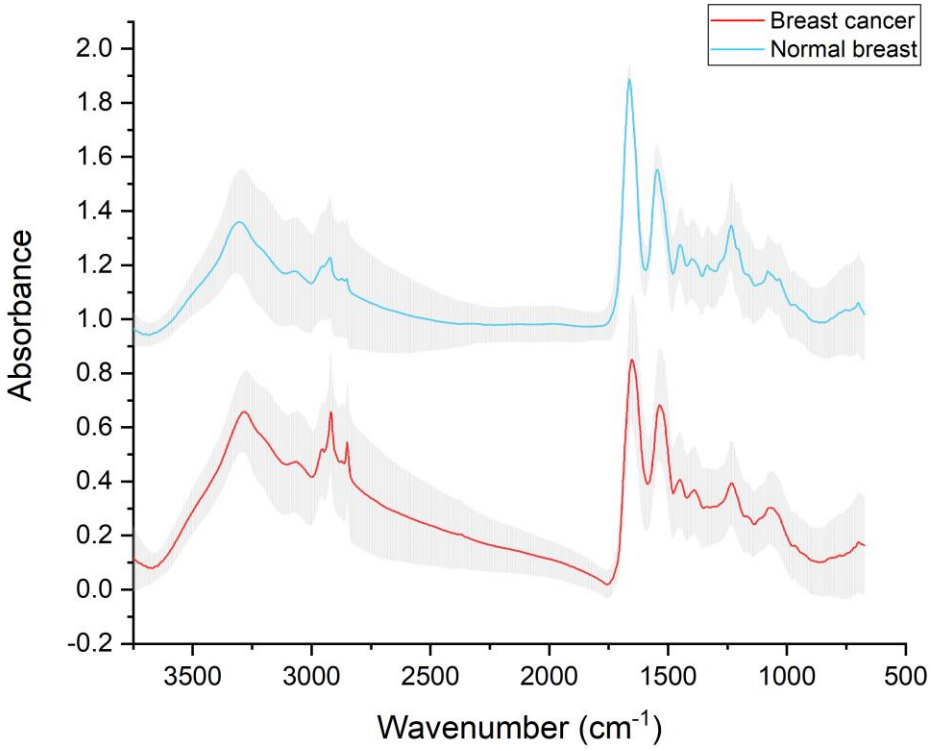

Figure 16. FTIR average spectra of Breast cancer (CA) and normal breast (NB), variance is represented with grey shadow.

Table 20. FTIR spectral bands of cancerous area (CA) and normal breast (NB) spectra averages with assignments. Assignments from Movasaghi et al and Rehman et al. [149]

CA		**NB**	
Wavenumber (cm^{-1})	Assignment	Wavenumber (cm^{-1})	Assignment
3282	νN–H (protein) N–H stretching of amide A and O–H stretching of carbohydrates	3305	Amide A band
3065	$C_2C–H_2$ aromatic stretching	3075	CH stretching band of the phenyl rings
2955	Asymmetric stretching vibration of CH_3 of acyl chains (lipids)	2955	Antisymmetric and symmetric stretching of CH_3
2919	Stretching C–H. Cholesterol ester. CH_2 asymmetric stretching: mainly lipids, with the little contribution from proteins, carbohydrates, nucleic acids	2922	Asymmetric stretching vibration of CH_2 of acyl chains (lipids)
2873	CH_3 symmetric stretching: protein side chains, lipids, with some contribution from carbohydrates and nucleic acids	2873	CH_3 symmetric stretching: protein side chains, lipids, with some contribution from carbohydrates and nucleic acids

2850	Stretching C–H νs CH_2, lipids, fatty acids. CH_2 symmetric stretching: mainly lipids, with the little contribution from proteins, carbohydrates, nucleic acids	2852	Symmetric stretching vibration of CH_2 of acyl chains (lipids) Stretching C–H. νs CH_2, lipids, fatty acids CH_2 symmetric stretching: mainly lipids, with the little contribution from proteins, carbohydrates, nucleic acids
1653	Amide I absorption (predominantly the C=O stretching vibration of the amide C = O) Protein amide I absorption C = O, stretching C=C uracil, NH_2 guanine Peptide amide I C=O stretching of amide I Amide I (protein) Amide I α-helix	1663	Amide I band. ν(C = C) cis, lipids, fatty acids. C=O Cytosine, uracil
1536	Stretching C = N, C = C. Amide II	1545	Protein band. Amide II (δN-H, νC–N). Peptide amide II
1451	Methylene deformation in biomolecules. Polyethylene methylene deformation modes. CH_3 symmetric deformations. Asymmetric CH_3 bending modes of the methyl groups of proteins. CH_2 bending: mainly lipids, with the little contribution from proteins	1450	Asymmetric CH_3 bending of the methyl groups of proteins. Methylene deformation in biomolecules. Polyethylene methylene deformation modes. CH_3 symmetric deformations. Asymmetric CH_3 bending modes of the methyl groups of proteins. CH_2 bending: mainly lipids, with the little contribution from proteins
1390	Stretching C–O, deformation C–H, deformation N–H	1400	Extremely weak peaks of DNA and RNA arise mainly from the vibrational modes of methyl and methylene groups of proteins and lipids and amide groups. Symmetric CH_3 bending modes of the methyl groups of proteins. Symmetric stretching vibration of COO^- group of fatty acids and amino acids. Symmetric bending modes of methyl groups in skeletal proteins

1335	δ(CH), ring (polysaccharides, pectin) CH_2 wagging. Collagen	1334	δ(CH), ring (polysaccharides, pectin) CH_2 wagging. Collagen
1308	Amide III		
1233	Overlapping of the protein amide III and the nucleic acid phosphate vibration. νPO_2^- asymmetric	1234	Composed of amide III as well as phosphate vibration of nucleic acids. PO_2^- asymmetric (Nucleic acid)
1172	CO–O–C asymmetric stretching: ester bonds in cholesteryl esters. Stretching modes of the C–OH groups of serine, threonine, and tyrosine residues of cellular proteins. C–O stretching(in malignant tissues)	1204	Collagen Vibrational modes of collagen proteins-amide III
1071	Symmetric phosphate [PO_2^- (sym)] stretching. Nucleic acid band	1080	Nucleic acid band
1066	C–O stretching of the phosphodiester and the ribose. PO_2^- symmetric stretching of nucleic acids	1033	ν(CC) skeletal cis conformation, ν(CH_2OH), ν(CO) stretching coupled with C–O bending. Stretching C–O ribose
969	C–N–C stretch: nucleic acids, C–N–C stretch: nucleic acids	972	νPO_4 of nucleic acids and proteins

Likewise, as seen in Figure 17C the shifting and increased intensity in the 1500-1700 cm^{-1} area associated with nucleic acids and amide stretchings might suggest changes in the conformational structure of proteins in the cancerous tissue. [106]

The band associated with the asymmetric stretching of CH_2 of acyl chains (lipids) was shifted to a lower wavenumber (2922.2 to 2918.7 cm^{-1}) on the cancerous area. This shift is related to an increment of trans- lipids conformations. A higher content of the trans- conformations translates into a higher order in lipids by decreasing the flexibility of the acyl chain. Studies by Spector et al. had reported the cell shape change based on the lipid order. Therefore, a change in the cell shape by modifications on the cell membrane fluidity can support and promote malignant activities as well as signalling pathways associated with carcinogenesis. [121], [150]

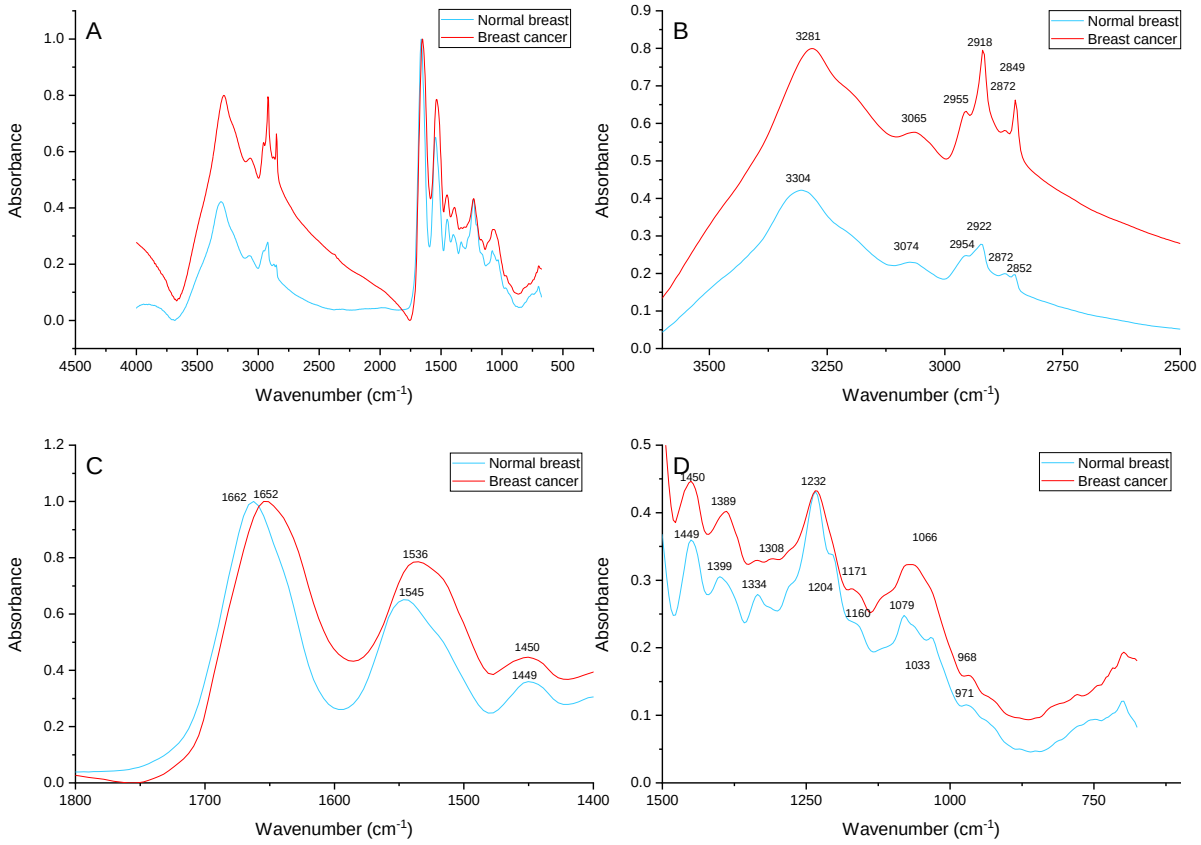

Figure 17. FTIR characterisation of cancerous area (CA) and normal breast (NB). (A) General region (4,000-675 cm^{-1}), (B) High wavenumber region (3,500- 2,550 cm^{-1}), (C) Amide region (1,700-1,500 cm^{-1}), and (D) Fingerprint region (1,450-600 cm^{-1}).

The intensity of peaks at ≈1334, ≈1204, and ≈1033 cm^{-1} is considerably higher in the normal breast in comparison with the cancer area. These peaks have been attributed to collagen vibrations and seem to appear sharper and stronger for normal tissues. Low content of collagen can be suggested as part of the FTIR spectral profile of breast cancer samples.

Additionally, the shift from 1160 cm^{-1} to 1171 cm^{-1} in the cancerous area has been reported as an increase of the cancerous cells in malignant tissue. This band is characteristic of C–O stretching of the parenchymal cells phosphorylated proteins. [105] On the other hand, this band is also associated with the CO–O–C asymmetric stretching of ester bonds in cholesteryl esters. The increased intensity at 1171 cm^{-1} on the cancerous area might be an affectation on the cell membrane fluidity in addition to the highest cholesterol requirement reported on malignant tissue. [121]

Different absorbance ratios have been used as biomarkers for different biochemical changes. For example, the A_{2958}/A_{2853} ratio can provide an insight on methylation, A_{1657}/A_{1204} and A_{1657}/A_{1278} indicate the collagen content relative to the protein content, A_{1640}/A_{1550} and A_{1160}/A_{1120} give information on proteins and its structure, A_{1640}/A_{1460} and A_{1550}/A_{1460} indicate a ratio of lipid to protein content, while A_{1310}/A_{1240} provide information on the nucleic acid content. When deconvolution of spectra is performed the following ratios can be obtained:

A_{3300}/A_{3075} gives an idea on how dense and close-packed structural the tissue is, whilst A_{1657}/A_{1635} reflect the content of α-helix and β-sheet domains in tissue proteins. The ratios for the cancerous area and the normal breast are presented in Figure 18.

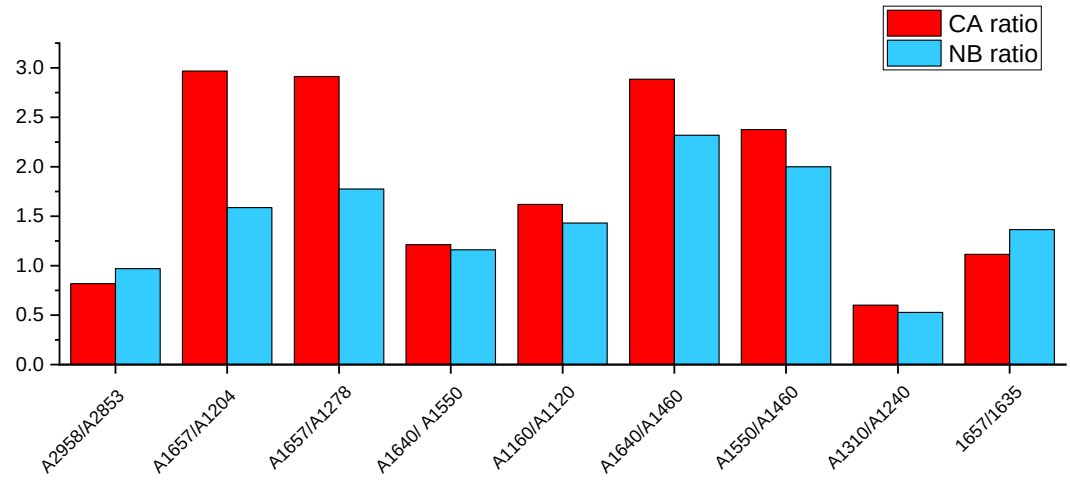

Figure 18. Selected absorbance ratios of cancerous and normal breast tissue.

The first ratio, A_{2958}/A_{2853} has a stretch relationship with the grade of methylation. The ratio of the methyl stretching band to the methylene band was smaller for the cancerous tissue in comparison with the normal breast. This reduction in the ratio from normal to malignant tissue has been explained through the DNA hypomethylation and the overall increase of malignant cellular content. In addition, it has been suggested that a reduction of protein fibres in the cancer tissue. [105] This finding can be supported by the ratios representing the content of collagen relative to the total content of protein, (A_{1657}/A_{1204} and A_{1657}/A_{1278}), which are higher for the normal breast tissue.

Once the deconvoluted IR spectra for the CA and NB amide area was obtained the deconvoluted peaks for α-helix (1657 cm^{-1}) and β-sheet (1635 cm^{-1}) protein structures were clearly identified as seen in Figure 19. In agreement with the downshift presented in the CA spectrum, both peaks presented a shift towards lower wavenumbers (1659 and 1632 cm^{-1}) in comparison with the normal breast (1666 and 1635 cm^{-1}). Moreover, the β-sheet cancer peaks presented a higher intensity in components with the normal breast. This increased intensity explains the CA broader Amide I peak. The A_{1657}/A_{1635} ratio showed a higher quantity of α-helix conformation in the normal breast in comparison with cancerous tissue. These behaviours indicated changes in the number of hydrogen intermolecular bonds in the secondary structural conformation. [102] It can be suggested that β-sheet structural proteins are more present in cancer tissue and that structural changes form an important part of the carcinogenesis process. Studies focused on the analysis of breast tumour microenvironment had reported β-sheet conformations surrounding the cancerous epithelia and changing towards an α-helix conformation as they move far from the tumour. [107]

As product of the deconvolution of the FTIR spectrum of normal breast, a band located at 1593 cm-1 was identified only in the normal tissue. This band is associated with the vibrations of the C-C stretch of the phenyl ring. [149] The identification of the deconvoluted band reinforces the idea of hydrogen intermolecular bonds changes during carcinogenesis. In

addition, the high content of collagen visible in the normal breast sections can contribute towards this vibration.

Based on the A_{1657}/A_{1204} and A_{1657}/A_{1278} ratios, a lower content of collagen relative to protein content has been found in the cancerous samples. Therefore, the presence of collagen in the normal breast samples seems to be dominant while a higher content of proteins constituting the malignant epithelia is present in the cancerous samples.

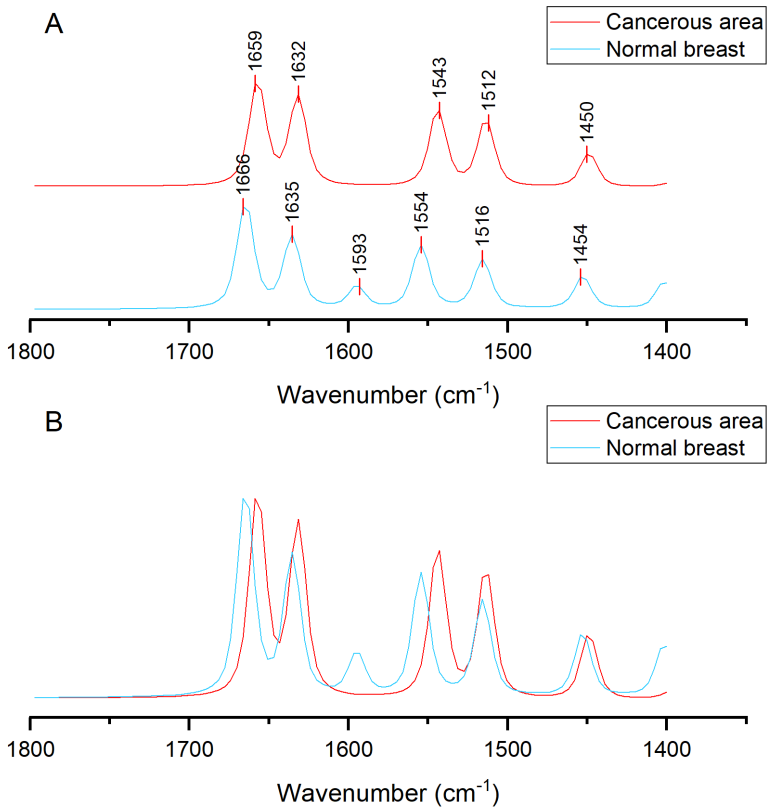

Figure 19. (A) Stacked deconvoluted peaks of FTIR spectra in the amide area (B) Overlaid comparison of spectra presented in (A).

The stretching band of C-OH in the cancerous tissue is shifted towards a higher wavenumber, 1171.5 cm^{-1}, while the band representing the same effect on the normal breast is presented at 1160 cm^{-1}. This shift is caused by the increased non-hydrogen bonded C-OH groups in the cancer tissue reflecting the phosphorylation of the cellular proteins. The phosphorylation of the C-OH groups in proteins is characteristic of the carcinogenesis process. In this process, the –OH groups are replaced by larger phosphate groups. By breaking this hydrogen bonds some of the structures like the β-sheets no longer exist leaving behind unordered structures. [103]

When the A_{1640}/A_{1460} (Amide I/Lipids) and A_{1550}/A_{1460} (Amide II/Lipids) ratios were compared both of the cancer ratios were higher. This result implies a lower content of lipids relative to the protein content for the cancer tissue, whereas the opposite trend is true for the normal breast epithelia.

The cancerous cohort represents a heterogeneous group of cases, therefore, a comparison of several spectra for both cases can provide a better insight of the malignant behaviour and all

the biochemical processes associated with the cancer progression. For this end, the results of the multivariate analysis are presented in the next sections.

4.1 General region (4,000-675 cm^{-1}) analysis of normal breast and breast cancer

Principal component analysis was applied to the spectral general region. Each dot in the score plot presented in A of Figure 20 represents the average spectrum of all the spatial random points taken from either a normal breast (blue dot) or a breast cancer tumour (red dot). Good separation was obtained with principal component 3 and 5, accounting for 15% of the variance as it can be seen in the score plot presented in Figure 20A.

A clear grouping can be seen for the cancerous area and the normal breast. CA and NB samples are presented at opposite sides of both principal components axis. The peaks responsible for the separation of cancerous and normal tissue on the score plots are 3275, 2963, 2916, 2849, 1669,1633, and 1530, cm^{-1}.

The first peak located at ≈3275 has been correlated to the N–H stretching of amide A in proteins and nucleic acids. [102] This peak seems to indicate a higher vibrational activity on the cancerous area in comparison with the normal breast. In the previous section, an overall downshift in the cancer area spectral profile was mentioned. This is the case for the amide A peak which went from 3304 cm^{-1} in NB to 3281 cm^{-1} in CA. This downshift might result from an increase on the hydrogen bonds of secondary structures in proteins. [151] In addition, important differences in intensity and broadness can be detected with the naked eye when spectral profiles are compared. These differences suggest that the cancerous areas contain a significant amount of extracellular components such as proteins and nucleic acids, which are necessary to promote and sustain the abnormal characteristic proliferation of cancerous cells. These findings are in agreement with previous reports. [102], [149]

The peak presented at ≈2963 cm^{-1} is associated with the asymmetric stretching vibration of CH_3 of acyl chains in lipids, but also with some contribution from nucleic acids and carbohydrates. [121], [149] Based on the positive peak found in the PC-2 loading and the positive location of the cancerous samples in the PC-2 axis in the score plot, a higher content of acyl chains in the cancerous area can be assumed. The increased content of unsaturated bonds in the lipid components in the cancerous tissue has been previously reported to have an effect on cell adhesion molecules in the cell membrane, as well as interactions between cell adhesion molecules and lipid components in the membrane. In addition, changes in the conformational structure of proteins have affectation on the physical behaviour of the cell membrane and therefore on the membrane fluidity. All these changes will modify the permeability of the membrane and how it will respond to its physical environment and radiation. [152], [153]

The peak at ≈2916/2918 cm^{-1} presented in both principal components –PC3 and PC5– has been associated with the asymmetric stretching of CH_2 bonds mainly in lipids. The next peak identified - 2849 cm^{-1}- is also attributed to stretching vibration of lipids but asymmetrically. Little contribution from proteins, carbohydrates, and nucleic acid has also been reported for both peaks. This vibration has been closely related to the stretching in cellular membrane

lipids. [106], [149] Based on the location of the majority of the CA point in the score plot a higher content of the lipids in the cancerous areas can be assumed. During carcinogenesis important lipid modification on the cellular membrane take place. These changes are reported to affect the permeation and the transport of metabolites considerably. Also, the accelerated and constant production of lipids is needed to fuel the malignant metabolism. [154]

Within the analysis of the whole spectral range, three peaks associated with amide vibrations and its different conformations were identified as biomarkers for the classification of breast cancer and normal breast. These peaks are 1669, 1633, and 1530 cm^{-1} and represent the amide I α-helix (1669 cm^{-1}) and amide I β-sheet (1633 cm^{-1}) domains in proteins and the Amide II vibration comprising the N–H bending and C–N stretching predominantly in α-helix (1530, cm^{-1}). [102] The scattering of cancerous samples on the score plots toward the negative axis in PC-3 suggested that the malignant areas presented more vibrations associated with the amide I β-sheet conformation and the amide II vibration in comparison with the normal breast. Furthermore, PC-5 reinforced the previous finding by confirming that normal breast presented less active amide I α-helix vibrations associated with the 1669 cm^{-1} peak in comparison with CA samples.

When the PCA model created using the whole spectral range was validated using Linear discriminant analysis (LDA) a sensitivity of 92% and specificity of 86% was confirmed.

Cluster analysis was obtained for the cancerous area and normal breast samples. The analysis was set up to recognize 10 clusters to facilitate separation as it can be seen in Figure 21. The fourth cluster presented in plot D is formed by most of the normal breast samples. However, some of the normal samples are located on B, E, and H. The infiltration of normal samples into cluster mainly formed by cancer samples confirmed the heterogeneity of our cohort. The cancerous samples analysed with this method represented different grades and subtypes, and as a result, present different characteristics that might resemble the normal tissue. Therefore, the clustering of both type of samples.

4.2 Analysis per region of normal breast and breast cancer

The previous analysis of the general region offered peaks responsible for the separation of cancer and normal samples. These peaks were located among the higher wavenumber, amide and fingerprint regions. Further analysis in each region was carried out to identify more biomarkers and these are presented in the following sections.

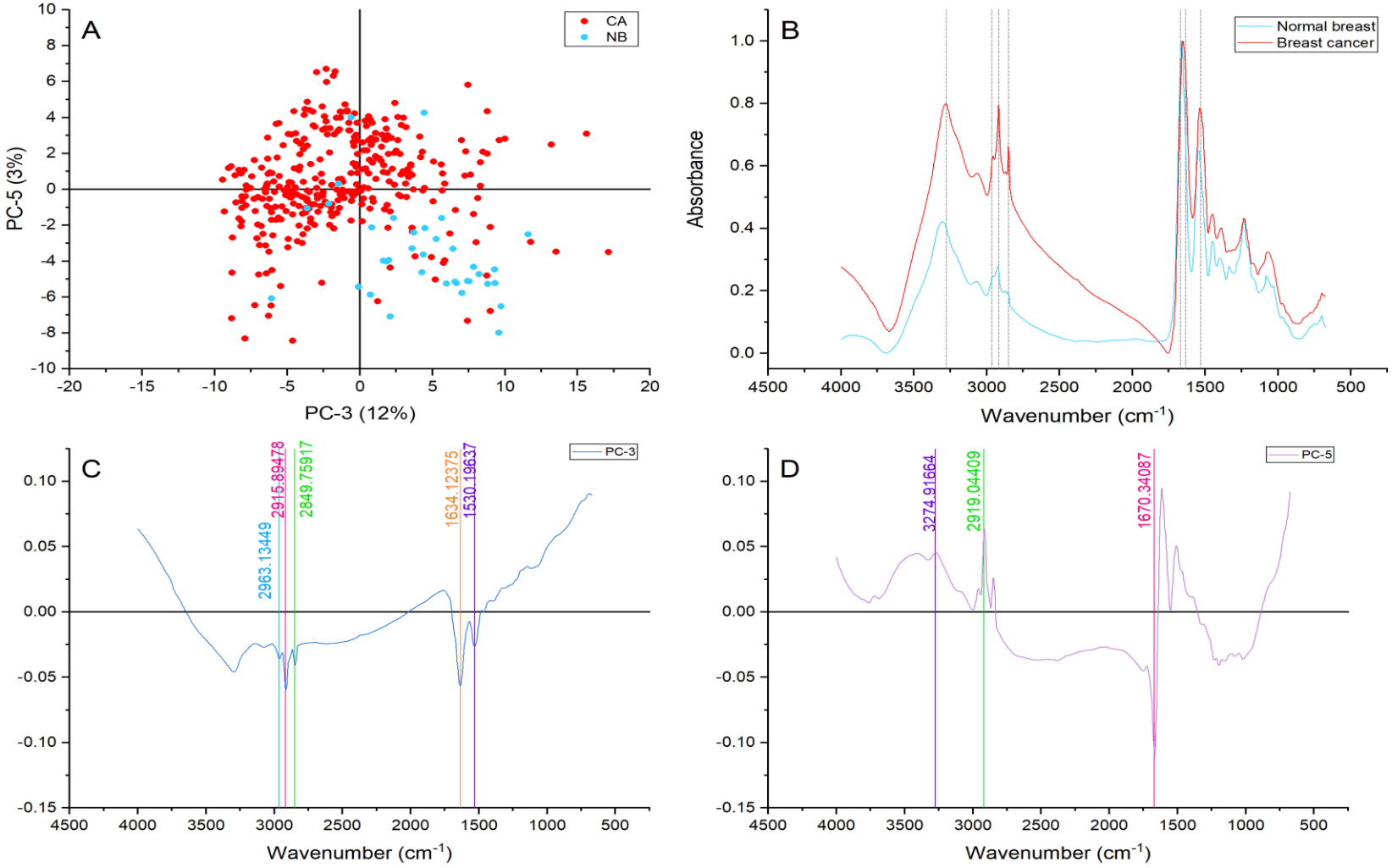

Figure 20. Principal component analysis of the general region (4,000-675 cm^{-1}). (A) Score plot using PC-3 and PC-5 accounting for 15% of the variance, (B) Average spectral profile for breast cancer and normal breast. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-3 Loading, (D) PC-5 loading.

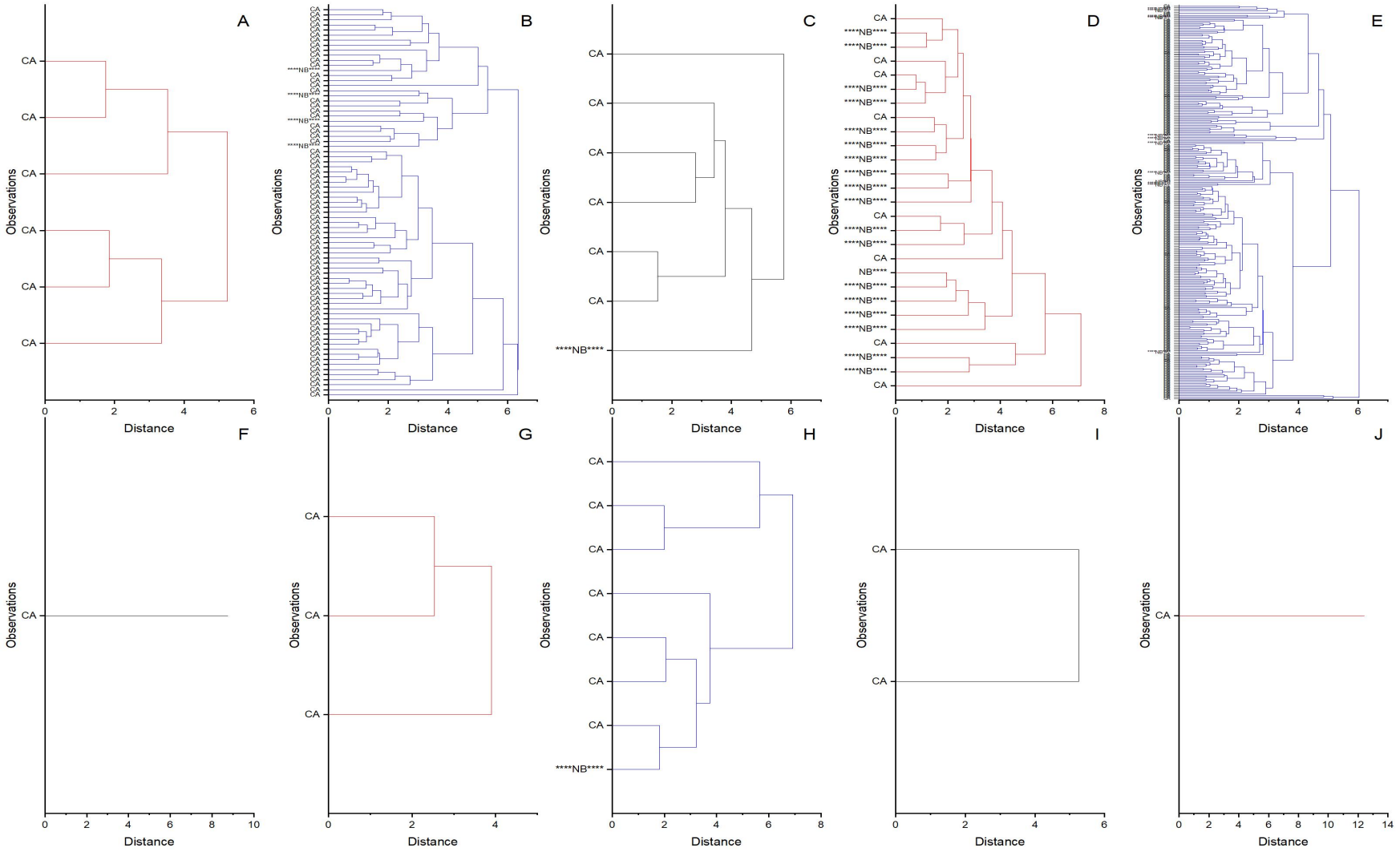

Figure 21. Cluster analysis of Cancerous area (CA) and Normal breast (****NB****). The analysis was set up to identify ten groups presented from A to J.

4.2.1 Higher-wavenumber region (3,500–2,550cm^{-1}) analysis of normal breast and breast cancer

Principal component analysis was used to identify changes and variations in the higher-wavenumber region. The best separation of cancerous and normal breast areas was found using PC-1 and PC-4 which were accountable for 82% of the variance as it can be seen in Figure 22.

The peaks located at 3264, 3173, 2930, 2903, 2861, 2842 cm^{-1} were identified as responsible for the grouping of the cancerous area and normal breast.

The first peak at ≈3264 cm^{-1} was reported in the general region analysis and attributed to the N–H stretching of amide A in proteins and nucleic acids. The band at ≈3173 has been reported as N-H stretching bands with cis configuration. The amide A band of proteins is presented around ≈3000 cm^{-1} with a characteristic shoulder at ≈3130 cm^{-1} representative of a secondary amide. This shoulder presents Fermi resonance from the combination of the N-H stretching with either the trans- or cis-conformation of the secondary amide overtone. [155] The normal tissue presented a lower contribution of the 3173 peak, whereas the cancerous area samples equally distributed among both PC-4 axes as it can be seen in Figure 22A. This spreading behaviour indicates a dynamic behaviour among the cancer samples in our cohort.

The bands located at ≈2930 and ≈2842 cm^{-1} are attributed to both antisymmetric and symmetric CH_2 stretching of hydrocarbon chains. These assignments are mainly correlated with long chains characteristic of lipids like ceramides, glycolipids and phospholipids. [156] Ceramides act as precursors for regulators of cell growth, proliferation and motility. [157] In the average spectra presented in Figure 22B, clear differences in position, intensity and width can be appreciated. It can be suggested that important changes in the conformation of lipids are taking place during carcinogenesis as the cancerous samples present higher contributions on these peaks in comparison with the normal breast samples. Changes in the lipid conformation and affectation on the lipid membrane fluidity during carcinogenesis had been widely reported by different research groups. Lipid modifications promote the signalling and stimuli to facilitate processes such as proliferation and cell adhesion. [158]–[160]

Furthermore, the peaks positioned at ≈2903 and ≈2861 cm^{-1} provided further information on the CH_3 symmetric stretch, and aliphatic C-H stretching in fatty acids and lipids. Both contributions were higher in the cancerous tissue confirming modification on the lipid metabolism. Breast cancer has shown high rates of production of novo lipids. Lipids play vital roles in different phases of tumorigenesis. Their involvement in the membrane synthesis and membrane saturation directly affect the cell growth, cell proliferation and the resistance to oxidative stress. Other important components are the lipid droplets. The lipid droplets are formed by triacylglycerides and cholesterylesters. Fatty acids are moved from the lipid store and are used as a source of energy through β-oxidation. Besides, cancer cells seem to present higher numbers of lipid droplets compared with normal tissue. The activated storage of triacylglycerides in lipid droplets has been suggested to work as fuel source after re-oxygenation during intermittent hypoxia. Fatty acids also promote redox balance supporting a high-glycolytic rate. In addition, lipids form the structural basis of paracrine hormones and

growth factors which promote tumour growth, neovascularisation, invasion, and metastatic spread. [160]

4.2.2 Amide I and amide II (amide I/II) region (1,700-1,500 cm^{-1}) analysis of normal breast and breast cancer

The principal component analysis of the spectral region from 1700-1500 cm^{-1} showed good separation and grouping when PC-2 and PC-3 are compared, as shown in Figure 23A. The average spectral profile of the analysed cancer area and the normal breast are presented in Figure 23B. These profiles present three peaks each. However, as in other regions these peaks are the product of the convolution of different spectral contributions as presented in Figure 19. When the loadings were analysed the following peaks were identified as responsible for the separation and grouping of both sample types: 1668, 1646, 1612, 1568, 1552, and 1528 cm^{-1}.

The first peak located at ≈1668 cm^{-1} seems to indicate the normal breast shifted peak assigned to α-helix found in proteins. Accordingly, the contribution of this band is higher for the normal breast. Half of the cancerous samples are located in the PC-3 negative axis. The location of these samples suggests a dynamic behaviour regarding conformational changes in proteins at different stages of carcinogenesis. The samples in the mentioned location presented a higher intensity on the band positioned around ≈1646 which is mainly associated with β structures and unordered random coils. This changes in intensity and the spread of samples on both axes suggests a transition from α-helix structures into β-sheet structures as cancer progress. [102]

Cancer epithelia presented higher ≈1612 cm^{-1} intensities in comparison to normal epithelia. This band is representative of adenine vibration in DNA. [149] Therefore, a higher content of nucleic acid content is assumed and is considered in agreement with the increased cellular content expected in abnormal proliferation. In addition, the peak identified at ≈1568 cm^{-1} seemed to have a higher contribution for all the normal breast samples and half of the cancer area samples.

The ≈1568 cm^{-1} peak has been assigned to amide II but also to C=N of adenine and guanine. Once again, the cancer samples demonstrate different behaviour having less and more content of amide II and pair bases than the average while the normal breast proved a constant increased behaviour probably because of the stromal collagen. High cell density characteristic of accelerated proliferation and invasive processes can be suggested for the samples located on the positive access of PC-3.

Amide II also is represented by the peak at ≈1552 cm^{-1} which is characteristic by the C-N stretching and the N-H bending of the amide II. [161] The contribution of this peak is higher in all the normal breast samples. This band is characteristic of the collagen component in the stromal area. It has been suggested that the stromal component has been included in the spectral analysis due to the low cell density (in comparison with CA) in the normal epithelium.

Stretching characteristic of C=C and C=N specific for guanine and adenine nucleic bases represented by the band found at ≈1528 cm^{-1} (deconvoluted spectra) presented higher

intensities in half of the cancerous samples. This behaviour suggested a higher cell density, but also a higher cell components content characteristic of active proliferation. [74]

4.2.3 Fingerprint region (1,450-600 cm^{-1}) analysis of normal breast and breast cancer

When principal component analysis was applied to the fingerprint region, a clear separation and grouping were achieved. Cancer and normal breast samples were placed on different axis on PC-4 and PC-5 accounting for <2% of the variance as it can be seen in Figure 24.

The peaks located at 1427, 1385, 1366, 1332, 1236, 1194, 1051, and 882 cm^{-1} were identified in the loading plots as responsible for the grouping and separation of CA and NB samples.

The band at ≈1427 cm^{-1} has been assigned to CH_2 vibrations in lipids, especially fatty acids. [149] The cancerous area analysed presented higher intensities for this band suggesting an active lipid metabolism and transformations. This behaviour is in agreement with the results previously reported in the higher wavenumber region.

The stretching of C-O, deformation of C-H and N-H is represented by the ≈1385 cm^{-1} peak. [149] The vibrations represented by this peak are mainly associated with proteins and nucleic acids. This band has a higher contribution in the cancerous samples suggesting either higher cell contents or higher cell density. [106] A higher cell density is characteristic of higher grades and can explain the spreading of the cancerous samples. Similar behaviour was detected with the peak at ≈1366 cm^{-1} which is representative of the same vibrations. A shoulder representative of this band can be appreciated in the spectral profiles shown in Figure 24B.

Figure 24B showed important differences in the ≈1332 cm^{-1} peak. This band has been reported as a collagen-specific peak and an increased intensity has been attributed to higher contents of collagen. [162] This peak presented higher intensities in the normal breast samples. Collagen-specific vibrations like this reinforced the idea of the stromal spectral acquisition when the normal breast was analysed. Furthermore, the band at ≈1194 cm^{-1} also represent a vibrational mode of collagen, amide III.

The assignment located at ≈1236 cm^{-1} has been correlated with the phosphodiester asymmetric stretch in nucleic acids, but also is representative of amide III and CH_2 wagging vibrations of collagen. [105] When the contribution of this peak was evaluated all the normal breast and half of the cancer samples presented higher intensities than the rest half of the cancerous samples and the average intensity. The collagen found and included in the normal breast analysis is responsible for the integration of the normal samples in this category due to the amide III vibration. This amide III peak arises from the C–N stretching and N–H deformation. [163] On the other hand, the ≈1236 cm^{-1} band is highly representative of nucleic acids content. The cancer samples that presented higher intensities of this peak are suggested to present higher cell density and cell component. This band is representative of two different contributions, and both sample types present different biochemical behaviour. Hence the difference in the spectral profile presented in Figure 24B.

The band located at ≈1051 has the highest contribution in the cancer area samples according to their positions in the score plot in Figure 24A. 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. [149] A clear visual representation of the relevant contributions of this band can be appreciated in the average spectral profile of the cancerous sample in Figure 24B. This peak suggested a higher content of nucleic acid components which characterise malignancy and aggressive behaviour. [164]

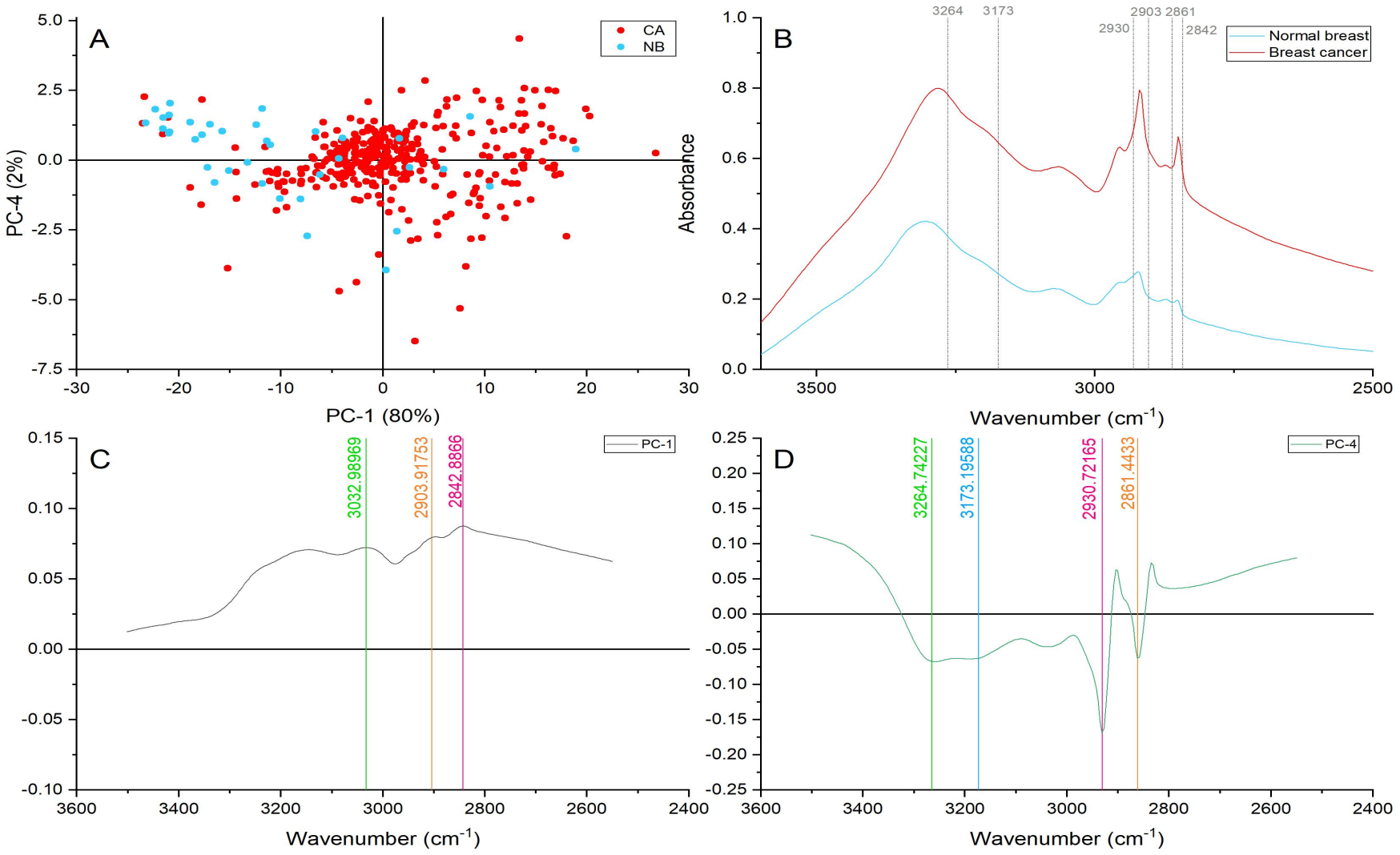

Figure 22. Principal component analysis of the higher wavenumber region (3,500 -2,550 cm^{-1}). (A) Score plot using PC-1 and PC-4 accounting for 82% of the variance, (B) Average spectral profile of the higher wavenumber region for breast cancer and normal breast. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-1 Loading, (D) PC-4 loading.

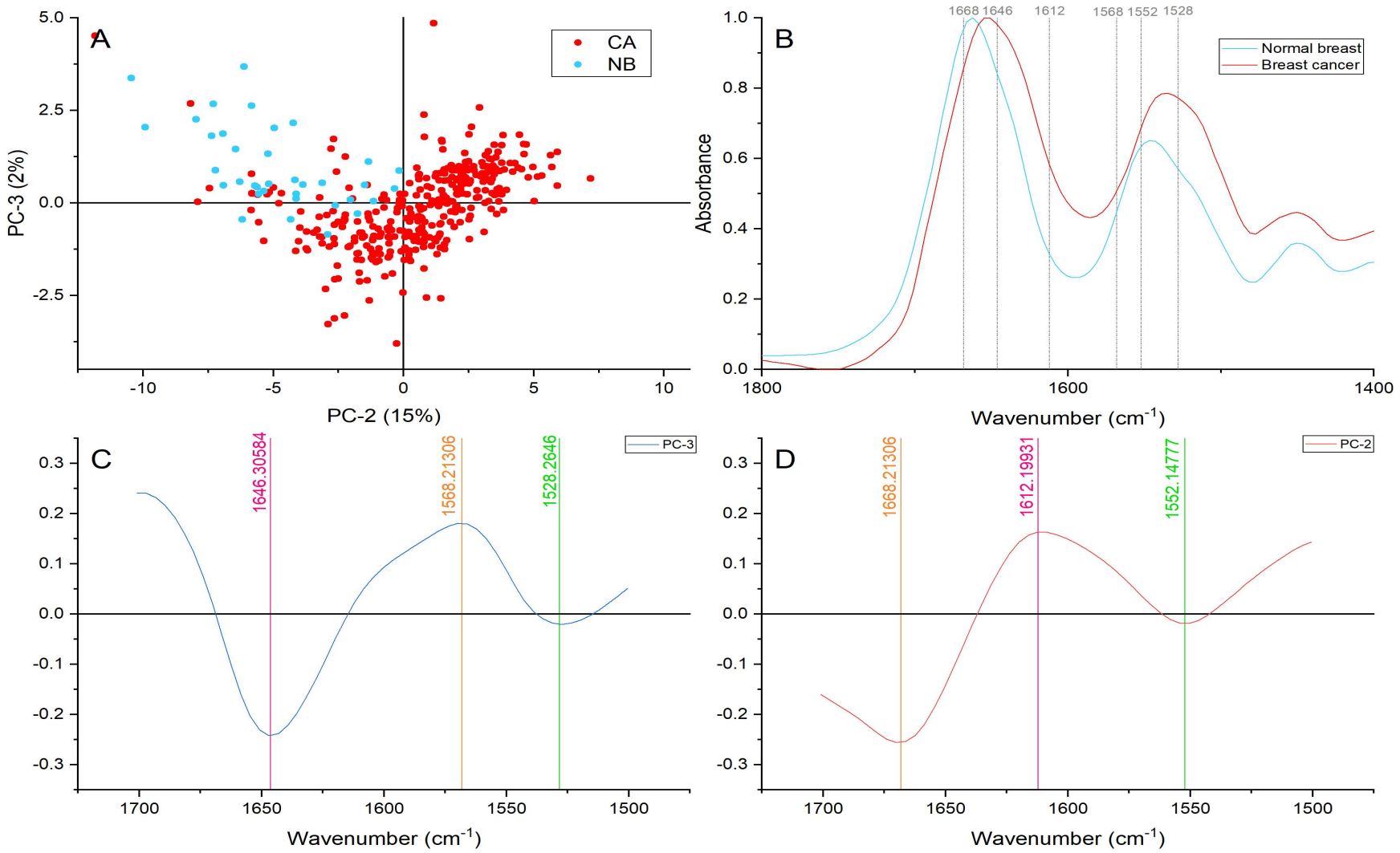

Figure 23. Principal component analysis of the amides region (1,700-1,500 cm^{-1}). (A) Score plot using PC-2 and PC-3 accounting for 17% of the variance, (B) Average spectral profile of the amides region for breast cancer and normal breast. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-3 Loading, (D) PC-2 loading.

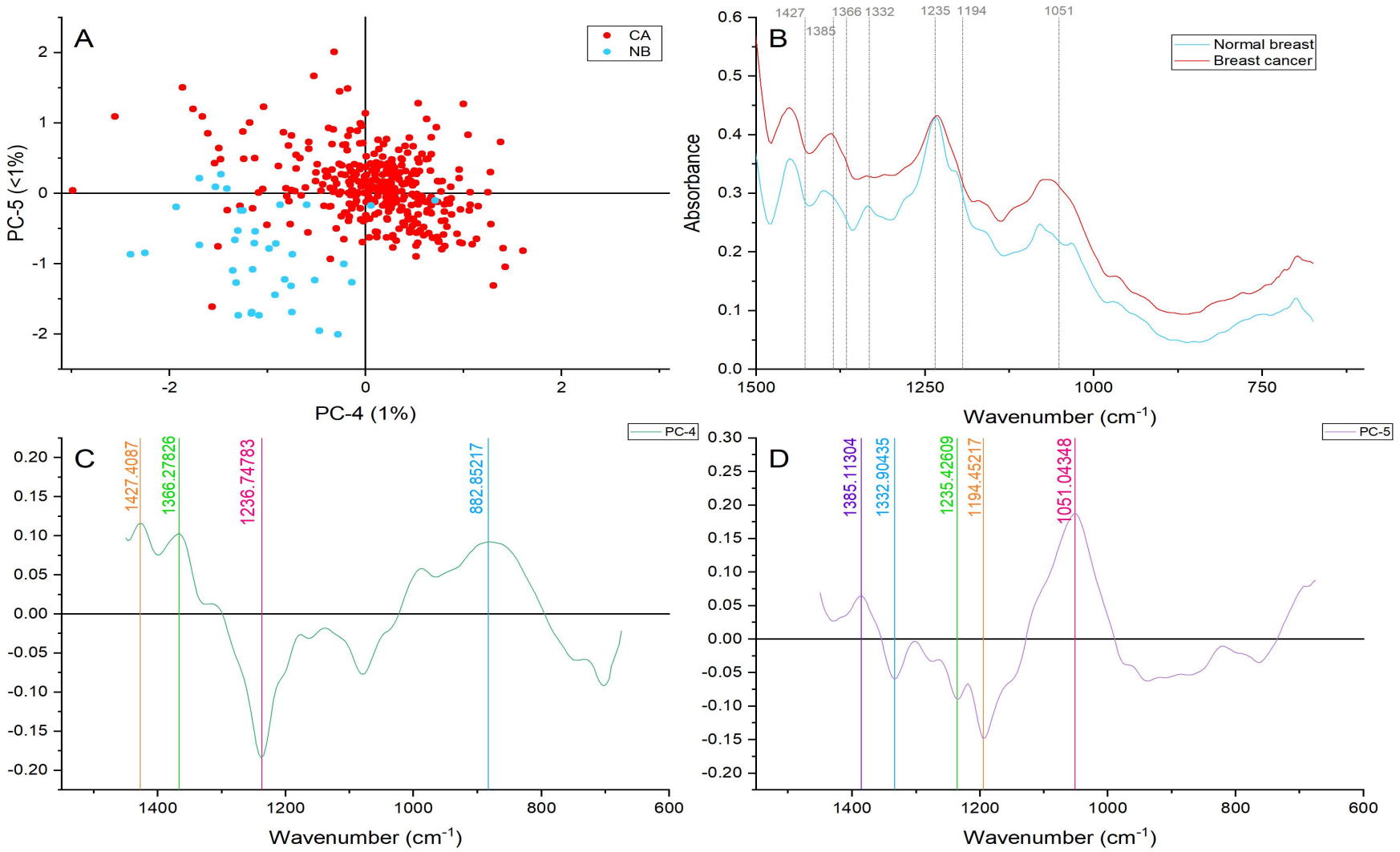

Figure 24. Principal component analysis of the fingerprint region (1,450-600 cm^{-1}). (A) Score plot using PC-4 and PC-5 accounting for <2% of the variance, (B) Average spectral profile of the amides region for breast cancer and normal breast. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-4 Loading, (D) PC-5 loading.

CHAPTER 5. FTIR ANALYSIS OF BREAST CANCER

5.1 COMPARISON OF THE SPECTRAL DATA OF DIFFERENT SUBTYPES WITHIN A SPECIFIC GRADE.

The results and discussion presented in this section include the FTIR spectral analysis of 2 tissue microarrays containing African breast tumours from similar tribal origins in Nigeria comparing different grades and subtypes.

A total of 359 cancerous samples were analysed with FTIR spectroscopy. Two comparison types were performed, the first one comparing different subtypes within one grade and the second one comparing the spectral data of different grades within a specific subtype.

5.1.1 GRADE 1 ALL SUBTYPES

When compared, the average spectral profile of grade 1 samples presenting luminal, HER2+ rich and triple negative breast cancer (TNBC) subtypes showed differences enough to see with the naked eye as presented in Figure 25 and Figure 26.

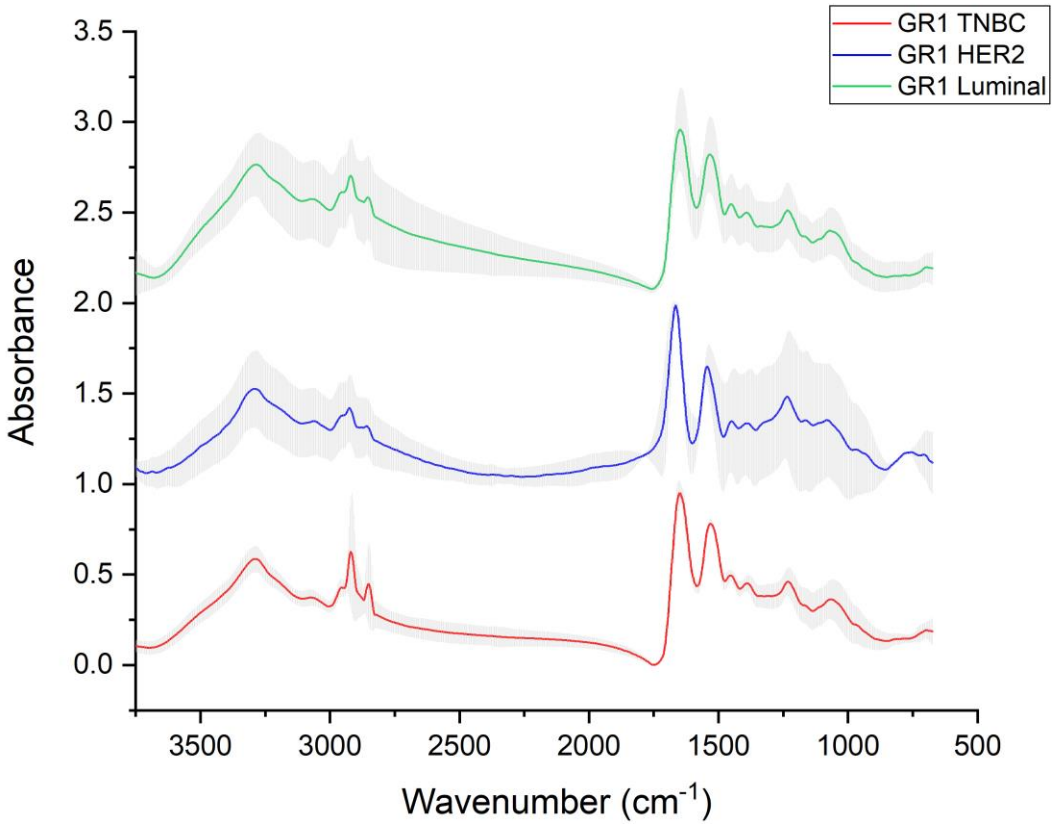

Figure 25. FTIR average spectral profile of grade 1 samples presenting triple negative (TNBC), luminal and HER2+ subtypes, variance is represented with grey shadow.

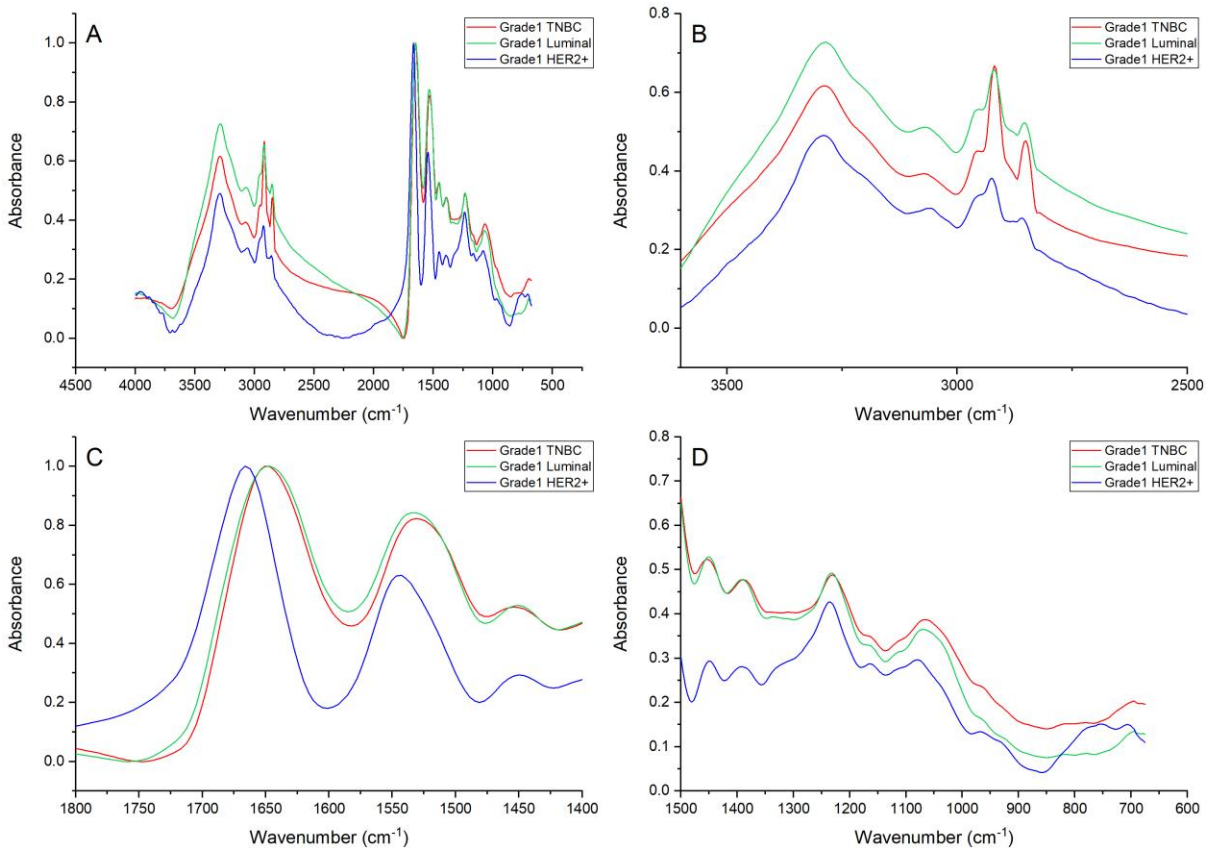

Figure 26. FTIR average spectral profile of grade 1 samples presenting triple negative (TNBC), luminal and HER2+ subtypes. (A) General region (4,000-675 cm^{-1}), (B) High wavenumber region (3,500-2,550 cm^{-1}), (C) Amide region (1,700 -1,500 cm^{-1}), and (D) Fingerprint region (1,450-600 cm^{-1}).

5.1.1.1 Higher wavenumber (3,500-2,550 cm^{-1}) analysis of Grade 1 samples all subtypes

The comparison of the different subtypes of grade 1 spectral profiles showed overall differences in absorbance and peak ratios. However, considerable shifts cannot be appreciated with a naked eye. Therefore, principal component analysis was used in the higher wavenumber (3,500-2,550 cm^{-1}) showing grouping and separation based on the subtype as it can be seen in Figure 27.

The peaks located at 3278, 3053, 3015, 2968, and 2920 cm^{-1} were identified as responsible for the sample separation.

The first peak located at ≈3278 cm^{-1} is characteristic of amide A stretching. [165] The HER2+ rich presented higher intensity than the average of this peak followed by TNBC and luminal grade 1 samples. Complementarily, the same trend is repeated with amide B. This peak is located at ≈3053 cm^{-1} and it is mainly associated with the N-H stretching of proteins.

In Figure 27B, the dotted line representing the peak located at ≈3015 cm^{-1} which showed a slight shoulder that became easily visible in the grade 1 HER2+ spectrum profile. This band has been associated with olefinic =CH stretching. [121] This shoulder became less clear in TNBC and luminal grade 1 samples. Higher content of the olefin =CH stretching was identified in

HER2+ grade 1 according to the score plot presented in Figure 27A followed by TNBC and luminal grade 1 samples.

The peak found at ≈2968 cm^{-1} associated with fatty acids and certain contribution of CH_3 symmetric stretching of side chain lipids, and nucleic acids. [149]The location of the samples on the PC-2axis revealed that TNBC showed a higher overall contribution of lipids based on this peak, whereas grade 1 luminal and HER2+ samples presented a more dynamic trend.

Finally, a peak at different wavenumber (2920 and 2927 cm^{-1}) but correlated with the same assignment was found in both loadings. For discussion purposes, the location will be referred to as ≈2920 cm^{-1}. This peak is representative of fatty acids by representing the asymmetric stretching of CH_2 and CH_3 in lipids. [121] The changing behaviour of this band in the spectral profile of all subtype grade 1 samples indicate that the role play by fatty acids is of vital importance in the early cancer cases. On the other hand, the differences in intensity and shape of these peaks indicate the varying content of fatty acids.

5.1.1.2 Amide region (1,700-1,500 cm^{-1}) analysis of Grade 1 samples all subtypes

When the amide region of grade 1 samples was analysed with principal component analysis showing good grouping for all three subtypes and good discrimination of HER2+ rich samples. An important shift can be appreciated in the spectral profiles HER2+ as it can be seen in Figure 28B. The bands responsible for the grouping of the subtypes are 1674, 1635, 1613, 1523 cm^{-1}.

The peak found at ≈1674 cm^{-1} representing C=O stretching, usually presented in guanine and deformation of N-H in-plane [149] evidenced similarities between luminal and TNBC grade 1 samples. However, the shift presented on the HER2+ subtype spectral profile is responsible for the distribution of those samples in the opposite axis of PC-2 suggesting higher absorbance of this peak. On the other hand, based on the location of the peak in the luminal and TNBC shifted spectral profile might suggest an increased content of unordered random coils and turns. [102]

Structural changes in the proteins presented on these cancer cases were responsible for the discrimination of the samples. The peak located at ≈1635 cm^{-1} has been associated with β-sheet structures of amide I and its presented as a shoulder. [166] The contribution of this peak is similar in the luminal and TNBC. The grade 1 HER2+ samples are spread on the opposite axis, suggesting lower absorbance of this peak. Hence, less content of β-sheet structures. Furthermore, the spectral profiles of the grade 1 luminal and TNBC showed wider amide I bands in comparison with the grade 1 HER2+ as a result of the low contribution of the β-sheet shoulder. The previous statement is validated by the peak at ≈1523 cm^{-1} which also represent the β-sheet but of amide II vibrational modes. [102] The higher contribution of this band on the hormonal subtypes can be easily appreciated with the naked eye as a shoulder that seems missing in the HER2+ subtype.

Peaks found in the amide region also provide insights into the nucleic acids biochemical profile of the cancer cases. This is the case for the peak located at ≈1613 cm^{-1} which represent the N-H adenine vibrations in DNA. [167] Luminal and TNBC grade 1 samples presented higher

intensities of this band, also affecting the width of the convoluted band, in comparison with HER2+ grade 1 samples. The similitude between the hormonal subtypes suggests a similar nucleic acid production behaviour at lower grades, whereas a slower overproduction of genetic material seems to take place in the HER2+ grade 1 samples.

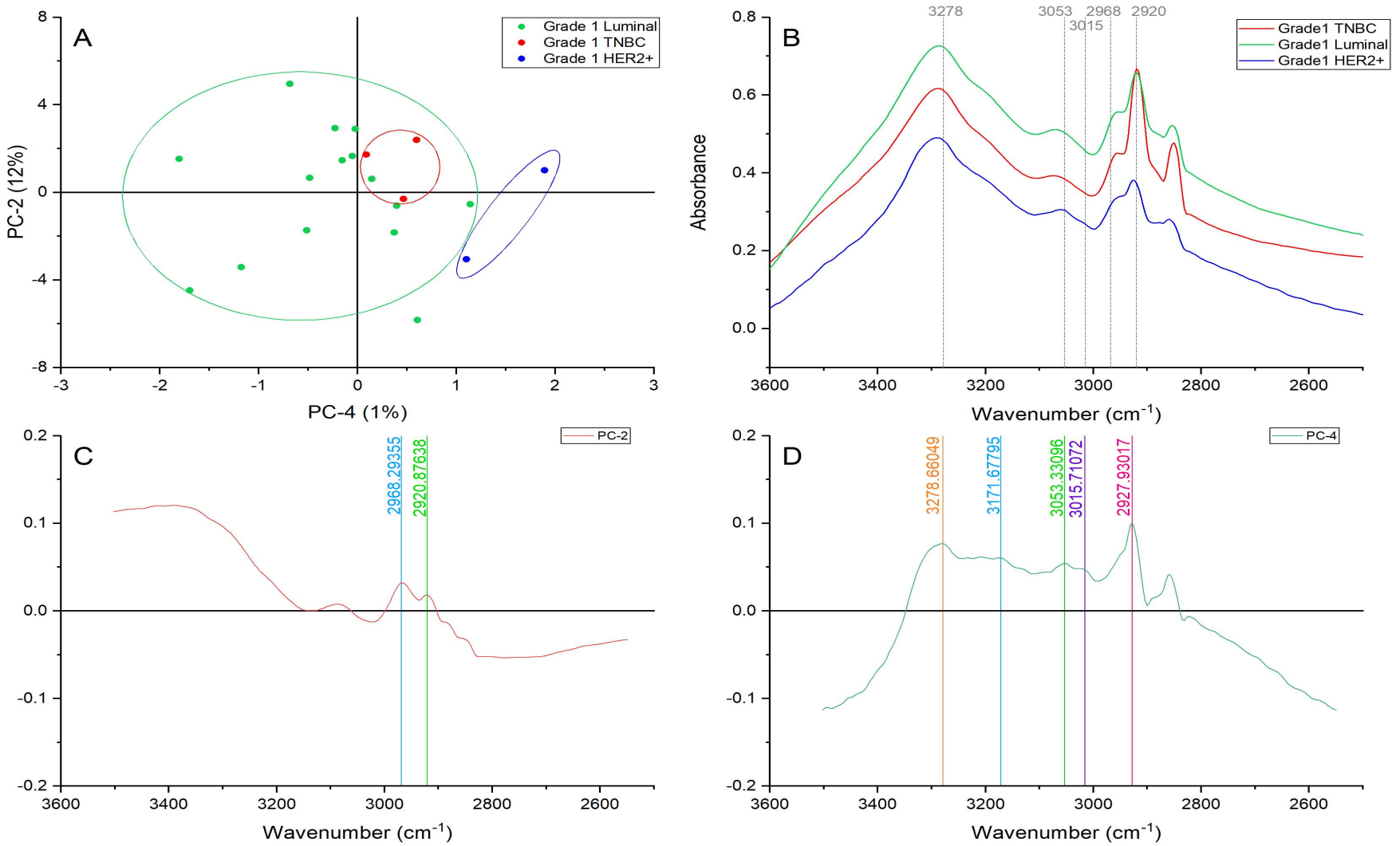

Figure 27. Principal component analysis of the higher wavenumber region (3,500-2,550 cm^{-1}). (A) Score plot using PC-2 and PC-4 accounting for 13% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the higher wavenumber region for grade 1 TNBC, luminal, and HER2+. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-2 Loading, (D) PC-4 loading.

5.1.1.3 Fingerprint region (1,450-600 cm^{-1}) analysis of Grade 1 samples all subtypes

The fingerprint region principal component analysis revealed differences among the three subtypes in the grade 1 samples. However, a relative separation of the samples was achieved (Figure 29) The listed peaks were identified as differentiators of all three subtypes: 1394, 1233, 1162, 1114, 1067, 952 cm^{-1}.

Important similarities between the luminal and TNBC subtype spectral profiles can be appreciated in Figure 29B. Most of the assignment associated with the differentiating peaks correspond to nucleic acids vibrations. Higher absorbance was detected on the samples representing the luminal and TNBC subtypes. For example, ≈1394 cm^{-1} representing the stretching of C-N in cytosine and guanine [167].The peak at ≈1233 cm^{-1} reflecting the PO_2^- asymmetric stretching of nucleic acids [164] presented higher intensities for the positive and negative hormonal subtypes. However, the peak presented in the HER2+ spectral profile is sharper and better defined suggesting important contributions of the nucleic acids in grade 1 HER2+ cancerous cases. Other nucleic acids peaks were also identified, the band at ≈1114 cm^{-1} associated with the phosphate groups in RNA [167], the peak at ≈952 cm^{-1} reflecting the C-O and C-C vibration in deoxyribose, which also has been reported as DNA vibrations of tumour cells. [164] Finally, the peak located at ≈1067 cm^{-1} corresponding to PO_2^- symmetric stretching of nucleic acids. The luminal and TNBC samples spectral profile presented higher intensities, but also a wider peak at this wavenumber.

The activity and contribution of proteins were also identified as an important differentiator on the grade 1 cancerous cases. The peak located at ≈1162 caused by the C-O stretch mode of C-OH groups of serine, threonine, and tyrosine of proteins presented in the HER2+ spectral profile revealed a better-formed peak in this wavenumber possibly reflecting structural changes for this biochemical component. Besides, essential eukaryotic protein serine and threonine phosphatase regulate important cellular functions involved in carcinogenesis such as cell division and the synthesis of proteins, and glycogen metabolism highly involved in survival during hypoxic conditions. [168] The activity of proteins was also evidenced in the luminal and TNBC subtype which presented higher intensities. However, this peak seemed less clear and diluted by neighbour contributions.

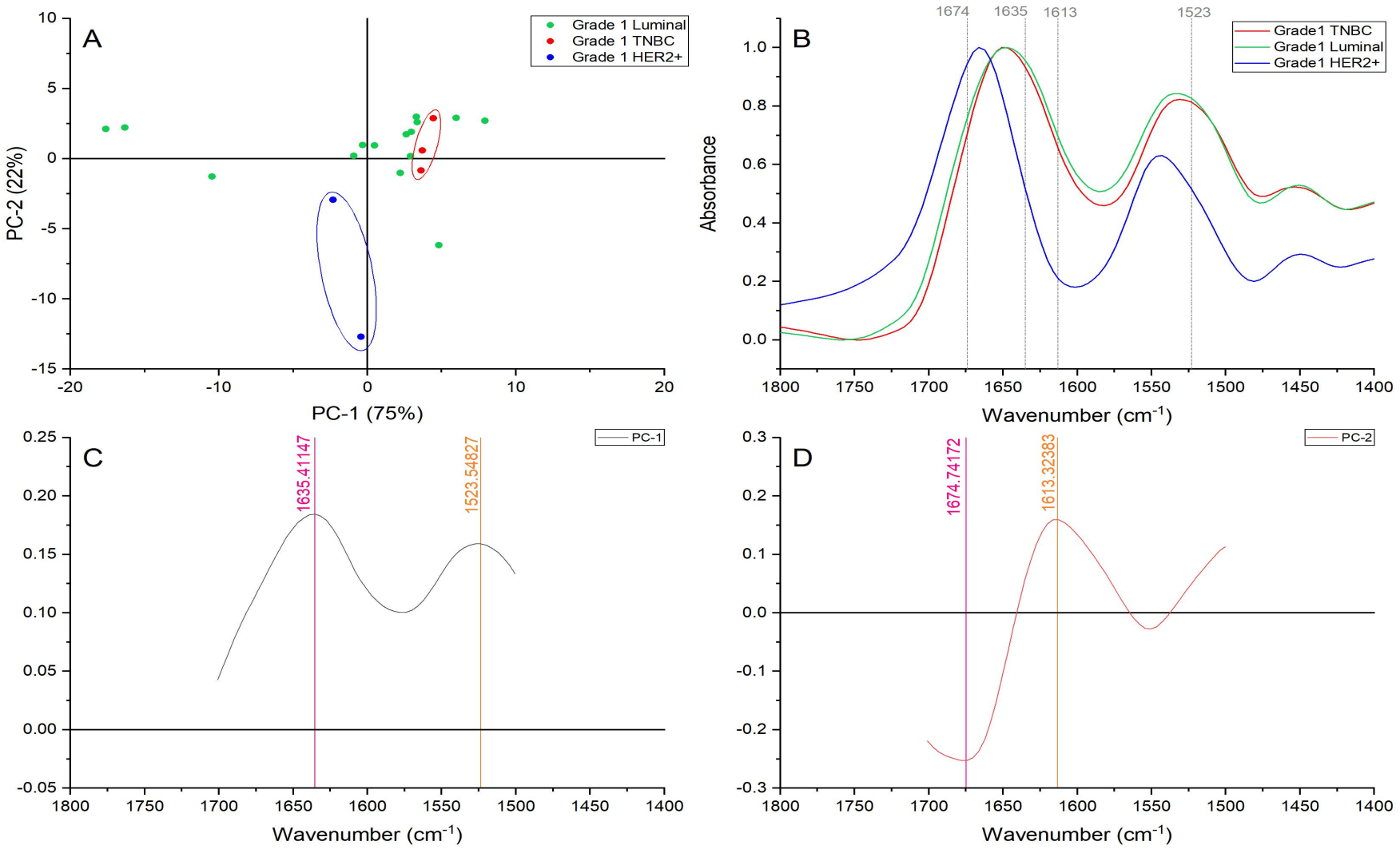

Figure 28. Principal component analysis of the amides region (1,700-1,500 cm^{-1}). (A) Score plot using PC-1 and PC-2 accounting for 97% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the higher wavenumber region for grade 1 TNBC, luminal, and HER2+. 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

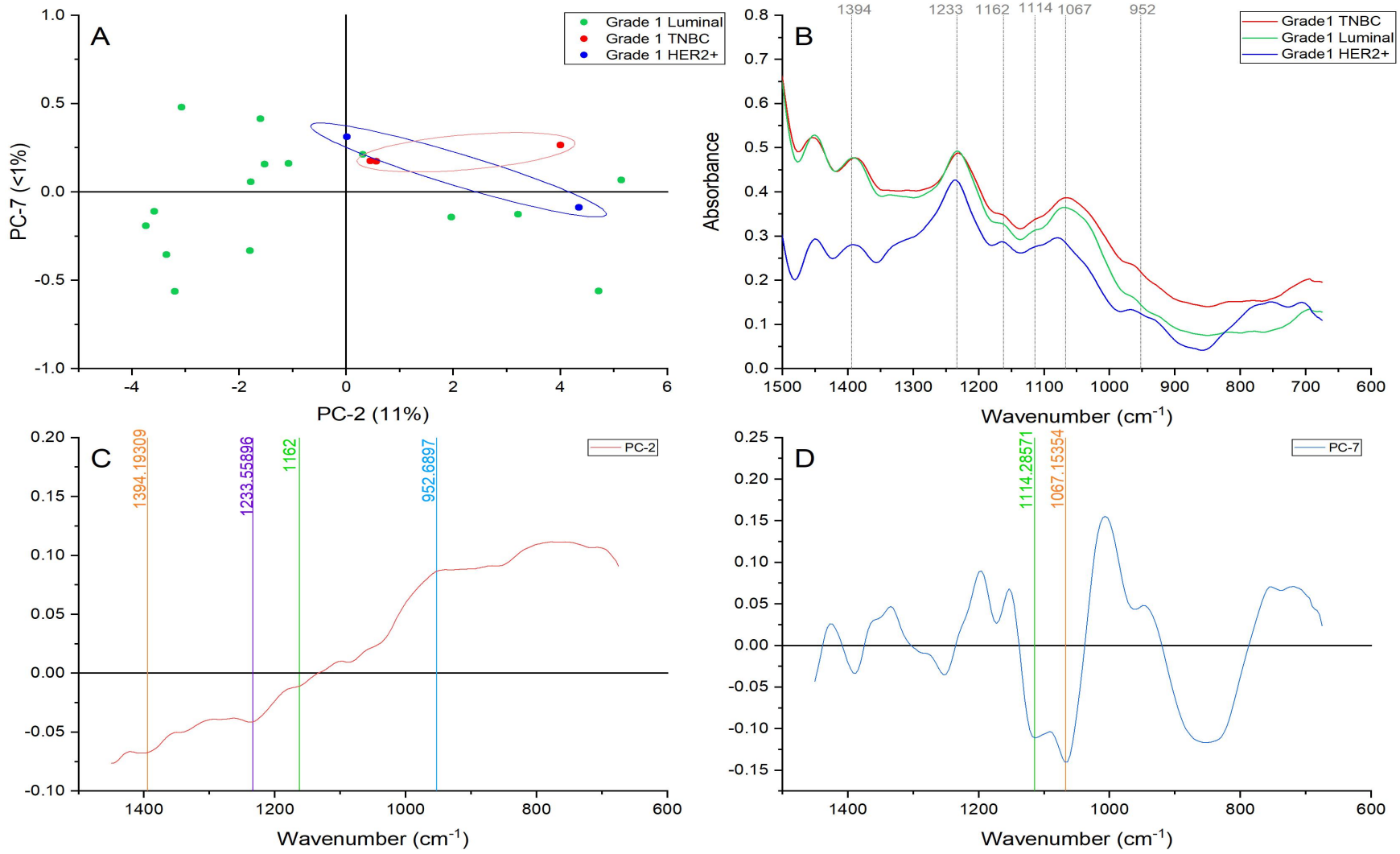

Figure 29. Principal component analysis of the fingerprint region (1,450-600 cm^{-1}). (A) Score plot using PC-2 and PC-7 accounting for 97% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the higher wavenumber region for grade 1 TNBC, luminal, and HER2+. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-2 Loading, (D) PC-7 loading.

5.1.2 Grade 2 all subtypes

When compared, the average spectral profile of grade 2 samples presenting luminal, HER2+ rich and triple negative breast cancer (TNBC) subtypes showed differences enough to see with the naked eye as presented in Figure 30 and Figure 31. Grade 1 luminal and TNBC samples presented similar spectral profiles and the same trend was detected in grade 2 samples in our cohort.

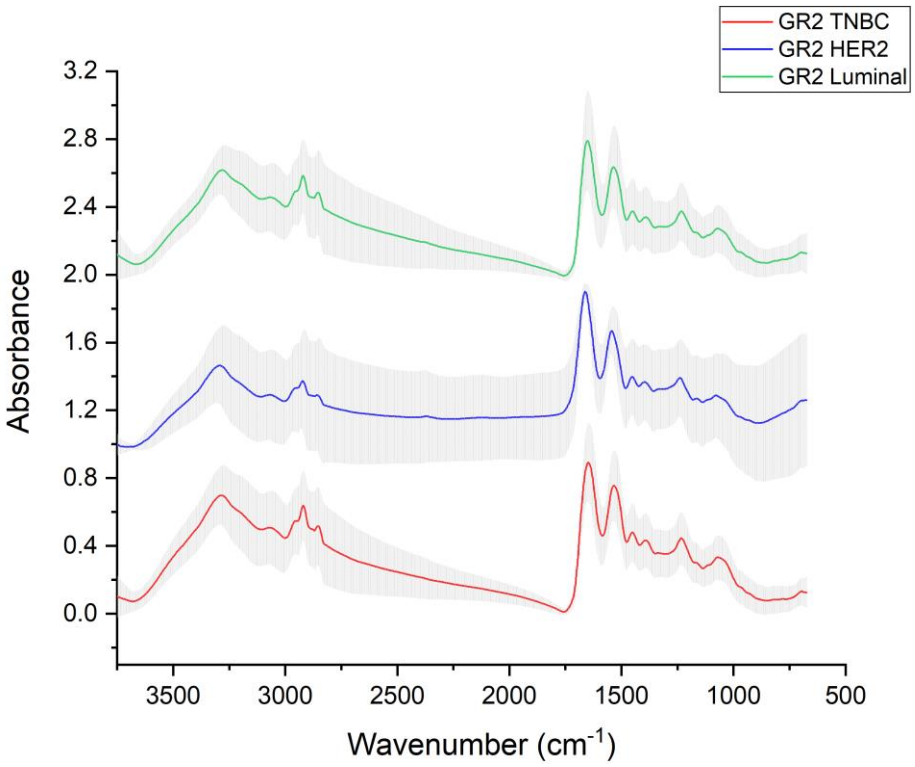

Figure 30. FTIR average spectral profile of grade 2 samples presenting triple negative (TNBC), luminal and HER2+ subtypes, variance is represented with grey shadow.

5.1.2.1 Higher wavenumber (3,500-2,550 cm^{-1}) analysis of Grade 2 samples all subtypes

The principal component analysis of the higher wavenumber region revealed groups but not a completely visible separation as can be seen in Figure 32. Five peaks were detected as a differentiator among groups. The peaks were located at ≈3151, 3024, 2948, 2914, and 2848 cm^{-1}.

The shoulder found at ≈3151 cm^{-1} associated with the symmetric stretching of N-H [149] is presented with higher intensities in the luminal and TNBC. The intensities presented in the HER2+ grade 2 samples indicate that overall the contribution of this peak is lower than the average in comparison with the other two subtypes suggesting that the hormonal subtypes have higher protein production at this grade.

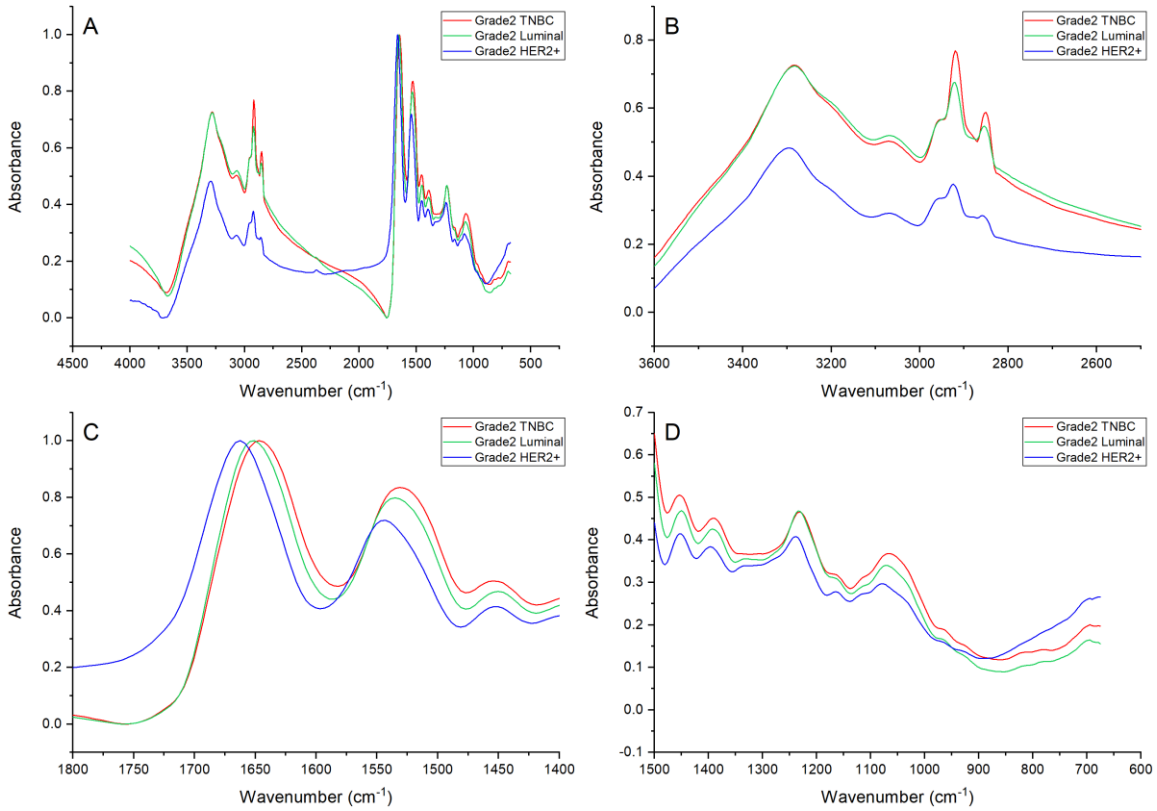

Figure 31. FTIR average spectral profile of grade 2 samples presenting triple negative (TNBC), luminal and HER2+ subtypes. (A) General region (4,000-675 cm^{-1}), (B) High wavenumber region (3,500-2,550 cm^{-1}), (C) Amide region (1,700-1,500 cm^{-1}), and (D) Fingerprint region (1,450 -600 cm^{-1}).

The band placed at ≈2948 cm^{-1} represent the C-H stretching that can be presented in a variety of biomolecules. [149] TNBC grade 2 samples presented a better grouping maintaining at the centre of the PC-1 axis and demonstrating more uniform biochemical changes in this subtype in grade 2 cases. Less C-H stretching vibrations were revealed by the lower intensities in the HER2+ grade 2 samples. Variations on this peak are more active in the luminal and TNBC reflecting more biochemical activity on these subtypes in comparison with HER2+ cases.

Vibrations related to lipids, cholesterol and creatine were identified twice with the peaks located at ≈2914 and ≈2848 cm^{-1}. These bands only showed higher intensities in the luminal and TNBC subtype, whereas the HER2+ subtype samples are all located in the lower than average area. This finding suggests that a more active lipid metabolism take place in the hormonal subtypes in grade 2 samples in comparison with the HER2+ subtype.

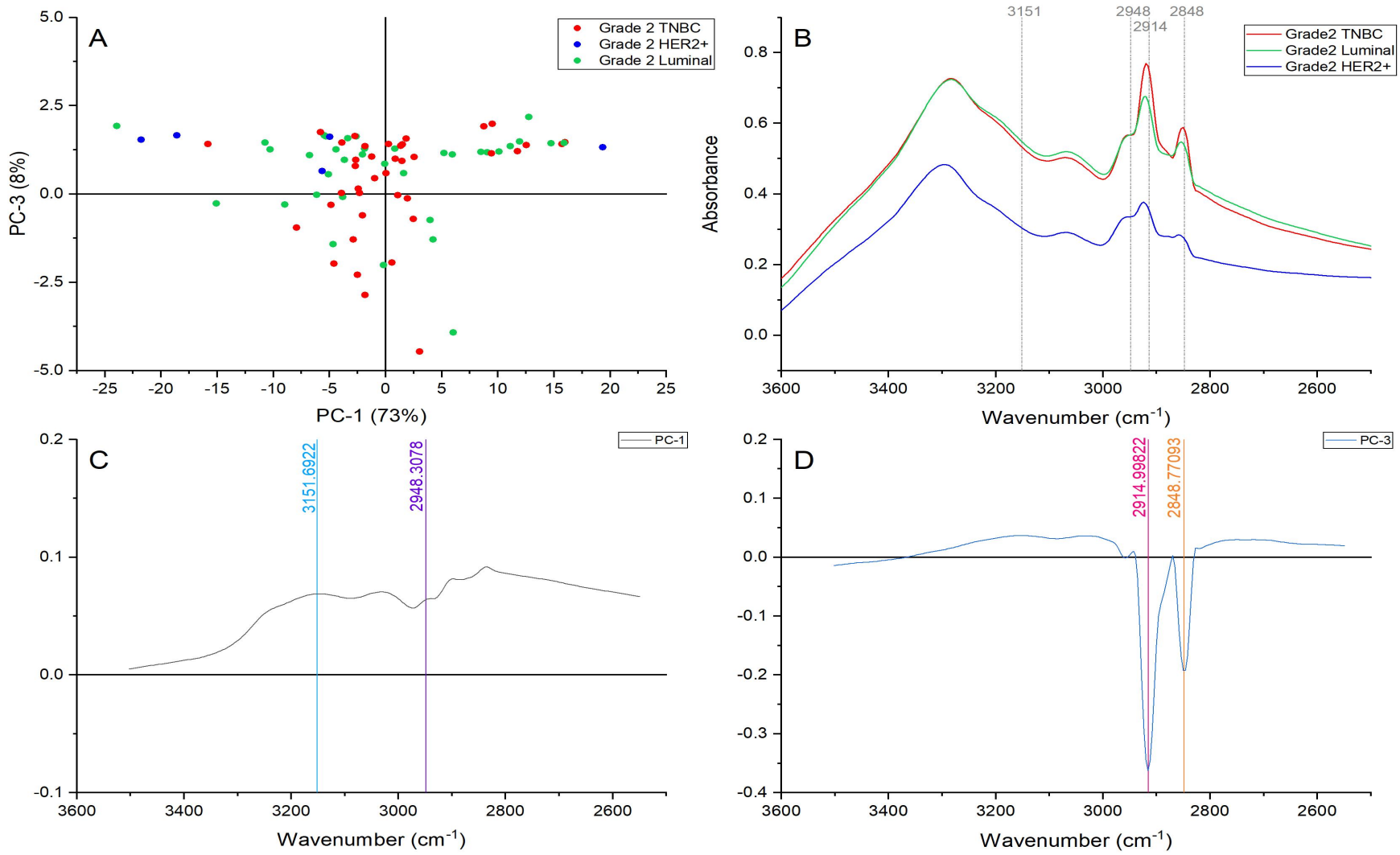

Figure 32. Principal component analysis of the higher wavenumber region (3,500-2,550 cm^{-1}). (A) Score plot using PC-1 and PC-3 accounting for 81% of the variance, (B) Average spectral profile of the higher wavenumber region for grade 2 TNBC, luminal, and HER2+. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-1 Loading, (D) PC-3 loadings.

5.1.2.2. Amide region (1,700-1,500 cm^{-1}) analysis of Grade 2 samples all subtypes

Clear grouping and separation were achieved among the grade 2 samples when the amide region was analysed with principal component analysis as presented in Figure 33A. The PC-1 seems to differentiate HER2+ samples by discriminating them on the positive axis.PC-1 helped identified the band at ≈1635 cm^{-1} associated with β-sheet structures (presented as a shoulder) [102] and ≈1524 cm^{-1} which represent the C=N stretching in guanine. [167] HER2+ rich samples presented higher intensities for these peaks, which might indicate more content of β-sheet secondary structures in comparison with the other subtypes and changes associated with the overproduction of nucleic acids. Additionally, an important upshift can be appreciated in the amide region of the HER2+ grade 2 spectral profile presented in Figure 33B.

On the other hand, PC-5 distinguished between TNBC samples, mostly in the positive axis, and luminal samples spread in the negative axis. The loading associated with this principal component identified the following peaks as responsible for the separation: 1667, 1612, and 1559 cm^{-1}.

The band located around ≈1667 and 1559 cm^{-1} presented higher intensities on the luminal grade 2 samples and half of the HER2+ grade 2 cases. These bands are representative of amide I β-turns of proteins and α-sheet amide II.[102] The differentiation of subtypes based on structural changes suggest structure alterations most likely due to the replacement of β-sheets with α- and random coils as suggested by Rehman et al. [106] Furthermore, the peak located at ≈1613 cm^{-1} associated with C=C and C=O vibration in uracil [167] presented higher intensities and therefore a major contribution in the TNBC grade 2 samples and half of the HER2+ grade 2 samples. These results suggest that TNBC grade 2 cases focus on the overproduction of genetic material to promote proliferation, whereas luminal grade 2 samples present considerable structural changes in their proteins. Structural changes in proteins are expected upon ligand binding in order to promote and facilitate different functions such as signalling. [169]

5.1.2.3. Fingerprint region (1,450-600 cm^{-1}) analysis of Grade 2 samples all subtypes

When principal component analysis was applied to the fingerprint region of the grade 2 samples a separation among luminal and TNBC was achieved using the PC-5, while the separation and grouping of the HER2+ grade samples were achieved using the PC-3 (Figure 34).

PC-5 identified the ≈1377 cm^{-1} and ≈1024 cm^{-1} peaks presented higher intensities in the TNBC and lower intensities in luminal and HER2+ grade 2 samples. These peaks are representative of C-H deformation and C-O stretching associated with glycogen.

In order to supply energy and provide glucose homeostasis, glucose is stored in cells as glycogen. Under stress processes such as hypoxia, cancer cells adapt to maintain and promote tumour growth. Glycogen plays a vital role in hypoxic conditions promoting cell survival. Furthermore, the glycogen metabolism has been proposed as the tumour and tumour

microenvironment energy source also having implication in treatment, chemotherapy and radiation, response. [168] The higher contribution of glycogen on the TNBC grade2 samples might be an indication of an accelerated proliferation process suggesting a more aggressive behaviour, which is characteristic of the TNBC subtype.

On the opposite, PC-5 helped to identify higher absorbance of the peaks located at ≈1242 and 1180 cm^{-1} on the luminal and HER2+ grade 2 samples. The first band is correlated with PO_2^- asymmetric stretching in nucleic acids [164], while the 1180 cm^{-1} peak is associated with cholesterol ester vibrations. [149] These findings suggest considerable changes and overproduction of genetic material in the luminal and HER2+ grade 2 samples. Whereas, the cholesterol ester vibrations might be related to changes in the cell membrane fluidity, which has been reported to require high contents of cholesterol on malignant tissue. [121] The lipid activity is highly relevant in several cellular activities. The variation in quantity and conformations affect the membrane fluidity having control of ligand-receptor interaction, endocytosis and antigen presentation which facilitate invasive processes. [157]

Principal component 3 provided an insight into the components presented with low intensities in the HER2+ grade 2 samples. These component are represented by the bands located at ≈1196, 1149, 1018, and 958 cm^{-1}. The 1018 cm^{-1} peak was interpreted as a shift of the previously described peak presented at ≈1024 cm^{-1} and therefore will not be discussed. The contribution at ≈1196 cm^{-1} is assigned to the phosphate (P=O) band on phospholipids, [121] and the lower intensity presented in the HER2+ grade 2 samples might indicate less active lipid changes in the membrane having less affectation in the fluidity of the membrane. The band located at ≈1149 cm^{-1} associated with the glycogen absorptions cause by the C-O and C-C stretching and the C-O-H deformation motions [149] confirmed a less active participation of the glycogen and its metabolism for the HER2+ subtype in grade 2 samples. This situation indicates less stress in the HER2+ cases, for example, no hypoxia or pre-hypoxia processes and overall less nutrient/energy requirement. [168] As a final point, the C-O deoxyribose vibrational mode [149] in the HER2+ grade 2 samples presented at ≈958 cm^{-1} showed noticeable differences in the spectral profiles in Figure 34B. This senescent contribution and in the spectra in combination with the findings presented in PC-3 suggested a less active proliferation in HER2+ grade 2 cases.

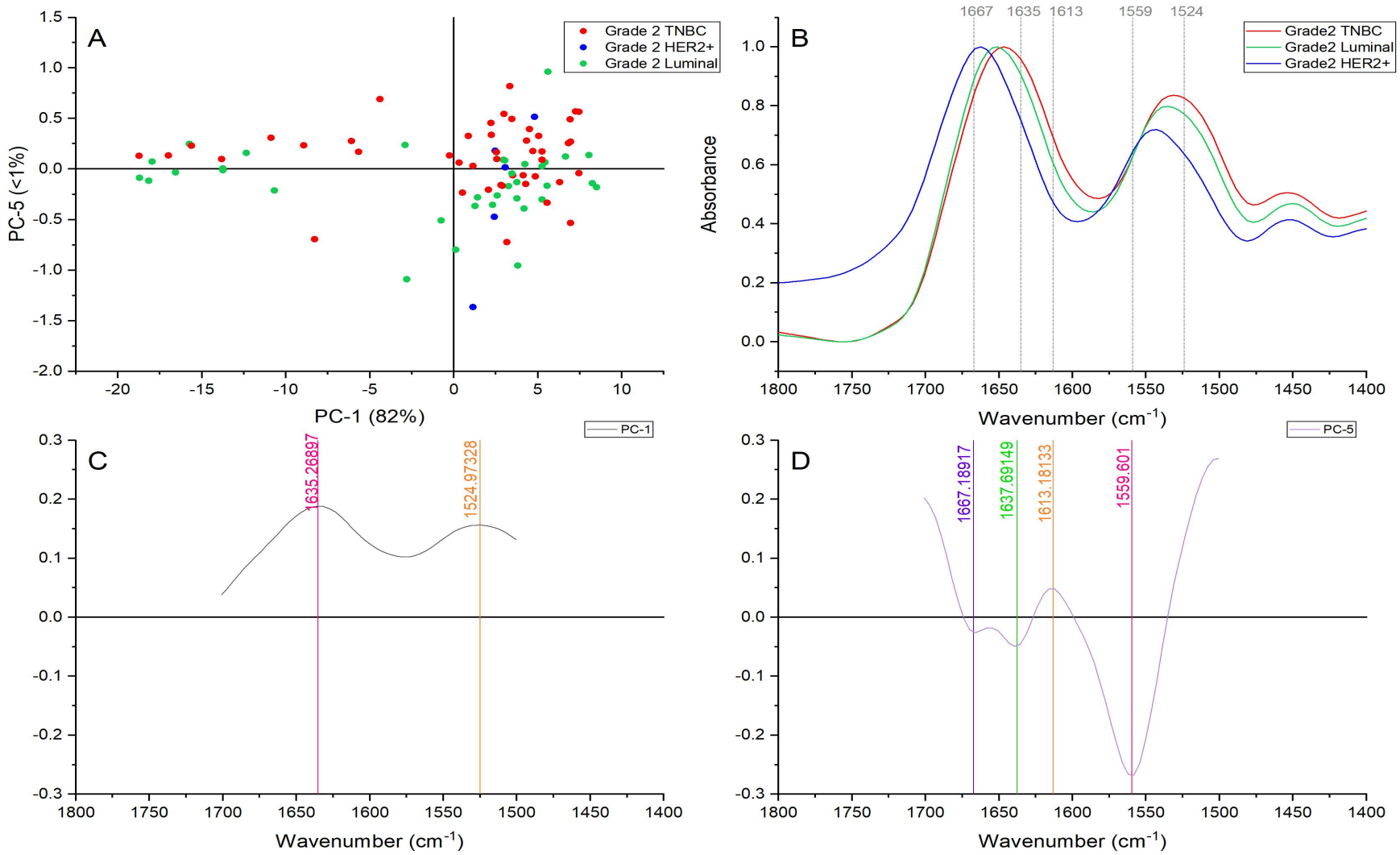

Figure 33. Principal component analysis of the amide region (1,700-1,500 cm^{-1}). (A) Score plot using PC-1 and PC-5 accounting for 83% of the variance, B) Average spectral profile of the higher wavenumber region for grade 2 TNBC, luminal, and HER2+. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-1 Loading, (D) PC-5 loading.

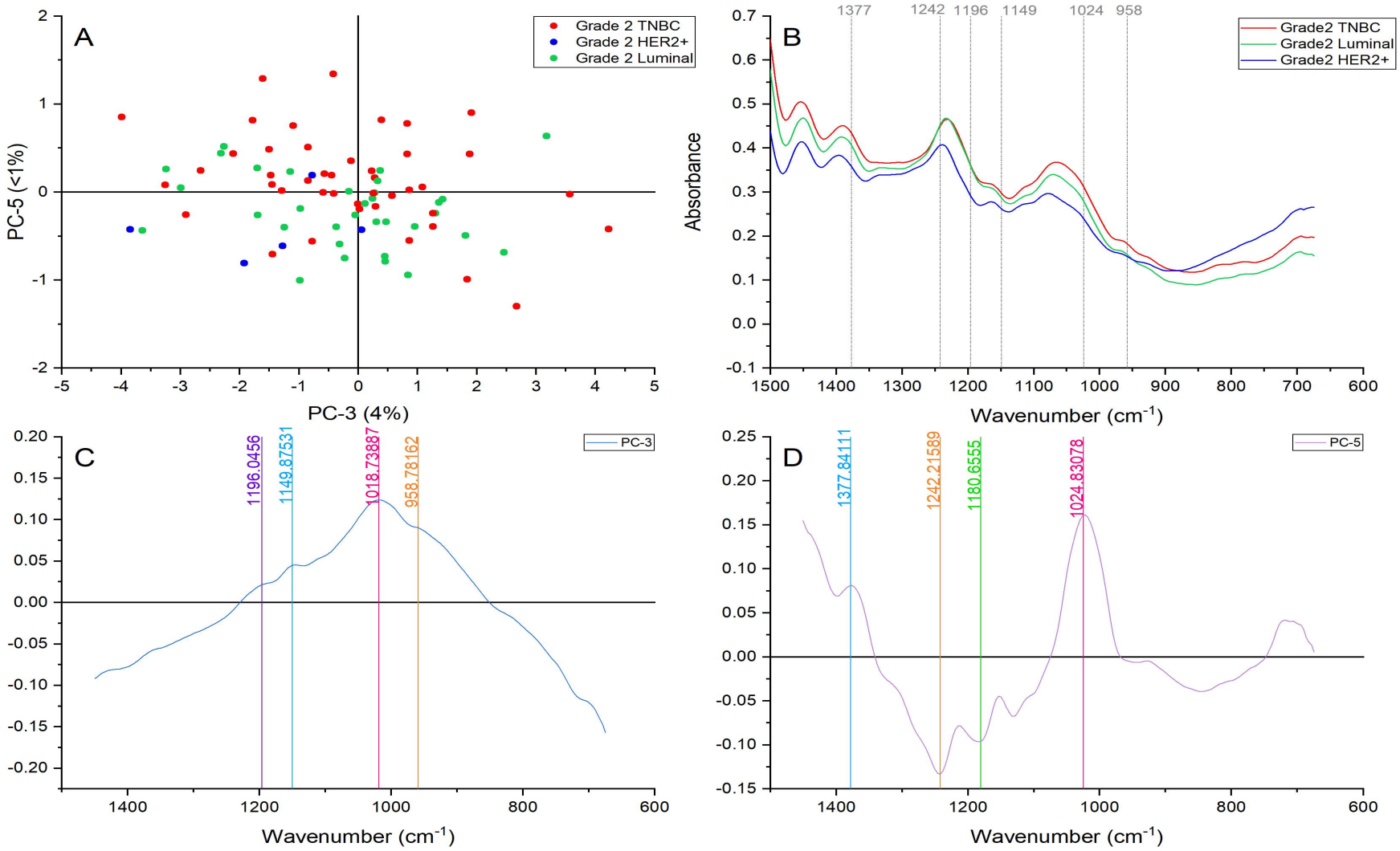

Figure 34. Principal component analysis of the fingerprint region (1,450-600 cm^{-1}). (A) Score plot using PC-3 and PC-5 accounting for 5% of the variance, (B) Average spectral profile of the higher wavenumber region for grade 2 TNBC, luminal, and HER2+. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-3 Loading, (D) PC-5 loading.

5.1.3 Grade 3 all subtypes

The average spectral profile of grade 3 samples presenting luminal, HER2+ rich and triple negative breast cancer (TNBC) subtypes showed clear differences that can be appreciated with the naked eye as presented in Figure 35and Figure 36. Grade 1 and grade 2 luminal and TNBC samples presented similar spectral profiles. However, the grade 3 samples in our cohort showed more similitudes on the TNBC and HER2+ rich subtypes. Differences among the three subtypes will be discussed in the following sections.

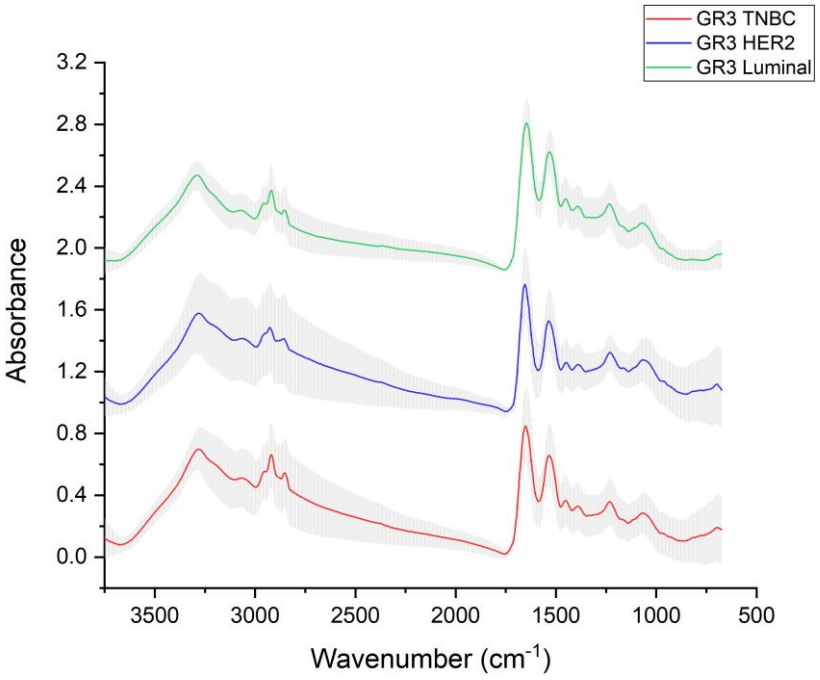

Figure 35. FTIR average spectral profile of grade 3 samples presenting triple negative (TNBC), luminal and HER2+ subtypes, variance is represented with grey shadow.

5.1.3.1 Higher wavenumber (3,500-2,550 cm^{-1}) analysis of Grade 3 samples all subtypes

The principal component analysis of the higher wavenumber revealed a separation between luminal and TNBC grade 3 samples. In addition, a clear grouping of sample points can be seen in all three subtypes as it can be seen in Figure 37.

Principal component 1 identified the following peaks as responsible for the separation of the hormonal subtypes: 3023, 2971, 2896, and 2896 cm^{-1}. For all these peaks, a higher contribution represented by higher intensities were identified in the TNBC placing the luminal in the opposite trend with the HER2+ samples lying in between.

The band at ≈3023 cm^{-1} representing aromatic C-H stretchings while the band at ≈2835 cm^{-1} is representative of the N-H stretching both accounting for amino acids and nucleic acids vibrations. [149] The peak located at ≈2971 cm^{-1} represents the asymmetric stretching of CH_3 from fatty acids, in combination with the symmetric stretching of the CH_3 located at ≈2896 cm^{-1} [170] demonstrate the highly active lipid activity in TNBC grade 3 samples. The accumulation of fatty acids in TNBC suggest aggressive malignant activities that need to be

fuel through these lipid components. When the spectral profile of the three subtypes are compared (Figure 37B), important contributions of the 2917 cm^{-1} band are recognised on all the subtypes with important differences in the sharpness of the shoulder-shaped peak. The differences in the acuteness of the shoulder suggest important changes in the chemical composition of the fatty acids which is indicative of their participation towards more invasive processes.

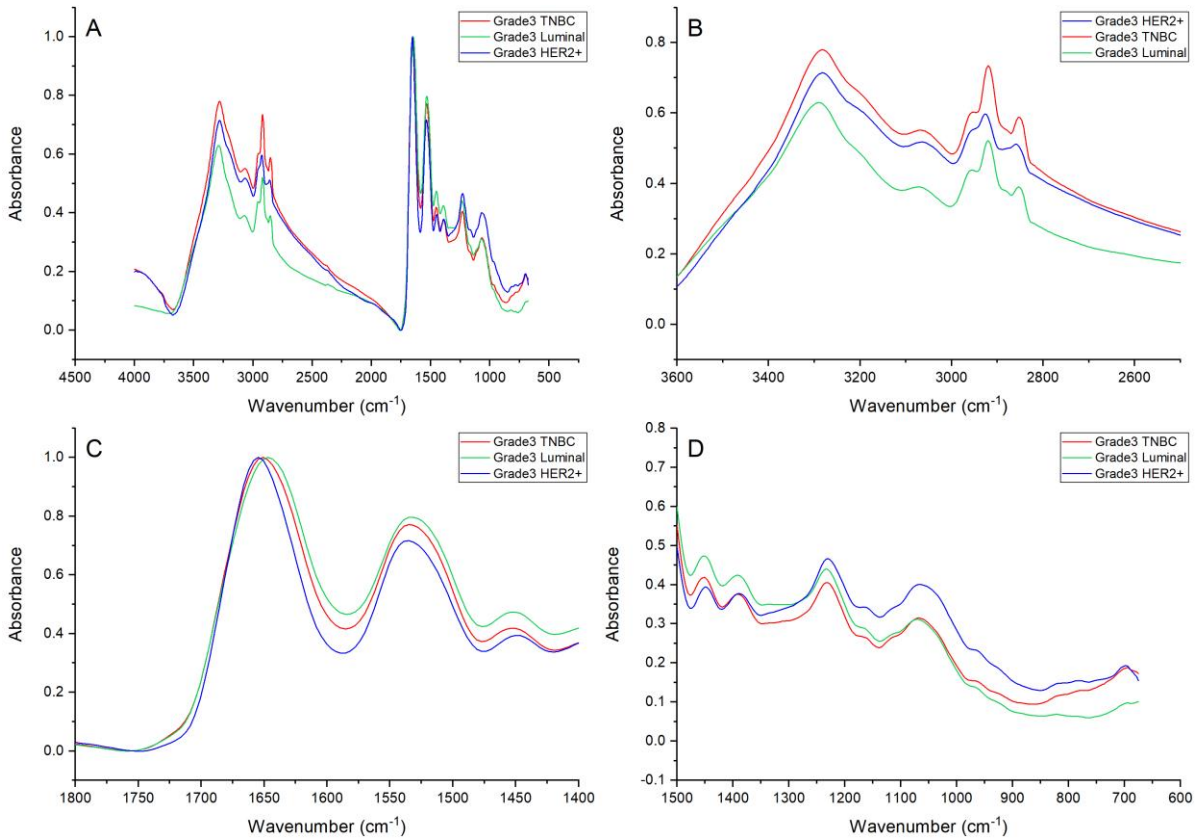

Figure 36. FTIR average spectral profile of grade 3 samples presenting triple negative (TNBC), luminal and HER2+ subtypes. (A) General region (4,000-675 cm^{-1}), (B) High wavenumber region (3,500-2,550 cm^{-1}), (C) Amide region (1,700-1,500 cm^{-1}), and (D) Fingerprint region (1,450-600 cm^{-1}).

Principal component 2 revealed a better dispersion of the TNBC and HER2+ samples suggesting a heterogeneous behaviour in the TNBC and HER2+ grade 3 samples. On the contrary, a clear grouping along the PC-2 axis can be seen for the luminal subtype in Figure 37A. The loading associated with PC-2 showed the peaks located at ≈3326, 2965, 2919, and 2870 cm^{-1} as responsible for the grouping.

The peak located at ≈3326 cm^{-1} has been reported as an asymmetric N-H stretching vibration corresponding to amide A. [102] The ≈2965 cm^{-1} has been interpreted as the asymmetric stretch CH_3 in fatty acids and was discussed previously as the 2971 cm^{-1} band. The peak located at ≈2919 cm^{-1} represents the cholesterol ester vibrations. [149] The TNBC samples showed a wide range of dispersion on the PC-2 axis confirming variable contribution of these peaks. The luminal samples presented intensities closer and higher to the average, whereas the HER2+ tended to distribute equally on both sides. The last peak was presented at ≈2870 cm^{-1} and

represent the CH_3 symmetric stretching of the protein side chain, and lipids with some contribution from carbohydrates and nucleic acids. [149]

5.1.3.2. Amide region (1,700-1,500 cm^{-1}) analysis of Grade 3 samples all subtypes

Principal component analysis of the amide regions showed good grouping for luminal and HER2+ grade 3 samples as presented in Figure 38 when PC-1 and PC-2 were used. The higher intensity in the peak located at ≈1636 in luminal grade 3 associated with structural changes suggested several adaptations towards β-sheets conformations. HER2+ grade 3 samples presented the lowest contribution of this secondary structure conformation blue shifted while the TNBC showed lower and higher than the average absorbance of this particular conformation, this can be appreciated in the spectral profile presented in Figure 38B. Thus, a reduction of the α-helix structure and increase in β-sheet structure can be assumed; this statement is consistent with the formation mechanism of β-sheet structures and are reflected in the amide peak shape. [171] Amide conformational changes have been proposed as part of the carcinogenesis pathological adaptations. [172]

Similar overall behaviour has been found for the 1526 cm^{-1} band corresponding to the C=N guanine vibration. This band suggests that luminal grade 3 samples present consistent higher than the average absorbance while TNBC showed higher and lower intensities and HER2+ samples kept in the lower absorbance axis. Changes in the conformation of guanine have been reported as part of the carcinogenesis process. These RNA changes affect primary, secondary and tertiary structures and therefore disturb the reparative mechanisms of the cell and tissue. [167]

Principal component 2 identified the peak representing Amide I located at ≈1661 cm^{-1} and the adenine vibration in DNA represented by the ≈1607 cm^{-1} peak. The vibration associated with the amide I is higher in absorbance for the TNBC and HER2+ grade 3 samples and can be appreciated in the spectral profiles presented in Figure 38B. The lowest contribution of amide I was presented in the luminal grade 3 samples; however, this can be explained by the red shift towards lower wavenumbers of this band in the luminal subtype. The second band associated with adenine vibrations presented higher intensities in the luminal grade 3 samples. This statement is in agreement with combine adenine and guanine adaptations reported by Dovbeshko et al. [167] TNBC samples showed a split behaviour with scattered samples among both axes, whereas the HER2+ showed the lowest intensities for this particular band suggesting lower nucleic pair bases modification in grade 3 samples

5.1.3.3 Fingerprint region (1,450-600 cm^{-1}) analysis of Grade 3 samples all subtypes

The analysis of the fingerprint region revealed a good grouping but poor separation of different subtypes of grade 3 samples. Principal component 1 and 4 showed the best results accounting for 58% of the variance. Nucleic acids were responsible for most of the peaks identified with these two PCs.

Principal component 1 revealed that luminal grade 3 samples showed higher intensities in the peaks located at ≈1382 cm^{-1} associated with δCH_3, stretching of C-O, and C-H and N-H deformation, 1222 and 1115 cm^{-1} representing asymmetric and symmetric stretching of phosphate groups in RNA, 1159 cm^{-1} associated with the C-O stretching of proteins and carbohydrates, and finally 1059 cm^{1} which is a representative peak of the triad nucleic acids bands by representing the C-O stretching of phosphodiester in ribose. [149] Half of the HER2+ grade 3 samples showed higher and lower intensities than the average. On the contrary, the TNBC grouping at the opposite side of luminal samples suggests lower contents of nucleic acids in comparison with the positive hormonal receptor subtype.

Similar biochemical behaviours were identified using the peaks highlighted by principal component 4 as responsible for the sample's grouping. The bands located at ≈1227 and 1065 cm^{-1} were previously discussed as 1222 and 1065. For these specific peaks, a better spreading of sample points was identified for HER2+ grade 3 samples affecting its spectral profile showed in Figure 39B. TNBC showed the best grouping by keeping closer to the Y-axis representing absorbance values closer to the average while luminal samples tended to distribute in the higher than the average intensity values region. Important contributions of nucleic acids components are expected on grade 3 samples, as important changes associated with poor differentiation and accumulation of genetic material for cell division and proliferation is considered a hallmark of high grades. In addition, changes in peak shape by the addition of a shoulder like the one presented at ≈1115 indicate a modification in the structural components of RNA which might facilitate invasive processes, but also affecting cell reparative mechanisms. [167]

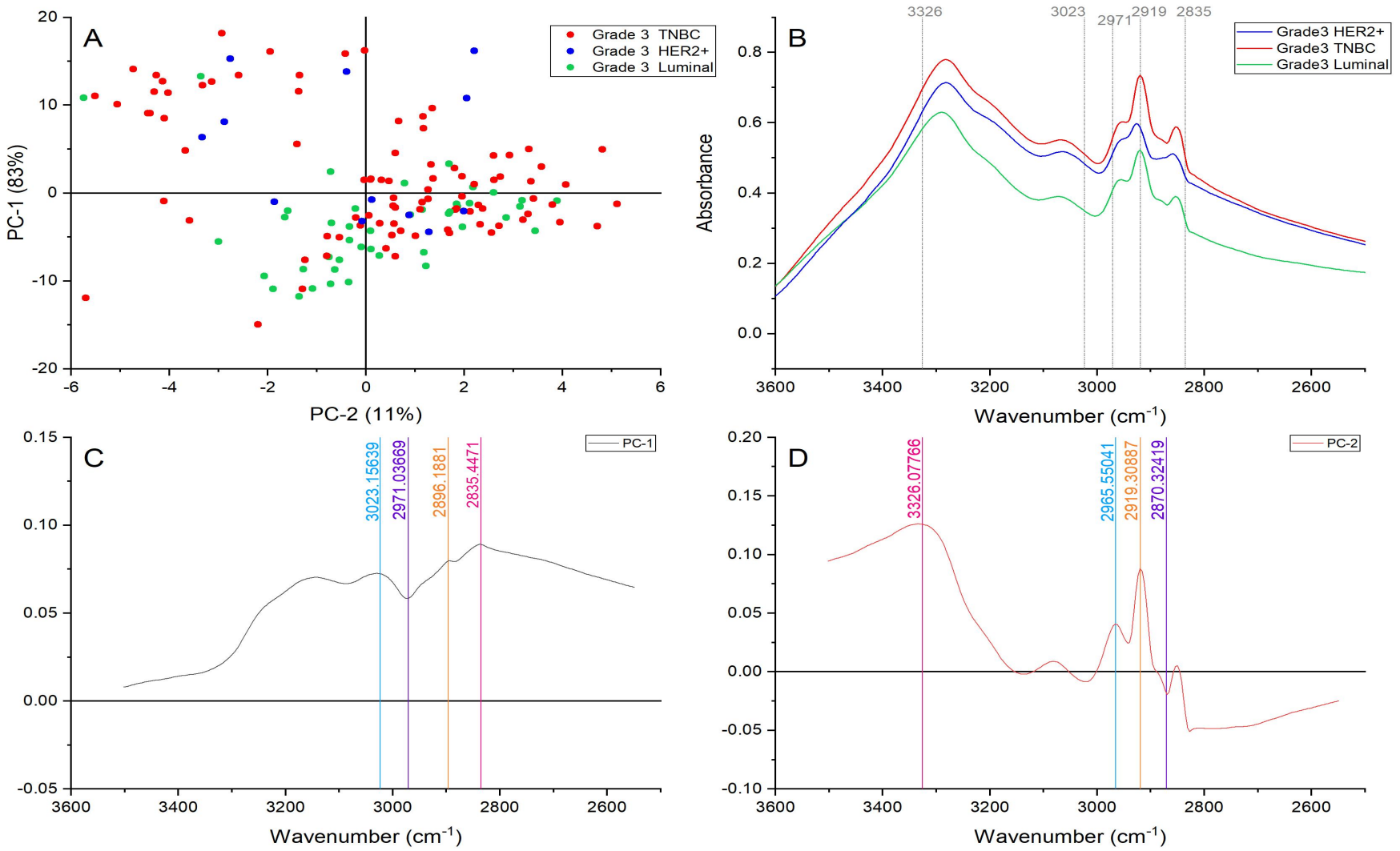

Figure 37. Principal component analysis of the higher wavenumber region (3,500-2,550 cm^{-1}). (A) Score plot using PC-1 and PC-2 accounting for 94% of the variance, (B) Average spectral profile of the higher wavenumber region for grade 2 TNBC, luminal, and HER2+. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-1 Loading, (D) PC-2 loading.

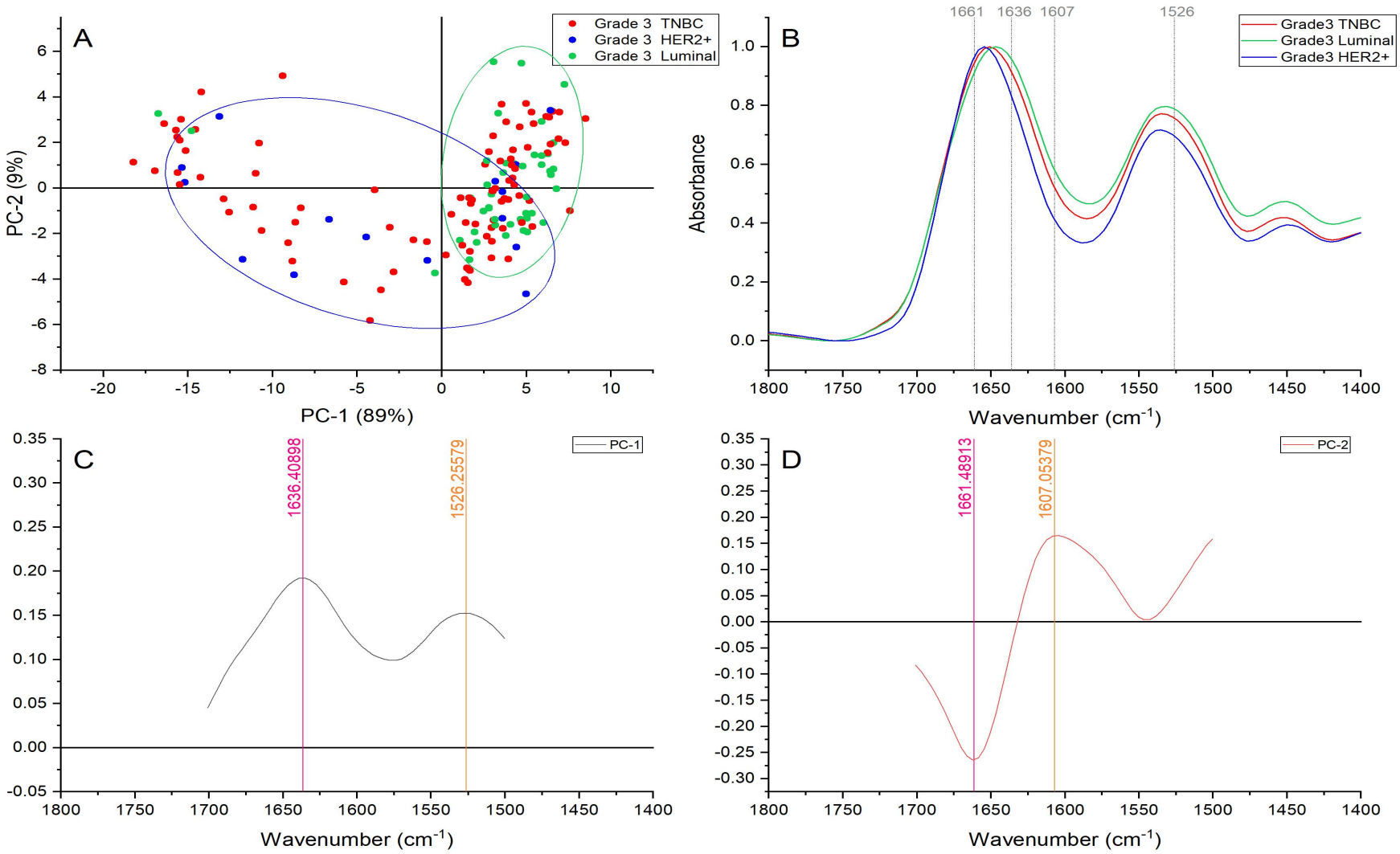

Figure 38. Principal component analysis of the amide region (1,700-1500 cm^{-1}). (A) Score plot using PC-1 and PC-2 accounting for 98% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the higher wavenumber region for grade 3 TNBC, luminal, and HER2+. 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 loadings.

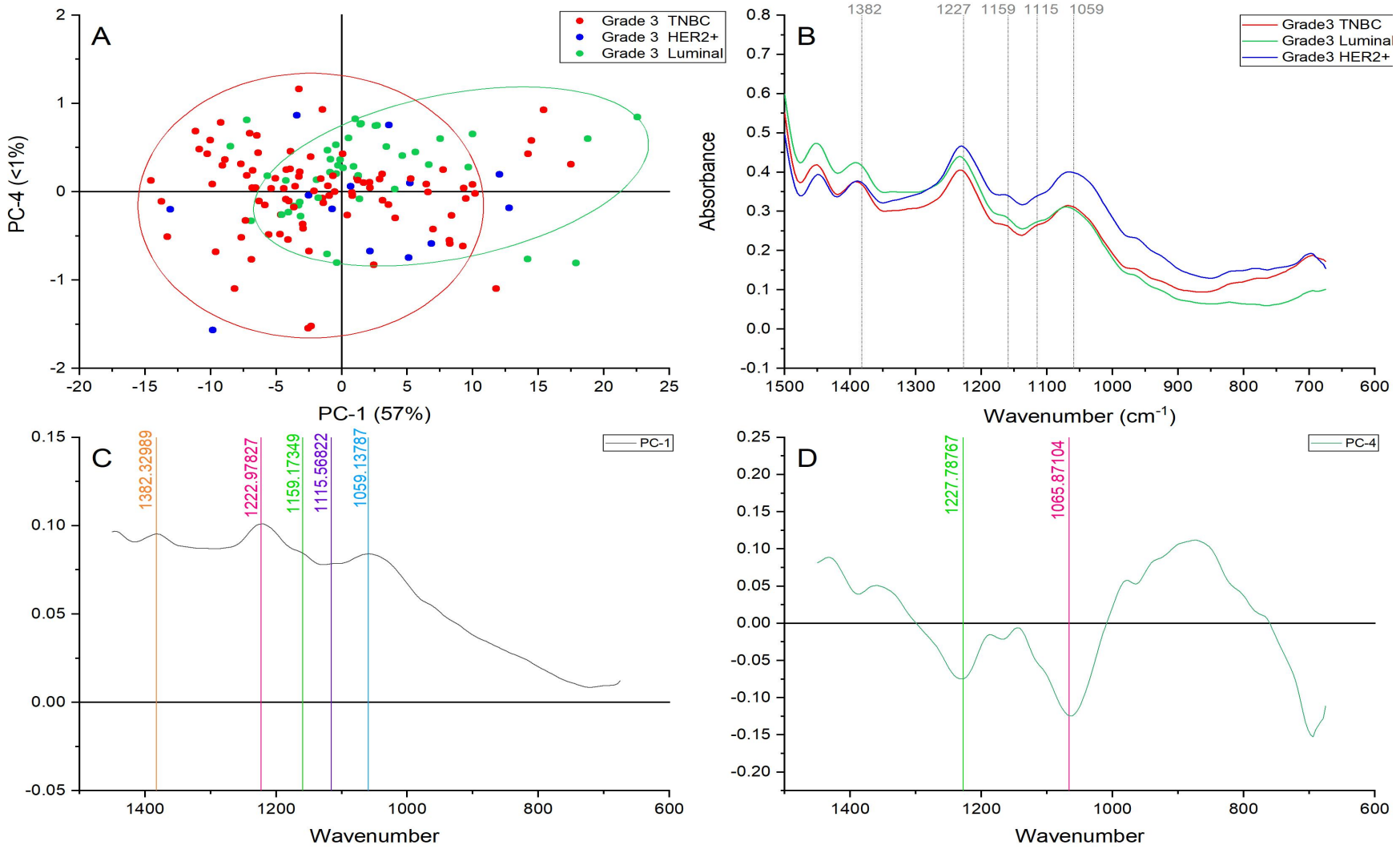

Figure 39. Principal component analysis of the fingerprint region (1,450-600 cm^{-1}). (A) Score plot using PC-1 and PC-4 accounting for 58% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the higher wavenumber region for grade 3 TNBC, luminal, and HER2+. 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-4 loading.

5.2 Comparison of the spectral data of different grades within a specific subtype.

5.2.1 Luminal samples all grades

When compared, the average spectral profile of luminal samples representing grade 1, 2, and 3 showed differences enough to see with the naked eye, as presented in Figure 40 and Figure 41. The higher wavenumber and fingerprint regions spectral profiles showed similitudes between grade 1 and 2 luminal samples, whereas the amide region showed that luminal grade 2 and 3 are more alike. The discussion of these specific areas is presented in the following sections.

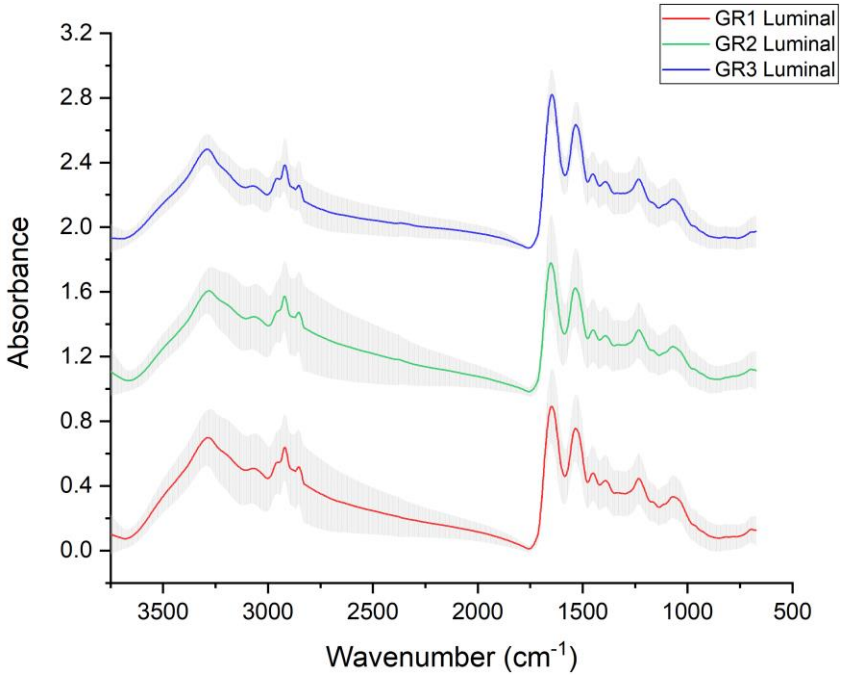

Figure 40. FTIR average spectral profile of luminal samples representing grade 1, grade 2, and grade 3 cases, variance is represented with grey shadow.

5.2.1.1. Higher wavenumber (3,500-2,550 cm^{-1}) analysis of Luminal samples all grades

Principal component analysis was used in the higher wavenumber (3,500-2,550 cm^{-1}) showing good grouping and separation based on the grade as it can be seen in Figure 42. The group of peaks located at 3027 and 3023 cm^{-1}, 2923 and 2925 cm^{-1}, and 2849 and 2854 cm^{-1} were identified in both PC-1 and PC-6 respectively. All these peaks are associated with lipid vibrations such as asymmetric stretching vibrations of CH_2 of acyl chains, CH_3 symmetric stretching, and stretching vibrations of CH_2 and CH_3 of phospholipids, cholesterol, and creatine. These bands showed higher intensities and contributions in grade 2 cases, followed by a stable grade 1 group placed closer to the average and reduced intensities in grade 3 cases. These results suggest a reduction in the lipid components as adopting an aberrant cell form with poor differentiation and reduced similarity to normal cells.

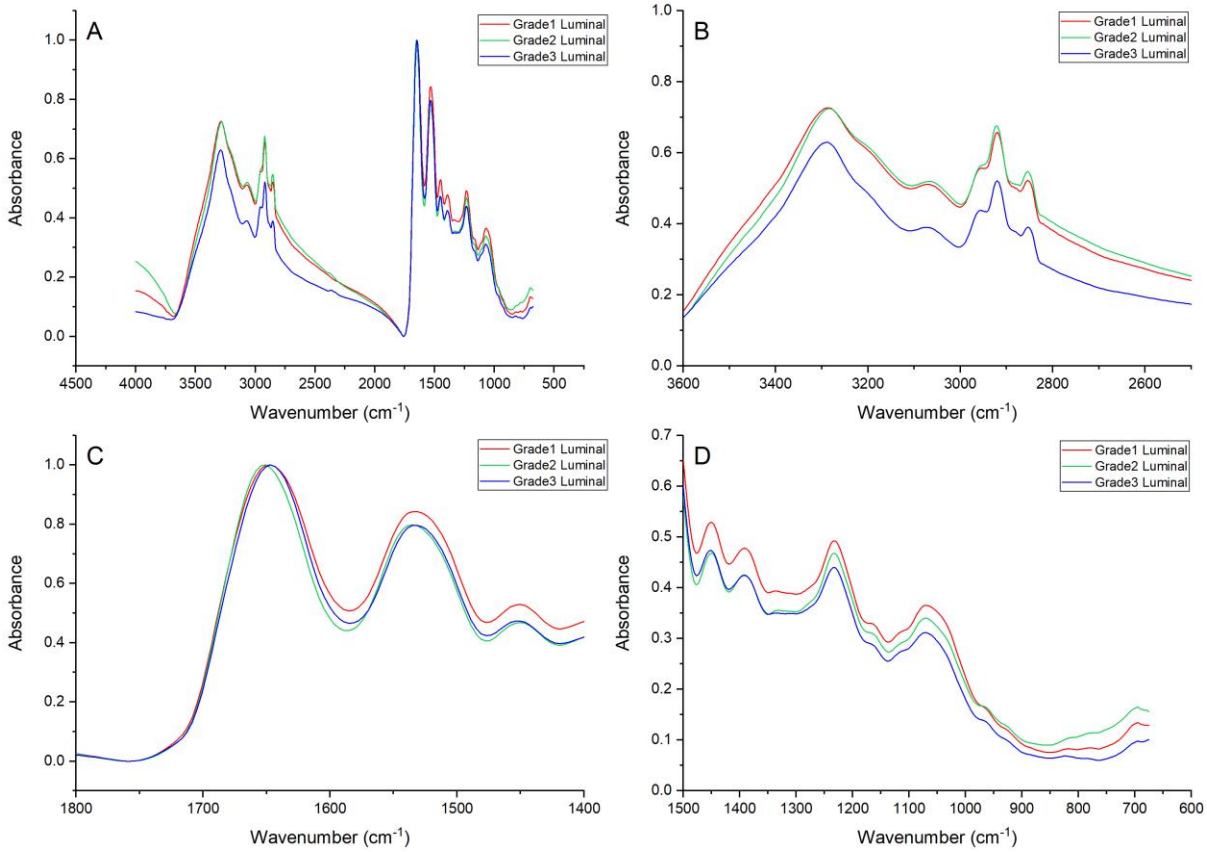

Figure 41. FTIR average spectral profile of luminal samples representing grade 1, grade 2, and grade 3 cases. (A) General region (4,000-675 cm^{-1}), (B) High wavenumber region (3,500-2,550 cm^{-1}), (C) Amide region (1,700-1,500 cm^{-1}), and (D) Fingerprint region (1,450-600 cm^{-1}).

Changes in the lipid components of grade 2 cases with luminal subtype might suggest cell membrane adaptations and changes which might affect the fluidity of the membrane in order to promote invasive processes. [172] Changes in the lipid metabolism have affectation not only in the structural aspect of membranes but also energy homeostasis, signalling and important cellular processes such as motility, proliferation, differentiation and cell growth. [160] Thus, lipid modifications are expected in grade 2 samples leading to a more aggressive behaviour than the grade 1 cells that resemble the closest to normal cells by changing the membrane fluidity and providing the energy resources for malignant activities. Tumours synthesise fatty acids de novo and actively look for its synthesis at early stages and its storage in lipid droplets. [173] Therefore, increased content of lipids is expected and has been confirmed in grade 1 and 2 luminal cases.

Complementarily, overexpression of antioxidant has been reported in several cancer cases accelerating and increasing the lipid peroxidation significantly in the malignant tissue. [174] Lipid peroxidation has been associated with the transformation and growth of tumours [175] to the point that drugs such as Tamoxifen have been used as a treatment by reducing the peroxidation in lipids. [174] As a consequence of the high oxidative processes, the lipid content tends to decrease as a product of its degradation as it has been identified in grade 3

luminal samples. Once a hypoxia state has been reached and there is a low content of lipids due to peroxidation cancer cells turn into glycogen as an energy source. [160]

Additionally, the peak located at ≈3293 cm^{-1} representing the amide A vibration associated with the N-H stretch was identified by PC-6. Grade 3 luminal cases presented a higher contribution of this band followed by grade 2 and 1. This vibration suggested that a higher content of proteins and nucleic acids were present in grade 3 samples probably representing the aberration and poor differentiation associated with higher grades.

5.2.1.2. Amide region (1,700-1,500 cm^{-1}) analysis of Luminal samples all grades

Principal component analysis was applied to the amide region of luminal samples representing grade 1,2, and 3 giving as a result a good grouping and differentiation using PC-3 and PC-4, as shown in Figure 43. The peaks located at ≈1689, 1595, and 1564 presented higher intensities in grade 1 luminal samples followed by grade 2 and 3 and were identified by PC-3. These bands are associated with base carbonyl stretching and ring breathing mode in nucleic acids, the methylation of nucleotides, and the ring base vibration in nucleic acids, respectively. The methylation of DNA has been reported in the early stages of carcinogenesis. This process can modify the genetic activity and has been proposed as a hypothesis for oncogenic transformations. [176]Therefore, its higher contribution in grade 1 luminal samples is highly relevant. The hypermethylation in breast cancer has affectation in the activity of genes in charge of steroid receptors, cell adhesion and matrix metalloproteinases inhibitors. The Oestrogen receptor (ER) gene in normal cells is unmethylated, whereas in ER positive cell lines present an important elevated methylation degree. [176]

The existence of methylated nucleotides in low grades serve as a representation of initial malignant behaviours as the methylation is presented as whole genomic hypomethylation and as local promoter hypermethylation in critical tumour suppressor and growth regulator genes. [177] The methylation plays a vital role on the hormonal expression of tumours, for example, the methylation of hormonal genes has been proposed as a hormonal status predictor, using the methylation of the PR gene (PGR) for ER status and as a predictor of PR status the methylation of the ER gene (ESR1). [178]

The peak located at ≈1640 cm^{-1} representing amide I band of proteins was identified by PC-3 and showed the same trend intensity wise with peaks located at ≈1624, 1592, and 1571 cm^{-1} which were identified with PC-4. This trend presented higher intensities in grade 3 luminal cases followed by grade 2 and 1. These peaks represent vibrations of amide I, C=N, and NH_2 from adenine and an increase in intensity in higher grades suggests higher contents of proteins and nucleic acids. [167] This overproduction of structural and genetic material behaviour is in agreement with the aggressive behaviour representative of higher grades.

In addition, the peak placed at ≈1659 cm^{-1} was also identified by PC-4. This peak correlates with the C=C cis conformation of fatty acids. [149] This peak presented greater contribution in grade 2 samples, which confirmed the highly active lipid metabolism previously discussed in the higher wavenumber region.

5.2.1.3. Fingerprint region (1,450-600 cm^{-1}) analysis of Luminal samples all grades

Principal component analysis of the fingerprint region revealed an unclear separation but relatively good grouping of the different grades of luminal samples as presented in Figure 44. PC-1 identified the following bands and assignments as differentiators of grade 1 luminal samples: ≈1383 cm^{-1} representing C-N and N-H deformations, ≈1227 cm^{-1} associated with phosphate I asymmetric PO_2^- stretching, ≈1165 cm^{-1} representing the C-O stretching mode of C-OH groups of serine, threonine, and tyrosine of proteins, ≈1062 cm^{-1} and ≈964 cm^{-1} which correlate with C-O stretching of phosphodiester and ribose. [167] Grade 1 samples presented absorbance closer to the average whereas grade 2 and 3 showed variable intensities with higher and lower than the average intensities.

On the contrary, a better separation was achieved using PC-4. Grade 3 luminal samples showed higher intensities in the peak representing CH_3 symmetric bending vibrational modes of lipids located at ≈1464 cm^{-1}. [149] This behaviour suggests that biochemical changes occur in later grades in cancer cells in the lipid component probably associated with lipid peroxidation. [174]

In addition, the bands located at ≈1237 and ≈1070 cm^{-1} representing phosphate I PO_2^- asymmetric stretchings and C-O vibration in disordered DNA structures [167] presented higher intensities in grade 2 samples while lower absorbance was found in a fraction of grade 3 samples. DNA methylation reported at the beginning of this section suggested changes in nucleic acids at early grades. The affectations on DNA detected through the peaks suggested by PC-4 might be responsible for the changes detected in grade 2.

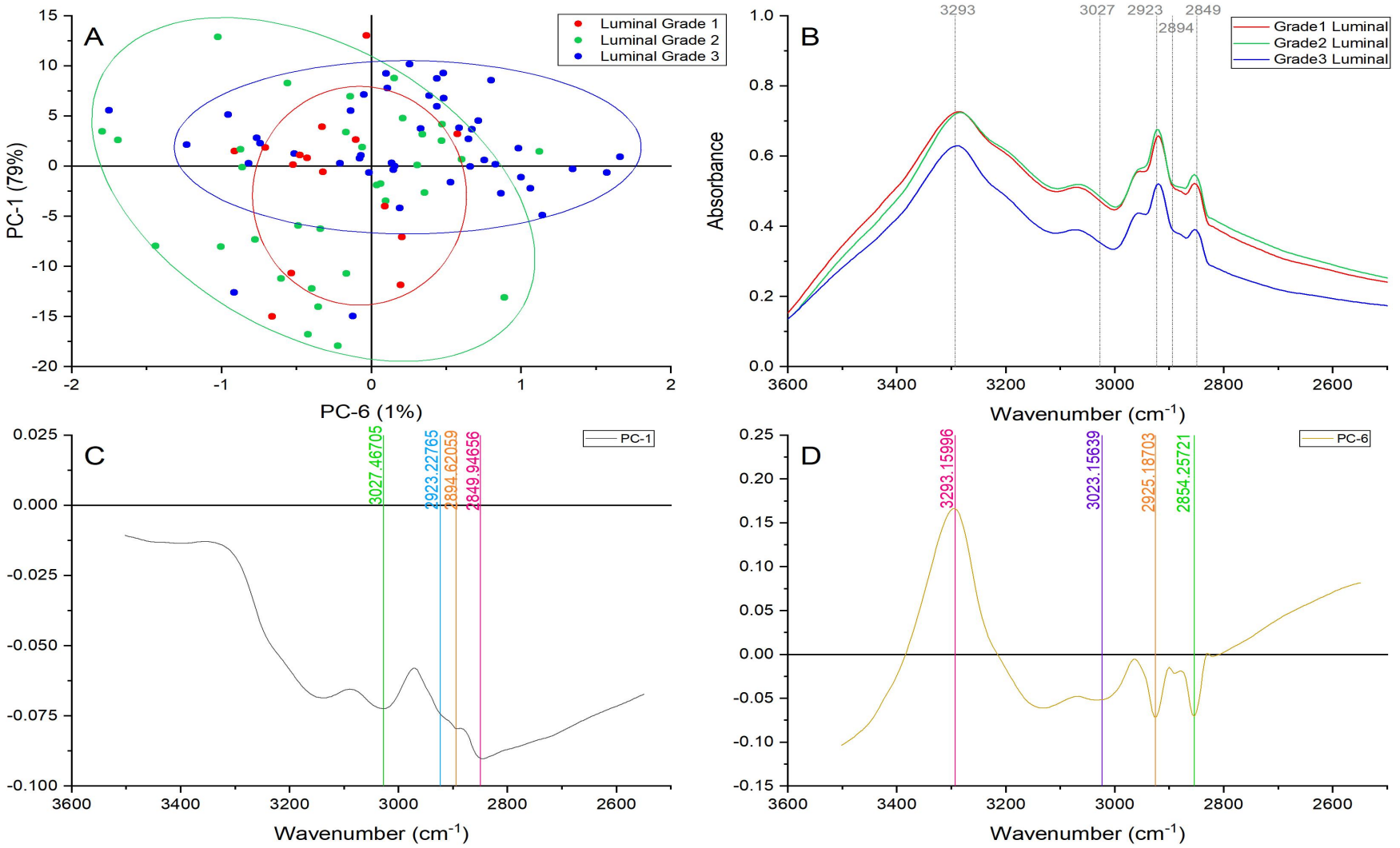

Figure 42. Principal component analysis of the higher wavenumber region (3,500-2,550 cm^{-1}). (A) Score plot using PC-1 and PC-6 accounting for 80% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the higher wavenumber region for luminal samples grade 1,2, and 3. 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-6 loading.

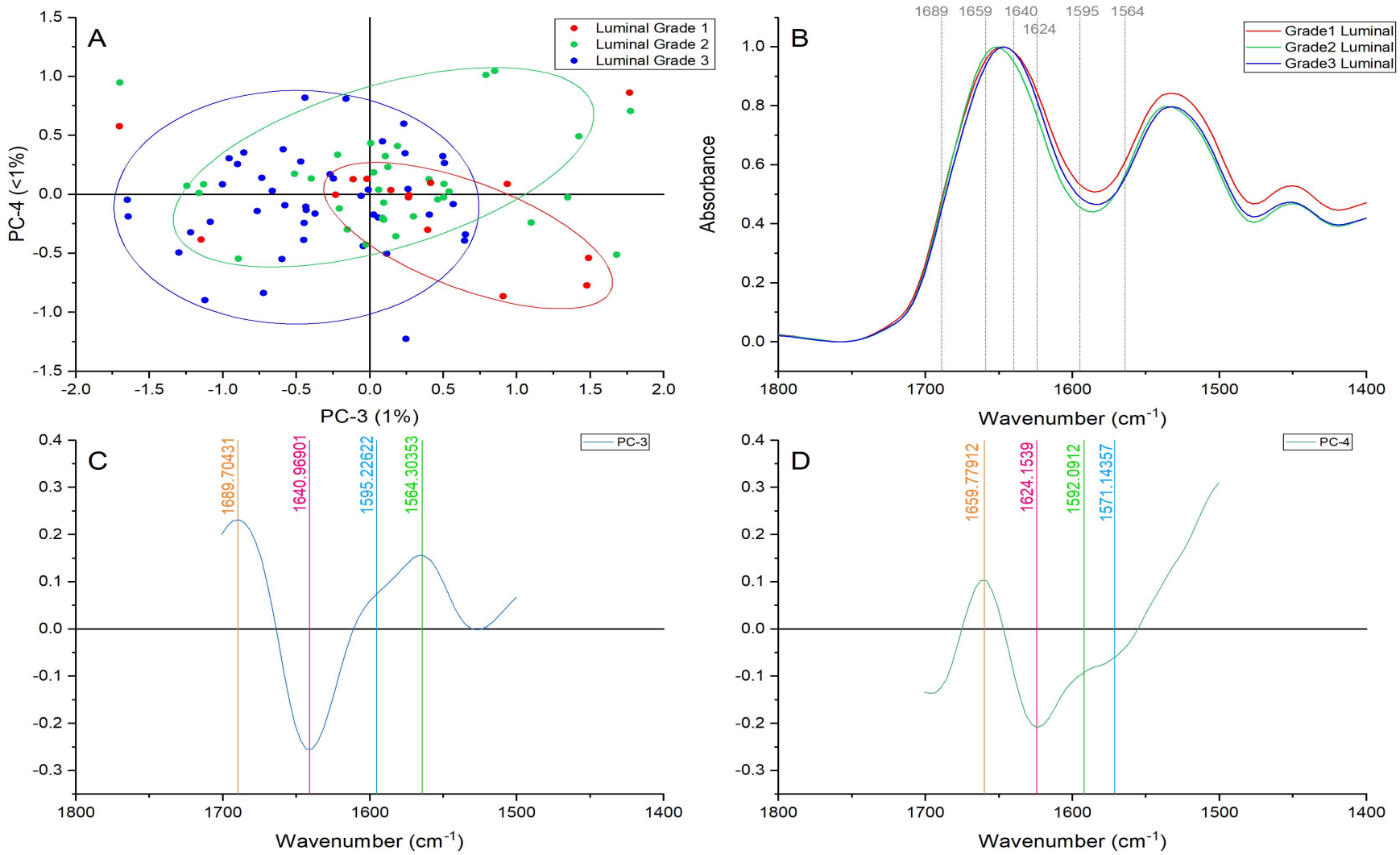

Figure 43. Principal component analysis of the amide region (1,700-1,500 cm^{-1}). (A) Score plot using PC-3 and PC-4 accounting for 2% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the higher wavenumber region for luminal samples grade 1,2, and 3. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-3 Loading, (D) PC-4 loading.

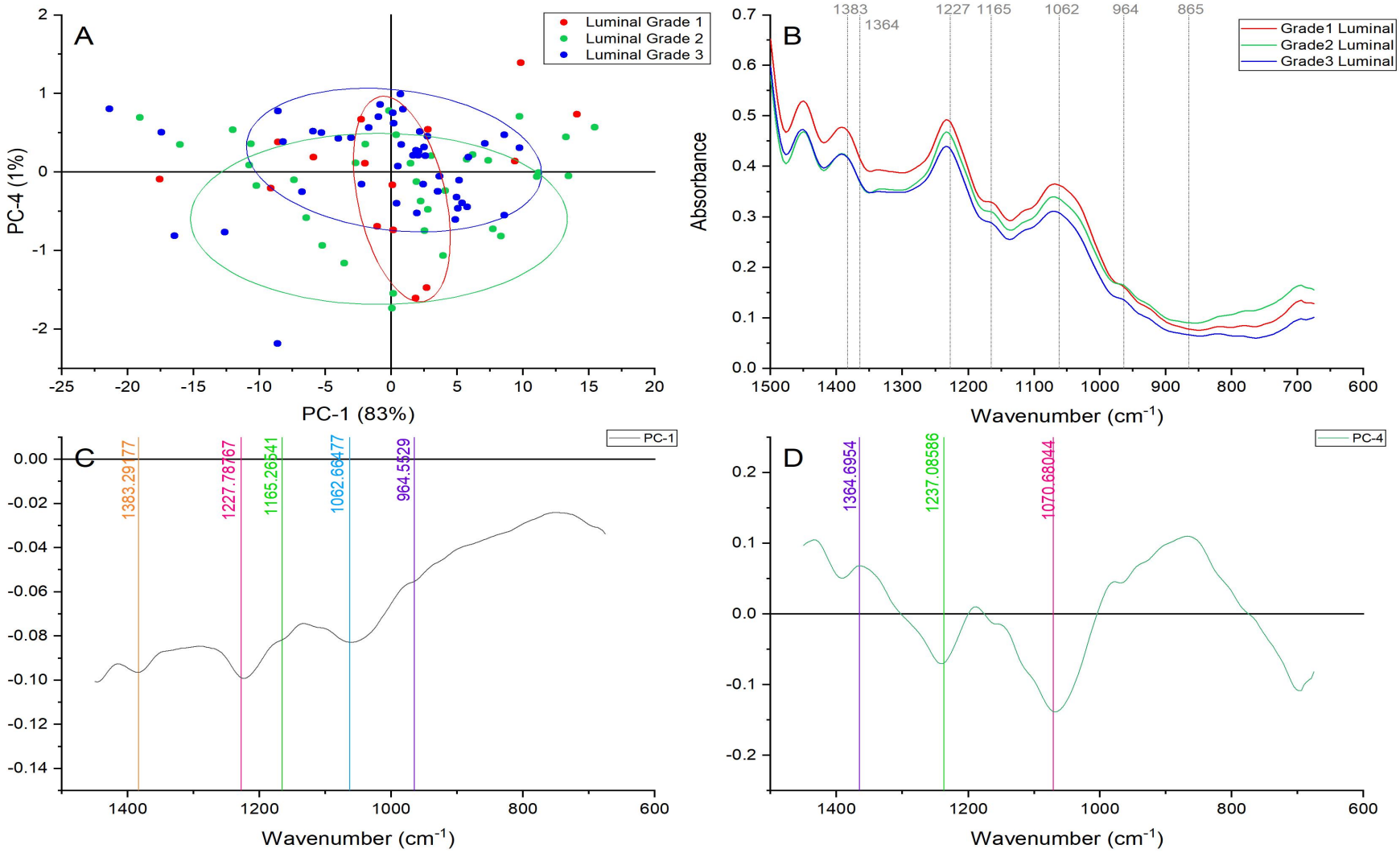

Figure 44. Principal component analysis of the fingerprint region (1,450-600 cm^{-1}). (A) Score plot using PC-1 and PC-4 accounting for 84% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the higher wavenumber region for luminal samples grade 1,2, and 3. 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-4 loading.

5.2.2. TNBC SAMPLES ALL GRADES

The average spectral profile of TNBC samples representing grade 1, 2, and 3 showed differences enough to see with the naked eye, as presented in Figure 45 and Figure 46. Different trends were detected in the different areas of the spectra, for example overall higher intensities can be appreciated for grade 3 TNBC samples in the higher wavenumber area, whereas in the amide and FP region the grade 3 TNBC spectral profile showed the lowest average intensities. In the amide region, grade 2 samples showed the most dominant spectra while grade 1 had the highest intensities in the spectral profile of the fingerprint region.

A complete analysis of these three regions will be included to identify biochemical changes associated with the cancer progression of triple negative breast cancer cases.

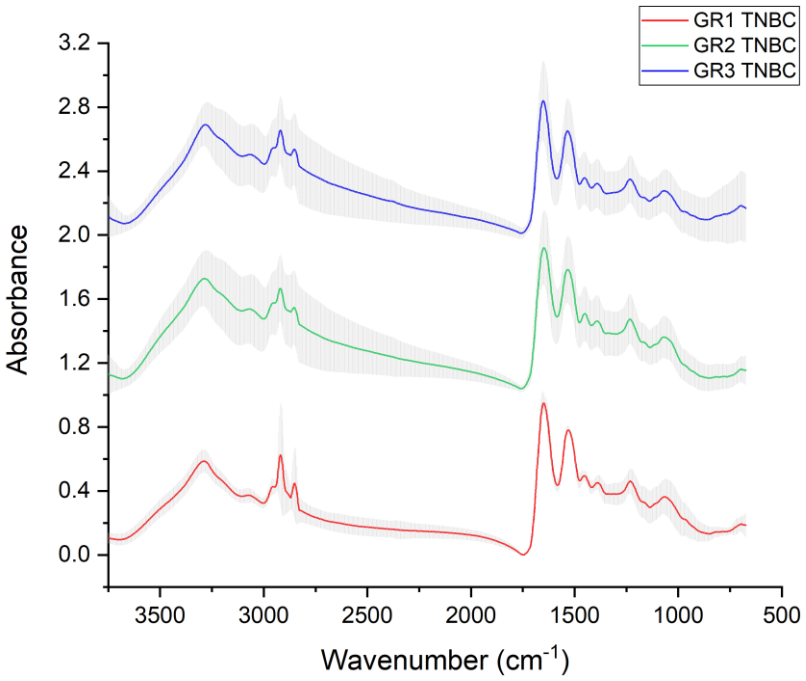

Figure 45. FTIR average spectral profile of TNBC samples representing grade 1, grade 2, and grade 3 cases, variance is represented with grey shadow.

5.2.2.1. HIGHER WAVENUMBER (3,500-2,550 CM^{-1}) ANALYSIS OF TNBC SAMPLES ALL GRADES

The principal component analysis of the higher wavenumber region revealed a good grouping of grade 1 and 2 cases. On the contrary, a clear dispersion of grade 3 TNBC cases was identified using PC-1 and PC-5 as it can be appreciated in Figure 47.

PC-1 identified the following peaks as responsible for the grade 2 grouping: ≈3026 cm^{-1} representing aromatic C-H stretchings, ≈2896 cm^{-1} accounting for CH_3 symmetric stretchings, ≈2867 cm^{-1} representing aliphatic C-H stretching, and ≈2831 cm^{-1} which correlates with the N-H stretching vibration. [149] Most of the grade 2 TNBC samples were distributed on the positive PC-1 axis, therefore lower intensities associated with the peaks previously reported can be suggested. Grade 3 TNBC samples were distributed among both axes showing a split behaviour which suggests important differences among grade 3 cases. The few grade 1 TNBC

cases reported in our cohort showed lower than the average intensities for peaks reported based on the PC-1 axis. However, this group showed intensities closer to the average for the peaks identified with PC-5.

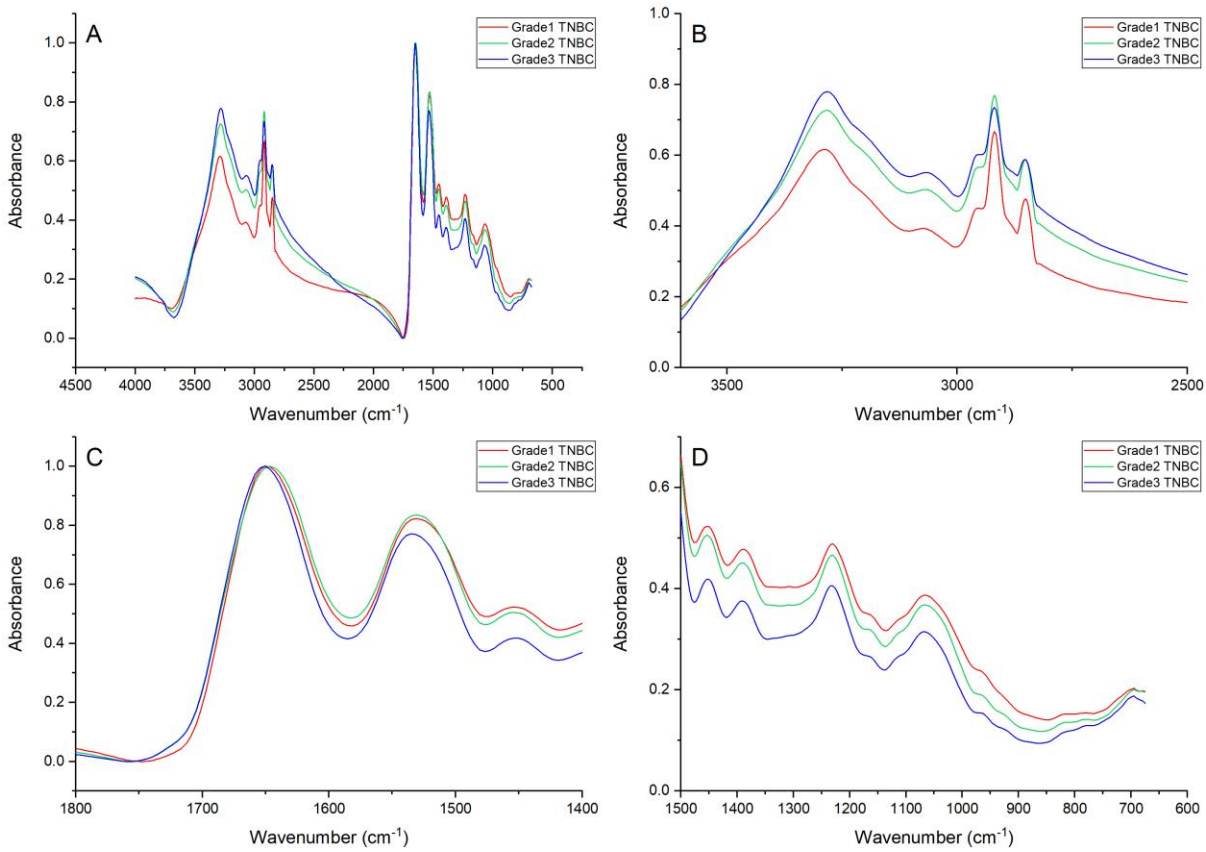

Figure 46. FTIR average spectral profile of TNBC samples representing grade 1, grade 2, and grade 3 cases. (A) General region (4,000-675 cm^{-1}), (B) High wavenumber region (3,500-2,550 cm^{-1}), (C) Amide region (1,700 -1,500 cm^{-1}), and (D) Fingerprint region (1,450-600 cm^{-1}).

PC-5 identified peaks associated with proteins represented by ≈3297 cm^{-1} representative of amide A and ≈3038 cm^{-1} correlating with Amide B (both representing N-H stretching), and fatty acids characterised by ≈2936 cm^{-1} and ≈2861 cm^{-1} which represent aliphatic and fatty acid vibrations. [149].

Grade 2 and 3 TNBC samples showed variable intensities including lower and higher than the average behaviour without showing a clear trend regarding these biochemical components. However, grade 1 samples kept intensities closer to the average which suggests a steady protein and fatty acid content in early grades of TNBC.

5.2.2.2. Amide region (1,700-1,500 cm^{-1}) analysis of TNBC samples all grades

Principal component analysis of the amide region showed a nice grouping and clear separation of grade1,2, and 3 of TNBC as can be appreciated in Figure 48. The analysis of the amide region revealed important overproduction of nucleic acids in grades 1 and 2 of the TNBC samples. The overproduction of this material is characterised by higher intensities in the peak located

at ≈1606 which represent adenine vibrations in DNA. [164]The overproduction of cell components is characteristic of cell division and proliferation processes which are expected in early grades of cancer.

The previous trend was also confirmed for C=C vibrations in fatty acids (≈1662 cm^{-1}). These increased intensities suggested that the fatty acids fuelled and were needed for the synthesis of membranes and signalling molecule[173] decreasing its content in higher grades. Greater content of fatty acid has been suggested as part of the carcinogenesis process and cancer progression, in fact, de novo lipogenesis in cancerous tissue has been accepted as a symbol of aggressive behaviour. [179] However, a decline in content has been suggested in advanced cases as a result of lipid peroxidation and oxidative stress [175]. In addition, structural conformations changed to facilitate certain malignant processes. [102]

In addition, grade 2 and 1 TNBC cases showed higher intensities for vibrations representing β-sheet conformations of amide II in proteins (≈1534 cm^{-1}) in comparison with grade 3 samples. These conformational changes from dominant α-helical structures to β-sheet conformation has been previously reported as transitions to malignant tissue. [102], [105] An important change in intensity and slightly change in shape can be appreciated in the spectral profile presented in Figure 48B. Considering the intensities and spectral profile, structural changes in proteins can be proposed in initial grades of the TNBC cases to facilitate invasive processes. On the contrary, overall overproduction and content of proteins can be assumed in grade 3 TNBC cases thanks to the peaks located at ≈1625 cm^{-1} which represent amide I vibrations. [74]

5.2.2.3. Fingerprint region (1,450-600 cm^{-1}) analysis of TNBC samples all grades

The fingerprint region principal component analysis showed a good grouping for grade 1 and 2 while grade 3 TNBC samples were dispersed as can be appreciated in Figure 49. Grade 1 TNBC samples seem to group based on their absorbance of the following peaks: ≈1382, 1225, 1158, 1114, and 1060 cm^{-1}. These peaks presented higher than the average intensities. These peaks were identified by PC-1.

The band located at ≈1384 cm^{-1} represent the deformation of CH_3, C-H, and N-H [149]which serves as the first indication of active modifications at cellular level. The ≈1225 and 1114 cm^{-1} peaks are representative of phosphate stretching in the phosphodiester linkages in DNA and RNA. [170] All grades presented samples with higher intensities than the average. However, grade 2 and 3 showed a split behaviour with half of the samples showing lower than the average intensities. In addition, similar changes were identified at ≈1158 and 1060 cm^{-1} associated with C-O of proteins and carbohydrates and the stretching of C-O characteristic of the nucleic acid triad respectively. The peak located at ≈1101 identified with PC-2 correlates with the symmetric stretching of P-O-C in nucleic acids. [167] Half of the grade 3 TNBC presented higher contributions of this component. These differences in intensity suggested that changes in the nucleic acids are dynamic among the different grades.

The PC-2 identified three peaks located at ≈1390, 1236, and 1101 cm^{-1}. The ≈1390 cm^{-1} assignments correlate with the CH_2 asymmetric bending and COO^- stretching presented in proteins and fatty acids. [149] Grade 3 samples showed higher intensities for this peak

followed by grade 2 and 1 TNBC samples suggesting accumulation of structural cell contents (proteins) and cell membranes (fatty acids). The same trend was identified with the band located at ≈1236 cm^{-1} associated with phosphate I (PO_2^-) asymmetric stretching mainly from nucleic acids with the little contribution from phospholipids [121] [167]Dovbeshko *et al.*, 2000)Furthermore, the overproduction of fatty acids can also be correlated with increase signalling favouring and stimulating invasive processes [173] which is associated with the aggressive behaviour expected in higher grades.

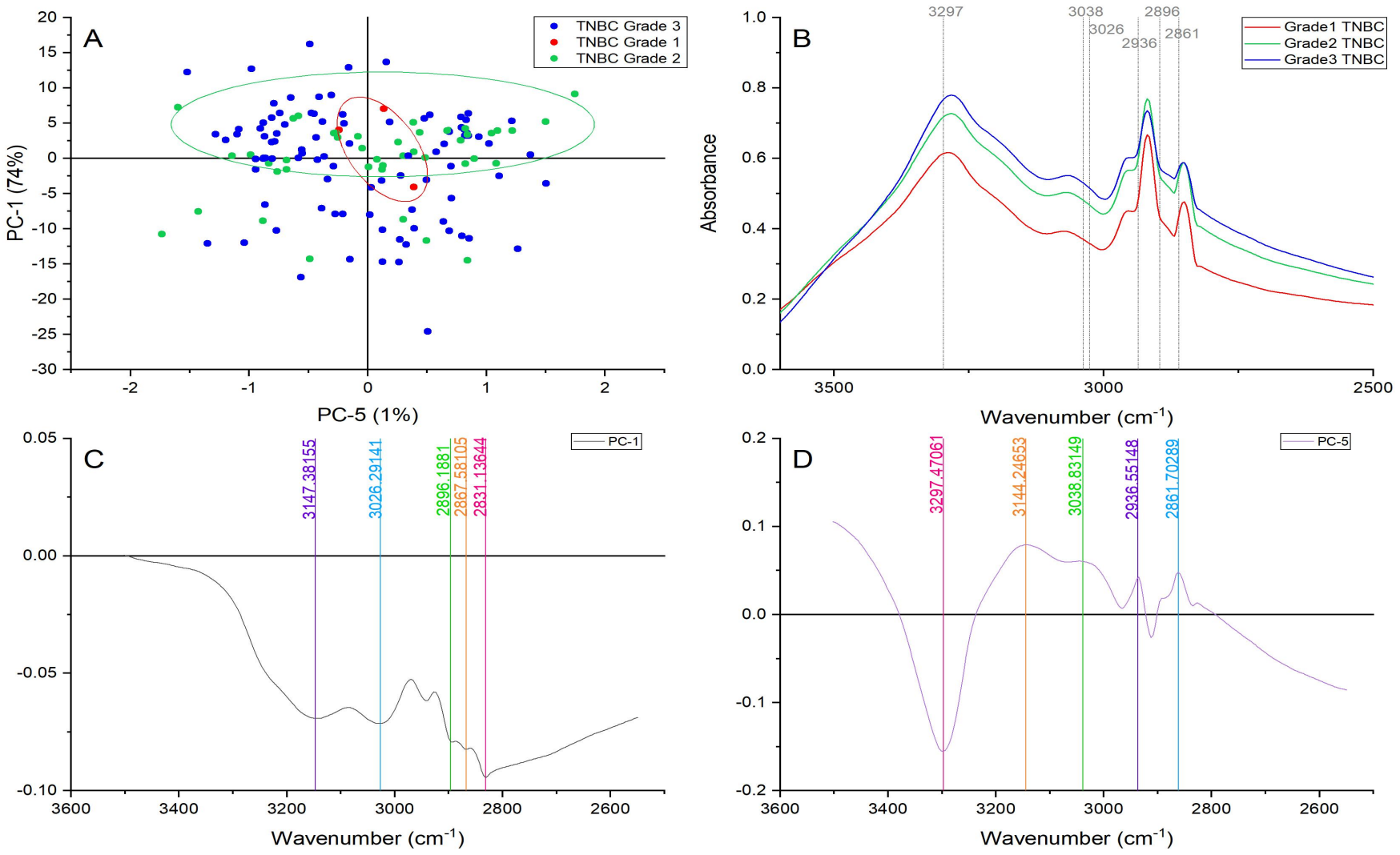

Figure 47. Principal component analysis of the higher wavenumber region (3,500-2,550 cm^{-1}). (A) Score plot using PC-1 and PC-5 accounting for 75% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the higher wavenumber region for TNBC samples grade 1,2, and 3. 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.

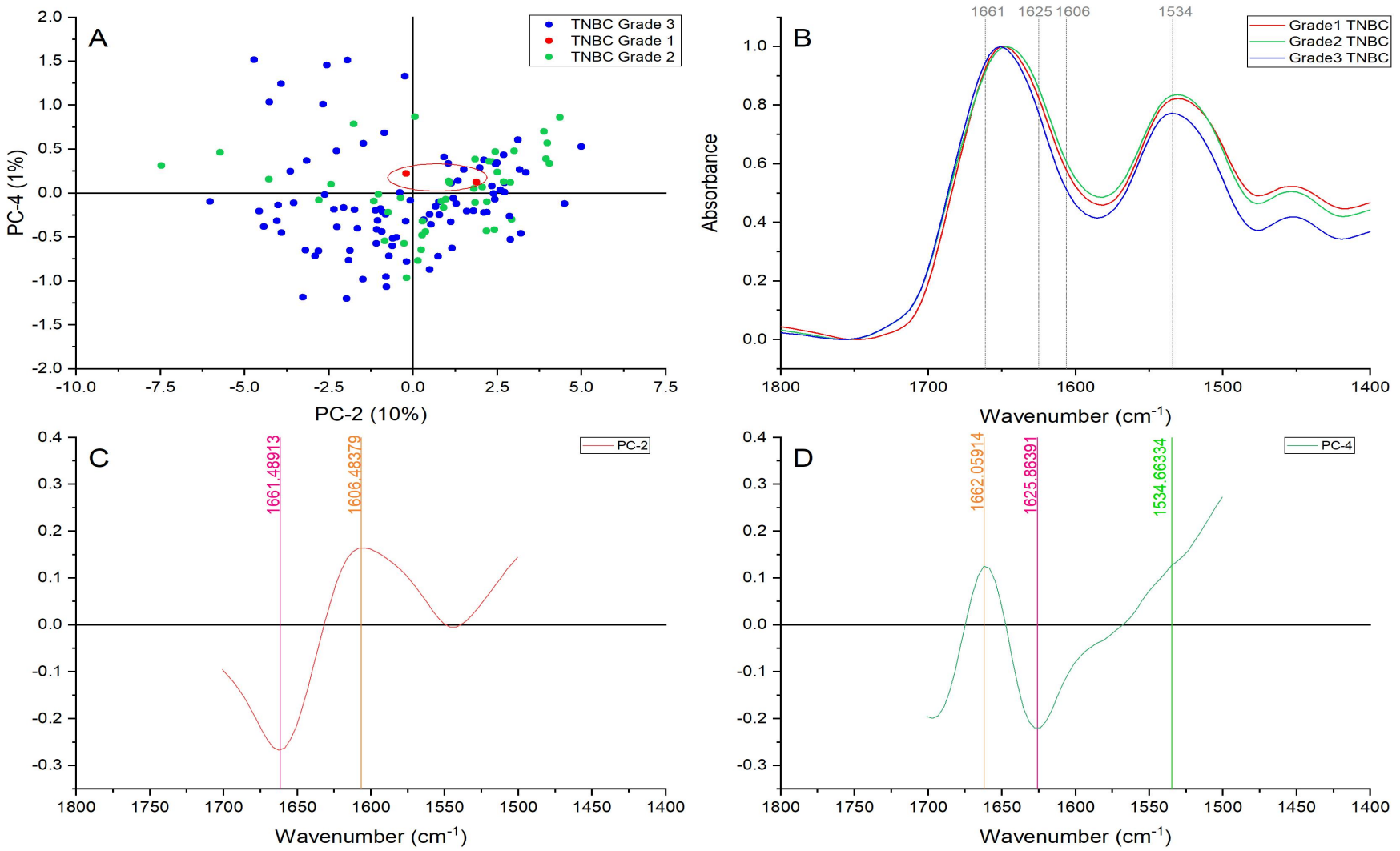

Figure 48. Principal component analysis of the amide region (1,700-1,500 cm^{-1}). (A) Score plot using PC-2 and PC-4 accounting for 11% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the higher wavenumber region for TNBC samples grade 1,2, and 3. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-2 Loading, (D) PC-4 loading.

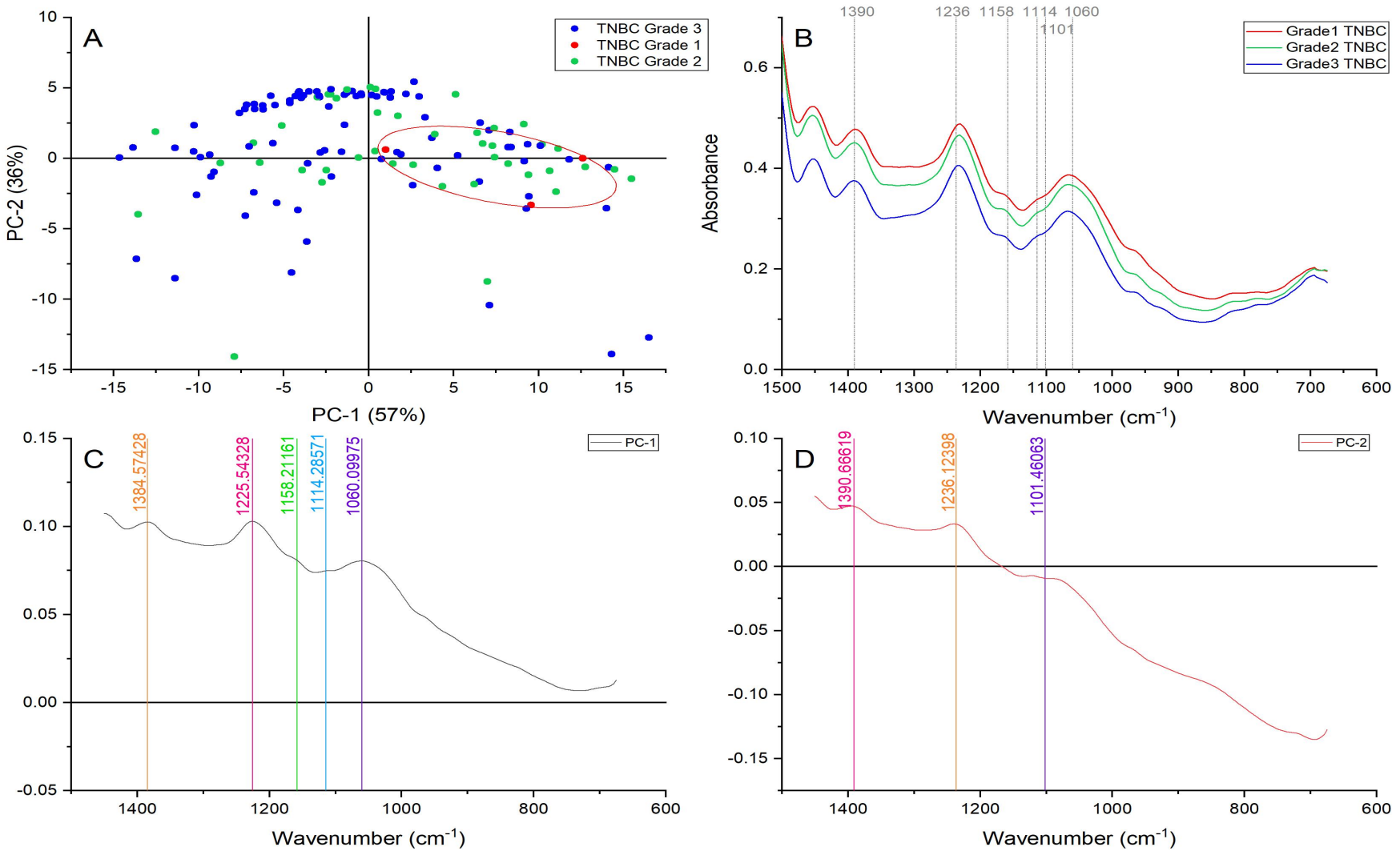

Figure 49. Principal component analysis of the fingerprint region (1,450-600 cm^{-1}). (A) Score plot using PC-1 and PC-2 accounting for 93% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the higher wavenumber region for TNBC samples grade 1,2, and 3. 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.

5.2.3. HER2+ ALL GRADES

The average spectral profile of HER2+ samples representing grade 1, 2, and 3 showed visible differences in intensity, shifts and peak shapes as presented in Figure 50 and Figure 51. Different trends were detected in the different areas of the spectra, for example, overall higher intensities can be appreciated for grade 3 HER2+ samples in the higher wavenumber and fingerprint area. Similar average intensities to grade 2 were appreciated in the amide region whereas grade 1 HER2+ samples showed the lowest average intensities in all regions.

A complete analysis of these three regions will be included to identify biochemical changes associated with the cancer progression of HER2+ breast cancer cases.

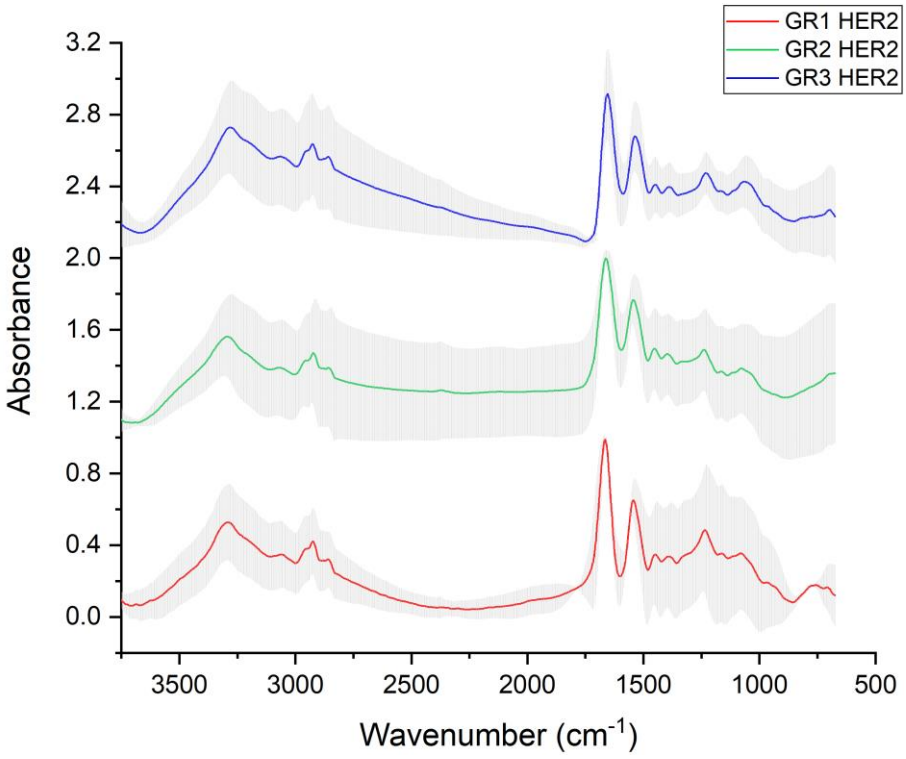

Figure 50. FTIR average spectral profile of HER2+ samples representing grade 1, grade 2, and grade 3 cases, variance is represented with grey shadow.

5.2.3.1. HIGHER WAVENUMBER (3,500-2,550 CM^{-1}) ANALYSIS OF HER2+ SAMPLES ALL GRADES

Good grouping and separation were achieved using PC-1 and PC-2 in the higher wavenumber area, as can be seen in Figure 52. Five peaks were identified as responsible for the separation and grouping of HER2+ samples of all grades. The first peak was located at ≈3168 cm^{-1} as has been associated with N-H symmetric stretching. The second peaks were identified at ≈3041 cm^{-1} representing the N-H amide B stretching. Additionally, a third band located at ≈2869 cm^{-1} was identified in PC-1 representing the CH_3 symmetric stretching mainly from proteins and lipids with little contribution from carbohydrates and nucleic acids. [149] These 3 peaks showed higher absorbance in grade 3 samples followed by grade1 and 2. This behaviour suggested higher protein content. Moreover, the increased content of proteins indicates a higher cell density in the high-grade samples.

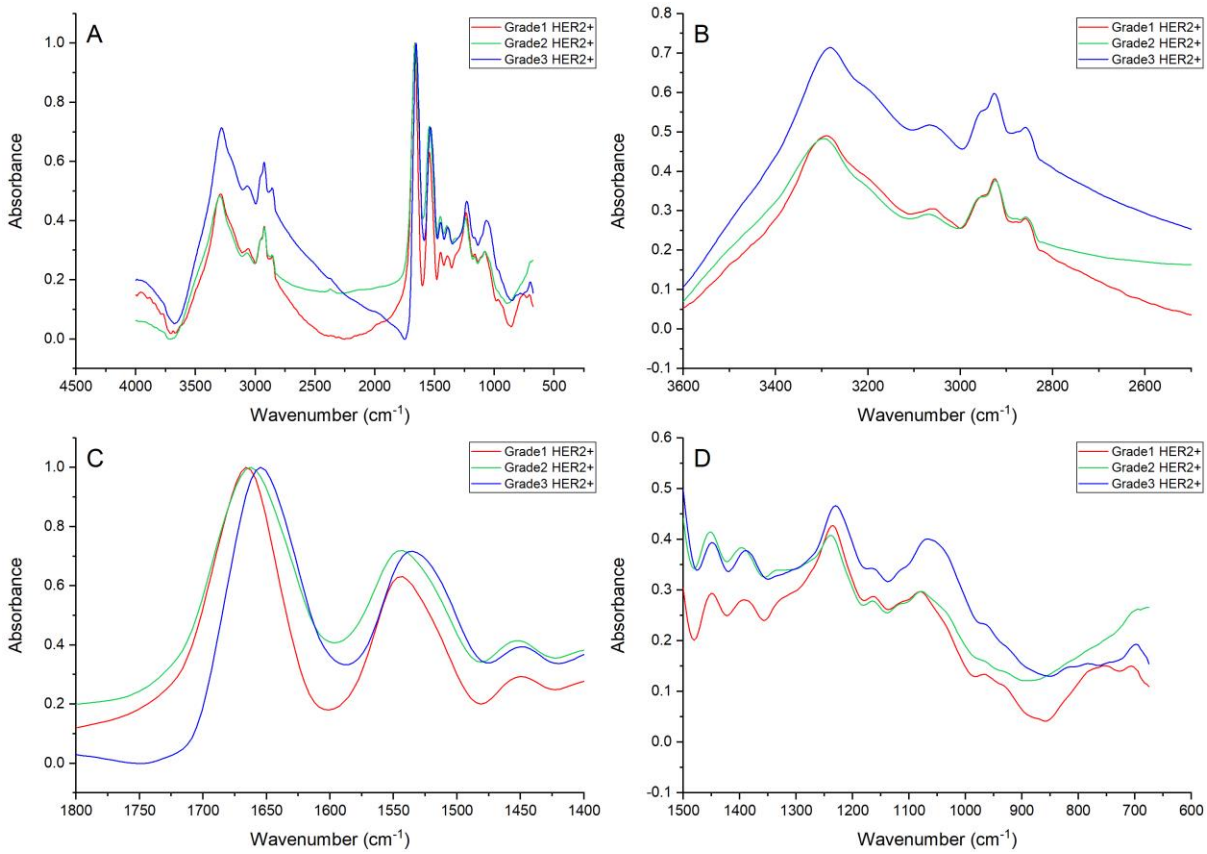

Figure 51. FTIR average spectral profile of HER2+ samples representing grade 1, grade 2, and grade 3 cases. (A) General region (4,000-675 cm^{-1}), (B) High wavenumber region (3,500-2,550 cm^{-1}), (C) Amide region (1,700-1500 cm^{-1}), and (D) Fingerprint region (1,450-600 cm^{-1}).

PC-2 confirmed the differences in the N-H stretching among different grades by identifying the peak at ≈3336 cm^{-1} which represents the asymmetric stretching of the N-H bond. [102] Most of the HER2+ samples grade 1 and 2 were located in the axis representing higher absorbance than the average. These differences in absorbance on all grades had an impact on the peak shape as it can be seen in the spectral profiles presented in Figure 52B. The same trend was recognised for the 2968 cm^{-1} peak which represents the CH_3 asymmetric stretching identified in lipids and fatty acids. [149] Both bands suggested higher contents of these components in lower grades. Fatty acids in lower grades might indicate the need for cellular building blocks for membrane formation, energy storage, and the production of signalling molecules which is required in early stages of cancer. [173]

5.2.3.2. Amide region (1,700-1,500 cm^{-1}) analysis of HER2+ samples all grades

Principal component analysis of the amide region resulted in good grouping and separation of HER2+ samples with different grades. The following peaks were identified by principal component 1 and 2. HER2+ samples grade 1 and 2 showed higher intensities in the peaks located at ≈1692, 1667, and 1543 cm^{-1}.

The 1692 cm^{-1} band represents the vibration of an antiparallel β-sheet of amide I, this vibration is caused by the in-plane C=O stretching, the weak coupling of C-N, and in-plane bending[149] The decrease in α-helix conformation and increase of β-sheet structures has been reported as

a biomarker for cancer identification. [102], [106] The differences in the contents are important and affect the width and shape of the amide I peak as can be seen in the spectral profiles presented in Figure 53B.

The contribution of β-turns represented by the peak located at ≈1667 cm^{-1} was also higher for grade 1 and 2 supporting the decrease in α-helix conformations and promoting β- structures. These changes represent structural fluctuations in proteins associated with its biochemical function and might be caused by mutations occurring at early steps of carcinogenesis. [169] A similar intensity trend was identified for the peak located at ≈1543 cm^{-1} representing an amide II stretching arising from the C-N stretching and the CHN bending vibrations. [149] However, a downshift of ≈10 cm^{-1} has been identified in the grade 3 HER2+average spectral profile suggesting a lower contribution of this assignment in high-grade samples.

Grade 3 HER2+samples showed a higher contribution of carbonyl stretch and ring breathing vibrations associated with nucleic acids [149] represented by the peak located at ≈1617 followed by grade 2 and grade 1 samples. This behaviour has been previously reported to increase in intensity as the grade increases as a result of higher cell density and poor differentiator of grade 3 cancer cells. [106]

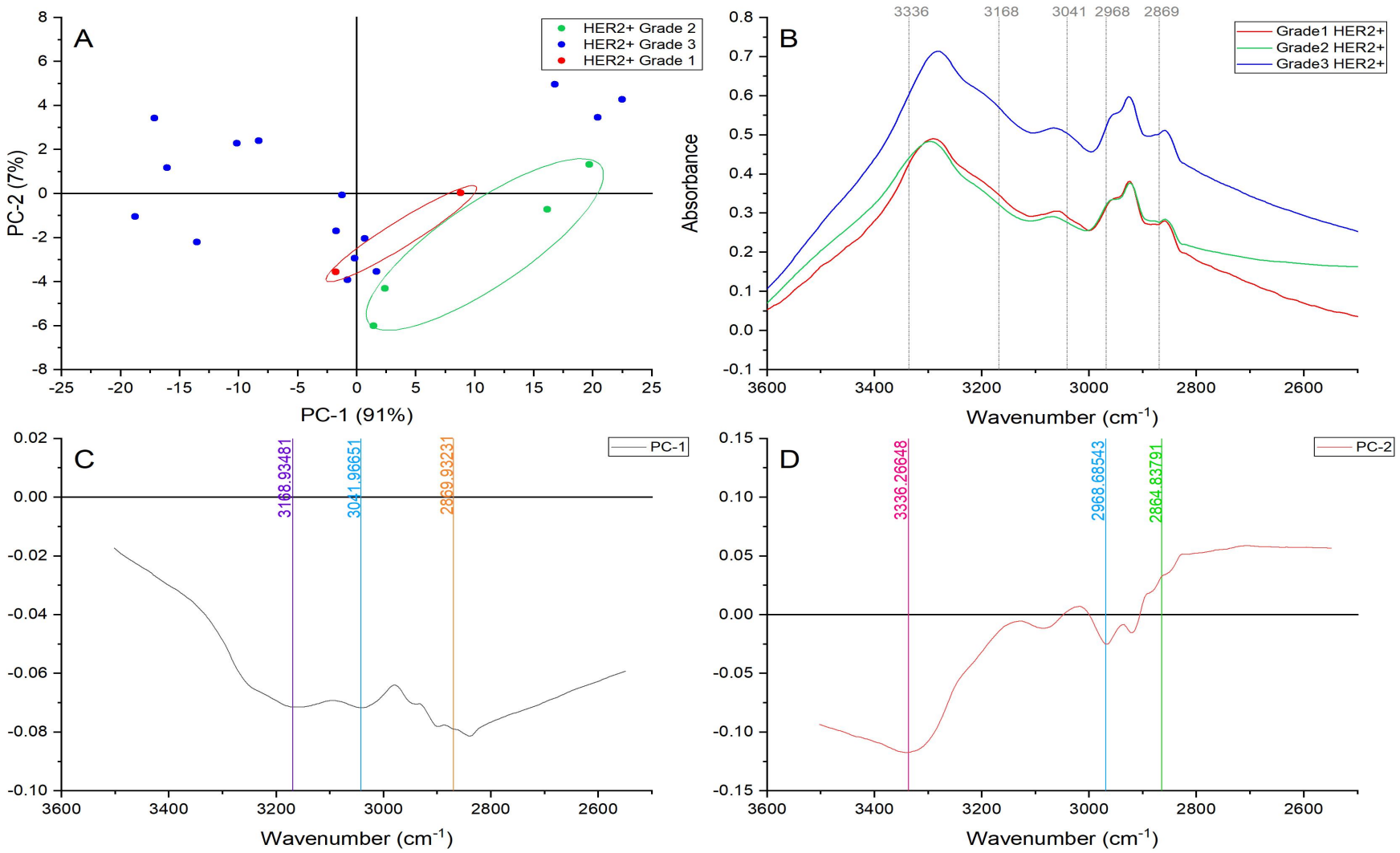

Figure 52. Principal component analysis of the higher wavenumber region (3,500-2,550 cm^{-1}). (A) Score plot using PC-1 and PC-2 accounting for 98% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the higher wavenumber region for HER2+ samples grade 1,2, and 3. 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.

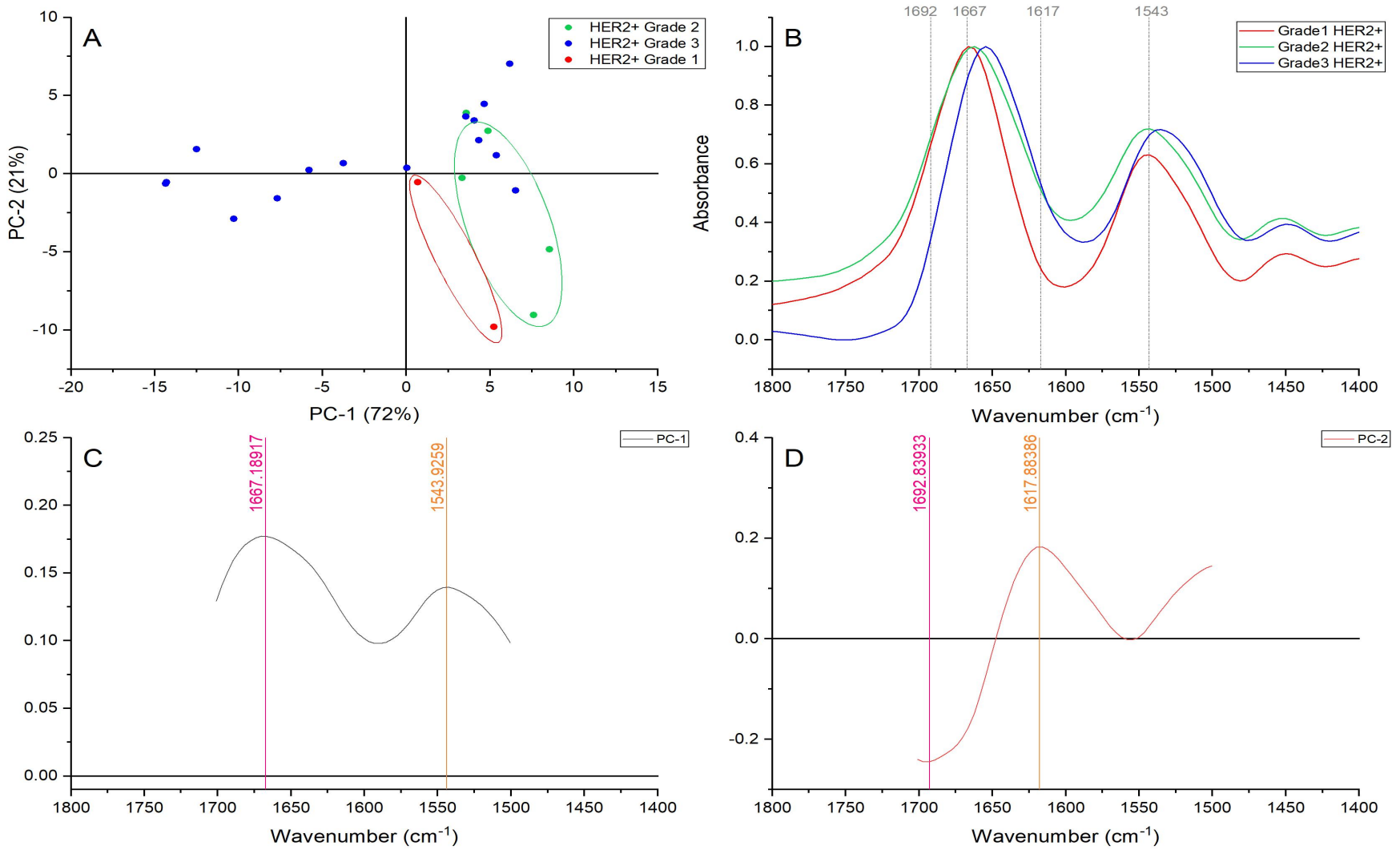

Figure 53. Principal component analysis of the amide region (1,700-1,500 cm^{-1}). (A) Score plot using PC-1 and PC-2 accounting for 93% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the higher wavenumber region for HER2+ samples grade 1,2, and 3. 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.

5.2.3.3. Fingerprint region (1,450-600 cm^{-1}) analysis of HER2+ samples all grades

The fingerprint principal component analysis of HER2+ samples showed good separation and grouping when PC-3 and PC-4 were used as presented in Figure 54. The CH_3 bending mode of proteins presented at ≈1404 cm^{-1} and the CH_2 wagging of phospholipids, fatty acids, triglycerides, and amino acid chains represented by ≈1342 cm^{-1} [149]showed higher intensities in grade 1 and 2.Grade 3 samples presented a split behaviour with half of the samples showing similar intensities to lower grade, but also intensities lower than the average. The increased intensity in low and intermediate HER2+ grade might indicate changes in proteins and accumulation of cellular material to promote proliferation. [106]

Grade 1 samples also showed high contributions of phosphate vibrations represented by ≈1241 cm^{-1} (asymmetric stretching of nucleic acid phosphate) and ≈1085 cm^{-1} (phosphate from nucleic acids, phospholipids and partially amide III in proteins).[149] This trend was shared with half of the grade 3 HER2+ samples, which can be explained by the downshift presented in the grade 3 spectral profile at those specific assignments. The downshift of 10 and 15 cm^{-1} for these peaks could have been caused by the alterations on the structural conformation of proteins previously discussed and the alteration of the membrane fluidity changes in invasive and aggressive cases. [158]

Grade 3 HER2+ samples showed higher absorbance in the peaks located at ≈1161 cm^{-1}, also recognised at ≈1155 cm^{-1} in PC-3, and 954 cm^{-1}. These two bands are representative of C-O stretching mode of C-OH groups of serine, threonine, and tyrosine, and the C-O vibration in deoxyribose. [149] The increased contribution of these bands in grade 3 is caused by the higher cell density and the higher cell material available for poor differentiation and accelerated proliferation characteristic of high-grade samples.

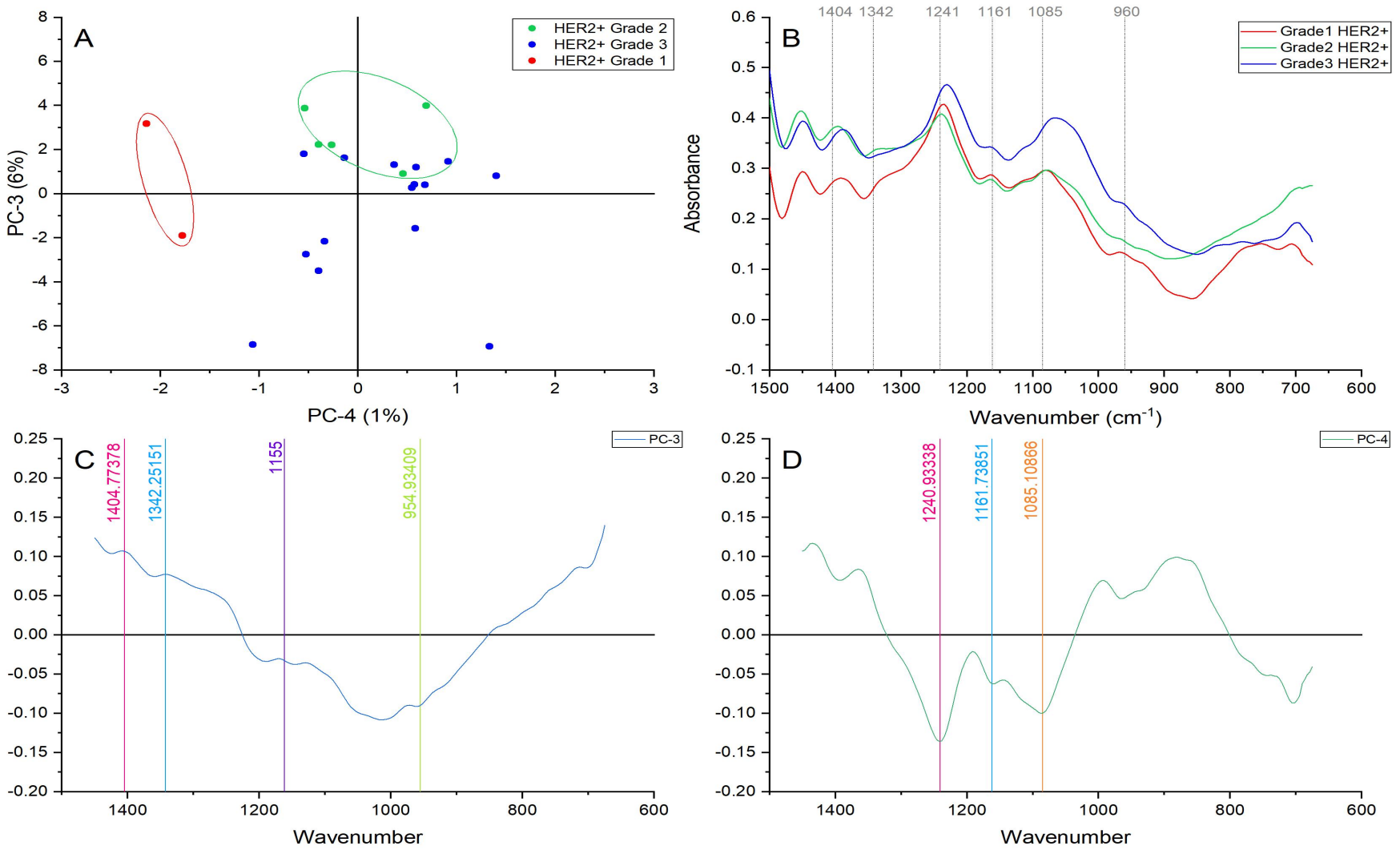

Figure 54. Principal component analysis of the fingerprint region (1,450-600 cm^{-1}). (A) Score plot using PC-3 and PC-4 accounting for 7% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the higher wavenumber region for HER2+ samples grade 1,2, and 3. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-3 Loading, (D) PC-4 loading.

CHAPTER 6. RAMAN ANALYSIS OF NORMAL BREAST AND BREAST CANCER

The results presented in this chapter include the Raman spectral analysis of 6 tissue microarrays containing African breast tumours from similar tribal origins in Nigeria including different grades and subtypes and one TMA containing normal breast sections.

Twenty spectra were collected from the epithelial tissue of each sample. A total of 499 cancerous tumours and 79 normal breast sections were studied with Raman spectroscopy. The average spectra of the cancerous epithelia named cancerous area and the normal epithelia called normal breast are presented in Figure 55 and Figure 56.

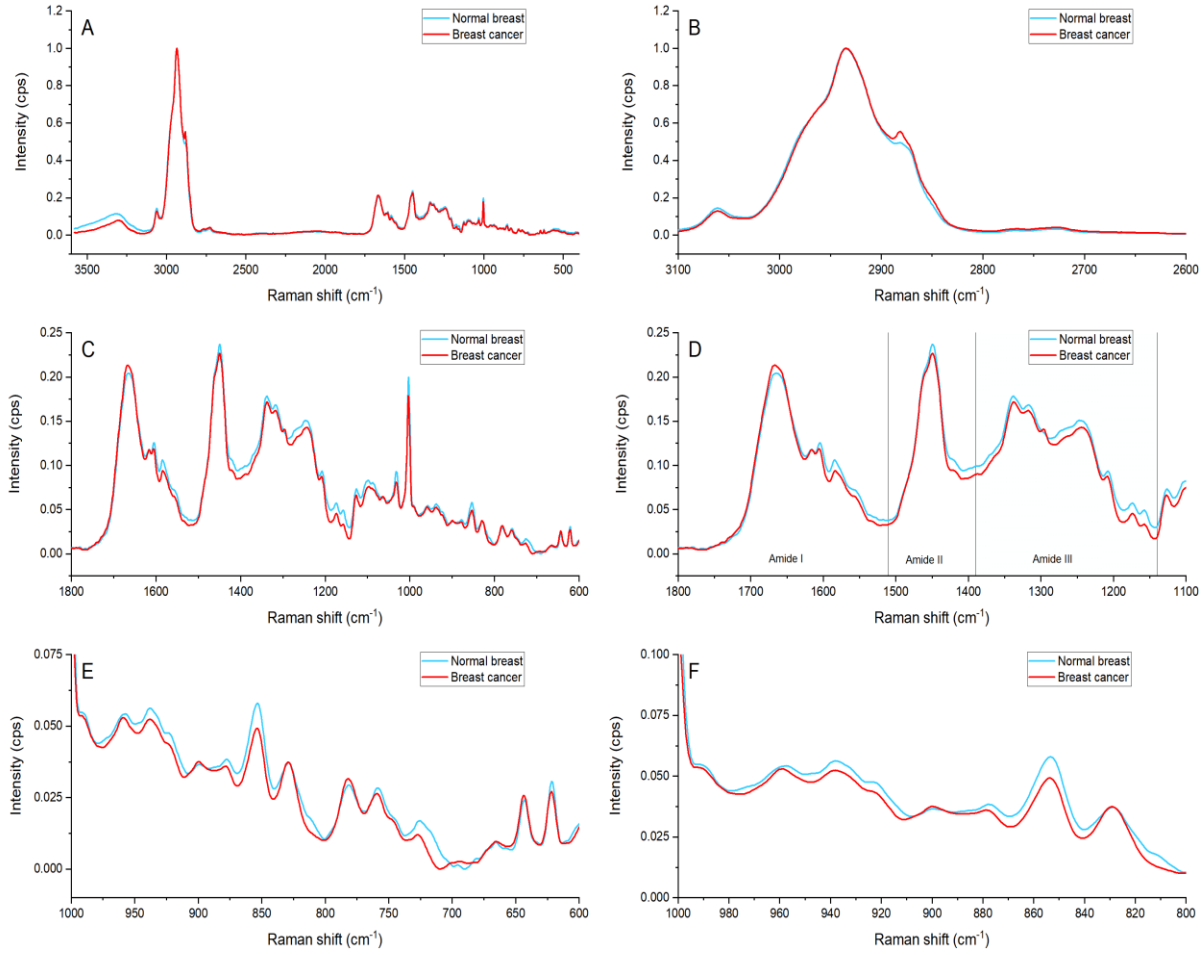

Figure 55. Raman average spectra of cancerous area and normal breast. (A) General region (3600-400 cm^{-1}), (B) Lipid region (3,100-2,680 cm^{-1}), (C) Fingerprint region (1,800-500 cm^{-1}), (D) Amides region (1,800-1,140 cm^{-1}) indicating: Amide I (1,800-1,510 cm^{-1}), Amide II (1,510-1,390 cm^{-1}) and Amide III (1,390-1,140 cm^{-1}) regions. (E) Amino acid and nucleic acids regions (980-600 cm^{-1}), and (F) Hydroxyproline and proline region (810-960 cm^{-1}).

The spectral profile of both tissue types showed subtle differences in intensity and peak location. Nonetheless, the differences were clear to the naked eye and suggested the presence of vital molecules such as carbohydrates, lipids, proteins, and nucleic acids on both tissue types. The spectral profile of breast cancer showed lower overall intensity (Figure 55A&C) except on the lipid area (Figure 55B) and the peaks located at ≈1667 and ≈1295 cm^{-1}.

Furthermore, noticeable changes in peak shapes and ratios can be appreciated in the amide region (Figure 55D) while subtle peak shifts are present on the amino and aminoacid region (Figure 55E).

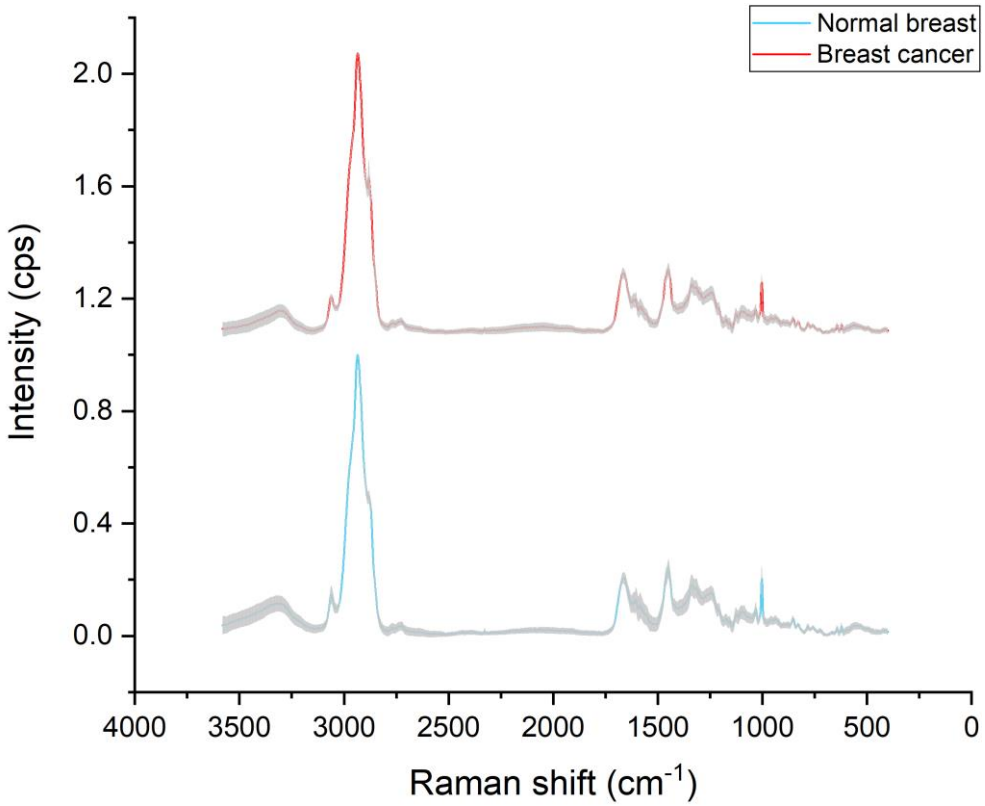

Figure 56. Raman average spectra of Breast cancer (CA) and normal breast (NB), variance is represented with grey shadow.

The main spectral features of both spectra are reported in Table 21 with the assignments taken from Rehman et al, Talari et al, and Lazaro-Pacheco et al unless otherwise specified. [148], [180], [181] As stated on this table, each peak represents the vibration of a functional group from a molecule. Vibrations corresponding to biomolecules such as proteins, lipids and carbohydrates were found as expected. Special attention should be paid to assignments related with carotenoids and cancer specific assignments such as ceramides. However, a proper comparison of the cancerous and normal breast tissue can only be achieved including the several readings obtained with Raman spectroscopy and the use of multivariate analysis.

Table 21. Raman spectral bands of cancerous area (CA) and normal breast (NB) spectra averages with assignments. Assignments were taken from Rehman et al, Talari et al, and Lazaro-Pacheco et al. [148], [180], [181]

Peak position in breast cancer (cm^{-1})	Assignment	Peak position in normal breast (cm^{-1})
3298	Attributed to OH stretch	3316
3061	CH stretch in lipids	3061
2934	Chain end CH_3 symmetric band	2935

3100-2800 2881	CH, CH_2, and CH_3 symmetric and antisymmetric stretching CH_2 and CH of lipids and proteins	2881
3300-2700 2764, 2726	C-H stretches	2767, 2724
1667	Protein band C=C stretching band α-Helical structure of amide I Structural protein modes of tumours Carbonyl stretch (C=O)	1664
	DNA Proteins, including collagen I C=C (amide I (C=O stretching mode of proteins, α-helix conformation)/C=C lipid stretch	
1616	Tyrosine, tryptophan, C-C (protein) Adenine	1615
1605	C=C in-plane bending mode of phenylalanine and tyrosine Ring C-C stretch of phenyl Cytosine (NH_2) Ring C-C stretch of phenyl Phenylalanine, tyrosine, C-C (protein)	1604
1584	C-C bending mode of phenylalanine	1584
1557	Tryptophan ν (CN) and δ (NH) amide II (protein assignment) ν(C C) porphyrin Tyrosine, amide II, COO^-	1556
1449	CH_2CH_3 deformation Collagen CH deformation C-H vibration (proteins and lipids) Lipids	1449
1418	C=C stretching in quinoid ring A, G (ring breathing modes of the DNA/RNA bases) CH deformation (DNA/RNA and proteins and lipids and carbohydrates)	1424
1337	CH_3CH_2 wagging, collagen (protein assignment) CH_3CH_2 wagging, nucleic acid CH_3CH_2 wagging mode of collagen and polynucleotide chain (DNA purine bases) CH_3CH_2 twisting and wagging in collagen Cellular nucleic acids CH_3CH_2 deforming modes of collagen and nucleic acids	1339
1317	G (ring breathing modes of the DNA/RNA bases); C H deformation (protein) Amide III (a-helix)	1317
1296	CH_2 deformation Ceramide	1296

1280	Differences in collagen content	1274
1267	Triglycerides (fatty acids)	1266
1260	Guanine, cytosine (NH_2) Amide III	1259
1243	Amide III (β sheet and random coils) Asymmetric phosphate [PO_2^- (asym.)] stretching modes (phosphate stretching modes originate from the phosphodiester groups of nucleic acids and suggest an increase the in nucleic acids in the malignant tissues). Phospholipids do not contribute to these bands	1247
1209	Hydroxyproline, tyrosine (collagen assignment)	1208
1173	Cytosine, guanine Tyrosine (collagen type I) δ(C H), tyrosine (protein assignment)	1174
1157	C-C, C-N stretching (protein) Carotenoids (absent in normal tissue) C-C (and C-N) stretching of proteins (also carotenoids) Glycogen	1157
1127	Paraffin ν(C-C) skeletal of acyl backbone in lipid (trans conformation) C-N stretching vibration (protein vibration) ν(C-O)+ ν(C-C), disaccharides, sucrose ν(C-N)	1126
1096	Phosphodioxy group (PO_2^- in nucleic acids)	1097
1092	Phosphodioxy Symmetric PO_2^- stretching vibration of the DNA backbone; phosphate backbone vibration as a marker mode for the DNA concentration; C-N of proteins	1093
1064	Skeletal C-C stretch of lipids Acyl chains ν(C C) trans Fatty acid C N stretching Ceramide	1066
1032	δ(C-H), phenylalanine (protein assignment) C-H in-plane bending mode of phenylalanine Carbohydrate residues of collagen Phenylalanine, C-N stretching of proteins C-H in-plane bending mode of phenylalanine	1032
1003	Phenylalanine (collagen assignment)	1002
959	Hydroxyapatite, carotenoid, cholesterol	958
938	Proline, hydroxyproline, ν(C C) skeletal of collagen backbone C C stretch backbone	938
922-921	C- C stretch Proline ring/glucose/lactic acid/praline ring	923
	Monosaccharides (β-glucose), (C-O- C) skeletal mode	900

899	Disaccharide (maltose), (C-O-C) skeletal mode Adenine	
878	Tryptophan, δ(ring)	877
875	Antisymmetric stretch vibration of choline group $N^{+}(CH_3)_3$, Characteristic of phospholipids Phosphatidylcholine, sphingomyelin	
870	Most probably due to single bond stretching vibrations for the amino acids proline and valine and polysaccharides C-C stretching, hypro (collagen assignment)	870
853	Ring breathing mode of tyrosine and C-C stretch of proline ring Glycogen	853
829	Out-of-plane ring breathing, tyrosine/O-P-O stretch DNA Phosphodiester O-P-O stretchings DNA/RNA Ring breathing tyrosine	829
782	Cytosine/uracil ring breathing (nucleotide)	782
758	Symmetric breathing of tryptophan (protein assignment) Ethanolamine group Phosphatidylethanolamine	758
746	T (ring breathing mode of DNA/RNA bases)	745
727	C-S (protein), CH_2 rocking, adenine C C stretching, proline (collagen assignment) Lipid	725
665	G, T (ring breathing modes in the DNA bases); tyrosine-G backbone in RNA	666
643	C-C twisting mode of tyrosine	643
621	C- C twisting mode of phenylalanine (proteins)	621
598	Phosphatidylinositol	603
548	Cholesterol	546
534	Cholesterol ester ν(S-S) gauche-gauche-trans (amino acid cysteine)	535
522	S-S disulfide stretching in proteins Phosphatidylserine	523
511	S-S disulfide stretching band of collagen ν(S-S) gauche-gauche-gauche (amino acid cysteine)	518
496	Glycogen	498
422	Cholesterol	421
411	Glucose	412

The fingerprint region comparison of both average spectral profiles showed higher intensities in the peaks located at ≈2881, 1667, and 1295 cm^{-1} These peaks are representative of CH_2 and CH asymmetric stretch of lipids and proteins, the C=O stretching characteristic of α-helix structural proteins , and ceramide vibrations respectively. [148], [180]

The first peak (≈2881 cm^{-1}) serves as a first indication of differences in the protein and lipid content of both tissue types. The peak located at ≈1667 cm^{-1} suggested that a higher content of proteins are found in the cancerous area in comparison with the normal breast confirming the differences stated by the first peak.

The peak placed at ≈1295 cm^{-1} represent ceramides and was only visible in the breast cancer spectral profile. Ceramides form part of the structural components of sphingolipids. Ceramides are form by fatty acids and sphingosine. Sphingolipids play an essential role in carcinogenesis as they are membrane lipids and work as signalling molecules that are involved in cell growth and differentiation, apoptosis, senescence and proliferation.[182] Therefore, sphingolipids participate and regulate various cellular processes linked to cancer development, progression, metastasis and therapy resistance. Ceramide synthases and ceramide levels are statistically significantly increased in breast cancer tissue in comparison with benign tumours and normal breast. When calculated, a 12-fold higher of the total ceramide level was identified in breast cancer samples in comparison with normal breast tissue. [183]

Frank et al proposed more than two decades ago the use of the I_{1667}/I_{1449} ratio as first differentiator of breast cancer over normal breast.[85] This ratio compares the protein to lipid content being represented by the C=C stretch in amide I and the CH_2 scissoring in lipids. When this ratio was calculated, the normal breast had a lower ratio than the breast cancer tissue (0.86 vs 0.93), confirming a higher content of lipids on the breast cancer sections. Lipid transformations during cancer have been widely reported, for example, a decrease of 65% triglycerides in cancerous breast compared to normal breast and a 4-fold increase in the phospholipid content for malignant tissue with a consistent increase in unsaturated fatty acids. [184]

Samples of 6 TMA containing cancerous samples and 1 normal breast tissue were analysed obtaining 11,560 spectra. The breast cancer cohort represents a diverse group of cases, thus, a comparison of numerous spectra for both cases can offer a better insight of the malignant behaviour and all the biochemical changes associated with the cancer development and progression. For this end, the results of the multivariate analysis are presented on the next sections.

Comparison of breast cancer and normal breast tissue in the general region, fingerprint, amides (I, II, III), lipids, aminoacid and nucleic acid, hydroxyproline and proline regions were performed using principal component analysis. Different scores and loadings are produced as PCA results. Both are complementary to each other in order to interpret the data. The scores show the dispersion of the samples based on the variance percentage along the analysed area. The loadings are a representation of the contribution of the Raman assignments, therefore it can be assumed that the higher the intensity the higher the contribution of that specific peak for the classification.

The PCA model created was validated using Linear discriminant analysis with a sensitivity of 90% and a specificity of 78%.

6.1 General region (4,000-400 cm^{-1}) analysis of normal breast and breast cancer

Principal component analysis was applied to the general spectral region. A clear separation and grouping was achieved with the PC-1 and PC-2 accounting for 57% of the variance as it can be seen in Figure 57. The loadings associated with these principal components identified the following peaks as responsible for the grouping and separation: 2884, 1583, 1450, 1336, 1245, 1159, 1122, and 1004 cm^{-1}.

The peak located at ≈2884 cm^{-1} has been assigned to the CH_2 and CH stretching of lipids and proteins[180] and was identified to contribute more in the cancer area in comparison with the normal breast. This difference in content might be cause by the accelerated and active cell component production characteristic of cancer development and progression.

The peak at ≈1583 cm^{-1} has been associated with tryptophan and the C=C bending mode of phenylalanine[185] and presented greater intensities on the cancer samples. This component form part of proteins and its contribution might be associated with the structural components in highly dense samples. This behaviour was confirmed by the peak at ≈1450 cm^{-1} also presenting higher intensities than the normal counterpart and which is representative of the CH_2 bending characteristic of malignant tissues. [180]

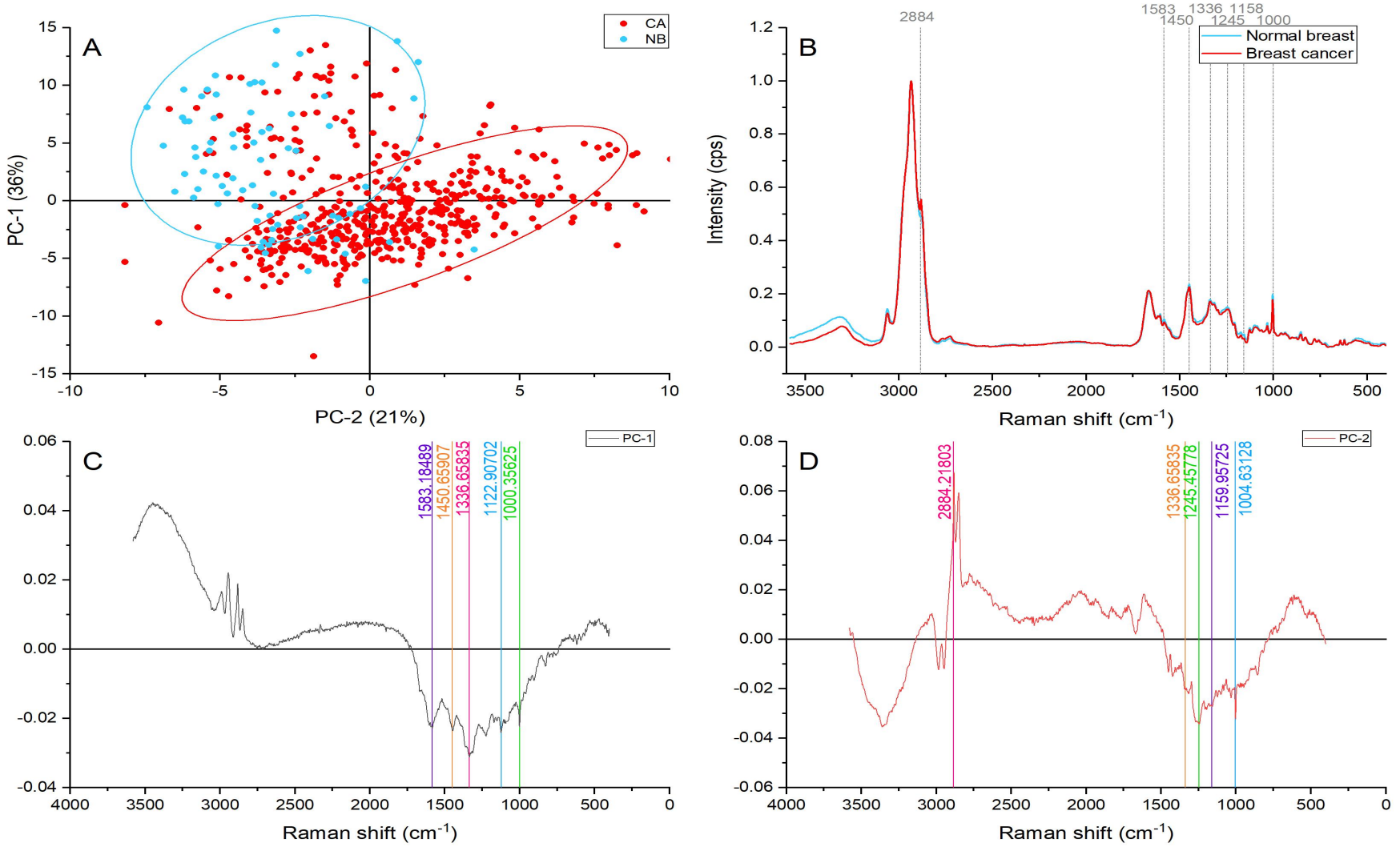

Figure 57. Principal component analysis on the general region (4,000-400 cm^{-1}). (A) Score plot using PC-1 and PC-2 accounting for 57% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average Raman spectral profile for breast cancer and normal breast. Dotted lines represent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

Breast cancer samples presented higher intensities in the peak located at ≈1336 cm^{-1}. This peak has been assigned to different components, and provide important information on the biochemical changes associated with the carcinogenesis process. This band is associated with the amide III, and polynucleotide chains, and CH_2 wagging vibrations from glycine backbone and proline side chain in collagen.[181] Higher intensities in this peak suggest higher content of proteins, nucleic acids, and collagen in the cancerous samples. The ≈1245 cm^{-1} peak representing an amide III assignment was identified by the PC-2 and also presented higher intensities in the cancerous samples.

The normal breast samples presented higher intensities than the average on the 1159 cm^{-1} peak. This band is associated with the C=C stretch mode characteristic of carotenoids and β-carotenoids. In addition, the peak at 1122 cm^{-1} representing C-C stretch on carotene and breast lipids also presented the same trend. Carotenoids play a vital role in different processes at cellular level; they offer oxidative damage protection, upregulate connexion junctional communication and prevent the lipid oxidation. Gap junctional communication (GJC) has been suspected to participate in carcinogenesis. Moreover, normal breast tissue proved to present a higher quantity of carotenoids than cancerous tissue, acting as a fatty acids reservoir. [112] Carotenoids role in normal breast tissue suggests protection against cancer development. [186] Normal tissue spectra had similarities with oleic acid, linoleic acid, and docosahexaenoic acid. Furthermore, carotenoids, oleic acid, and docosahexaenoic acid are well-known to play a protective role in breast cancer. [187] The decrease intensity of carotenoids peaks suggested a lower content of these components in the malignant tissue. Brożek-Płuska's publications suggested that the oxidation of aging-related pigments might be the reason of the carotenoids reduction.[188] Furthermore, the carotenoids reduction has been also explained through the oxidation of lipofuscin material, and the lipid reduction was attributed to lipid peroxidation.[88]

Furthermore, the peak at 1000-1004 cm^{-1} presented greater contributions in cancerous samples and was assigned to phenylalanine. [180] Different authors focus on this peak as a collagen specific assignment. The literature has reported collagen fibre deposition, rearrangement of fibres and overproduction occurring when tumours develop and grow. [189] Consequently an increase of intensity in peaks associated with stromal overproduction were expected as classifiers of cancerous samples. Nonetheless, stromal components were an important and dominant part of the normal breast sections and they could have been included in the Raman acquisition. Besides, a phenylalanine assignment can be correlated to proteins and structural components of the epithelial cells.

6.2 Fingerprint region (1,800-500 cm^{-1}) analysis of normal breast and breast cancer

The analysis of the spectral region was narrowed down to the fingerprint region resulting in a good separation and grouping of the cancer and normal breast as seen in Figure 58. The following peaks were identified as responsible for the separation and grouping of samples: 1668, 1586, 1451-1449, 1275, 1090, 1004, 784, and 546 cm^{-1}. Some of the listed peaks were already discussed on the general region findings and will not be discussed if repeated. The new peaks are identified in Figure 58B with dotted lines.

The peak located at ≈1586 cm^{-1} is assigned to a phenylalanine and hydroxyproline vibrations. Both biological component suggests that this peaks is caused by the collagen content. The PCA suggested that the normal breast samples and the cancerous samples located on the negative axis of PCA presented higher intensities on this collagen peak. The cancerous epithelia were visibly denser than the normal epithelia, therefore the inclusion of collagen in stromal can be assumed in the normal breast data collection. The ECM plays an important role in tumour progression and metastasis. Talari et al reported collagen production and remodelling caused by hypoxia processes through their involvement in the transcriptional status of procollagen genes.[190] On the other hand, collagen content changes, have a close relationship with stromal desmoplasia, which is a malignancy response.[80]

Representing nucleic acids bands, the peaks located at ≈1090 and 784 cm^{-1} representing symmetric phosphate stretching vibrations and ring breathing modes in DNA and RNA bases[148], [191] showed higher intensities in cancerous areas. This fact suggests that the nucleic acids are present in greater quantity in the malignant tissue. Higher cell density and the overproduction of genetic material is characteristic of early stages of carcinogenesis. In addition, the phosphate band can also serve as indicator of higher metabolic activity.[74]

The peak found at ≈1275 cm^{-1} represent an amide III contribution and is a protein characteristic vibration including collagen.[180]. This band presented higher contributions in the normal breast due to the higher content of collagen suggested previously. Amide III represents α helix conformations [192] providing information of the structural conformation of proteins and collagen. The lower intensity in α helix characteristic peaks suggested conformational changes in the cancerous tissue. This finding is in agreement with our FTIR results where higher relative quantity of α-helix conformation was proposed in the normal breast in comparison with cancerous tissue. These changes indicate modifications in the number of hydrogen intermolecular bonds in the secondary structural conformation in the cancerous samples.[102]

A cholesterol related peak located ≈546 cm^{-1} presented greater contributions in the normal breast readings. Higher content of oxidize and modified cholesterol has been reported as sign of histological progression from normal epithelium to hyperplasia. [193] However, a starting stable cholesterol content is expected in normal breast. [194]

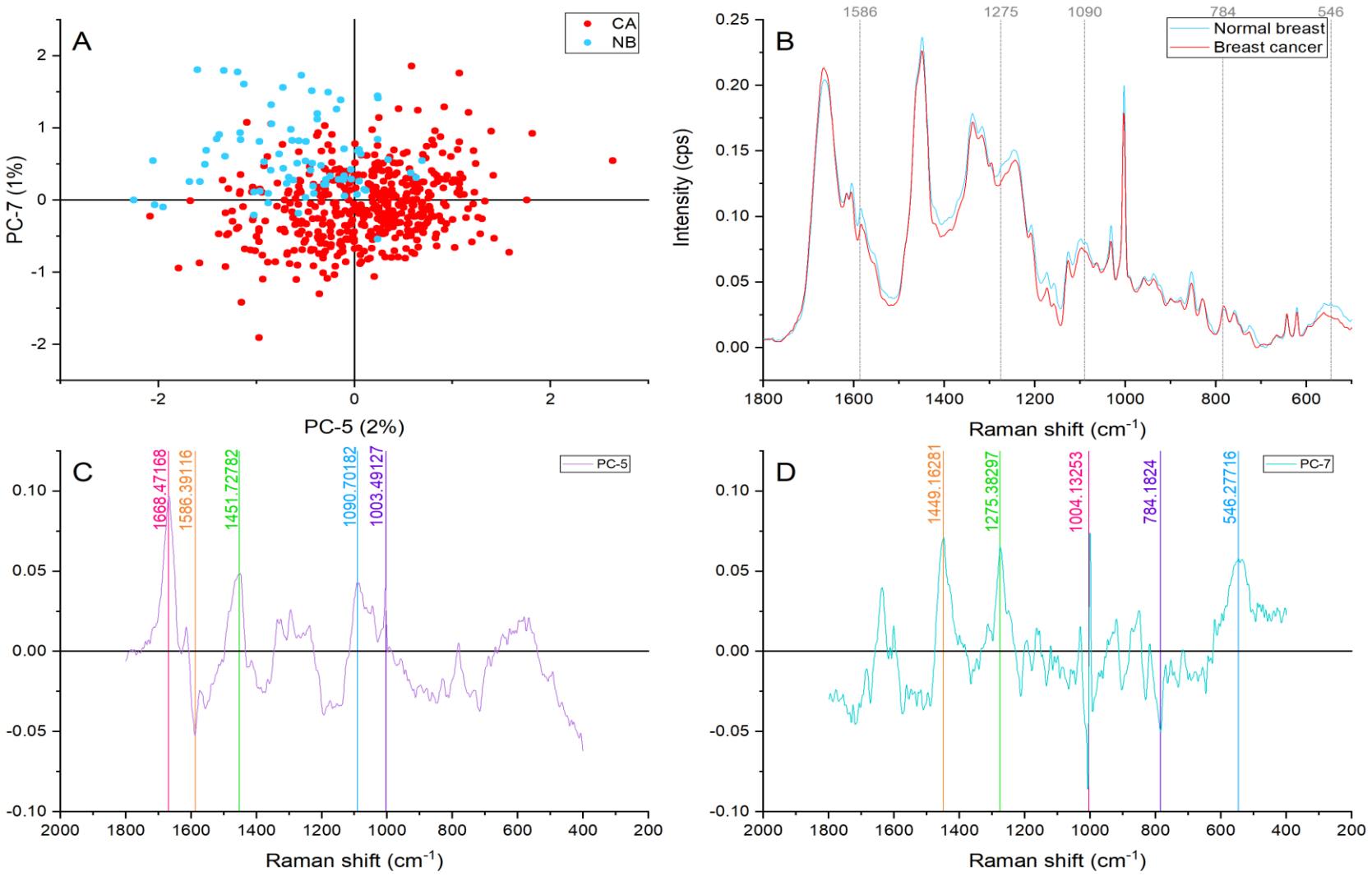

Figure 58. Principal component analysis on the fingerprint region (1,800-500 cm^{-1}). (A) Score plot using PC-5 and PC-7 accounting for 3% of the variance, (B) Average Raman spectral profile for breast cancer and normal breast. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-5 Loading, (D) PC-7 loading.

6.3 Amides region (1,800-1,510 cm^{-1}) analysis of normal breast and breast cancer

Principal component analysis of the amide spectral region resulted in a decent separation and good grouping of normal and cancerous breast samples when the PC-2 and PC-3 were used as shown in Figure 59. Peaks identified by the general and fingerprint region were also identified in this analysis as follow: 1615, 1590-1588. New peaks were found as responsible for the grouping of samples and are presented in Figure 59B. These bands are located at ≈1667, 1605, 1448, 1291, 1243, and 1204 cm^{-1}.

The first peak located at ≈1667 cm^{-1} is cause by the vibrations of amide I. The amide I peak is the result of the carbonyl (C=O) stretching vibrations of the amide group, but it is also representative of the α-helix structures of proteins.[89] The first visible evidence of the change in this peak is a upshift in frequency from 1663 cm^{-1} in the normal breast to 1667 cm^{1} in the cancerous area. Manoharan has previously reported this shift towards higher wavenumbers in addition to a widening of the peak. [80] In our case, a minimal widening can be appreciated in the spectral profile, but a higher intensity of this band is visible in the cancerous peak and has been confirmed by the score plot. Greater contribution of proteins with dominant α-helix structures is suggested by the higher intensities in breast cancer samples.

Higher content of proteins suggested by increased intensity of phenylalanine and tyrosine peak located at ≈1605 cm^{-1} was also present in the malignant tissue. In the same form, changes linked to protein and lipid content was suggested by variable intensities in the band located at ≈1448, which is assigned to CH deformations.

Nucleic acids variation were also suggested in the amide analysis with the ≈1291 cm^{-1} peak. This peak is linked to cytosine vibrations [148] and its dominance over the normal breast analogue peak can be appreciated with the naked eye in the spectral profiles (Figure 59B). Higher intensities than the average were confirmed for the cancerous areas in comparison with the normal breast. The production of genetic material and cellular content is expected as part of the cell division which form parts of the uncontrolled proliferation in malignancies.

The bands found at ≈1243 and ≈1204 cm^{-1} are both related to collagen and proteins. These assignments represent the amide III vibration, which is caused by the CH_2 wagging coupled with the C-N stretching and pyrimidine bases (C-T). [180]. These bands are considered collagen assignments as the CH_2 wagging of the glycine backbone and proline side changes contribute considerably to the total intensity of this peak. Normal breast presented higher contributions of these peaks due to the inclusion of collagen in the Raman data collection. On the other hand, some cancerous samples presented considerable contributions of these two peaks. As previously discussed, the ECM present adaptation changes in order to prepare the cancer cells for further invasion and hypoxia survival. [107], [190] Among the adaptation changes, the overproduction of collagen is present by fibroblast stimulation.

Further analysis focused on the amide I (1800-1510 cm^{-1}), amide II (1510-1390 cm^{-1}), and amide III (1390-1140 cm^{-1}) regions resulting in partial separation and good grouping of normal and cancer samples as presented in Figure 60.

The amide peaks as observed in proteins are characteristically broad as result of superposition of peptide contributions.

The peaks identified in the loadings associated with the score plots presented in Figure 60A, C, and E are presented in the average spectral profiles in Figure 60B, D, and F. Some of these peaks have been discussed in the previous sections and only the new peaks found will be discussed.

6.3.1 Amide I region (1,510-1,800 cm^{-1})

Peaks in the amide I region are highly sensitive to conformational changes in the secondary structure of proteins.[89] The most important band representing amide I is formed by the overlapping of three bands. However, the amide I β-turn consists of two overlapping bands located around ≈1676 and ≈1652 cm^{-1}. The former represent the antiparallel β-sheet structure, whilst the second band represent the α-helix structure.[195] Conformational differences were responsible for the grouping of normal breast samples. The peak located at≈1648 cm^{-1} related to random coil vibrations presented higher intensities in comparison with breast cancer samples. Complementarily, the band placed at ≈1676 cm^{-1} caused by amide I β-sheet conformations showed higher intensities in the cancerous spectra. The change in conformational changes has been reported to be caused by ECM adaptation in hypoxia,[190] and differentiator of normal and cancerous cells in in-vitro studies [74]

Also, the peak located at ≈1615 cm^{-1} had higher intensities in the cancerous area spectra. This band is caused by the tyrosine, tryptophan, and C=C vibrations in proteins.[148] This vibration indicates the overproduction of cellular components leading to abnormal proliferation.

6.3.2 Amide II region (1,510-1,390 cm^{-1})

The amide II region is less sensitive for conformational changes. Nucleic acids have representative bands in this region located at ≈1461 and 1489 cm^{-1}. The first band is a deoxyribose vibration. However, it has also been reported as a CH_2 wagging and CH_3 deformation of lipids and collagen. The later spectral peak is attributed to DNA vibrations. These two bands presented greater intensities in the cancerous samples. The dominant intensities in the cancerous area suggest an accelerated genetic material production leading to cell division.

Normal breast samples presented higher intensities in the 1423 cm^{-1} band. This peak is caused by the N-H in plane deformation mainly representative of proteins. [180]. This deformation freedom might be caused by the vibrational modes found in strict strongly packed structures in normal breast in comparison with the much relaxed β-structures in cancerous areas. [196]

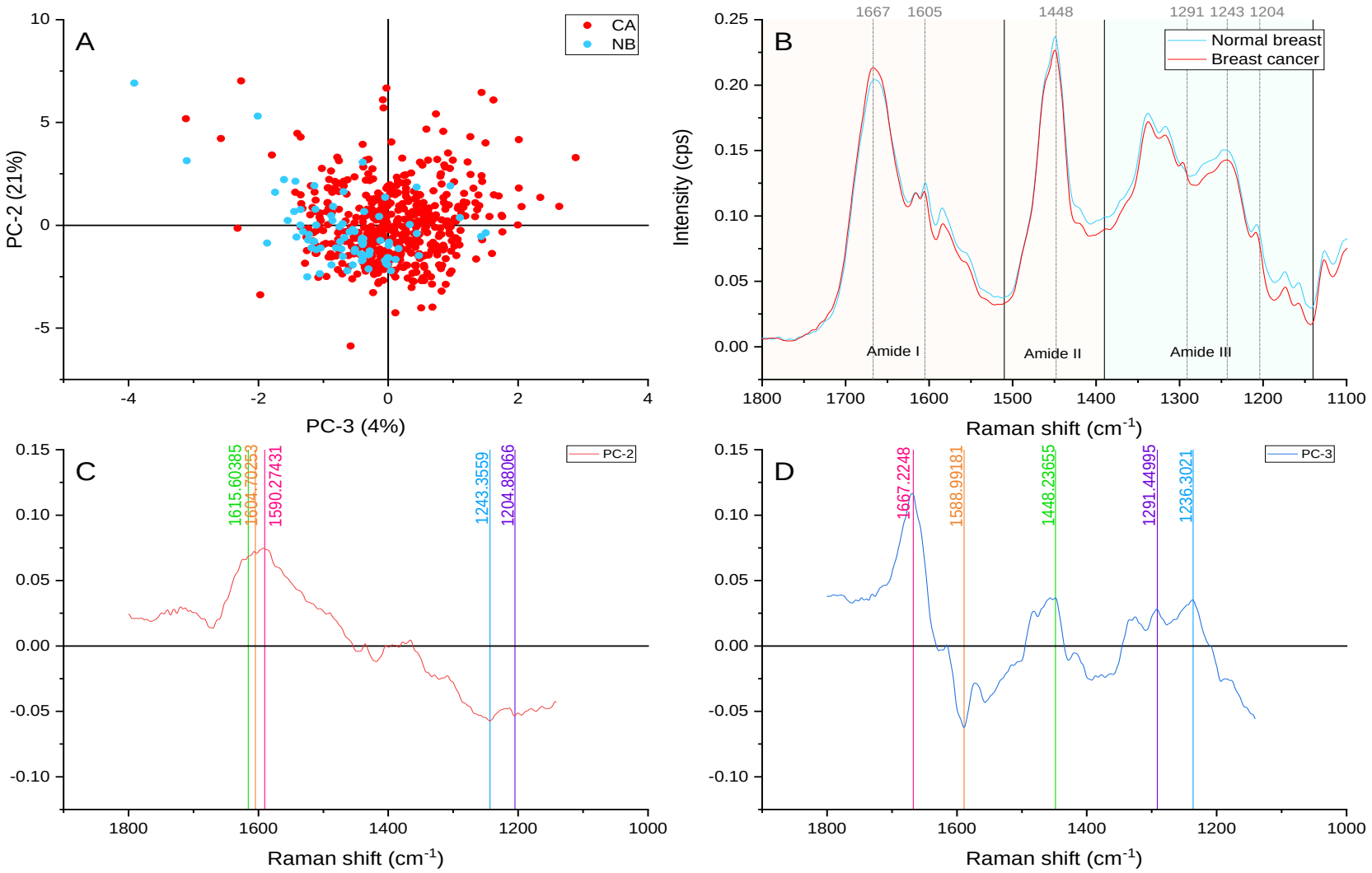

Figure 59. Principal component analysis on the amide region (1,800-1,510 cm^{-1}). (A) Score plot using PC-2 and PC-3 accounting for 25% of the variance, (B) Average Raman spectral profile for breast cancer and normal breast. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-2 Loading, (D) PC-3 loading.

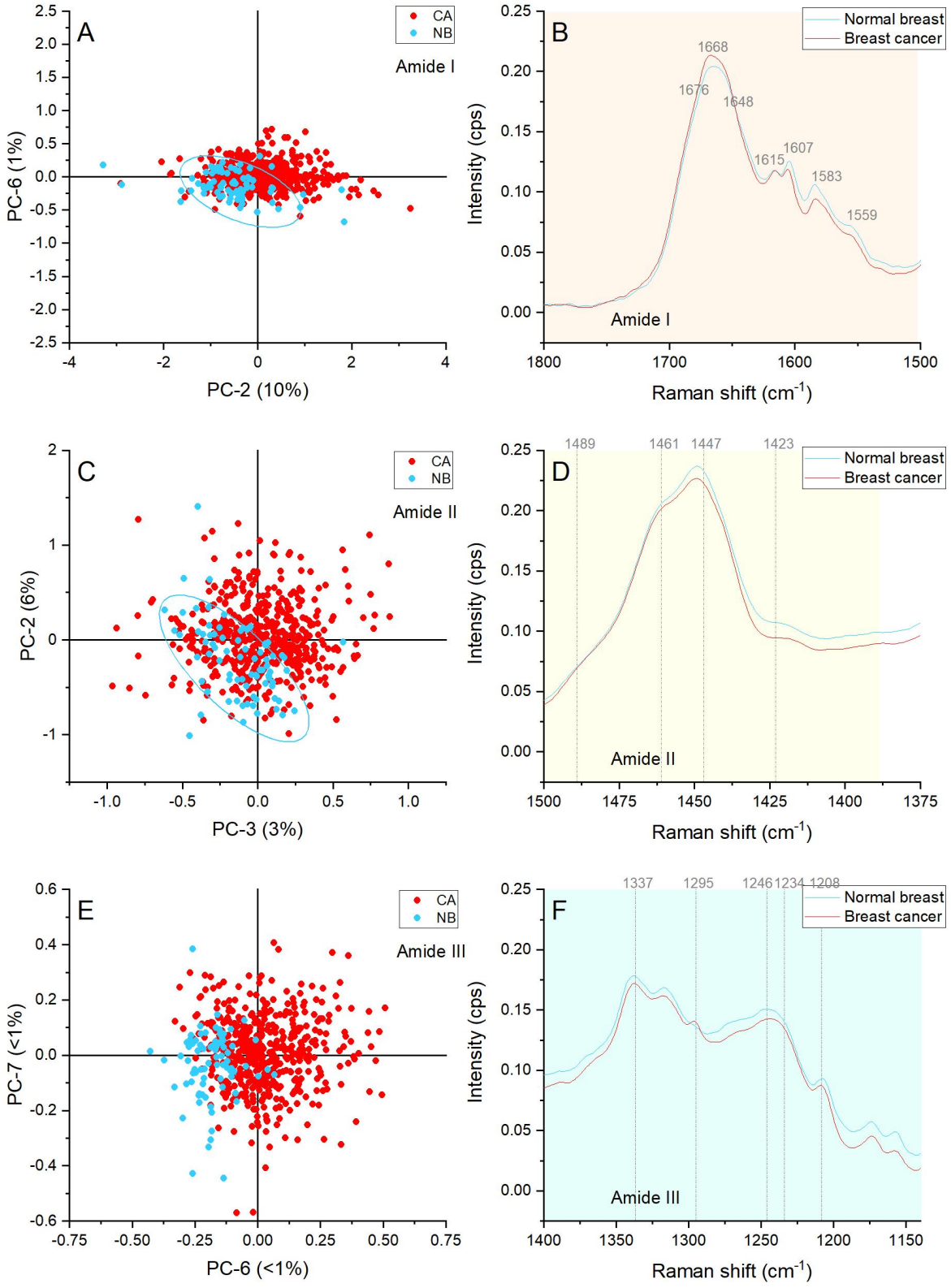

Figure 60. Principal component analysis of the amide region. (A) Score plot of the amide I region (1,800-1,510 cm^{-1}), (B) Average Raman spectral profile for breast cancer and normal breast with peaks responsible for the separation in the amide I region, (C) Score plot of the amide II region (1,510-1,390 cm^{-1}), (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^{-1}), (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

6.3.3 Amide III region (1,390-1,140 cm^{-1})

The amide III region is associated with coupled C–N stretching and N–H bending vibrations of peptide groups.[196] The peaks found as responsible for the grouping of cancerous and normal breast were caused by ceramides (≈1295 cm^{-1}), Amide III from a concerted ring mode(≈1234 cm^{-1}), and tryptophan, phenylalanine, and amide III vibrations characteristic of proteins (≈1208 cm^{-1}). All this peaks presented greater intensities in the cancerous areas in comparison with the normal breast. These differences suggested higher metabolic activity and cell density in the malignant tissue. In addition, the ceramide peak serve as an important biomarker due to its participation as structural component of sphingolipids. Sphingolipids are highly relevant in carcinogenesis as they are linked to various cellular processes such as cancer development, progression, metastasis and resistance to therapy.[183]

6.4 Amino acids and nucleic acids region (980-600 cm^{-1}) analysis of normal breast and breast cancer

Principal component analysis of the amino acids and nucleic acid region resulted in excellent grouping of the cancerous samples and relatively good separation of normal breast and breast cancer samples as presented in Figure 61. Several new peaks were identified in this area responsible for the grouping of sample points.

Among the identified peaks there were different collagen associated bands such as ≈921 cm^{1} and ≈939 cm^{-1}. The former band is cause by the C-C stretch of proline ring, while the later band represent the C-C stretch of the skeletal collagen backbone (proline and hydroxyproline). These peaks were presented in both sample types with important intensities, however, this can be explaining with the dominant stromal component in the normal breast samples and the overproduction and hyperactive activity of carcinoma-associated fibroblasts (CAFs) in the malignant tissue. The stromal microenvironment actively contributes to the development of breast tumours. Malignant cells recruit stromal cells, (inflammatory cells, vascular cells, and fibroblasts) into the tumour tissue, in order to promote its growth. Desmoplastic tumour stroma or reactive tumour stroma produce granulation tissue mainly composed by accumulated extracellular matrix. Desmoplastic stroma has been found in breast epithelial malignancies.[197] Besides, the fibroblasts in malignant masses have different biological characteristics in comparison with normal fibroblasts.

Breast cancer samples presented more intense peaks located at ≈830 cm^{-1} produced by PO_2^- stretch of nucleic acids, ≈644 cm^{-1} caused by the C-C twisting mode of tyrosine and phenylalanine characteristic of proteins, and ≈691 cm^{-1} which represent cholesterol ester vibrations in comparison with normal breast. The dominant cancerous peaks in these assignments are caused by a higher metabolic activity as reported in previous works. [74], [89]

Normal breast spectra showed greater intensities in the ≈719 cm^{-1} band that represents the C-N vibration in membrane phospholipid head. [180] A decrease in phospholipids in malignant breast tissue has been proposed by different authors. [84], [198]–[200] Different explanations

such as synthesis and breakdown of choline phospholipids caused by specific genetic alterations and phospholipids alterations affecting the membrane fluidity.

6.5 Lipid region (3,100-2,680 cm^{-1})

Differentiation and grouping of normal breast and breast cancer samples could been seen when the lipid region was analysed (Figure 62). Several peaks constituting the 3 clear visible bands were responsible for the differentiation of normal and malignant tissue.

Aromatic CH stretching modes characterised by the ≈3061 cm^{-1} peak and CH2 asymmetric stretching of lipids and fatty acids assigned to the band located at ≈2946 cm^{1} presented higher intensities in normal breast spectra. Different studies reported dominance of lipids, specifically fatty acids in normal tissue spectra. Whereas, protein dominates abnormal tissue, both benign and malignant.[74], [78], [79] The ≈2961 cm^{-1} band also showed important differences in the normal breast and cancer spectra. Out-of-plane chain end antisymmetric CH_3 stretch band in lipids and fatty acids vibrations are responsible for this band.[180]

Breast cancer spectra showed more intense bands associated with protein vibrations such as ≈2928, 2881, and 2849 cm^{-1}. These peaks were produced by the symmetric CH_3 stretch due primarily to protein, CH_2 asymmetric stretch of lipids and proteins, and symmetric stretch of C-H respectively. A greater contribution of proteins can be attributed to; uncontrolled and abnormal cell proliferation, cell division, and migration in the malignant tissue. [201], [202]

6.6 Hydroxyproline and proline region (960-810 cm^{-1})

Hydroxyproline and proline region analysis was included due to its informative nature on the collagen component of the breast samples. Excellent grouping was achieved when PC-4 and PC-7 as presented in Figure 63. Normal breast samples showed dominant intensities in the bands located at ≈956 and 851 cm^{-1}. The former band is assigned to carotenoids.[180] Oleic acid and carotenoids have been identified to dominate normal breast tissue. [186] Carotenoids have antioxidant properties reducing DNA damage and mutations, in addition to offering immune protection. Carotenoids play an important role in cellular processes that avoid the development of pathologies in the breast tissue. [112]

The ≈851 cm^{-1} band is caused by proline, hydroxyproline and glycogen vibrations. Collagen is protein consisting of repetitive proline-rich tripeptide domains composed by three main amino acids: glycine, proline and hydroxyproline (Gly-X-Y). The fibre-like structure of collagen allows it to produce a highly interconnected network throughout the ECM, providing the structural integrity and biochemical properties of the connective tissue.[203]

On the other hand, breast cancer samples presenter greater contribution of the following peaks: ≈941, 899, and 828 cm^{-1}. The stretching of C-C skeletal backbone of collagen represented by the band located at ≈941 cm^{-1} represented adaptation in the ECM on the cancerous samples.[107] Collagen can go through structural rearrangement resulting in the exposure of proline and hydroxyproline amino acids affecting the intensities of this band. [203] The ≈828 cm^{-1} band had also being assigned to hydroxyproline and proline, however tyrosine vibrations and stretching of PO_2^- in nucleic acids also had been attributed to this peak.

In combination with the adenine peak placed at ≈899 cm^{1} the nucleic acids vibrations are evidence of higher metabolic activity and cell density in the cancerous samples.

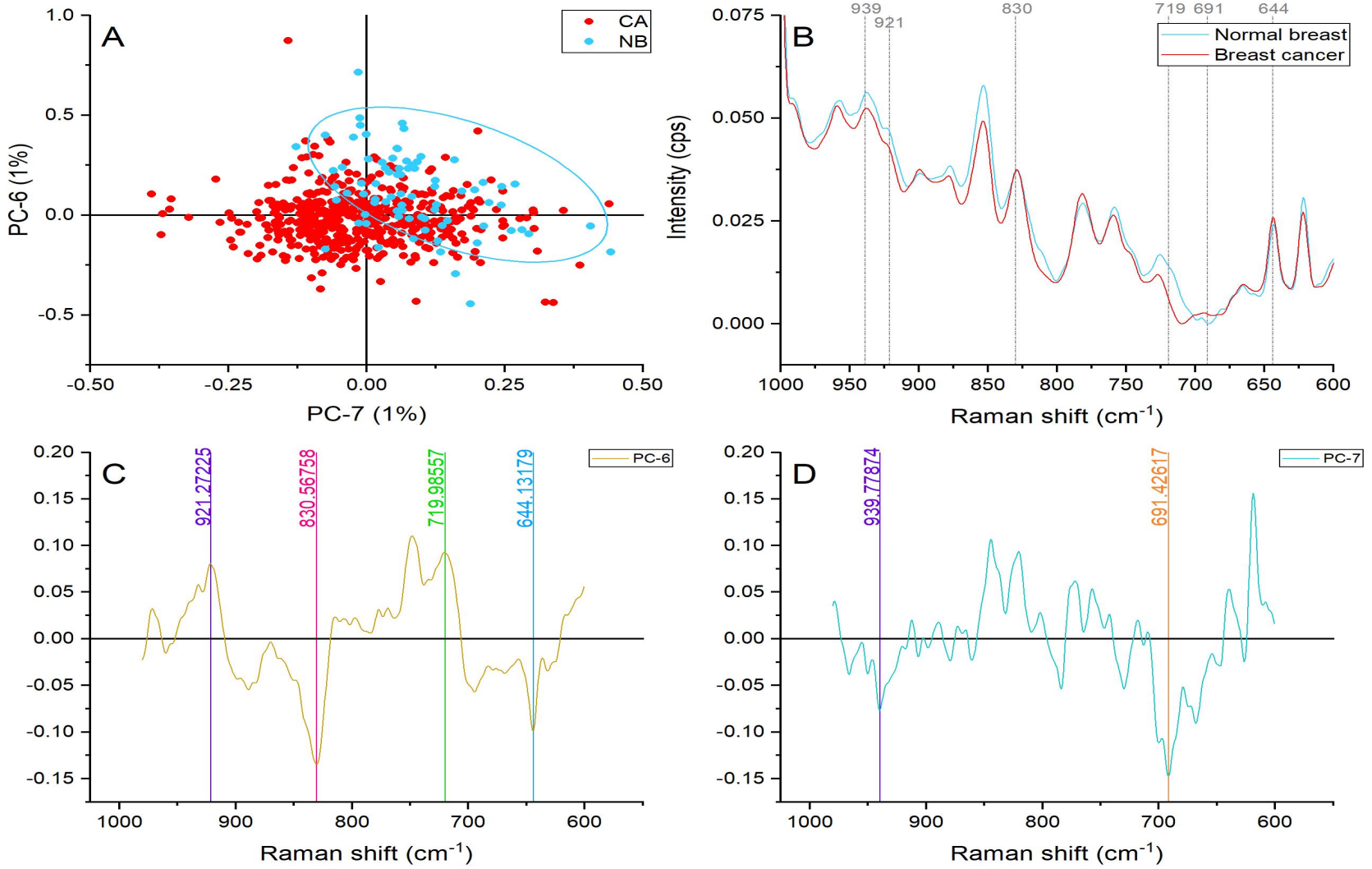

Figure 61. Principal component analysis on amino acids and nucleic acids region (980-600 cm^{-1}). (A) Score plot using PC-6 and PC-7 accounting for 2% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average Raman spectral profile for breast cancer and normal breast. Dotted 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.

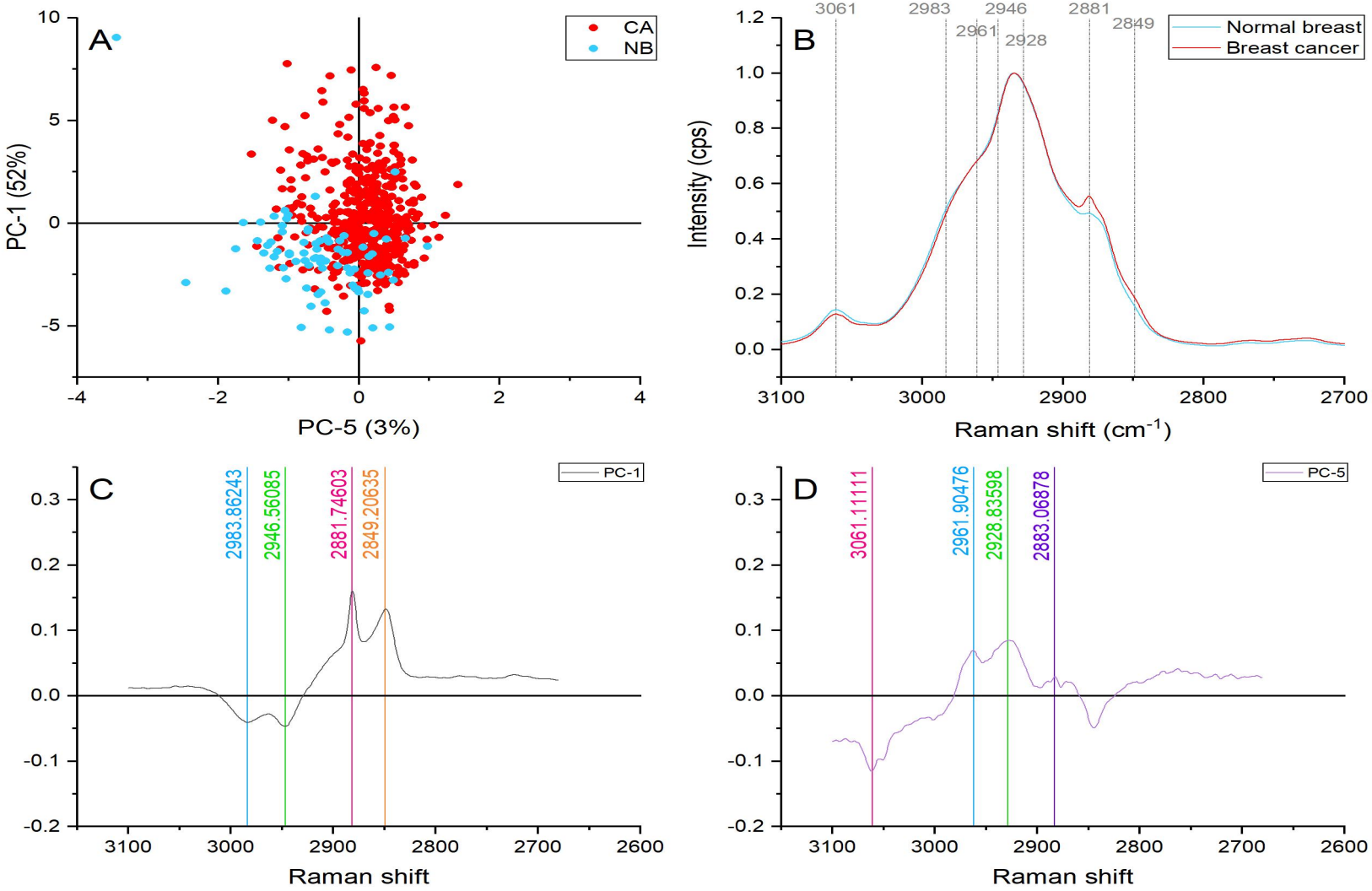

Figure 62. Principal component analysis on the lipid region (3,100-2,680 cm^{-1}). (A) Score plot using PC-1 and PC-5 accounting for 55% of the variance, (B) Average Raman spectral profile for breast cancer and normal breast. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-1 Loading, (D) PC-5 loading.

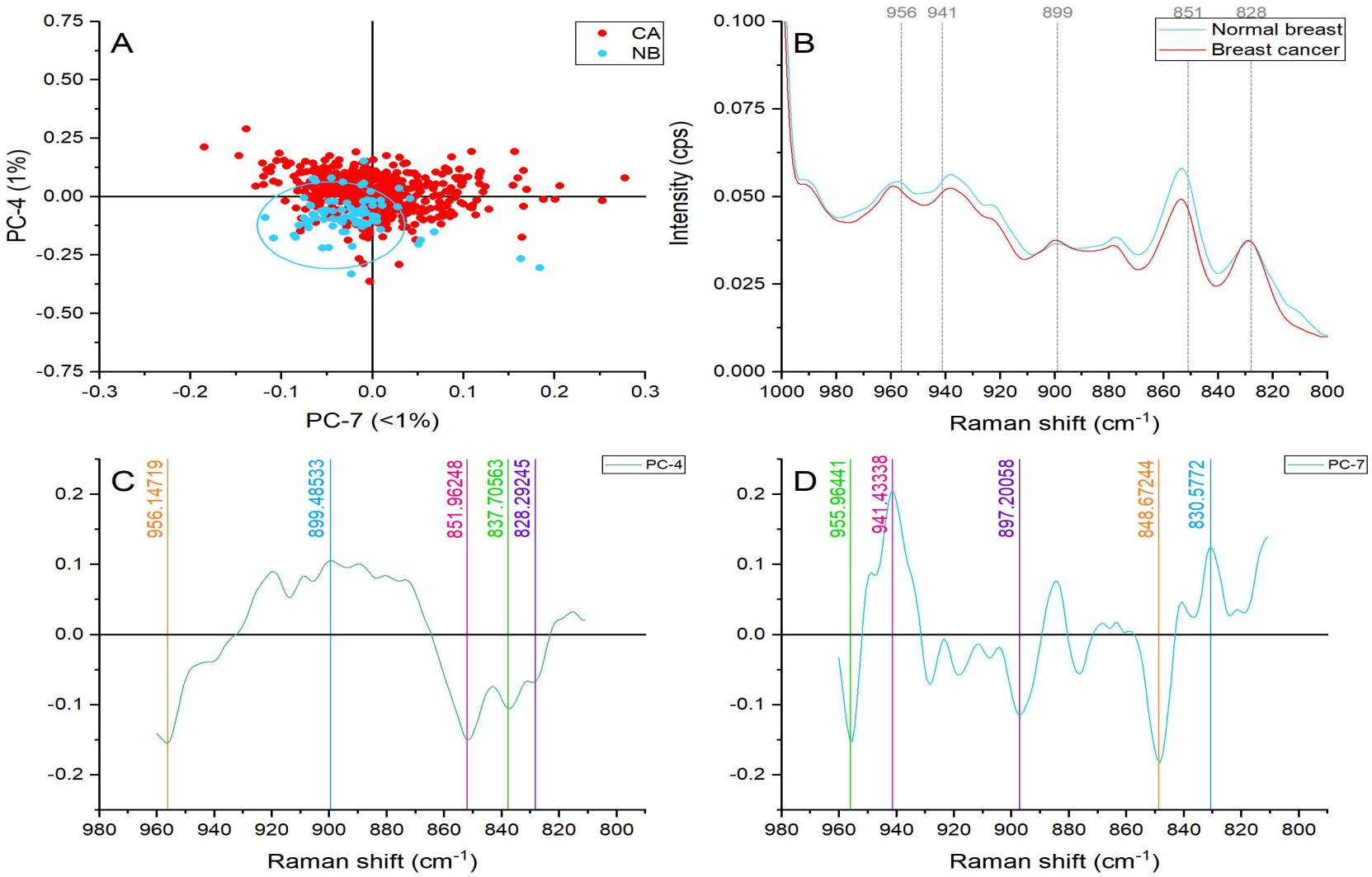

Figure 63. Principal component analysis on the hydroxyproline and proline region (960-810 cm^{-1}). (A) Score plot using PC-4 and PC-7 accounting for 2% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average Raman spectral profile for breast cancer and normal breast. Dotted 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.

CHAPTER 7. RAMAN ANALYSIS OF BREAST CANCER

7.1 COMPARISON OF THE SPECTRAL DATA OF DIFFERENT SUBTYPES WITHIN A SPECIFIC GRADE.

The results and discussion presented in this section include the Raman spectral analysis of 6 tissue microarrays containing African breast tumours from similar tribal origins in Nigeria comparing different grades and subtypes.

A total of 499 cancerous samples were analysed with Raman spectroscopy. Two comparison types were performed, the first one comparing different subtypes within one grade and the second one comparing the spectral data of different grades within a specific subtype.

7.1.1 GRADE 1 ALL SUBTYPES

When compared, the average spectral profile of grade 1 samples presenting luminal, HER2+ rich and triple negative breast cancer (TNBC) subtypes showed differences enough to see with the naked eye as presented in Figure 64 and Figure 65.

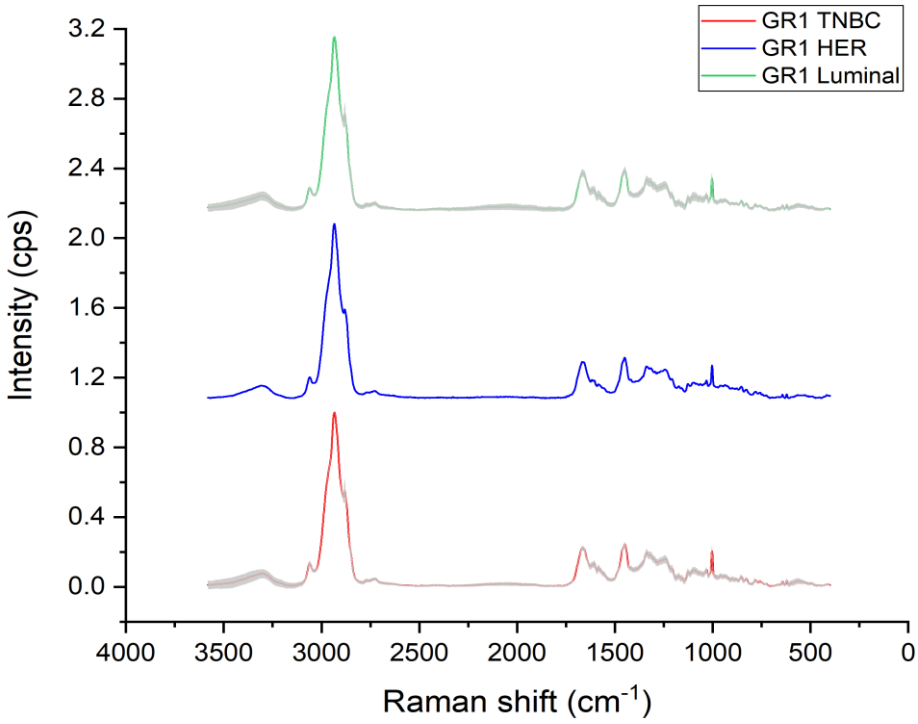

Figure 64. Raman average spectral profile of grade 1 samples presenting triple negative (TNBC), luminal and HER2+ subtypes, variance is represented with grey shadow.

7.1.1.1 LIPID REGION (3,100-2,680 CM^{-1}) ANALYSIS OF GRADE 1 SAMPLES ALL SUBTYPES

The comparison of the different subtypes of grade 1 spectral profiles showed overall differences in intensity and peak ratios. A good trend showing decent grouping and partial separation was obtained when the lipid region was analysed in the grade 1 breast cancer samples as seen in Figure 66.

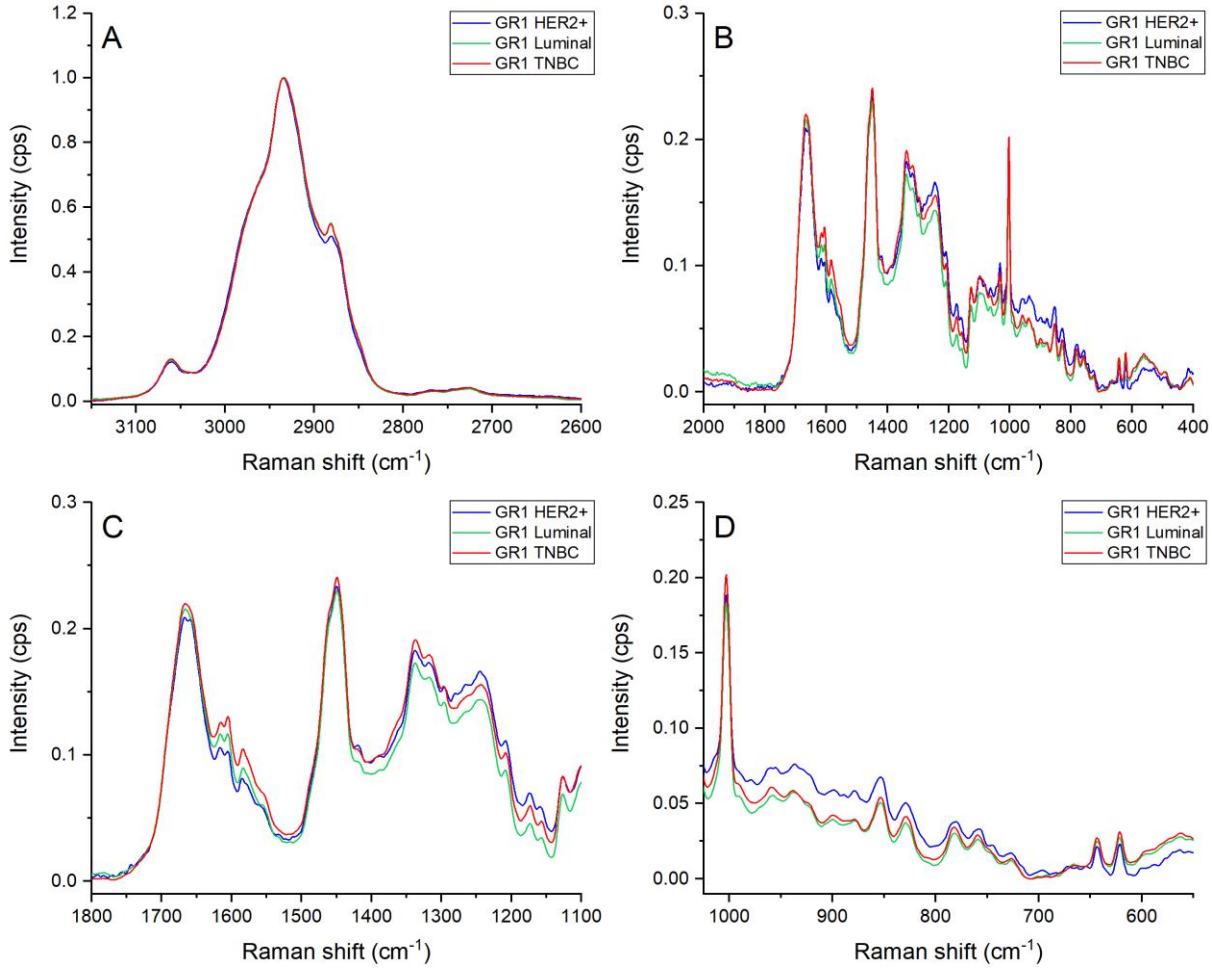

Figure 65. Raman average spectra of breast cancer (grade 1) samples. (A) Lipid region (3,100-2,680 cm^{-1}), (B) Fingerprint region (1,800-500 cm^{-1}), (C) Amides region (1,800-1,140 cm^{-1}) (D) Amino acid and nucleic acids regions (980-600 cm^{-1}).

The following bands were identified using the loadings presented in Figure 66C and D: 2987, 2946, 2928, 2883, and 2844 cm^{-1}. These peaks showed differences according to the subtype they presented. The peak located at ≈2987 represents a cholesterol ester contribution and presented higher content on the luminal samples and HER2+ followed by the TNBC samples. The intratumour accumulation of cholesteryl ester is closely connected to proliferation and aggressive potential of breast cancer tumours.[204]

The same trend was detected in the ≈2946 and ≈2883 cm^{-1} peaks. Both bands represented CH and CH_2 vibrations in lipids and proteins.[180] Additionally, a downshift can be seen from ≈2883 cm^{-1} in the luminal and triple negative breast cancer spectral profile to ≈2878 cm^{-1} in the HER2+ samples (Figure 66B). The ≈2880 cm^{-1} band has demonstrated sensitivity to identify intermolecular packing interactions and serve as indication of ordered lateral chain–chain interactions in lipids.[205], [206] Therefore, this shift might suggest rearrangement associated with lipids structural changes, which can be related to cell membrane adaptations.

The bands identified by PC-4 suggested greater contributions of the ≈2928 and 2844 cm^{-1} peaks in the luminal samples with lower intensities in the TNBC and HER2+ samples. The first band is associated with the symmetric stretch caused primarily due to proteins.[180] Hence, the presence of higher content of proteins in positive hormonal breast cancer cases can be suggested. Analysing the similar trend of the band located at ≈2844 cm^{-1} caused by the symmetric stretching of CH_2 in lipids especially fatty acids, lipid peroxidation seemed to have an early active participation in TNBC in comparison with its positive counterpart (Luminal subtype).

7.1.1.2 Fingerprint region (1,800-500 cm^{-1}) analysis of Grade 1 samples all subtypes

When the fingerprint region of grade 1 samples was analysed with principal component analysis it showed decent grouping for all three subtypes (Figure 67). Noticeable variations on peaks intensities and sharpness can be appreciated among all the spectral area. The bands responsible for the grouping of the subtypes are indicated in Figure 67B.

Luminal grade 1 spectra presented more intensity in the bands located at ≈1668, 1333, 1312, 1203, and 1000 cm^{-1}. Therefore, it can be suggested that the biomolecules associated with those peaks are present in higher amounts in the luminal grade 1 samples in comparison with TNBC and HER2+. The only HER2+ grade 1 sample showed intensities closer to the average for all these assignments. On the contrary to the luminal subtype, the TNBC sample points were located on the opposite axis suggesting less contribution of the listed peaks.

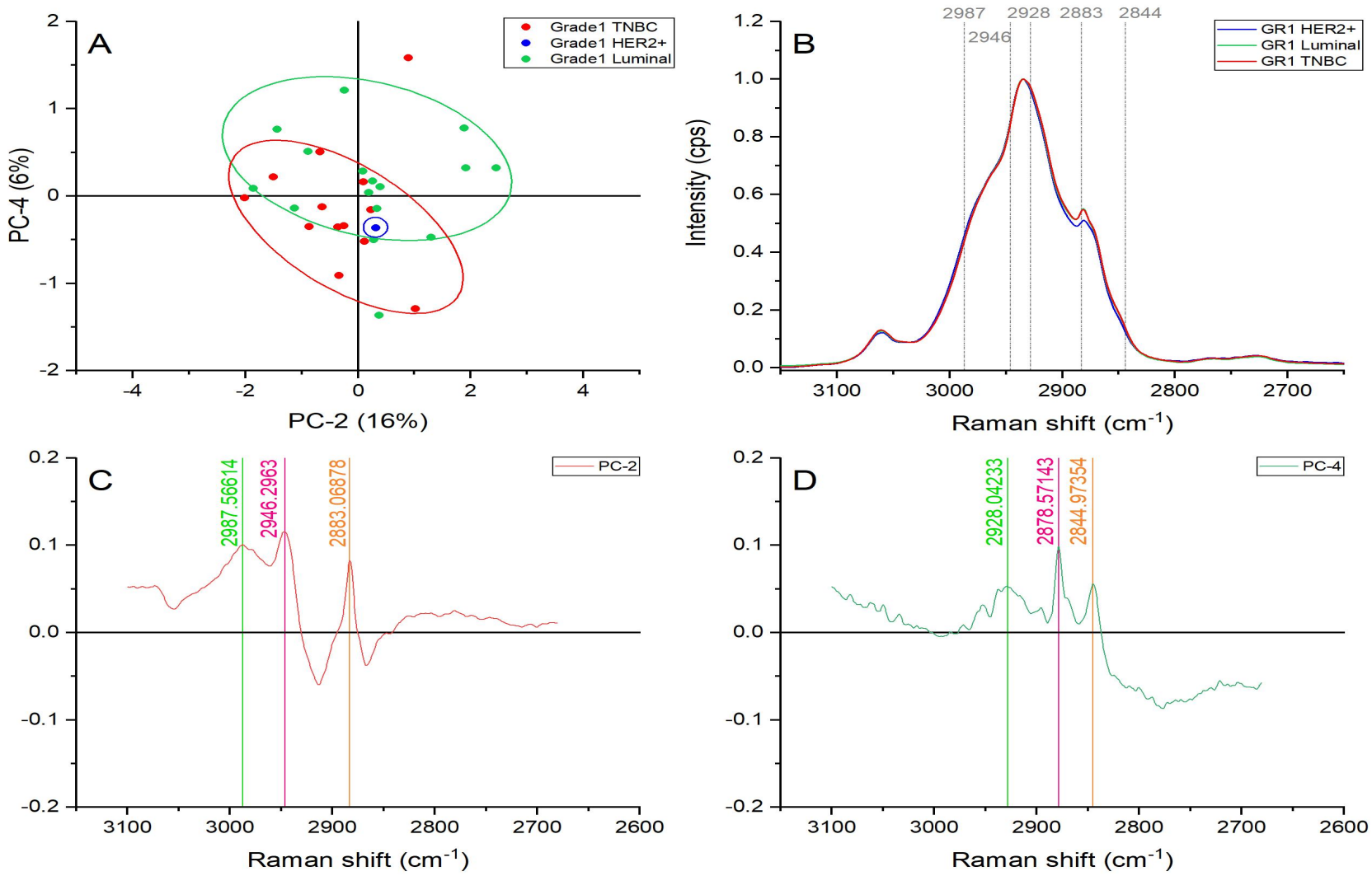

Figure 66. Principal component analysis of the lipid region (3,100-2,680 cm^{-1}). (A) Score plot using PC-2 and PC-4 accounting for 22% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the lipid region for grade 1 TNBC, luminal, and HER2+. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-2 Loading, (D) PC-4 loading.

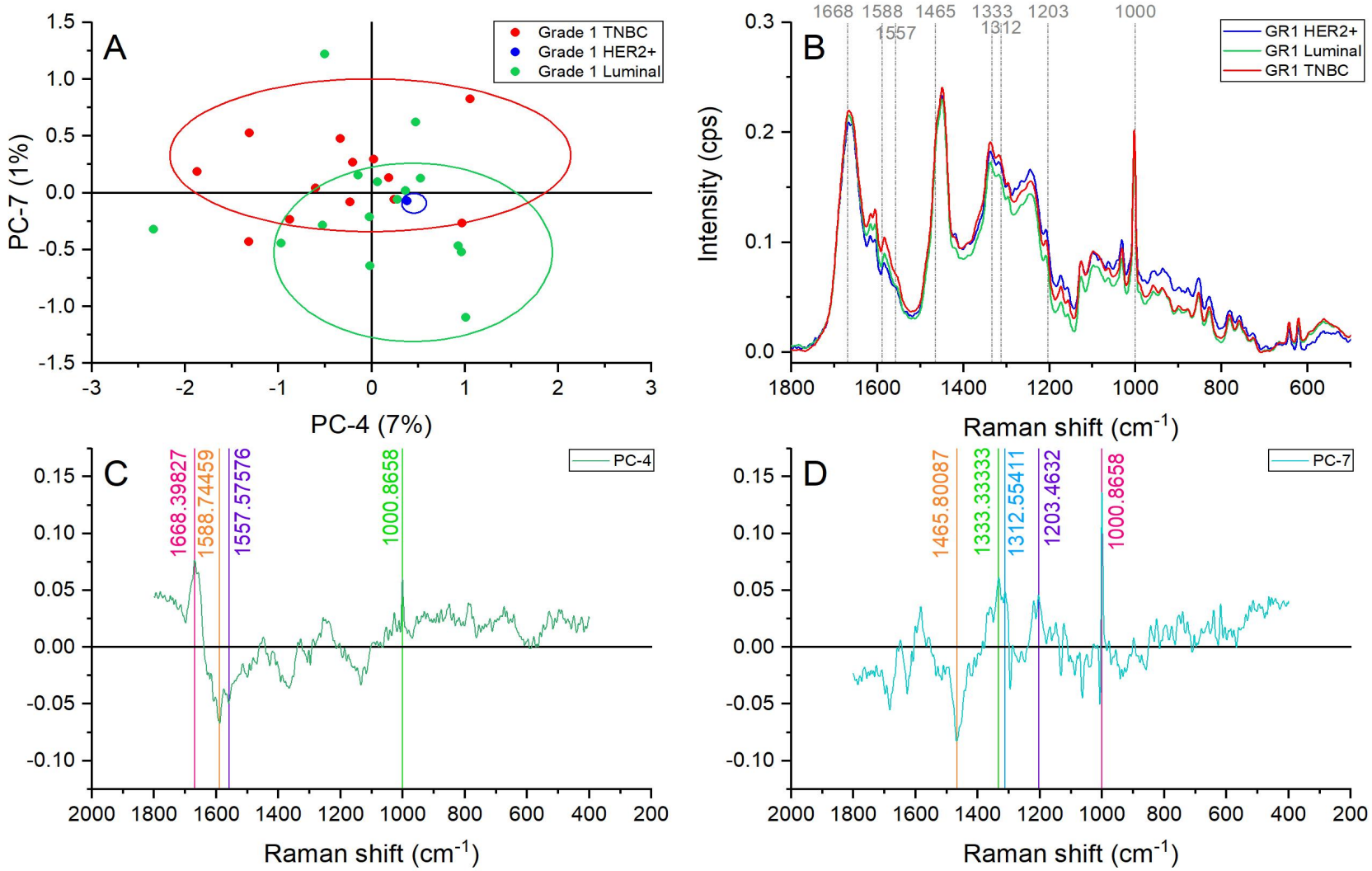

Figure 67. Principal component analysis of the fingerprint region (1,800-500 cm^{-1}). (A) Score plot using PC-4 and PC-7 accounting for 8% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the fingerprint region for grade 1 TNBC, luminal, and HER2+. 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.

The ≈1668 cm^{-1} peak is caused by the stretching of the carbonyl group (C=O) in the α-helical structure of proteins. Moreover, the next bands located at ≈1333 and 1312 cm^{-1} represent the CH_3CH_2 wagging and twisting of nucleic acids, and collagen. [180]

In addition, the band placed at ≈1203 cm^{-1}also represented a collagen and protein peak representing the amide III and CH_2 wagging vibrations from glycine backbone and proline side chains. [180] Other collagen and protein assignment located at ≈1000 cm^{-1} representing the symmetric stretching of phenylalanine presented the same trend.

The TNBC grade 1 samples presented greater contribution on the ≈1465 cm^{-1} peak assigned to lipid vibrations. Likewise, The peak placed at ≈1557 cm^{-1} representing the stretching of C-N and deformation of N-H which are representative of amide II also showed higher intensities than the luminal counterpart with the HER2+ only sample point placed closer to the average. This behaviour is consistent with the findings presented on the FTIR section.

7.1.1.3 Amides region (1,800-1,140 cm^{-1}) analysis of Grade 1 samples all subtypes

Clear but subtle differences can be appreciated in the different regions of the amide spectral range of the average spectral profile of grade 1 samples as presented in Figure 68. Variations in intensity, shape, sharpness, and ratios are reflections of changes occurring at the molecular level.

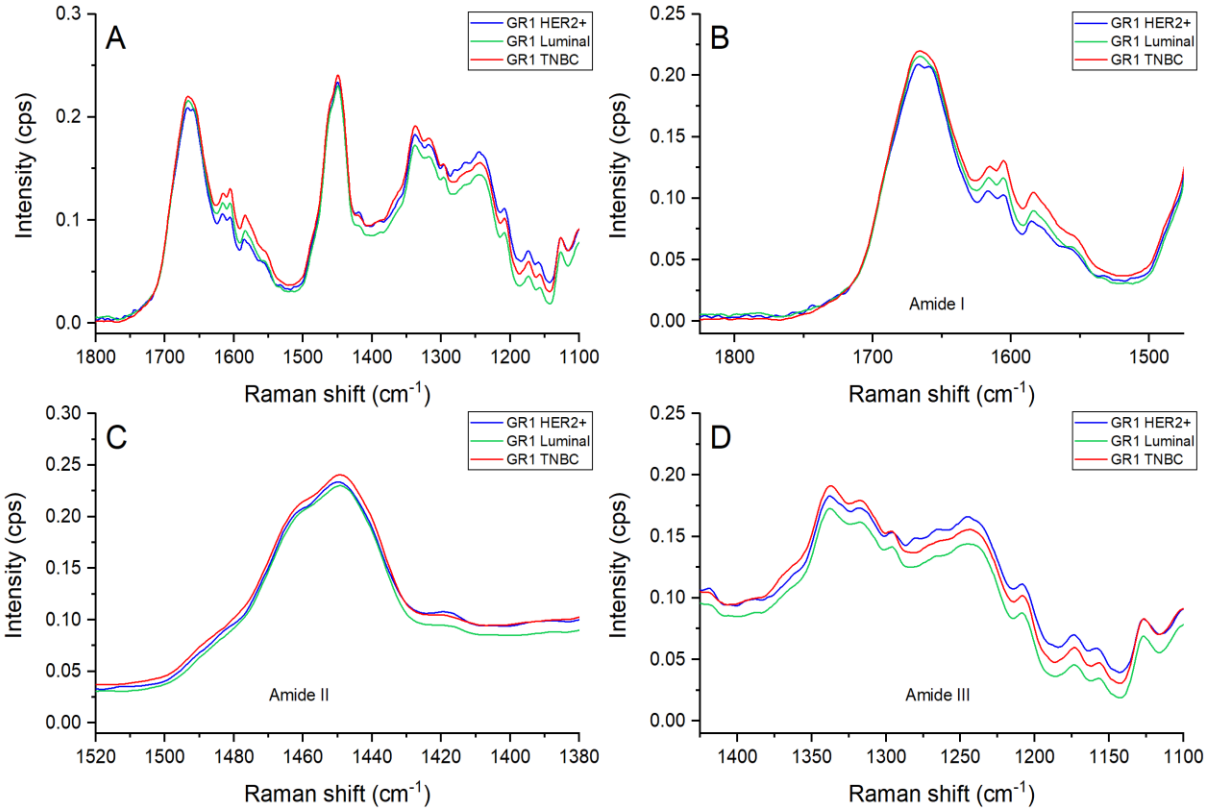

Figure 68. (A) Average spectral profile of the amide region (1,800-1,140 cm^{-1}) for grade 1 TNBC, luminal, and HER2+, (B) amide I region (1,800-1,510 cm^{-1}), (C) amide II region (1,510-1,390cm^{-1}), (D) amide III region (1,390-1,140 cm^{-1}).

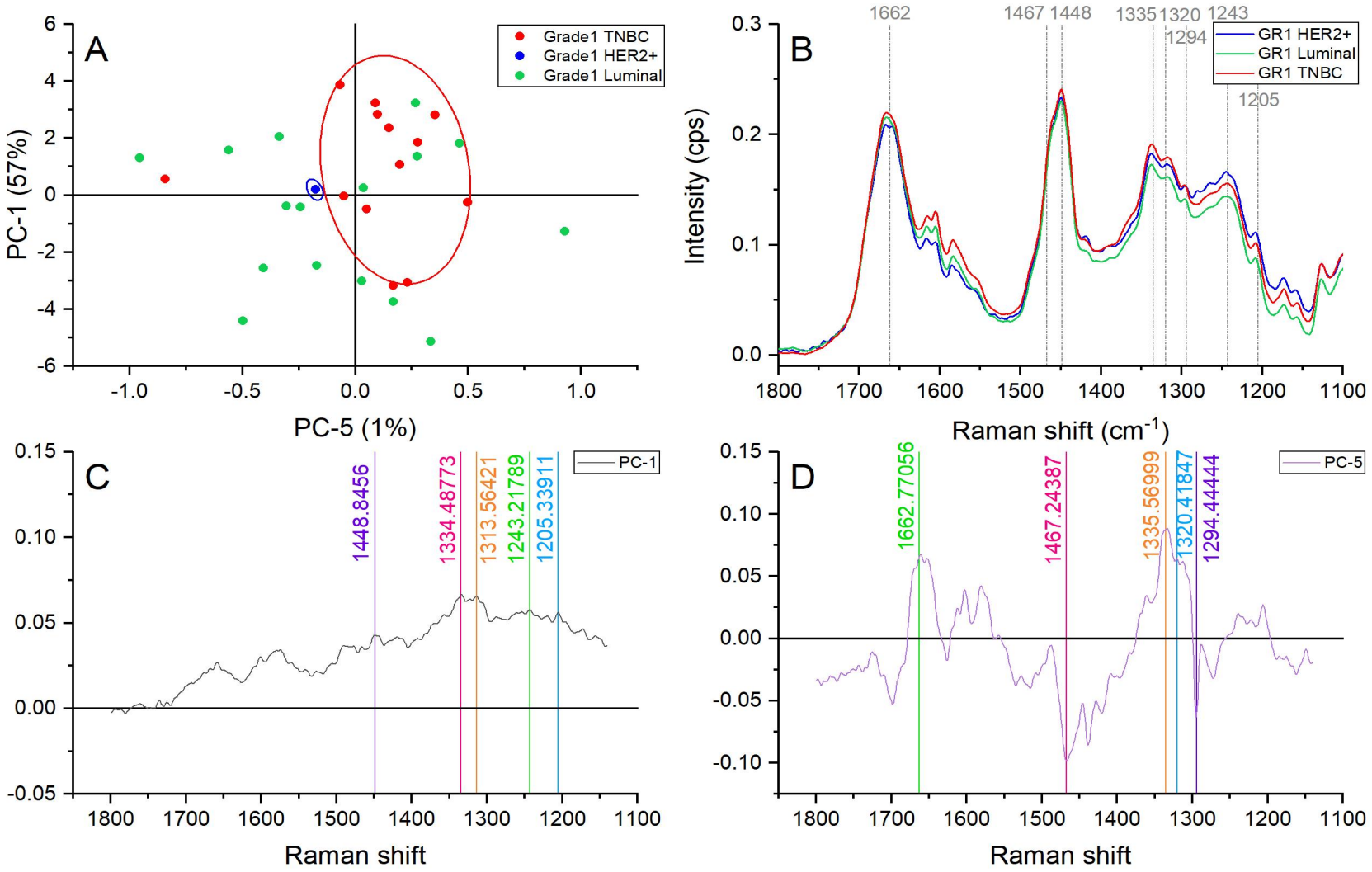

Figure 69. Principal component analysis of the amide region (1,800-1,140 cm^{-1}). (A) Score plot using PC-1 and PC-5 accounting for 58% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the amide region for grade 1 TNBC, luminal, and HER2+. Dotted 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.

The principal component analysis of the whole amide region showed a good grouping for the TNBC grade1 samples. However, the luminal subtype presented a scattered behaviour reflecting a dynamic behaviour. HER2+ only sample was found close to the average intensity of the peaks to be discussed as seen in Figure 69.

The band located at ≈1662 cm^{-1} is considered an amide I vibration [180], however, it has also been reported as ceramide backbone [207] This band presented higher intensities and therefore contributions in the TNBC, followed by the luminal subtype and HER2+ samples.

Besides, TNBC grade 1 samples also showed the highest contributions of nucleic acids and collagen bands located at ≈1448, 1334, 1320, 1243, and 1205 cm^{-1}. The ≈1448 and 1334 cm^{-1} bands correspond to the CH_3CH_2 deformation and wagging in collagen and polynucleotide chain.[180] The ≈1320 and1243 cm^{-1} peaks represent amide III vibrations. [148]The former band is caused by CH deformation in proteins and DNA and RNA, while the later bands characterise the CH_2 wagging and C-N stretching in collagen and pyrimidine bases (C-T). Lastly, the ≈1205 cm^{-1} peak representing the hydroxyproline structural component in collagen [203] confirmed the ECM adaptation in the TNBC subtype. Due to the increased intensity of these bands in the TNBC grade 1 samples it can be suggested a higher contribution and content of collagen and nucleic acids in comparison with luminal and HER2+ subtypes.

The HER2+ grade 1 sample showed greater contribution in the ≈1467 cm^{-1} band which is assigned to lipid vibrations, and the ≈1294 cm^{-1} peak that is assigned to methylene twisting. [180], [208] The greater intensities on the HER2+ spectra suggest higher content of lipids, and methylation processes. Early carcinogenesis stages are characterised by DNA methylation. The methylation has effects on the genetic activity, but also oncogenic transformations. [176] Half of the luminal samples presented similar intensities, however the other half showed higher than the average intensities suggesting a split behaviour. On the contrary, the TNBC had the lowest contributions for these peaks in comparison with the other subtypes in grade 1 samples.

7.1.1.4 Amino acid and nucleic acids regions (980-600 cm^{-1}) analysis of Grade 1 samples all subtypes

The amino acid and nucleic acid principal component analysis revealed differences among the three subtypes in the grade 1 samples. Good separation and grouping of the samples was achieved (Figure 70A) The listed peaks were identified as differentiators of all three subtypes: 963, 877, 849, 783, 717, 641, and 621 cm^{-1}.

The TNBC grade 1 samples grouped in the PC-2 positive and PC-4 negative quadrant in the score plot (Figure 70A). The location of this group indicated higher intensities on lipid vibrations represented by 963 cm^{-1}, which agrees with our FTIR results. Furthermore, important and higher contribution of protein and nucleic acids vibrations were identified for this subtype in comparison with luminal and grade 1 HER2+ samples. These contributions were represented by the peaks located at se contributions were represented by the peaks located at se contributions were represented by the peaks located at ≈849 cm^{-1} caused by

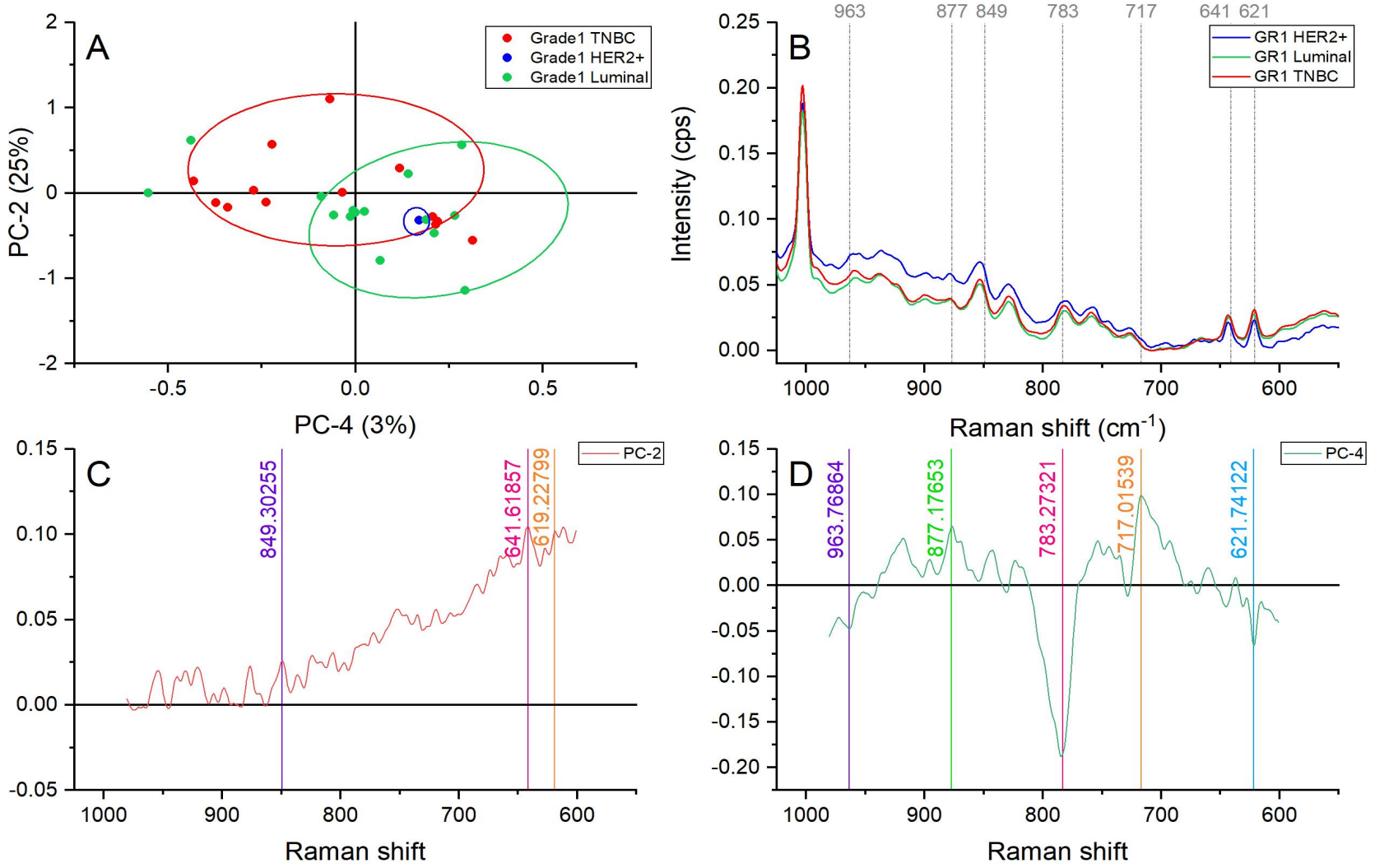

Figure 70. Principal component analysis of the amino acid and nucleic acid region (980-600 cm^{-1}). (A) Score plot using PC-2 and PC-4 accounting for 28% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the amino acid and nucleic acid region for grade 1 TNBC, luminal, and HER2+. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-2 Loading, (D) PC-4 loading.

tyrosine stretch, ≈641 cm^{-1} representing the C-S stretching and C-C twisting of proteins, and ≈621 cm^{-1} which is assigned to C-C twisting mode of phenylalanine. [180]

Luminal and HER2+ grade 1 samples showed a similar behaviour which was suggested by their same quadrant location in the score plot (Figure 70A). They both presented higher intensities in the peaks placed at ≈877 and 717 cm^{-1}. Both peaks can be assigned to phosphodiester vibrations in nucleic acids. [180]However, the 717 cm^{-1} band has been suggested to represent C-N group vibrations in membrane phospholipid head [148] which might imply cellular membrane changes and affectation is membrane fluidity.

7.1.2 Grade 2 all subtypes

When compared, the average spectral profile of grade 2 samples presenting luminal, HER2+ rich and triple negative breast cancer (TNBC) subtypes showed subtle differences presented in Figure 71 and Figure 72. Grade 2 HER2+ and TNBC samples presented similar spectral profiles.

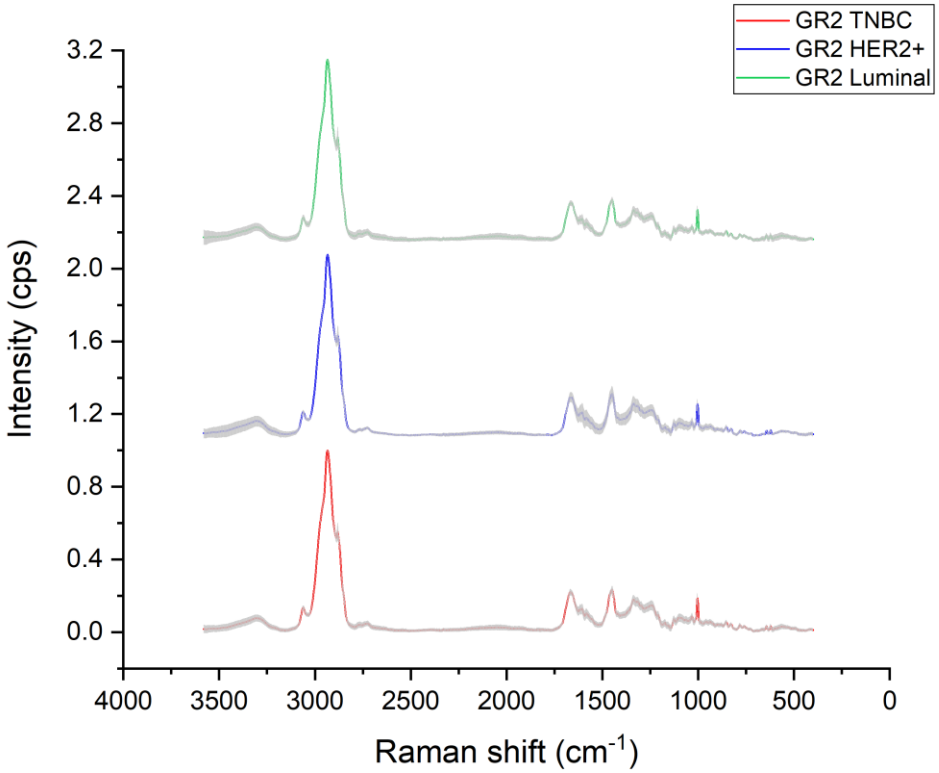

Figure 71. Raman average spectral profile of grade 2 samples presenting triple negative (TNBC), luminal and HER2+ subtypes, variance is represented with grey shadow.

7.1.2.1 Lipid region (3,100-2,680 cm^{-1}) analysis of Grade 2 samples all subtypes

Relatively good grouping was identified when the lipid region was analysed with principal component analysis as presented in Figure 73. Four bands were identified as responsible for the grouping depending on subtype of the grade 2 breast cancer samples: ≈2944, 2933, 2880, and 2846 cm^{-1}.

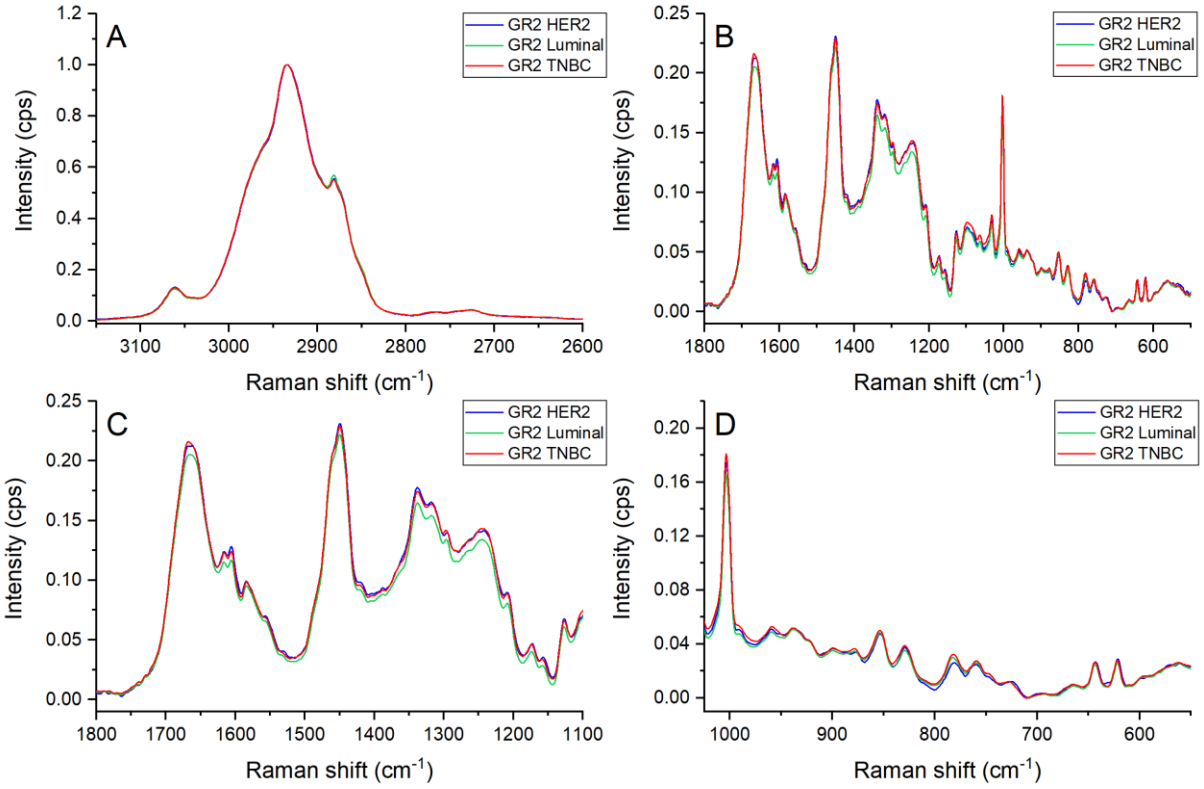

Figure 72. Raman average spectra of grade 2 breast cancer samples. (A) Lipid region (3,100-2,680 cm^{-1}), (B) Fingerprint region (1,800-500 cm^{-1}), (C) Amides region (1,800-1,140 cm^{-1}) (D) Amino acid and nucleic acids regions (980-600 cm^{-1}).

The first peak (≈2944 cm^{-1}) represents the CH_2 asymmetric stretch in lipids chains, especially fatty acids.[180] This band presented a split behaviour among the samples with half of the samples of each subtype presenting lower intensities than the average and the other half higher intensities.

Principal component 5 showed the best separation for both hormonal subtypes. The luminal grade 2 samples presented higher intensities in the ≈2933, 2880, and 2762 cm^{-1} bands. The first band is cause by the CH_2 asymmetric stretch, while the ≈2880 cm^{-1} peak is representative of CH_2 and CH vibrations in lipids and proteins. [148]The band located at ≈2762 cm^{-1} is assigned to general C-H stretches. [180] Besides, the TNBC samples showed higher intensities of the peak located at ≈2843 cm^{-1} that suggest a higher lipid methylation (CH_3). This suggestion was due to the CH_3 symmetric stretch of lipids assigned to this peak.[180] However, HER2+ grade 2 sample points stayed the closest to the average suggesting a more consistent biochemical behaviour.

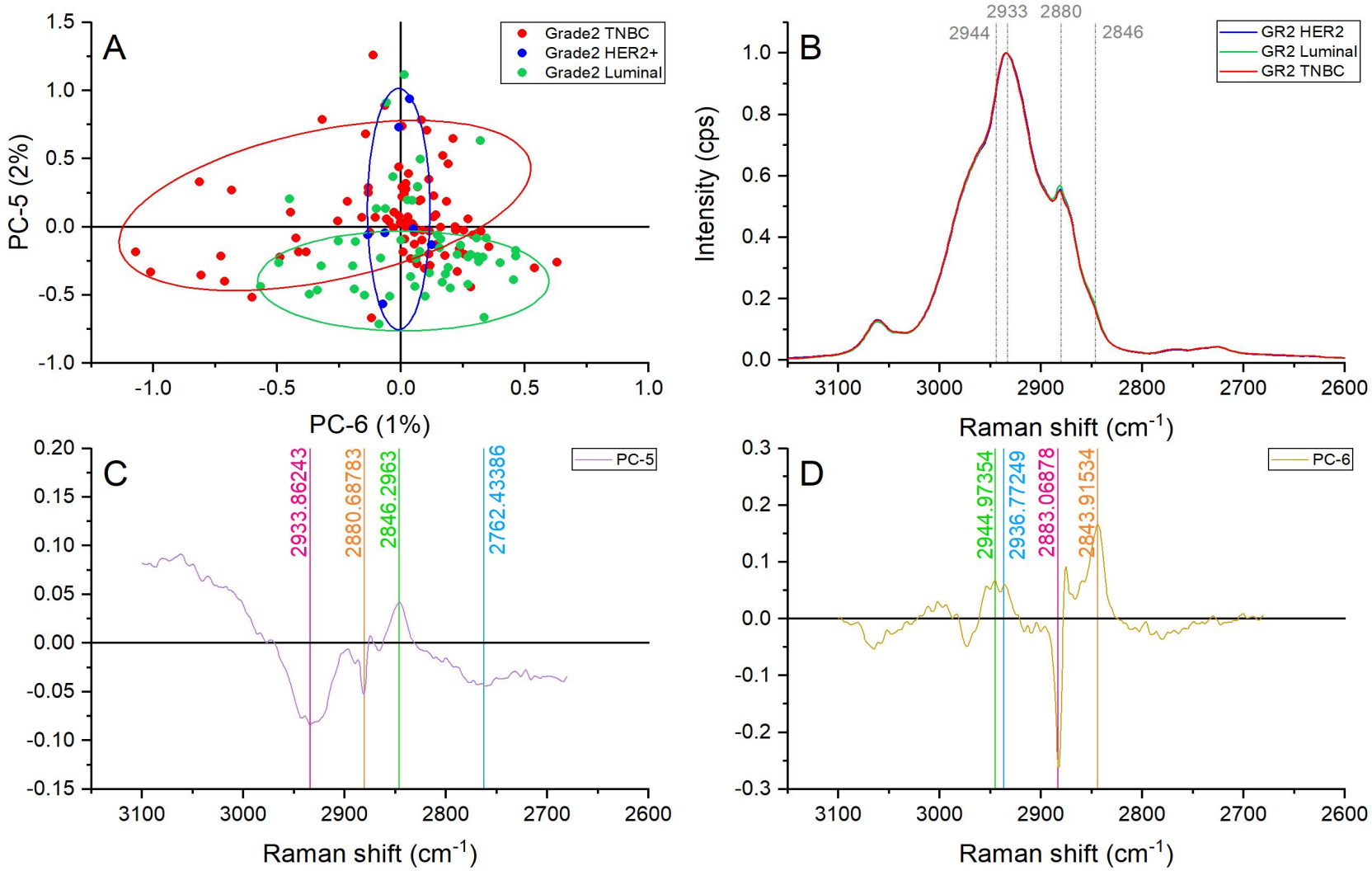

Figure 73. Principal component analysis of the lipid region (3,100-2,680 cm^{-1}). (A) Score plot using PC-5 and PC-6 accounting for 3% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the lipid region for grade 1 TNBC, luminal, and HER2+. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-5 Loading, (D) PC-6 loading.

These increased intensities in the luminal and TNBC subtypes suggest that a more active lipid metabolism take place in the hormonal subtypes in grade 2 samples in comparison with the HER2+ subtype. The lipid changes are reflected in the intensity variations, and suggest fatty acid adaptations. Fatty acids support different mechanisms that can influence on breast cancer, as they might may control cell division. This processes include: membrane fluidity, macromolecule mobility, receptor availability, and prostaglandin and cAMP/cGMP biosynthesis. Moreover, unsaturated fatty acids might transform by being metabolised into epoxides and peroxides and act as initiators for cell proliferation. [209] Therefore, lipid transformations serve as evidence of biochemical adaptations during the cancer evolution.

7.1.2.2 Fingerprint region (1,800-500 cm^{-1}) analysis of Grade 2 samples all subtypes

The principal component analysis of the fingerprint region revealed a clear grouping of luminal and TNBC grade 2 sample points following opposite trends. In addition, HER2+ grade 2 samples showed a good grouping. However, certain overlapping of groups was identified as it can be seen in Figure 74.

Grade 2 HER2+ samples classification peaks were defined by PC-2. This subtype showed higher intensities than the average on the band located at ≈1243 cm^{-1} representing a higher contribution of Amide III vibrations, caused by CH_2 wagging and C-N stretching.[180]. This particular assignment can correlate with collagen presence and with pyrimidine bases representing a nucleic acid vibration. Extensive changes and overproduction of genetic material in the HER2+ grade 2 samples can be suggested due to the increased ≈1243 cm^{-1} band contribution. Furthermore, HER2+ grade 2 samples also showed lower intensities in peaks associated with C=C in plane bending mode of phenylalanine (≈1604 cm^{-1}) and symmetric stretching of CH_3 groups in phospholipids (≈1366 cm^{-1}). The first band, 1604 cm^{-1}, confirmed that the increased intensity in ≈1243 cm^{-1} was caused by nucleic acid dominant content instead of collagen overproduction, whereas the phospholipids band (≈1366 cm^{-1}) suggest lipid peroxidation and changes in membrane fluidity.

The separation among luminal and TNBC grade 2 samples was achieved using the PC-5. TNBC grade 2 spectra showed greater intensities in the bands located at ≈1669 (amide I), 1613 (tyrosine), 1447(CH_2 bending in proteins), and 1002 cm^{-1}(phenylalanine) in comparison with luminal samples. All the listed bands represent protein related vibrations suggesting that TNBC grade 2 cases focus more on the overproduction of structural material to promote proliferation in comparison with the other subtypes.

The analysed luminal grade 2 samples showed higher contribution of C=C olefinic stretch represented by the ≈1585 cm^{-1} peak. [196]This change translates to rearrangement of hydrogen bonds within proteins and collagen resulting in structural changes. [203]

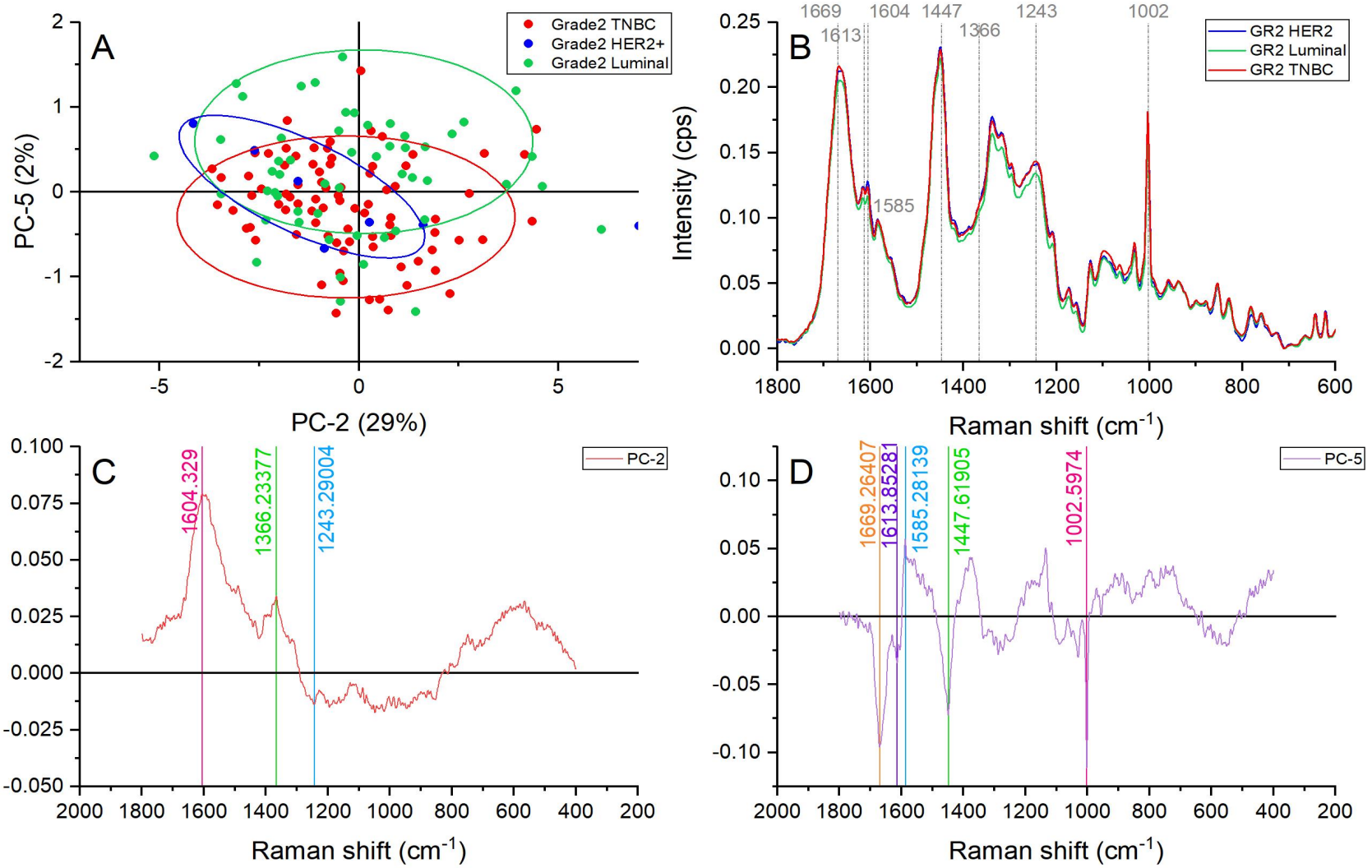

Figure 74. Principal component analysis of the fingerprint region (1,800-500 cm^{-1}). (A) Score plot using PC-2 and PC-5 accounting for 31% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the fingerprint region for grade 2 TNBC, luminal, and HER2+. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-2 Loading, (D) PC-5 loading.

7.1.2.3 Amides region (1,800-1,140 cm^{-1}) analysis of Grade 2 samples all subtypes

Grade 2 breast cancer samples showed slight subtle differences in the different regions of the amide spectral range of the average spectral profile as presented in Figure 75. HER2+ samples seem to have a similar spectral profile to TNBC. In comparison with grade 1 samples the amide region of grade 2 cancerous samples had less evident differences.

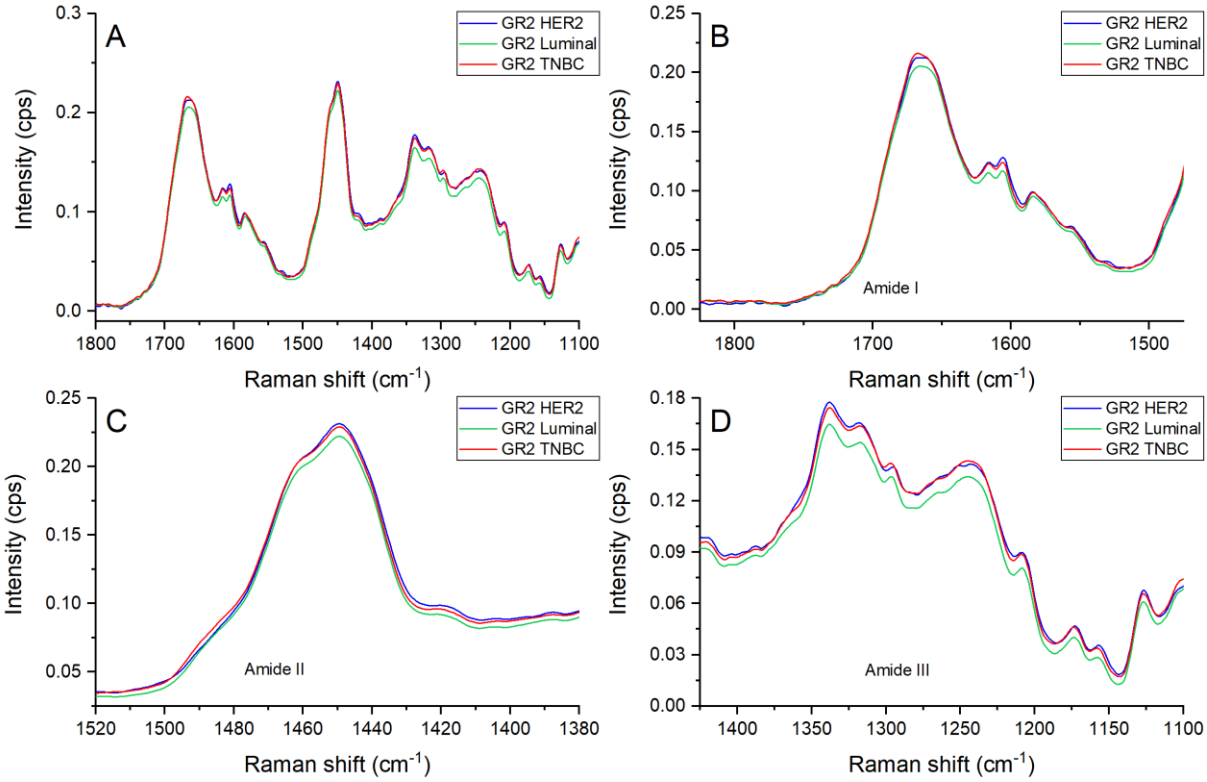

Figure 75. (A) Average spectral profile of the amide region (1,800-1,140 cm^{-1}) for grade 2 TNBC, luminal, and HER2+, (B) amide I region (1,800-1,510 cm^{-1}), (C) amide II region (1,510-1,390cm^{-1}), (D) amide III region (1,390-1,140 cm^{-1}).

Luminal grade 2 breast cancer samples were located in the positive PC-4 axis with the TNBC samples with the same grade following the opposite trend as shown in Figure 76. Three peaks associated with nucleic acids vibrations represented by the bands located at ≈1629, 1406, and 1177 cm^{-1} showed greater contributions in the luminal samples in comparison with TNBC. Amide vibrations caused by C=O stretching for β-form of polypeptides (≈1629 cm^{-1})[180], histidine vibration (≈1406 cm^{-1})[210], and cytosine and guanine stretching (≈1177 cm^{-1})[148] were identified as responsible for these contributions. In addition, the majority of HER2+ grade 2 samples also presented greater contributions on the ≈1629 cm^{-1} band suggesting dominant β-form structural conformations.

TNBC grade 2 samples presented more intense peaks at ≈1333, and 1313 cm^{-1}.[180] The former band is caused by nucleic acids and phosphates vibrations which related to phospholipids, while the later peak is assigned to CH_3CH_2 twisting mode of collagen and lipids.

This finding suggests that a more active lipid metabolism take place in the negative hormonal subtypes in grade 2 samples in comparison with the HER2+ and luminal subtype.

PC-6 helped to identify the peaks responsible for the HER2+ grouping. For example, lower intensities of ≈1467 and 1296 cm^{-1} were identified for this subtype representing lower contributions of lipids, and acyl chains in fatty acids respectively. On the contrary, HER2+ grade sample presented higher intensities in amide I α-helix conformations (≈1656 cm^{-1}), and deformation of C=C in phenylalanine (≈1601 cm^{-1}).[180] The clustering of HER2+ subtypes based on structural changes suggest structure alterations most possible due to the replacement of β-sheets with α- and random coils. [106]

7.1.2.4 Amino acid and nucleic acids regions (980-600 cm^{-1}) analysis of Grade 2 samples all subtypes

The principal component analysis of the amino acid and nucleic acid region revealed groups but not a completely visible separation as can be seen in Figure 77. When the spectral profile of each subtype for grade 2 samples were compared similar intensities and peak shape were visible. Therefore, the PCA offered a better insight of the biochemical changes occurring on each subtype for the same breast cancer grade.

Five peaks were detected as differentiator of HER2+ grade 2 samples. The peaks were located at ≈920, 858, 829, 786, and 756 cm^{-1}. The band located at ≈920 and 858 cm^{-1} are associated with collagen assignments (ring and amino acid side chain vibrations of collagen backbone, proline and hydroxyproline)[180] and presented lower intensities than the average in the HER2+ spectra. Greater intensities were detected in nucleic acids related peaks located at ≈829 and 756 cm^{-1} assigned to phosphodiester stretching in nucleic acids and symmetric breathing of tryptophan respectively. This dominant peaks suggest an active accumulation of genetic material leading towards cell proliferation and invasive processes.

Luminal and TNBC grade 2 samples were group following opposite trends in PC-6. Luminal spectra showed more intense contributions of ring breathing mode of DNA/RNA bases (≈748 cm^{-1}), and C-N vibrations in membrane phospholipid heads (≈712 cm^{-1}).[148] These contributions with increased intensities suggest changes in membrane fluidity and overproduction of cellular material. [150] Besides, TNBC grade 2 samples presented increased contributions of ≈837 cm^{-1} representing deformative vibrations of amide groups, ≈826 cm^{-1} caused by phosphodiester stretch of nucleic acids, and ≈642 cm^{-1} assigned to C-S stretching and C-C twisting of proteins.[180] All these contributions suggests structural adaptations in proteins, and accumulation of cellular material, structural and genetic material, indicating higher metabolism and cell density. [74]

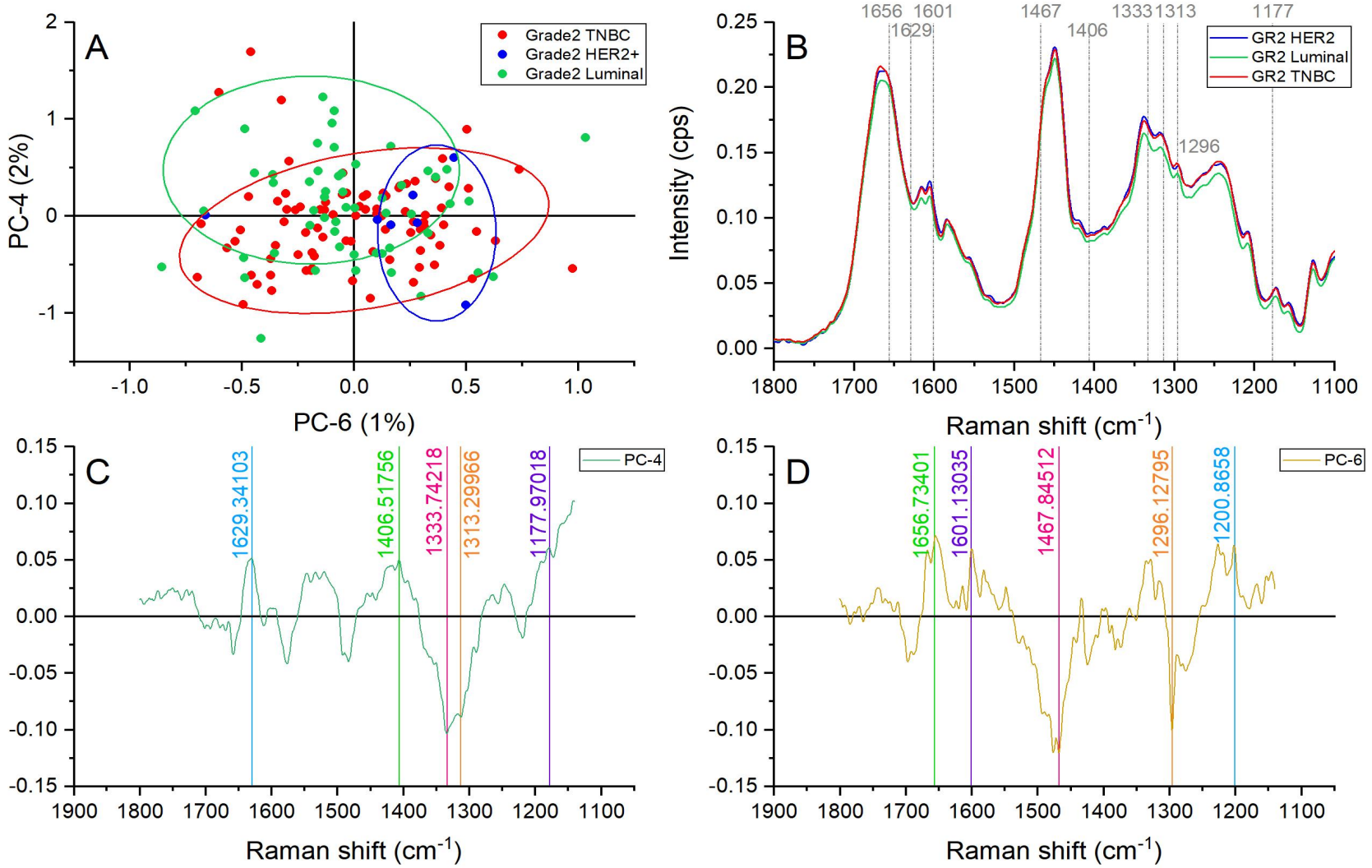

Figure 76. Principal component analysis of the amide region (1,800-1,140 cm^{-1}). (A) Score plot using PC-4 and PC-6 accounting for 3% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the amide region for grade 2 TNBC, luminal, and HER2+. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-4 Loading, (D) PC-6 loading.

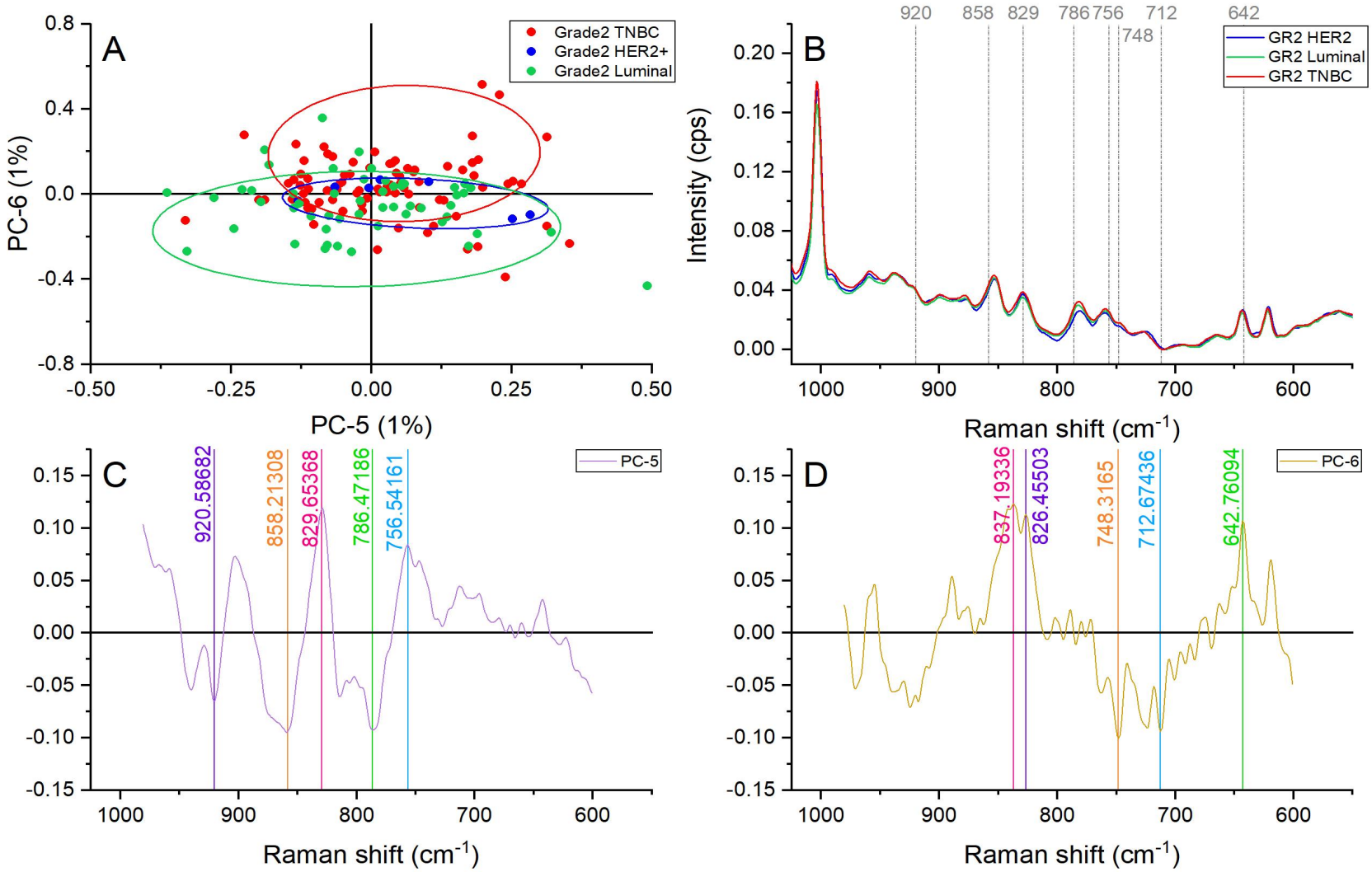

Figure 77. Principal component analysis of the amino acid and nucleic acid region (980-600 cm^{-1}). (A) Score plot using PC-5 and PC-6 accounting for 2% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the amino acid and nucleic acid region for grade 2 TNBC, luminal, and HER2+. 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-6 loading.

7.1.3 GRADE 3 ALL SUBTYPES

When compared, the average spectral profile of grade 3 samples presenting luminal, HER2+ rich and TNBC subtypes showed differences in intensity able to be seen with the naked eye as presented in Figure 78 and Figure 79. Grade 3 HER2+ rich subtypes seemed to have the most different spectral profile in comparison with the hormonal positive and negative grade 3 spectral profiles. The differences among the three subtypes will be discussed in the following sections.

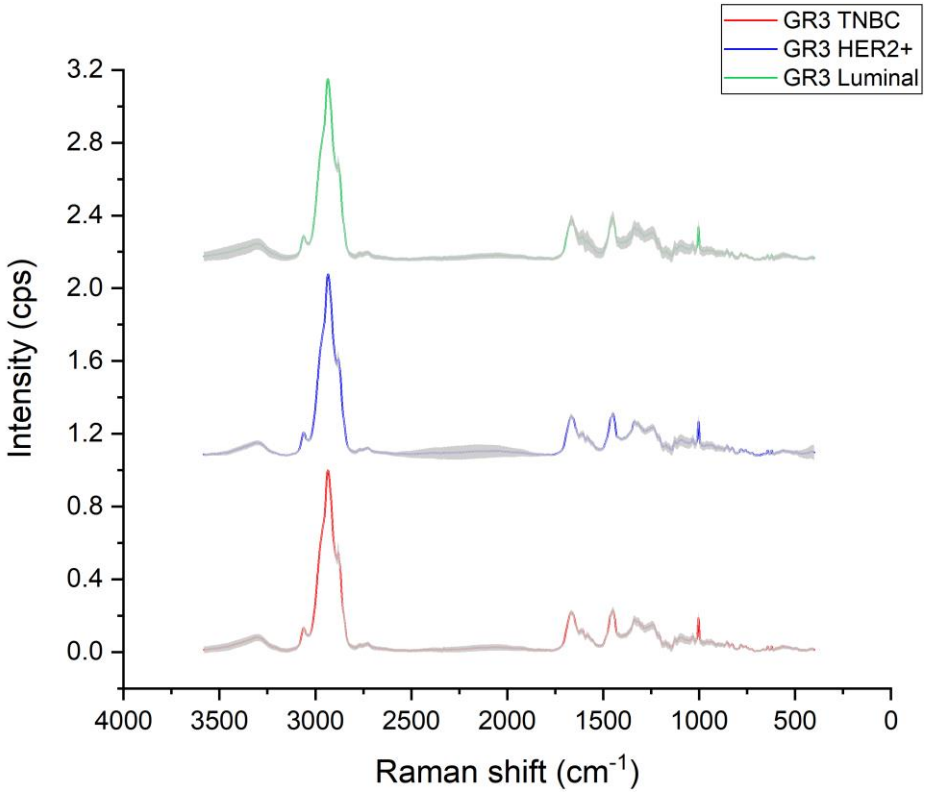

Figure 78. Raman average spectral profile of grade 3 samples presenting triple negative (TNBC), luminal and HER2+ subtypes, variance is represented with grey shadow.

7.1.3.1 LIPID REGION (3,100-2,680 CM^{-1}) ANALYSIS OF GRADE 3 SAMPLES ALL SUBTYPES

Clear grouping and mildly separation was achieved among the grade 3 samples, when the higher wavenumber region, representing lipids, was analysed with principal component analysis as presented in Figure 80A. PC-5 seems to differentiate luminal from TNBC samples by discriminating them on the positive axis. HER2+ grade 3 spectra points were located the closest to the average suggesting average concentrations of lipid component in comparison with the hormonal subtypes.

The loadings associated with the score presented (Figure 80C&D) identified the following peaks as responsible for the separation of the hormonal subtypes: 2963, 2841, 2881, and 2848 cm^{-1}. For all these peaks, a higher contribution represented by higher intensities were identified in the TNBC placing the luminal in the opposite trend with the HER2+ samples lying in between.

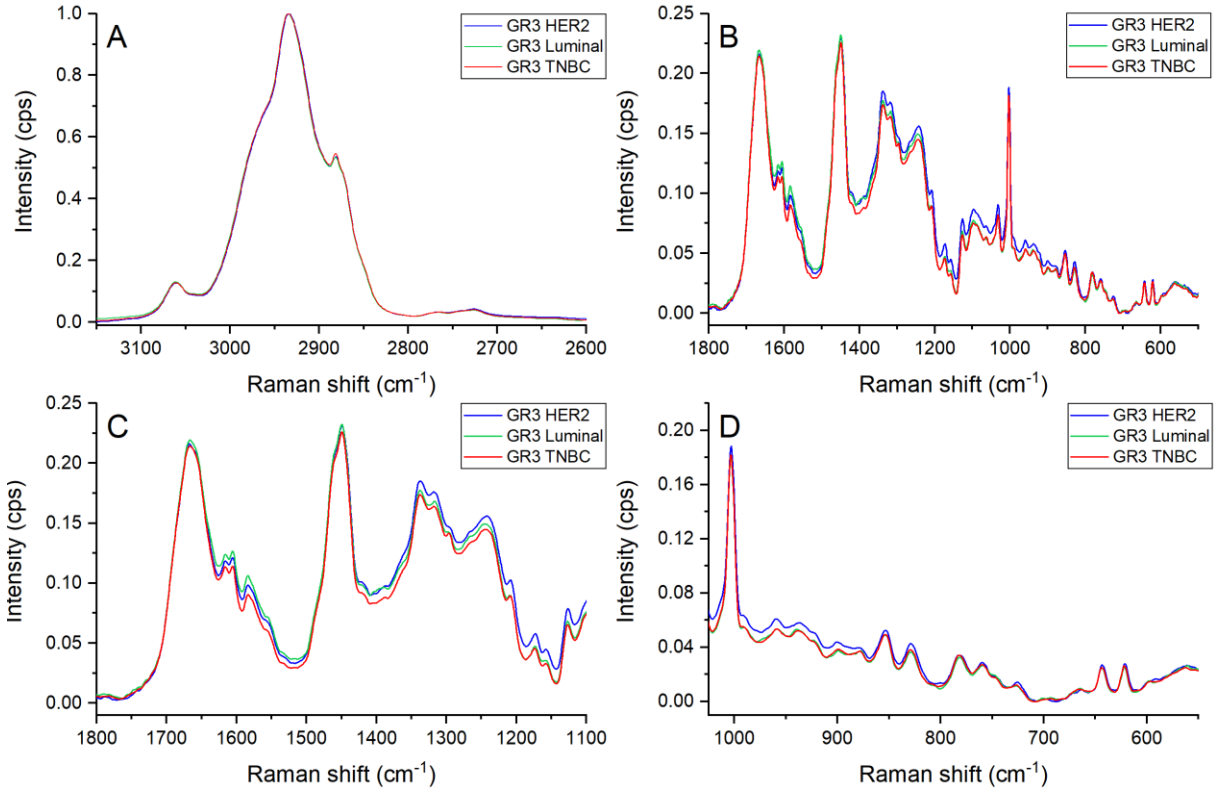

Figure 79. Raman average spectra of grade 3 breast cancer samples. (A) Lipid region (3,100-2,680 cm^{-1}), (B) Fingerprint region (1,800-500 cm^{-1}), (C) Amides region (1,800-1,140 cm^{-1}) (D) Amino acid and nucleic acids regions (980-600 cm^{-1}).

The listed peaks represent out-of-plane end antisymmetric CH_3 stretch (≈2963 cm^{-1}), CH vibrations in lipids and proteins (≈2941 cm^{-1}), CH_3 vibrations in fatty acids (≈2881 cm^{-1}), and symmetric stretching of lipids (≈2848 cm^{-1}).[211] Differences in intensity and therefore content suggest the effect of lipid metabolism in cellular signalling, storage of energy, and building and maintenance of cellular membranes.

7.1.3.2 Fingerprint region (1,800-500 cm^{-1}) analysis of Grade 3 samples all subtypes

The principal component analysis of the fingerprint region revealed a clear grouping of sample points in all three subtypes. In addition, overlapping of luminal, HER2+, and TNBC grade 3 samples can be seen in Figure 81A.

Once again, the luminal and TNBC spectra points are spread on opposite axis following opposite trends with the HER2+ samples lying in between and closer to the average in the PC-2 axis. Luminal grade 3 samples showed greater contribution of C=C bending in phenylalanine (≈1582 cm^{-1}) and phosphate stretching in guanine (≈1363 cm^{-1}).[148]

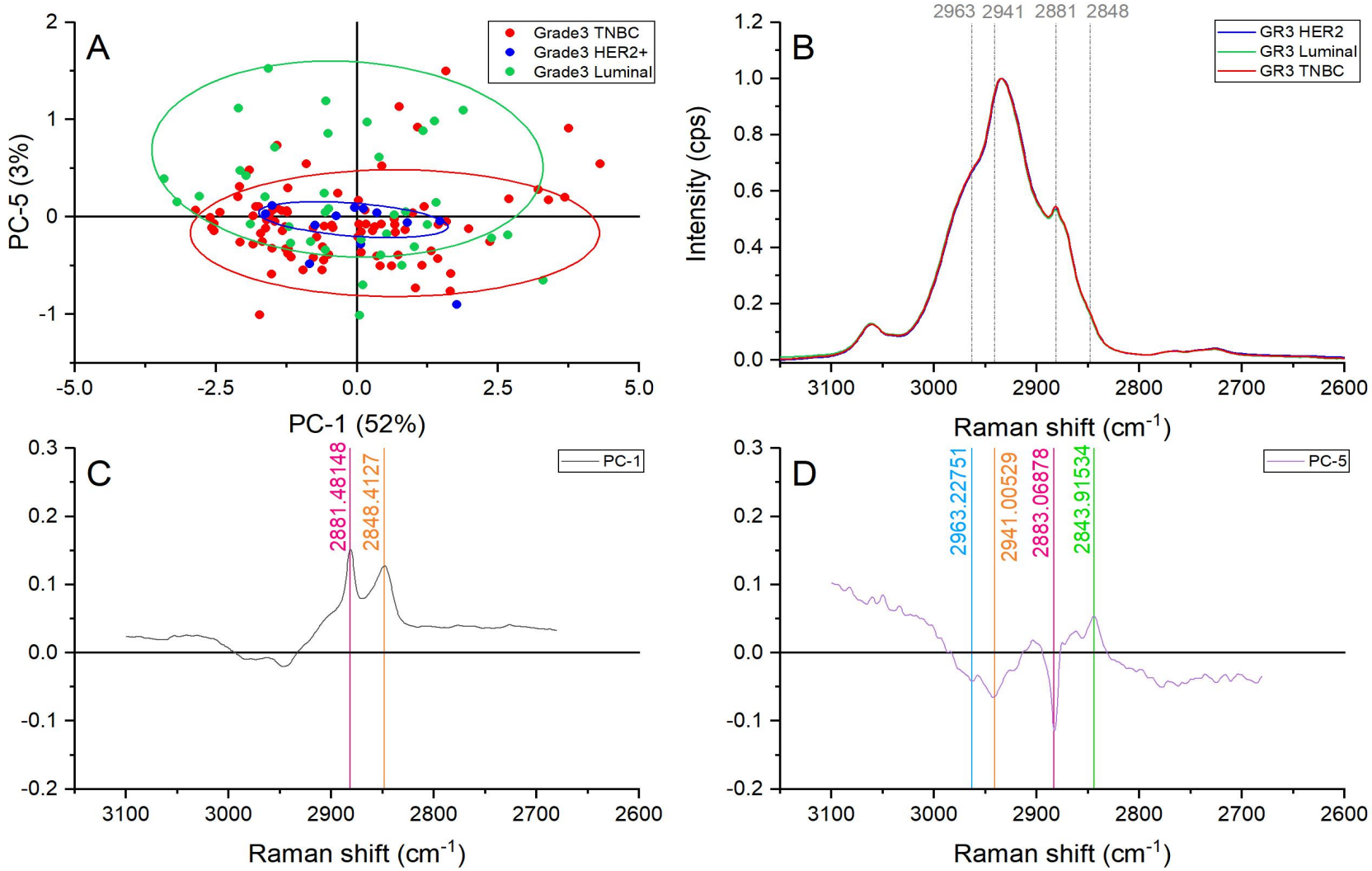

Figure 80. Principal component analysis of the lipid region (3,100-2,680 cm^{-1}). (A) Score plot using PC-1 and PC-5 accounting for 55% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the lipid region for grade 3 TNBC, luminal, and HER2+. Dotted 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.

TNBC samples presented lower intensities than the average on both mentioned peaks. However, higher intensities of collagen-related peaks were present in the TNBC grade 3 samples in comparison with the luminal subtype. These peaks represented CH_2, CH_3 bending modes of collagen and phospholipids (≈1032 cm^{-1})[180], rocking motions of methyl groups in carotenoids and ring breathing in phenylalanine (≈1005 cm^{-1})[212], and C-C stretch backbone mainly in proline and hydroxyproline [203]. These differences hinted to specific behaviours in the hormonal subtypes, with an accumulation of genetic and structural (proteins) material in the luminal subtype, while the TNBC subtype focuses not only on this, but also in lipid transformations and collagen overproduction. The suggested lipid alterations include variations in phospholipids and carotenoids. The first one affects the membrane fluidity and therefore cell signalling,[158] while the carotenoid vibrations in neoplastic tissue correlates with carcinogenic mechanisms. Hormone activity and cell processes have a close relation with breast cancer and tumour growth. Hormone secretion and cell differentiation is regulated by the gap junction communication (GJC), which is upregulated by carotenoids.[187]

Principal component 4, revealed peaks assigned to amide I (≈1669 cm^{-1}), nucleic acids and phosphate vibrations (phospholipids)(≈1331 cm^{-1}), amide III β-sheet and random coils (≈1242 cm^{-1}) , and uracil-based ring breathing mode (≈778 cm^{-1}) [180] with variable intensities among the positive and negative hormonal subtypes. Furthermore, PC-4 offered the ≈1086 cm^{-1} band which presented higher intensities in the HER2+ samples. This assignment represents the C-C skeletal stretch of acyl backbone in lipid gauche conformation.[213] A weak band was observed in the 1080-1088 cm^{-1} region and was assigned to a random liquid-like conformation, which suggest lipid conformational adaptations. [214] Akhtar et al described from a model phospholipid systems reflecting intramolecular trans/gauche conformational changes within the hydrophobic acyl chain matrix of the bilayers of the aforementioned region.

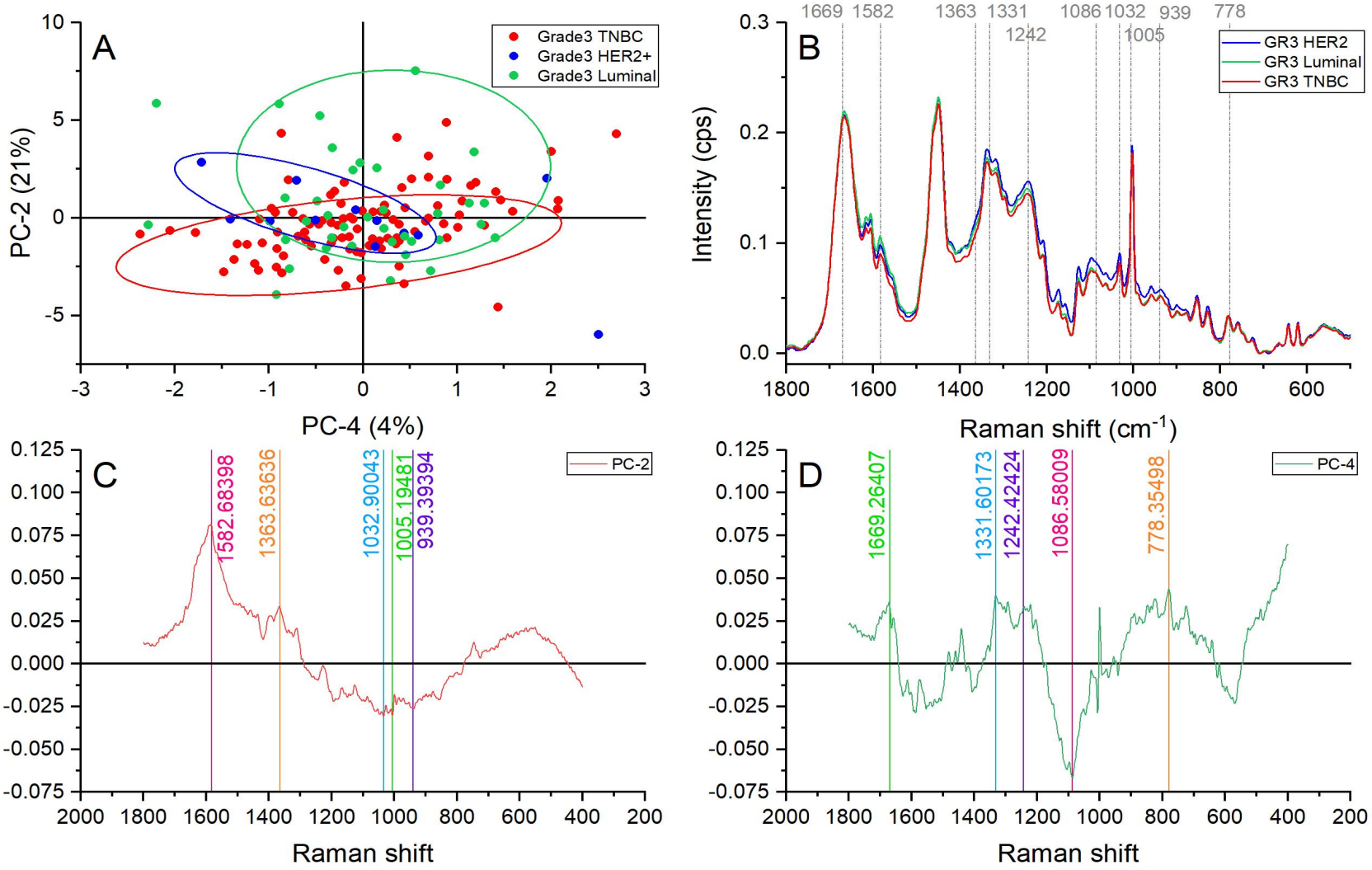

Figure 81. Principal component analysis of the fingerprint region (1,800-500 cm^{-1}). (A) Score plot using PC-2 and PC-4 accounting for 25% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the fingerprint region for grade 3 TNBC, luminal, and HER2+. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-2 Loading, (D) PC-4 loading.

7.1.3.3 Amides region (1,800-1,140 cm^{-1}) analysis of Grade 3 samples all subtypes

Slight, but noticeable differences were identified in the three different regions of the amide spectral range as presented in Figure 82. The amide I spectral region showed greater contrast between TNBC and luminal spectra. Amide II average spectral profile suggest similarities between HER2+and luminal grade 3 samples, while amide III seems to differentiate the best the HER2+ spectral profile.

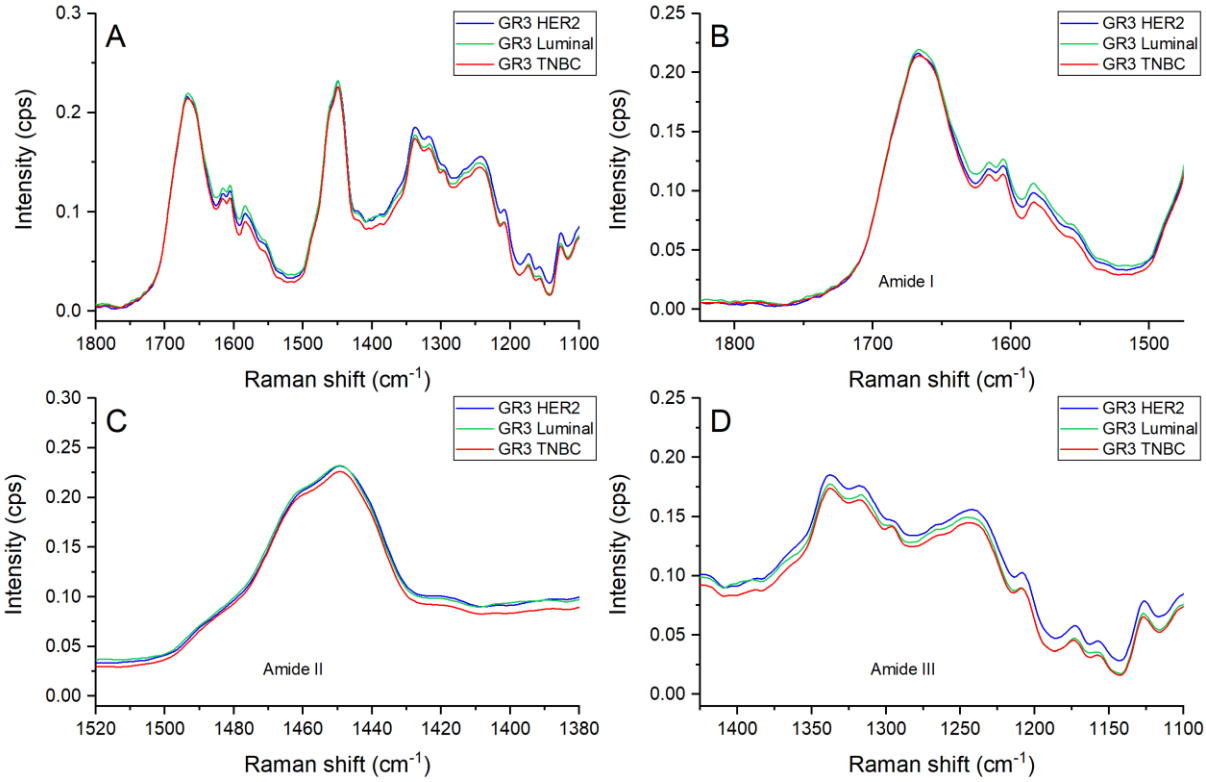

Figure 82. (A) Average spectral profile of the amide region (1,800-1,140 cm^{-1}) for grade 3 TNBC, luminal, and HER2+, (B) amide I region (1,800-1,510 cm^{-1}), (C) amide II region (1,510-1,390cm^{-1}), (D) amide III region (1,390-1,140 cm^{-1}).

Principal component analysis of the amide regions showed good grouping for all three subtypes of grade 3 samples as presented in Figure 83A where PC-2 and PC-7 were used. PC-2 offered the identification of peaks responsible for the differentiation of luminal and TNBC samples. Luminal grade 3 spectra showed greater contributions of C=C bending mode of phenylalanine represented by the peak located at ≈1583 cm^{-1}. [180] Phenylalanine vibrations indicate protein overproduction leading to proliferation of cancerous cells or collagen overproduction. Either way, both processes have important roles leading to more aggressive behaviour and invasion to surrounding tissues. TNBC spectra presented greater contributions of amide III vibrations according with PC-2 loading (Figure 83C). The first peak representing CH_2 wagging vibrations from glycine backbone and proline side chains (≈1204 cm^{-1}), while the second band represented CH_2 wagging and C-N stretching on collagen and pyrimidine bases (C and T).[180] The identified peaks suggest accumulation of genetic material and ECM production confirming a higher metabolism and targeting higher cell densities through

proliferation.[74] The same principal component revealed intensities closer to the average for the aforementioned bands in the HER2+ grade 3 samples.

Besides, PC-7 identified a variety of bands mainly associated with proteins and nucleic acids overproduction suggesting proliferation, but also fatty acids content variations. The band located at ≈1667 cm^{-1} caused by amide I α-helical structure seemed to have more contribution in the majority of HER2+ and luminal samples. On the other hand, the peak located at ≈1609 cm^{-1} assigned to cytosine (NH_2),and ≈1470 cm^{-1} representing C=N stretching presented greater contributions in TNBC confirming the focus of this subtype in overproducing genetic content for replication and accelerated proliferation. The band located at ≈1297 cm^{-1} caused by acyl chain and fatty acids [180], [211] also presented increased intensities in the TNBC grade 3 spectra. Good correlation with the FTIR results previously reported for grade 3 reaffirm the suggestion of accumulation of fatty acids in TNBC to work as fuel for malignant activities.

7.1.3.4 Amino acid and nucleic acids regions (980-600 cm^{-1}) analysis of Grade 3 samples all subtypes

The analysis of the amino acid and nucleic acid region revealed a decent grouping and unclear separation of different subtypes of grade 3 samples. The separation was avoided by overlapping of all three subtypes in the analysed spectral region as presented in Figure 84A. Principal Component 5 and 6 showed the best results accounting for 2% of the variance. Protein vibrations were responsible for most of the peaks identified with these two PCs.

Luminal grade 3 samples were mostly distributed among the negative PC-6 axis which suggested a higher contribution of C-C skeletal stretch of collagen backbone, proline and hydroxyproline [203].

HER2+ grade 3 spectral points location suggested higher intensities than the average in several peaks including; aminoacid side chain vibrations in collagen (≈857 cm^{-1})[203], out-of-plane ring breathing in proteins (≈823 cm^{-1})[148], phosphodiester in nucleic acids (≈784 cm^{-1})[148], breathing of tryptophan and ring mode in DNA and RNA bases (≈754 and 745 cm^{-1})[180], and C-S stretching and C-C twisting in proteins (≈ 640 and 618 cm^{-1}).[180]. An active proliferation triggered by protein and nucleic acid content and ECM preparation for further invasive processes can be proposed as part of the HER2+ cancer progression. Furthermore, TNBC spectral points showed a good clustering closer to the average of both axis suggesting a dynamic behaviour on the protein and collagen overproduction represented by the aforementioned bands in this section.

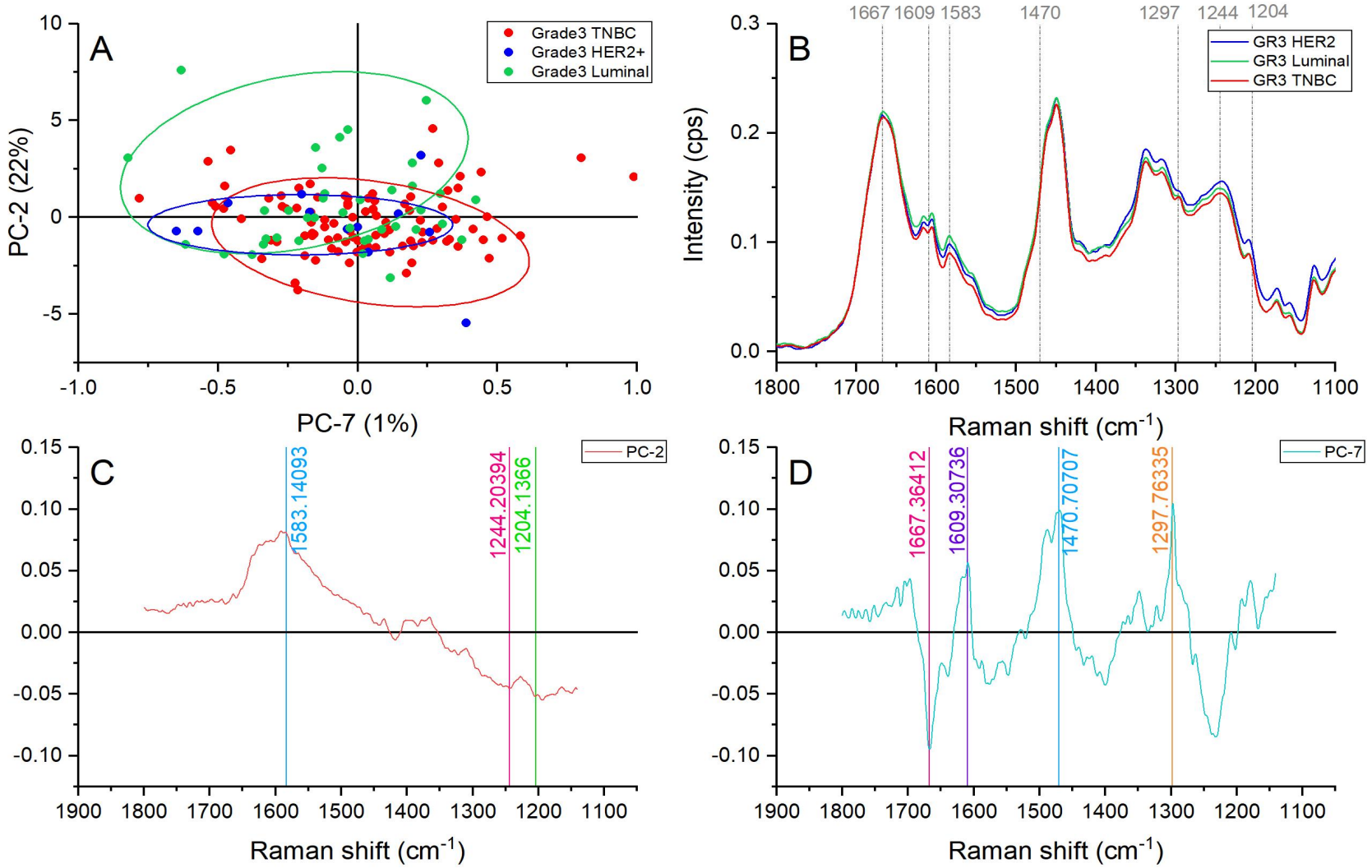

Figure 83. Principal component analysis of the amide region (1,800-1,140 cm^{-1}). (A) Score plot using PC-2 and PC-7 accounting for 23% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the amide region for grade 3 TNBC, luminal, and HER2+. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-2 Loading, (D) PC-7 loading.

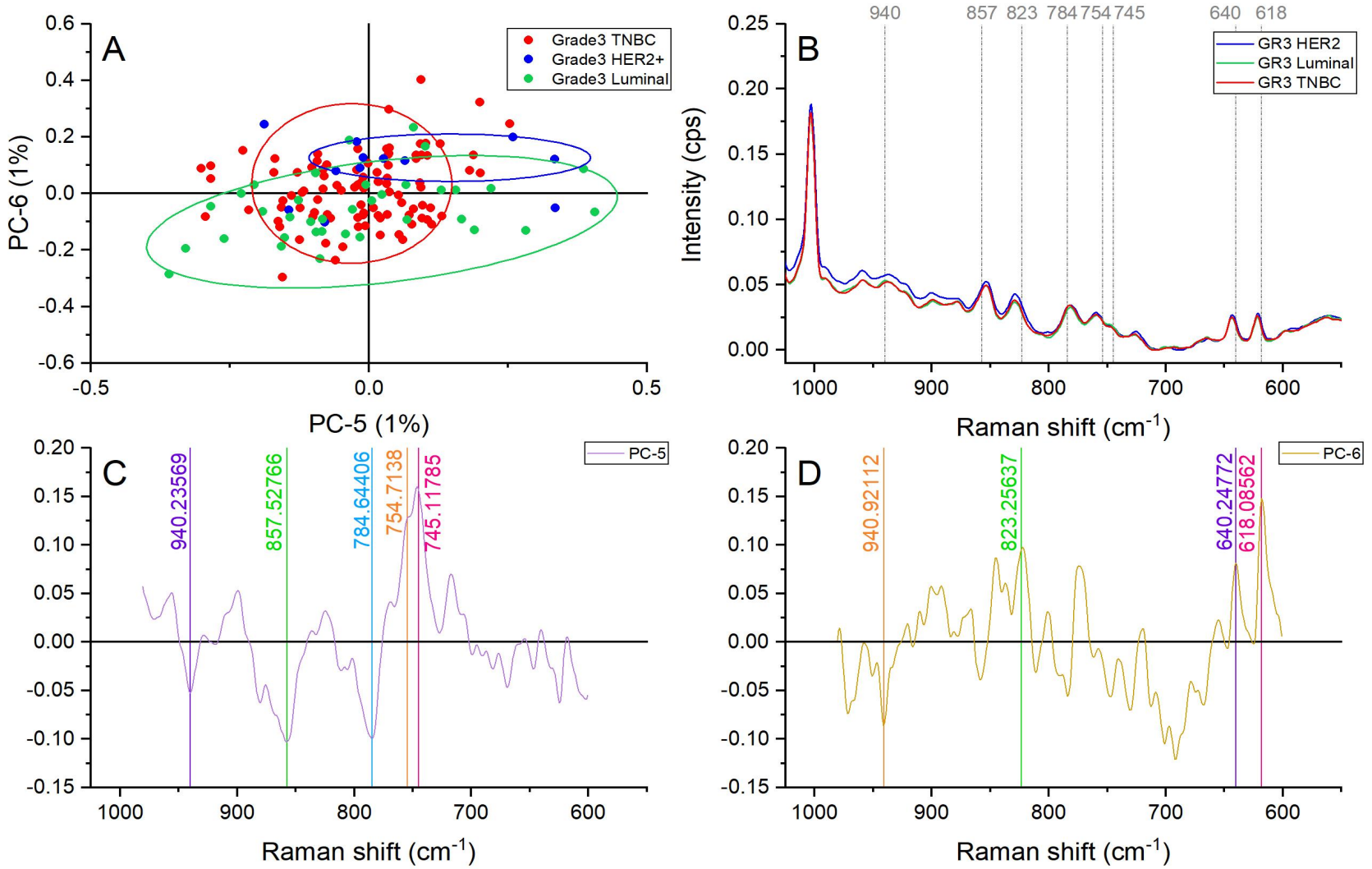

Figure 84. Principal component analysis of the amino acid and nucleic acid region (980-600 cm^{-1}). (A) Score plot using PC-5 and PC-6 accounting for 2% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the amino acid and nucleic acid region for grade 3 TNBC, luminal and HER2+. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-5 Loading, (D) PC6 loading.

7.2 Comparison of the spectral data of different grades within a specific subtype.

7.2.1 Luminal samples all grades

When compared, the average spectral profile of luminal samples representing grade 1, 2, and 3 showed subtle differences in intensity as it can be appreciated in Figure 85 and Figure 86. In all regions, spectral profiles showed similitudes between grade 1 and 3. The discussion of these specific areas is presented in the following sections.

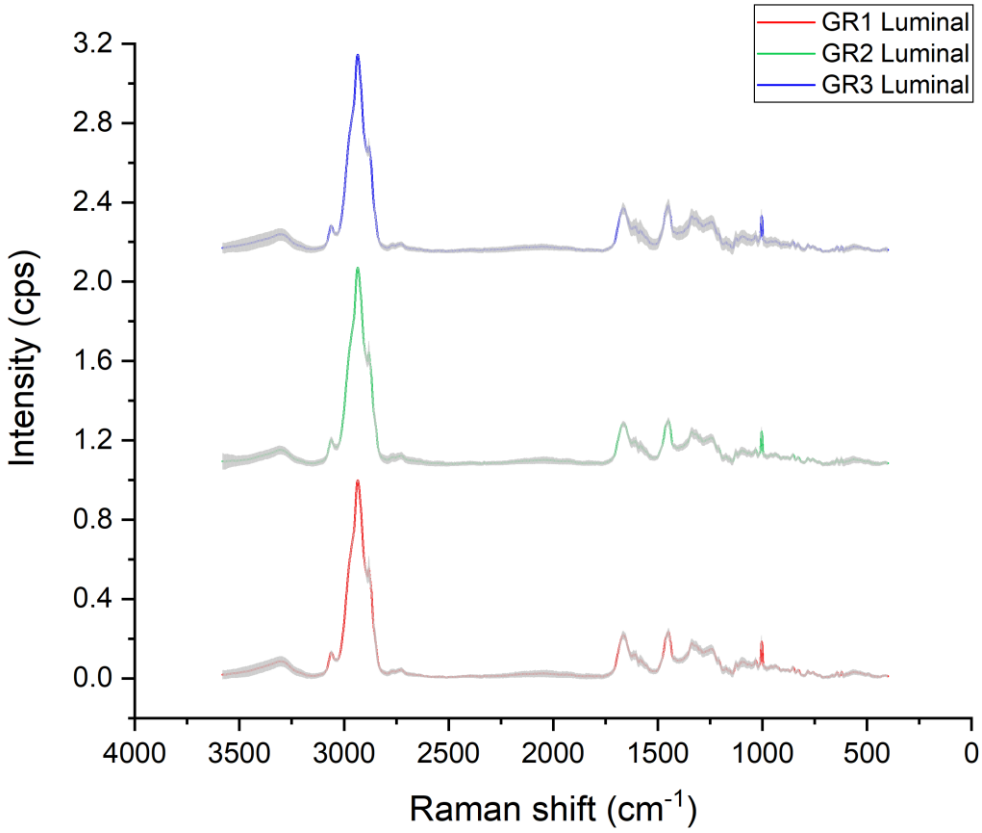

Figure 85. Raman average spectral profile of luminal samples representing grade 1, grade 2, and grade 3 cases, variance is represented with grey shadow.

7.2.1.1 Lipid region (3,100-2,680 cm^{-1}) analysis of luminal samples all grades

Principal component analysis in the lipid region of luminal samples showed good grouping and good separation of grade 2 and 3 samples as it can be seen in Figure 87. The group of peaks located at ≈2988, 2946, 2913, 2881, and 2844cm^{-1} were identified with PC-2 and PC-6. All these peaks are associated with lipid and protein vibrations caused by CH vibrations.

Grade 2 and 1 luminal samples presented an increased intensities in the band located at ≈2881 cm^{-1} and ≈2844 cm^{-1} which represent CH_2 asymmetric stretch of lipids and fatty acids[180] in comparison with grade 3 samples. Moreover, the band placed at ≈2913 cm^{-1} representing CH band of proteins [180]showed higher intensities than the average in the majority of grade 2 samples suggesting important protein overproduction in those cases.

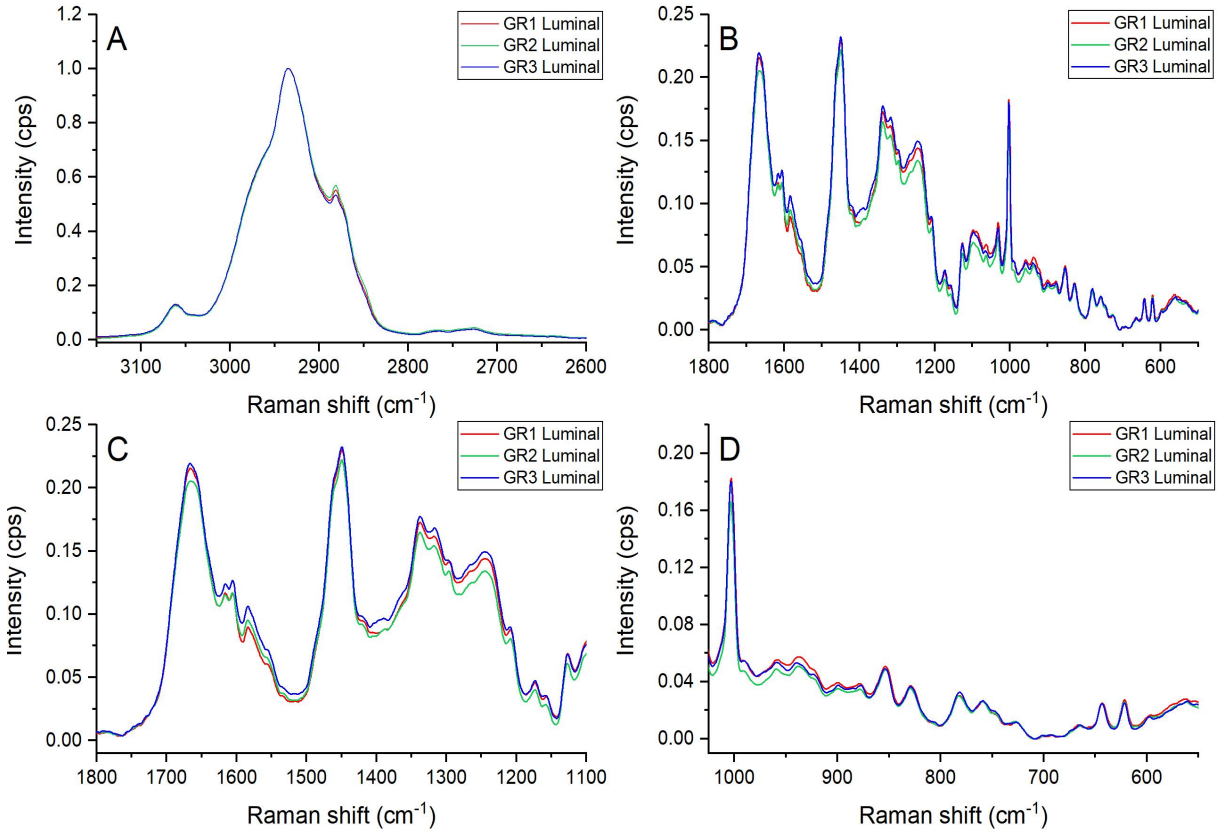

Figure 86. Raman average spectral profile of luminal samples representing grade 1, grade 2, and grade 3 cases. (A) Lipid region (3,100-2,680 cm^{-1}), (B) Fingerprint region (1,800-500 cm^{-1}), (C) Amides region (1,800-1,140 cm^{-1}) D) Amino acid and nucleic acids regions (980-600 cm^{-1}).

Grade 3 luminal spectra showed greater contribution of cholesterol ester represented by the band located at ≈2988 cm^{-1} and CH vibration in proteins represented by the protein characteristic lower energy level band placed at ≈2946 cm^{-1}. [180] Breast tumours rich in cholesterol ester have been associated with higher histologic grade in addition to be linked to proliferation and aggressive potential.[204] Cholesterol ester affect cellular activity, for example the stabilization and fluidity of the cell membrane as well as controlling the membrane permeability.[211] However, Cholesterol ester plays a relevant role when it comes to luminal breast cancer. A cholesterol metabolite presented as 27-hydroxycholesterol (27HC), has been proposed to function as oestrogen agonist, promoting and accelerating the proliferation of oestrogen receptor (ER)-positive (luminal) breast cancer cells.[215]

7.2.1.2 Fingerprint region (1,800-500 cm^{-1}) analysis of luminal samples all grades

Principal component analysis of the fingerprint region revealed an unclear separation but relatively good grouping of the different grades of luminal samples as presented in Figure 88A. PC-6 identified the following bands and assignments as differentiations of grade 1 and 3 from grade 2: ≈1650, 1605, 1332, 1123, and 1002 cm^{-1}. Grade 1 and grade 3 samples presented more intense peaks than grade 2. However, grade 2 luminal samples showed higher and lower intensities than the average.

The peak located at ≈1650 cm^{-1}assigned to amide I with random coils conformations [180] suggest conformational changes in proteins as the histological grade increases. The bands located at ≈1605 and 1535 cm^{-1} associated with cytosine NH_2 vibrations and CH_2CH_3 wagging mode of collagen and polynucleotide chains in DNA purine bases [148] had higher intensities in grade 1 and 3 cases. This fact suggests the accumulation of genetic material on abnormal and abhorrent poor differentiated cancer cells, which serves the invasive purpose of cancer.

Following the same trend, grade 1 and grade 3 samples showed collagen related peaks with increased intensities including ≈1335, 1313, and 1000 cm^{-1}. Fibroblasts play an important role in cancer progression, due to their production of growth factors, chemokines and extracellular matrix which assists the angiogenic recruitment of endothelial cells and pericytes.[216] Elevated collagen density rises tumorigenesis, local invasion, and metastasis, which translates in higher content of stromal collagen affecting tumour formation and progression. [217] Thus, the active production of fibroblast products serves as an indication of cancer progression through poor differentiation of the cancer cells.

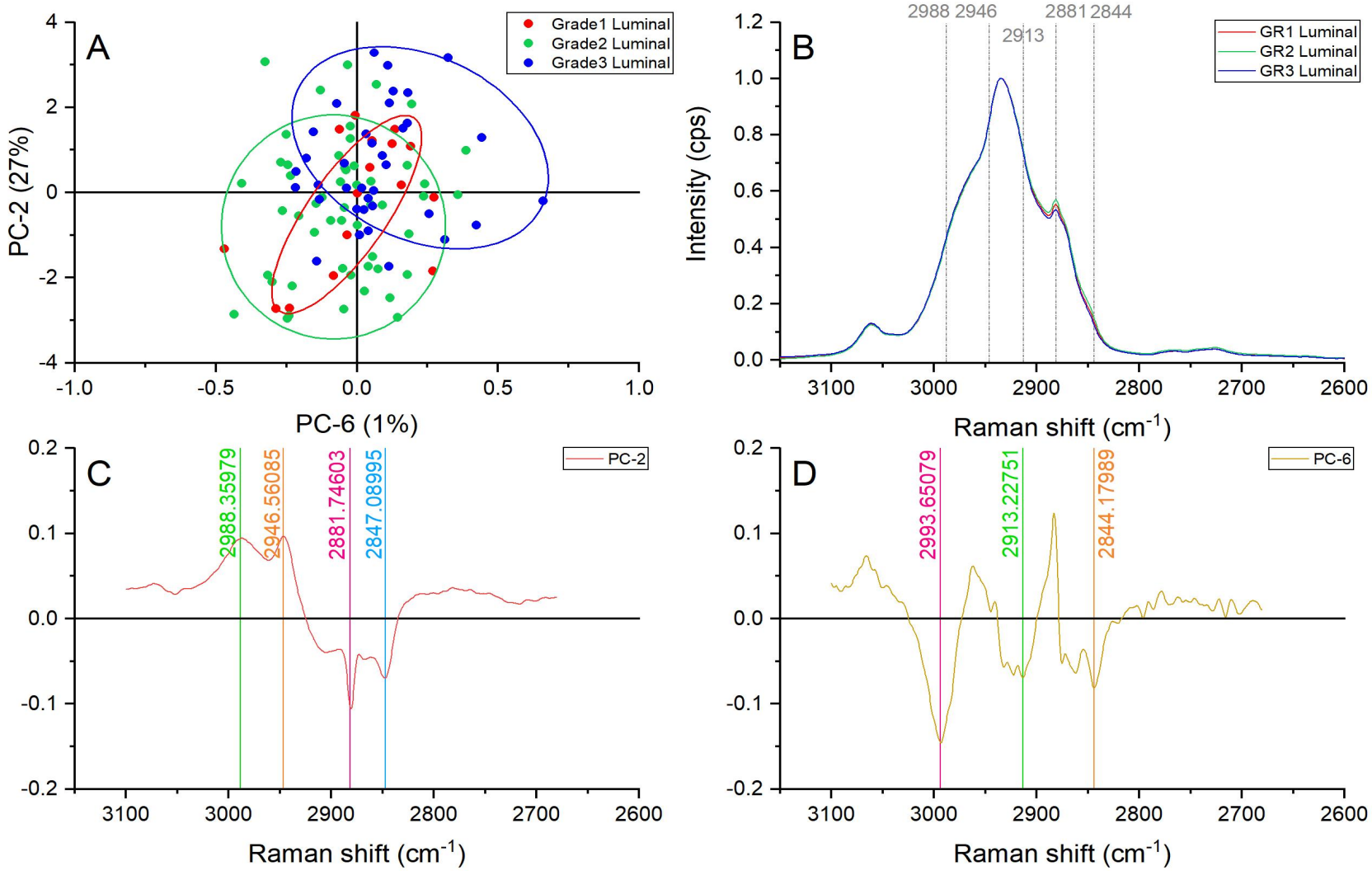

Figure 87. Principal component analysis of the lipid region (3,100-2,680 cm^{-1}). (A) Score plot using PC-2 and PC-6 accounting for 28% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the lipid region for luminal samples grade 1,2, and 3. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-2 Loading, (D) PC-6 loading.

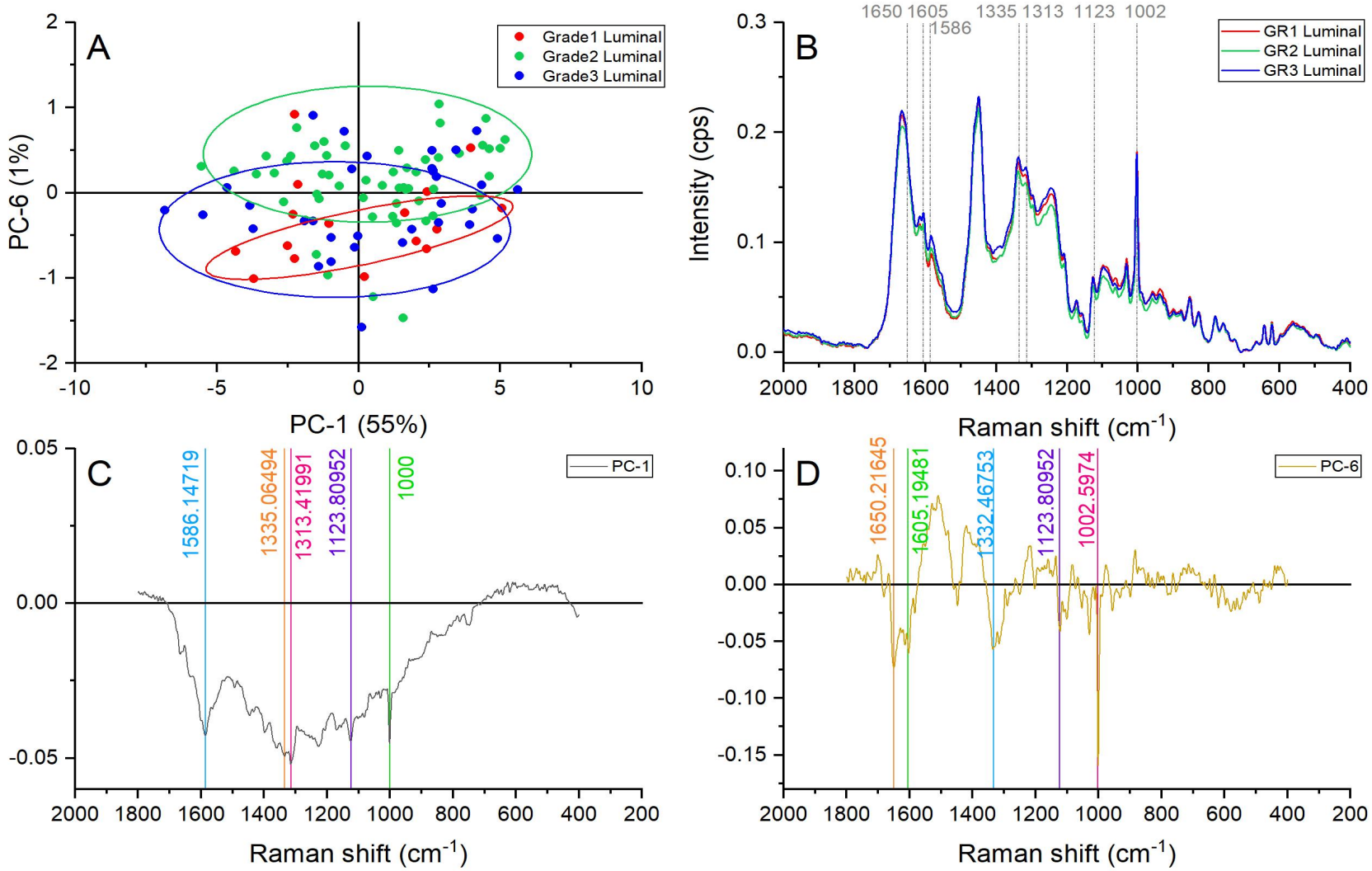

Figure 88. Principal component analysis of the fingerprint region (1,800-500 cm^{-1}). (A) Score plot using PC-1 and PC-6 accounting for 56% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the fingerprint region for luminal samples grade 1,2, and 3. 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-6 loading.

7.2.1.3 Amides region (1,800-1,140 cm^{-1}) analysis of luminal samples all grades

The different regions within the amide spectral range showed visible differences in intensity as it can be appreciated in Figure 89. Grade 1 and 3 in the luminal cases showed similar intensity behaviours in the average spectral profiles with more noticeable differences in comparison with grade 2.

Principal component analysis was applied to the amide region of luminal samples representing grade 1,2, and 3 giving as a result a good grouping and differentiation using PC-4 and PC-7 as shown in Figure 90. Grouping and differentiation among grades was achieved using PC-7 as reference. Five peaks were identified with this principal component: ≈1669, 1586, 1438, 1295, and 1243 cm^{-1}.PC-4 helped identified bands were grade 1 samples presented dominant contributions. These peaks included amide I random coils (1650 cm^{-1}), tyrosine (1614 cm^{-1}), guanine (1333 cm^{-1}), and CH_3CH_2 twisting mode in collagen (1312 cm^{-1}).[180]

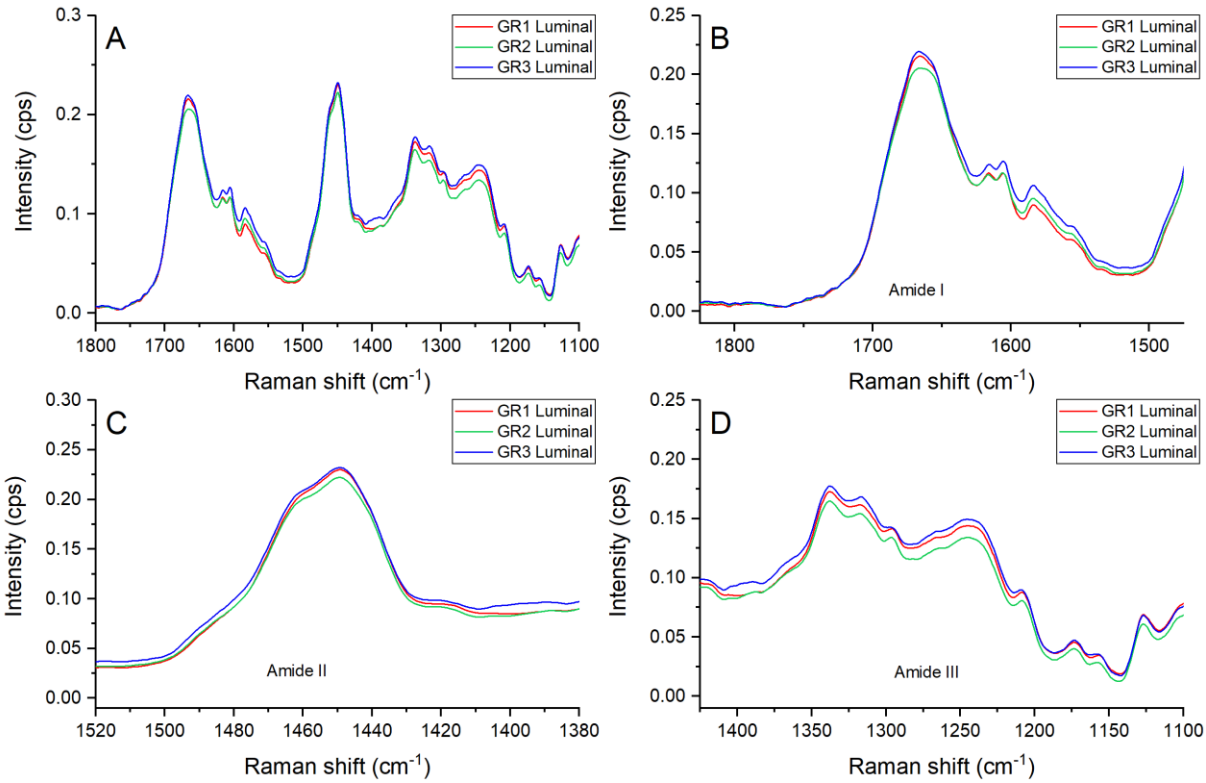

Figure 89. (A) Average spectral profile of the amide region (1,800-1,140 cm^{-1}) for luminal samples grade 1,2, and 3, (B) amide I region (1,800-1510 cm^{-1}), (C) amide II region (1,510-1,390 cm^{-1}), (D) amide III region (1,390-1140 cm^{-1}).

Greater contributions of amide I with α-helix configuration (≈1669 cm^{-1}), phenylalanine and hydroxyproline vibrations (≈1586 cm^{-1}), and amide III vibrations linked to asymmetric phosphate stretching modes in nucleic acids (≈1243 cm^{-1})[181] were identified in grade 3 luminal samples followed by grade 1 and grade 3. This behaviour suggested that higher luminal breast cancer grades focus on the proliferation though generation of cellular component and stimulation of the ECM. [89]

On the other hand, grade 2 spectra presented a dominant contribution of lipid related bands presented at ≈1438 cm^{-1} (CH_2 deformation and scissoring in lipids) and ≈1295 cm^{-1}(CH_2 deformation and ceramide vibrations)[180]. Grade 2 presented higher intensities for the aforementioned peaks followed by grade 1 and grade 3 samples respectively. Ceramides are a family of lipid molecules and are product of sphingolipid synthesis. Ceramides participate in several cellular signalling processes such as regulation of differentiation, proliferation, and programmed cell death (PCD) of cells.[218] The active lipid transformation occurring in grade 2 samples suggest de novo lipogenesis focusing on sphingolipids components as an essential part of cancer development and dedifferentiation towards higher histological grades.

7.2.1.4 Amino acid and nucleic acids regions (980-600 cm^{-1}) analysis of luminal samples all grades

Relatively good grouping and overlapping of cluster was obtained when the amino acid and nucleic acid region was analysed with principal component analysis as presented in Figure 91A. Within the PC-5 axis grade 2 and grade 3 samples were clustered following opposite directions with grade 1 laying in between. As result of the peak identification using PC-5 loading the following bands were assigned as responsible for the samples grouping: ≈ 783, 754, 746, 715, and 704 cm^{-1}.

The band located at ≈783 cm^{-1} showed greater contribution in grade 3 spectra of luminal samples followed by grade 2 and 1. This peak is assigned to phosphodiester vibrations and therefore corresponds to nucleic acids contributions. Its higher contribution and dominant behaviour as the histological grade goes up indicates a more active metabolism and cell density at higher grades. [74]

Grade 2 luminal samples followed by grade 1 and 3 presented more intense peaks representing symmetric breathing of tryptophan (≈754 cm^{-1}), ring breathing modes of DNA and RNA bases (≈746 cm^{-1}), $CN^{+}(CH_3)_3$ lipid vibrations (≈715 cm^{-1}), and cholesterol ester (≈704 cm^{-1}). All these bands represent a wide collection of biomolecules including proteins, nucleic acids, and lipids. Grade 2 seemed to behave more dynamically during the abnormal production of cellular components leading to cell division in comparison with the lower and higher histological grade. Additionally, once again and correlating with the presented FTIR results, adaptations of the lipid components possibly affecting membrane fluidity, and energy storage reservoirs have been identified.

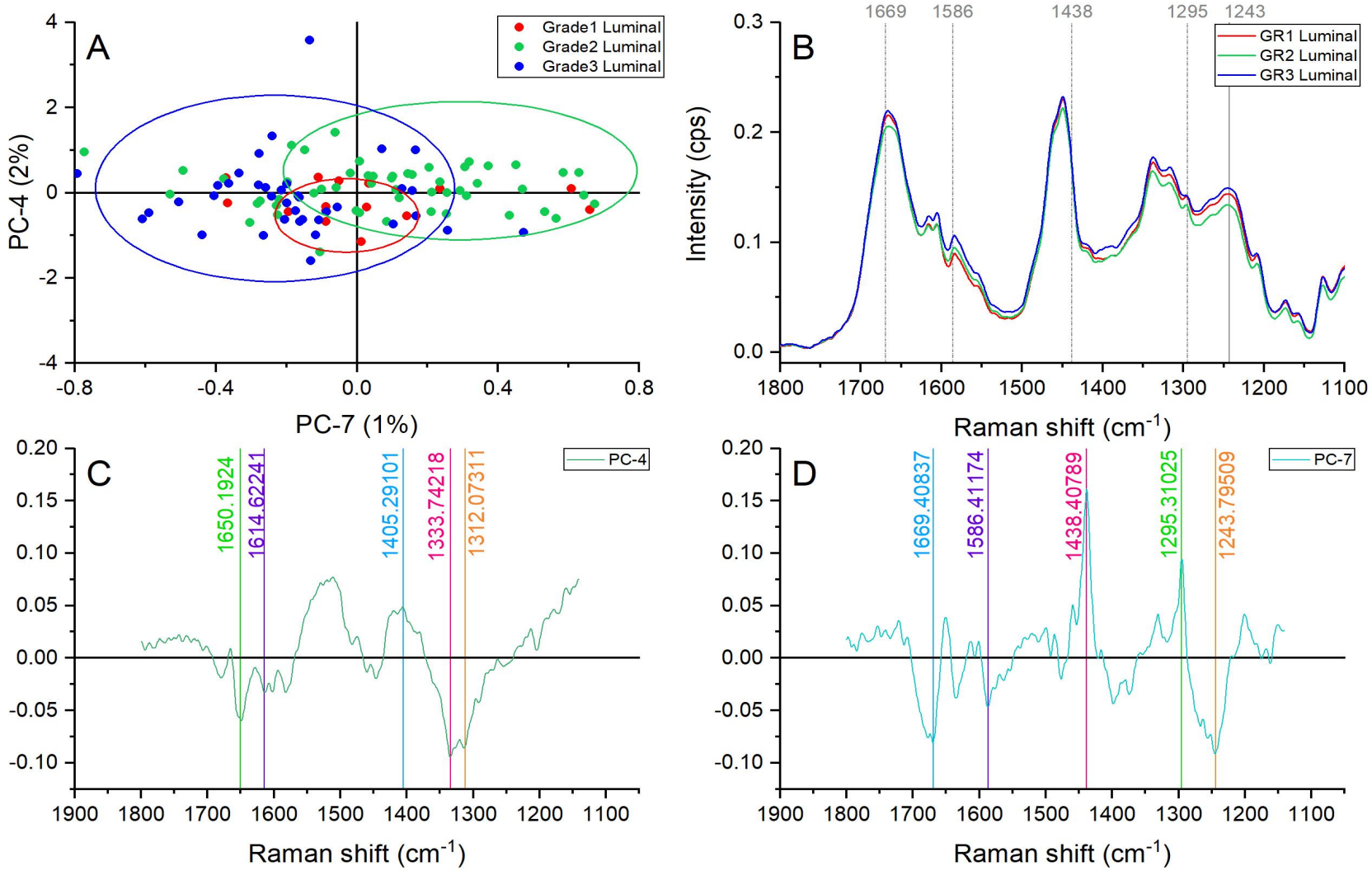

Figure 90. Principal component analysis of the amide region (1,800-1,140 cm-1). (A) Score plot using PC-2 and PC-7 accounting for 23% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the amide region for luminal samples grade 1,2, and 3. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-2 Loading, (D) PC-7 loading.

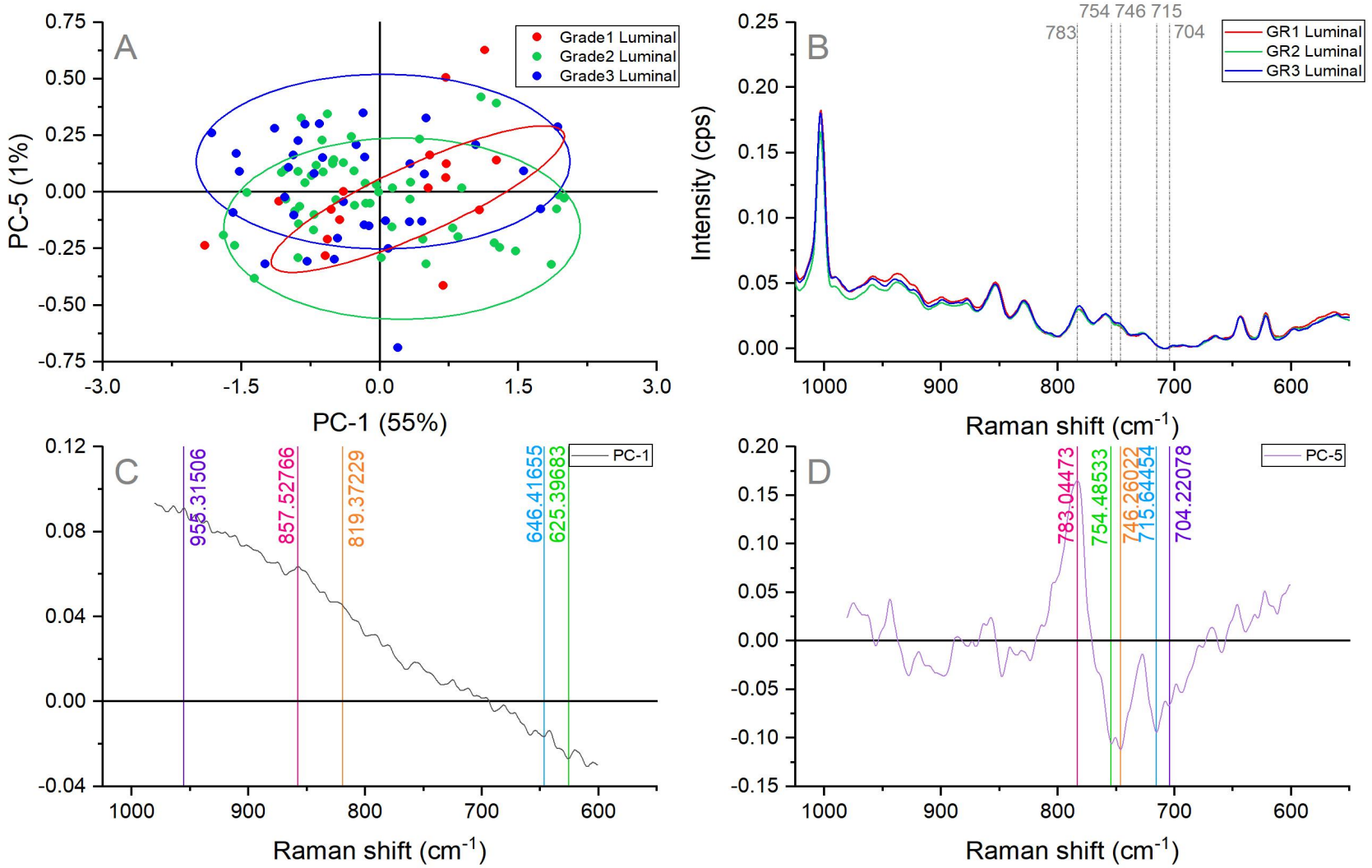

Figure 91. Principal component analysis of the amino acid and nucleic acid region (980-600 cm^{-1}). (A) Score plot using PC-1 and PC-5 accounting for 56% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the amino acid and nucleic acid region for luminal samples grade 1,2, and 3. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-1 Loading, (D) PC5 loading.

7.2.2 TNBC SAMPLES ALL GRADES

The average spectral profile of TNBC samples representing grade 1, 2, and 3 showed differences enough to see with the naked eye as presented in Figure 93. In all of the spectral profile regions similitudes between grade 2 and 3 TNBC samples whereas the grade 1 showed more differences with the rest of the cases. The discussion of these specific areas is presented in the following sections.

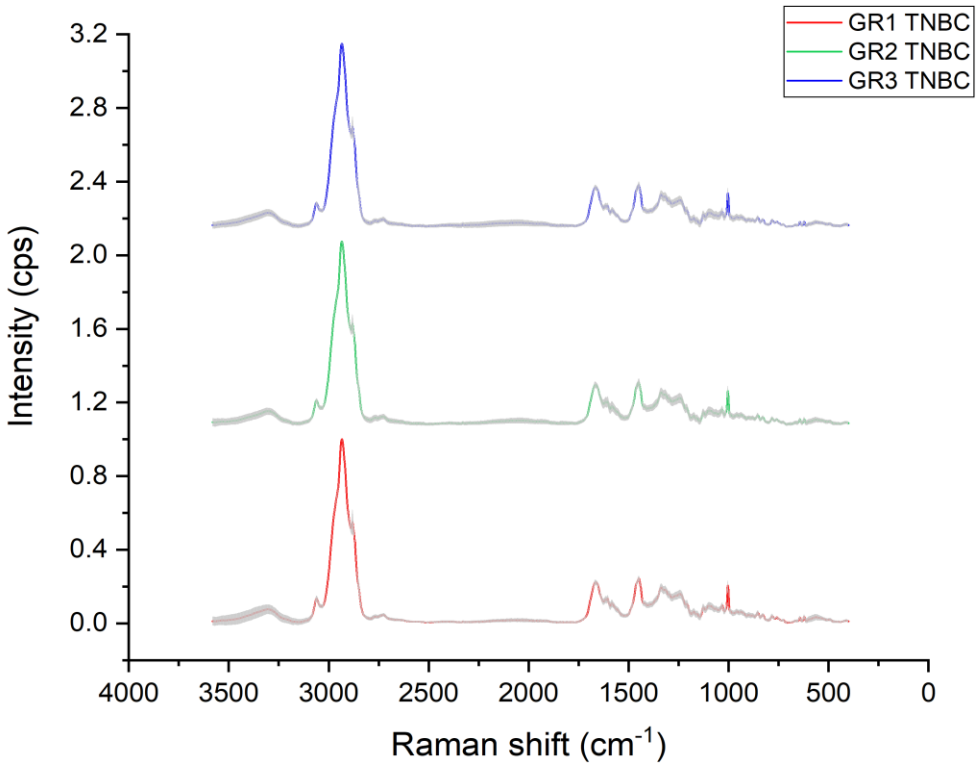

Figure 92. Raman average spectral profile of TNBC samples representing grade 1, grade 2, and grade 3 cases, variance is represented with grey shadow.

7.2.2.1 LIPID REGION (3,100-2,680 CM^{-1}) ANALYSIS OF TNBC SAMPLES ALL GRADES

Principal component analysis of the lipid region revealed an unclear separation but acceptable grouping of the different grades of TNBC samples as presented in Figure 94. PC-1 and PC-5 were used as they presented the best separation and clustering.

PC-1 achieved the best separation identifying the following bands and assignments as responsible for the clustering of all three TNBC cases: ≈2947 cm^{-1} assigned to asymmetric stretch of CH_2 in lipids, ≈2881 cm^{-1} caused by CH_2 asymmetric stretch of lipid and proteins, and 2849 cm^{-1} representing CH_3 symmetric stretch of lipids.[180] All these bands presented greater contribution in grade 2 samples followed by grade 1 and 3. Grade 1 samples were clustered in the origin of the score plot, which confirm average intensities for the previously listed peaks.

As discussed before, the evolution of breast cancer is characterised by lipid peroxidation. [88] Therefore, a worst differentiated cancer cell is suggested to have a lower content of lipids. Additionally, grade 3 cancer cells are associated with more aggressive behaviour, which might

indicate hypoxic events in advanced cases. The lack of oxygen supply in hypoxic areas in breast tumours lead to higher rates of oxidative stress, resulting in considerable lipid peroxidation and oxidative degradation of lipids. [190]

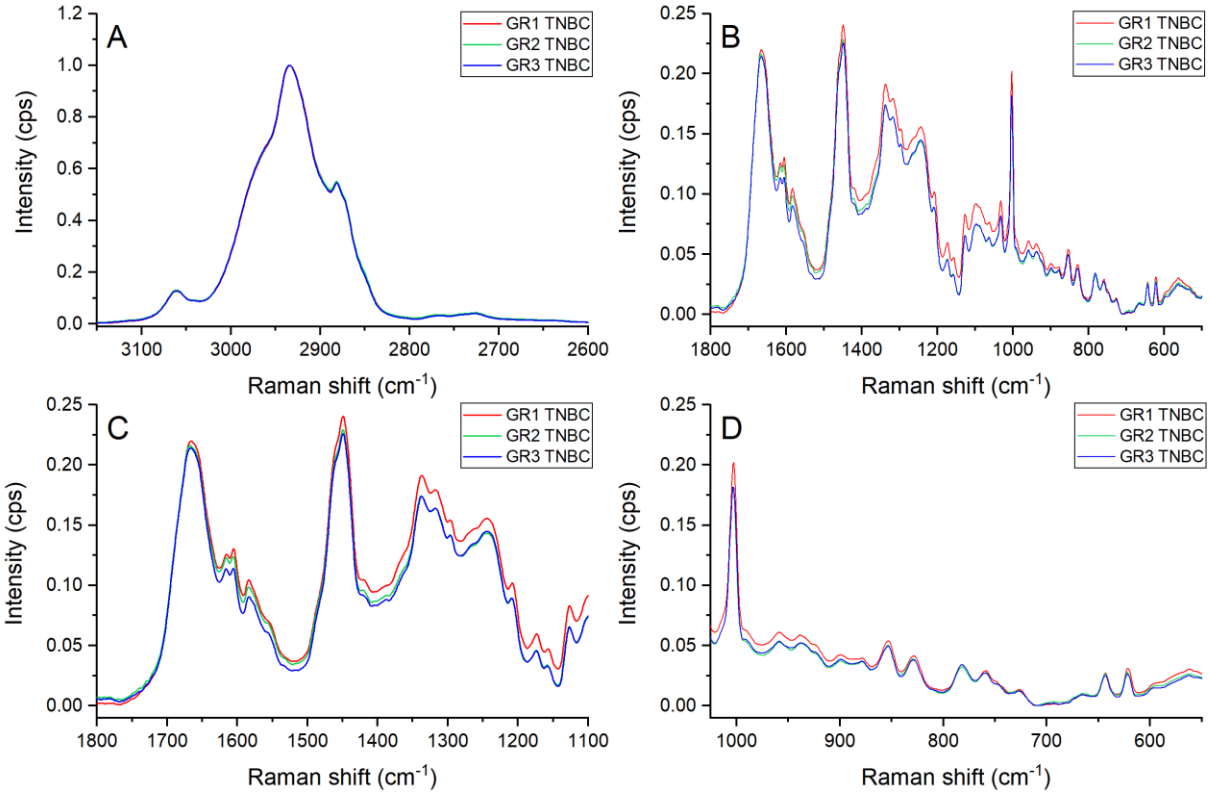

Figure 93. Raman average spectral profile of TNBC samples representing grade 1, grade 2, and grade 3 cases. (A) Lipid region (3,100-2,680 cm^{-1}), (B) Fingerprint region (1,800-500 cm^{-1}), (C) Amides region (1,800-1,140 cm^{-1}) (D) Amino acid and nucleic acids regions (980-600 cm^{-1}).

Besides, PC-5 revealed higher intensities in the ≈2925 cm^{-1} band in grade 3 samples, followed by grade 2 and 1. This band is assigned to the CH_3 symmetric stretch due primarily to proteins.[89] This situation confirms the varying protein contents with respect to the grades.

7.2.2.2 Fingerprint region (1,800-500 cm^{-1}) analysis of TNBC samples all grades

The fingerprint region principal component analysis showed a good grouping for grade 3 TNBC cases while grade 1 and 2 samples were dispersed as can be appreciated in Figure 95. Grade 1 TNBC samples were spread equally in the PC-4 axis. The grade 3 cluster was identified in the positive axis of PC-4, while grade 2 samples were mainly dispersed in the negative axis of the same principal component.

The peaks located at ≈1669, 847, 780 cm^{-1} presented higher intensities in the grade 3 TNBC samples. These bands were assigned to amide I with α-helix conformation, C-O-C skeletal mode in monosaccharides, and uracil-based ring breathing mode respectively. [180]. The greater contribution of proteins and nucleic acids in grade 3 samples might indicate higher metabolic activity, but also higher cell density in the collection volume.

Furthermore, lower intensities of the ≈1605 and 1108 cm^{-1} peaks in the grade 3 TNBC samples. These two band are representative of collagen assignments.[180]. This fact suggests a lower content of collagen in grade 3 samples in comparison with the lower histological grades.

The spreading of grade 1 and grade 2 samples of TNBC cases between both axes in both principal components suggest highly dynamic behaviours in this subtype in the less aggressive histological grades.

7.2.2.3 Amides region (1,800-1,140 cm^{-1}) analysis of TNBC samples all grades

The spectral average of all three histological regions presenting TNBC subtype showed visible differences in the three amide regions as presented in Figure 96. Grade 1 samples showed the most differences with grade 2 and 3 that showed similar profiles.

Principal component analysis of the amide region showed a nice grouping and partial separation of grade1,2, and 3 of TNBC as can be appreciated in Figure 97A. A clear clustering for grade 1 and grade 2 were achieved with partial clustering of grade 3 samples. PC-3 was responsible for the separation of all samples.

The analysis of the amide region revealed important overproduction of nucleic acids and proteins in grade 3, followed by grade 2 and 1 of the TNBC samples. The overproduction of proteins and nucleic acids was characterised by higher intensities in the bands located at ≈1669, 1247, and 1236 cm^{-1} which represent amide I α-helix conformation, amide III , and a combination of amide III and CH_2 wagging vibrations from nucleic acids and glycine backbones and proline chains respectively.[180] The first finding complemented the presented FTIR results which suggest a higher β-sheet conformation in grade 1 and 2 in TNBC. The overproduction of cell components is characteristic of cell division and proliferation processes which is expected with more contributions in more aggressive and higher grades of cancer

Grade 1 samples showed greater contributions of the following bands found at ≈1589 and 1557 cm^{-1} which are considered collagen characteristic peaks (phenylalanine and hydroxyproline vibrations).[148]

7.2.2.4 Amino acid and nucleic acids regions (980-600 cm^{-1}) analysis of TNBC samples all grades

The amino acid and nucleic acid region principal component analysis showed a good dispersion and overlay of the grade 1 ,2, and 3 TNBC partial clusters as can be appreciated in Figure 98. Grade 1 TNBC samples seem to cluster on the positive PC-3 axis. This indicate a greater contribution of collagen represented by the ≈940 and 624 cm^{-1} peaks assigned to collagen backbone vibrations and C-C twisting mode of phenylalanine.[203] In agreement with previously reported results lower histological grades seem to present higher collagen and ECM components.

Other protein related bands were identified with greater contributions in the grade 3 samples followed by grade 2 TNBC. These peaks were located at ≈921 and 750 cm^{-1} representing proline ring vibrations and breathing modes of tryptophan.[196] This elevated content of proteins reflect higher metabolism and cell density in the most abnormal cancer cells, which

is expected in breast cancer cells with aggressive behaviours. However, the overlay of cluster suggested a variety of biochemical changes among different histological grades. Therefore, and overproduction of proteins and nucleic acids characterised by ≈853 and 830 cm^{-1} assigned to the vibration caused by the ring breathing mode of tyrosine [180] was constant on all the grades.

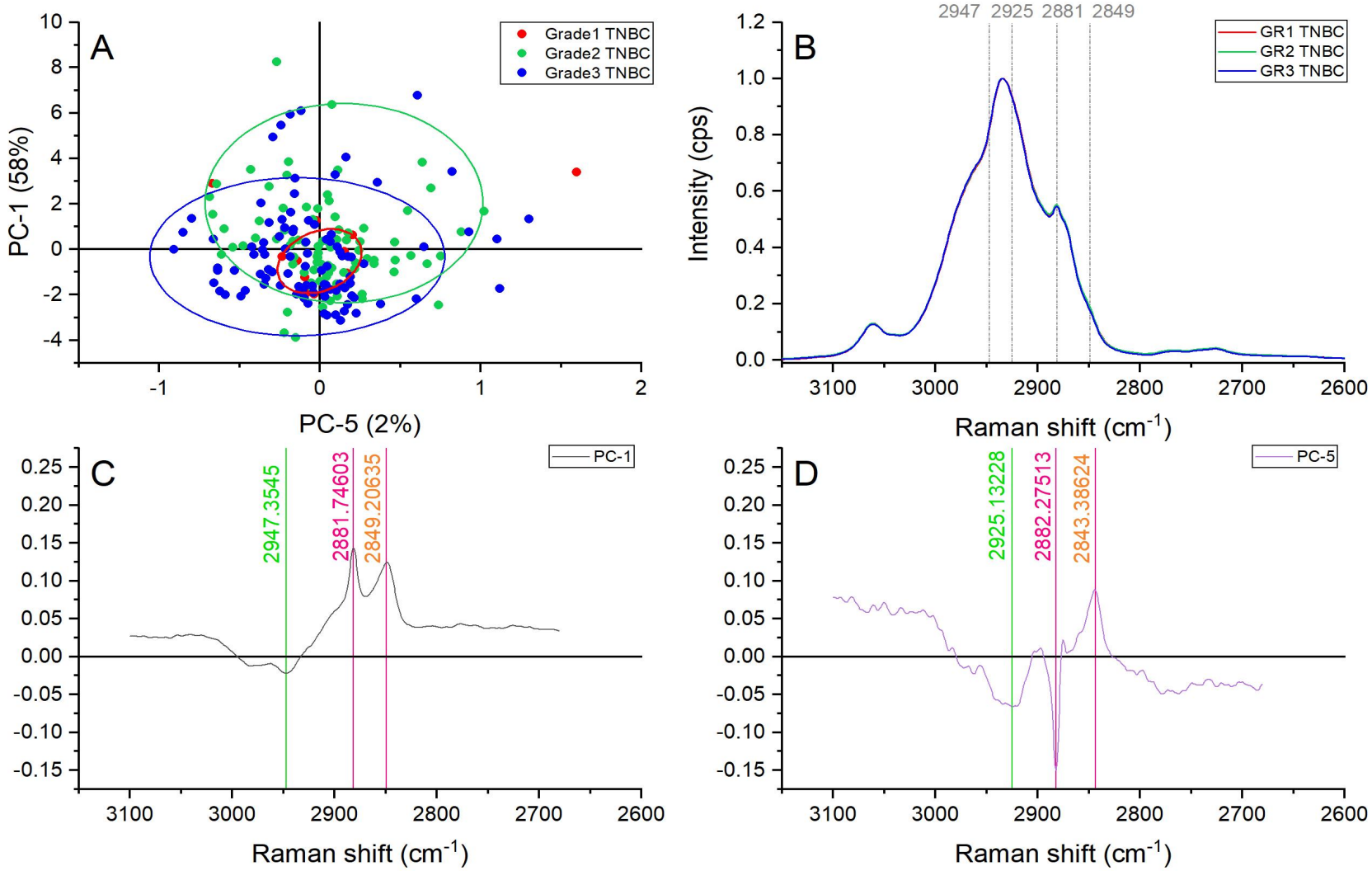

Figure 94. Principal component analysis of the lipid region (3,100-2,680 cm^{-1}). (A) Score plot using PC-1 and PC-5 accounting for 60% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the lipid region for TNBC samples grade 1,2, and 3. Dotted 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.

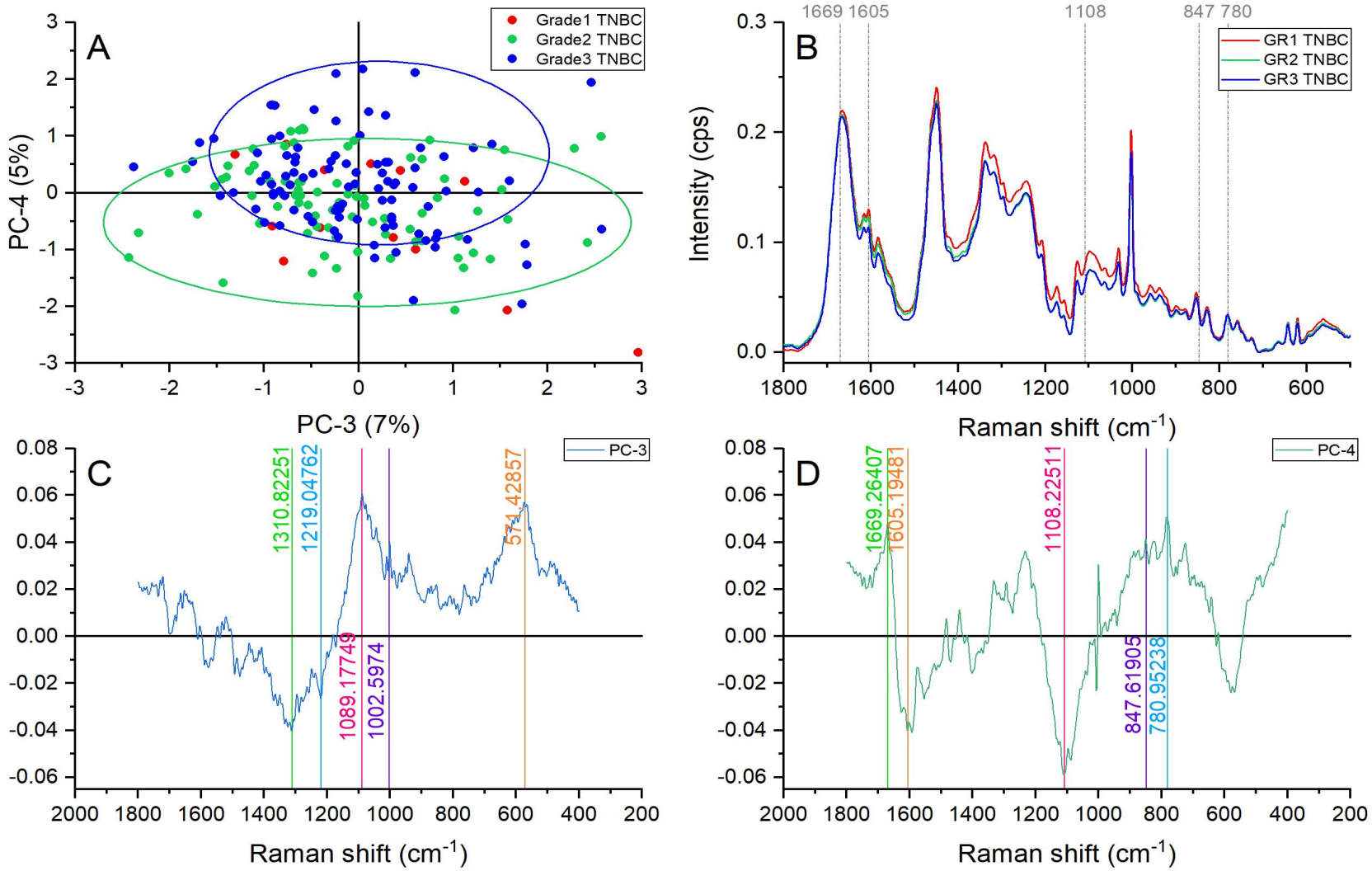

Figure 95. Principal component analysis of the fingerprint region (1,800-500 cm^{-1}). (A) Score plot using PC-3 and PC-4 accounting for 12% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the fingerprint region for TNBC samples grade 1, 2, and 3. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-3 Loading, (D) PC-4 loading.

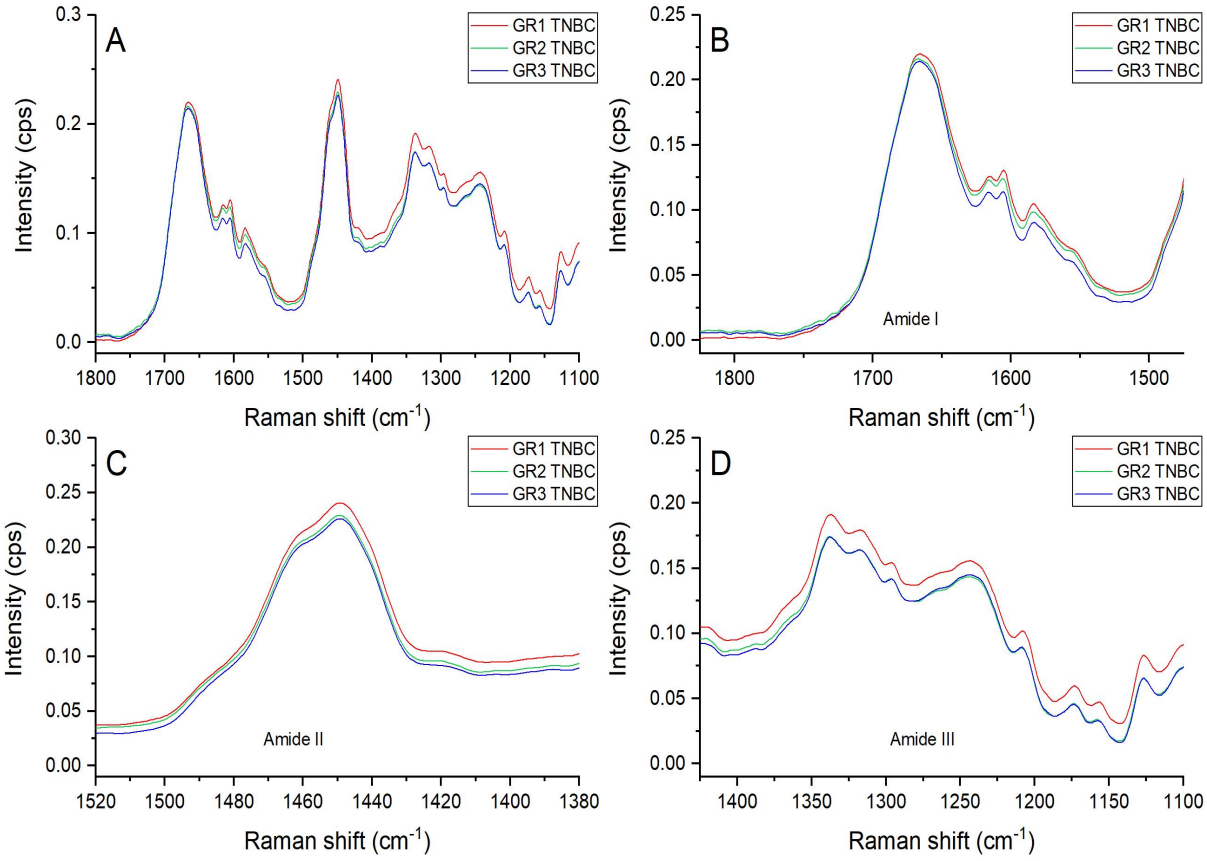

Figure 96. (A) Average spectral profile of the amide region (1,800-1,140 cm^{-1}) for TNBC samples grade 1,2, and 3, (B) amide I region (1,800-1,510 cm^{-1}), (C) amide II region (1,510-1,390 cm^{-1}), (D) amide III region (1,390-1,140 cm^{-1}).

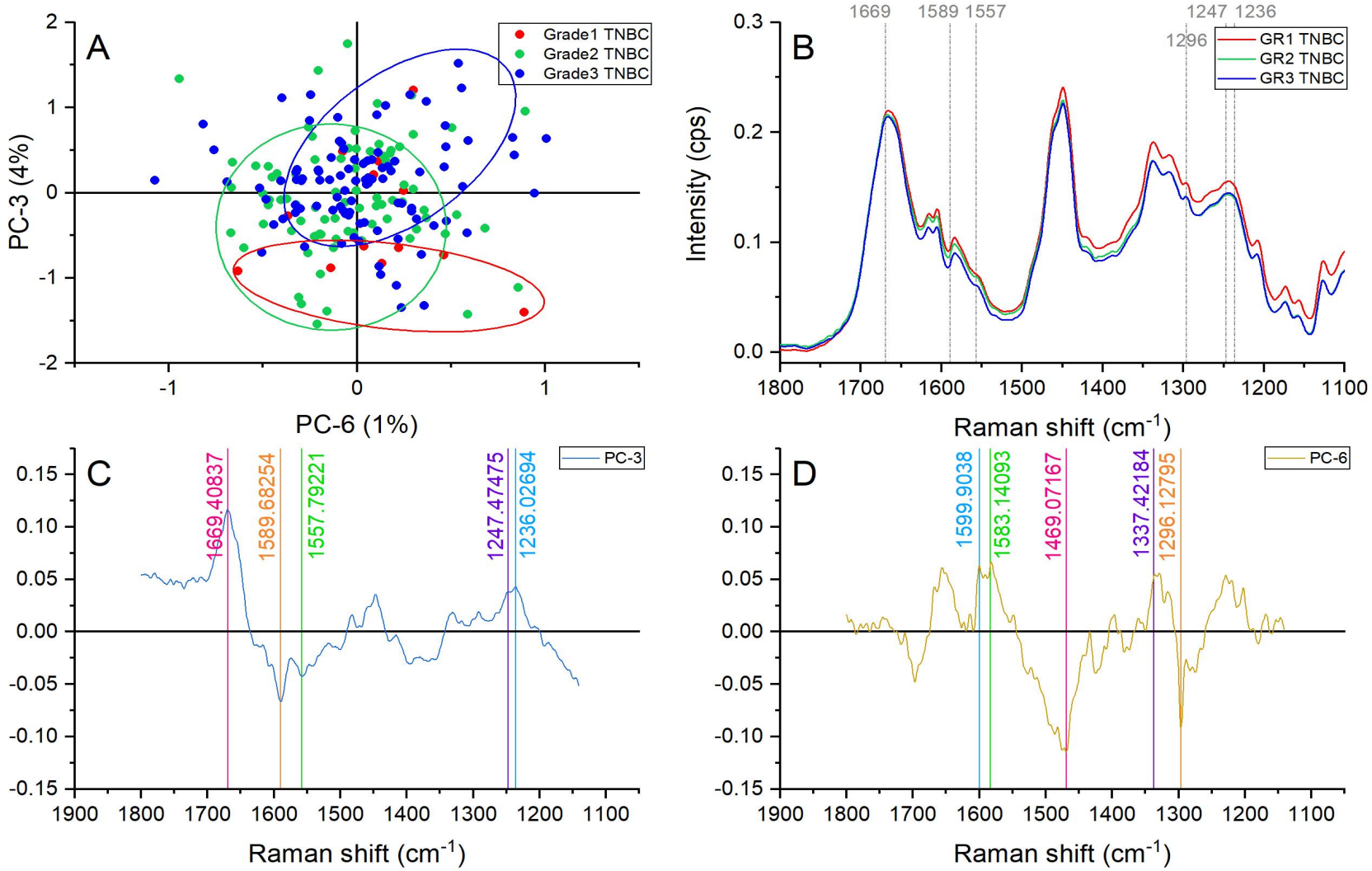

Figure 97. Principal component analysis of the amide region (1,800-1,140 cm^{-1}). (A) Score plot using PC-3 and PC-6 accounting for 5% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the amide region for TNBC samples grade 1, 2, and 3. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-3 Loading, (D) PC-6 loading.

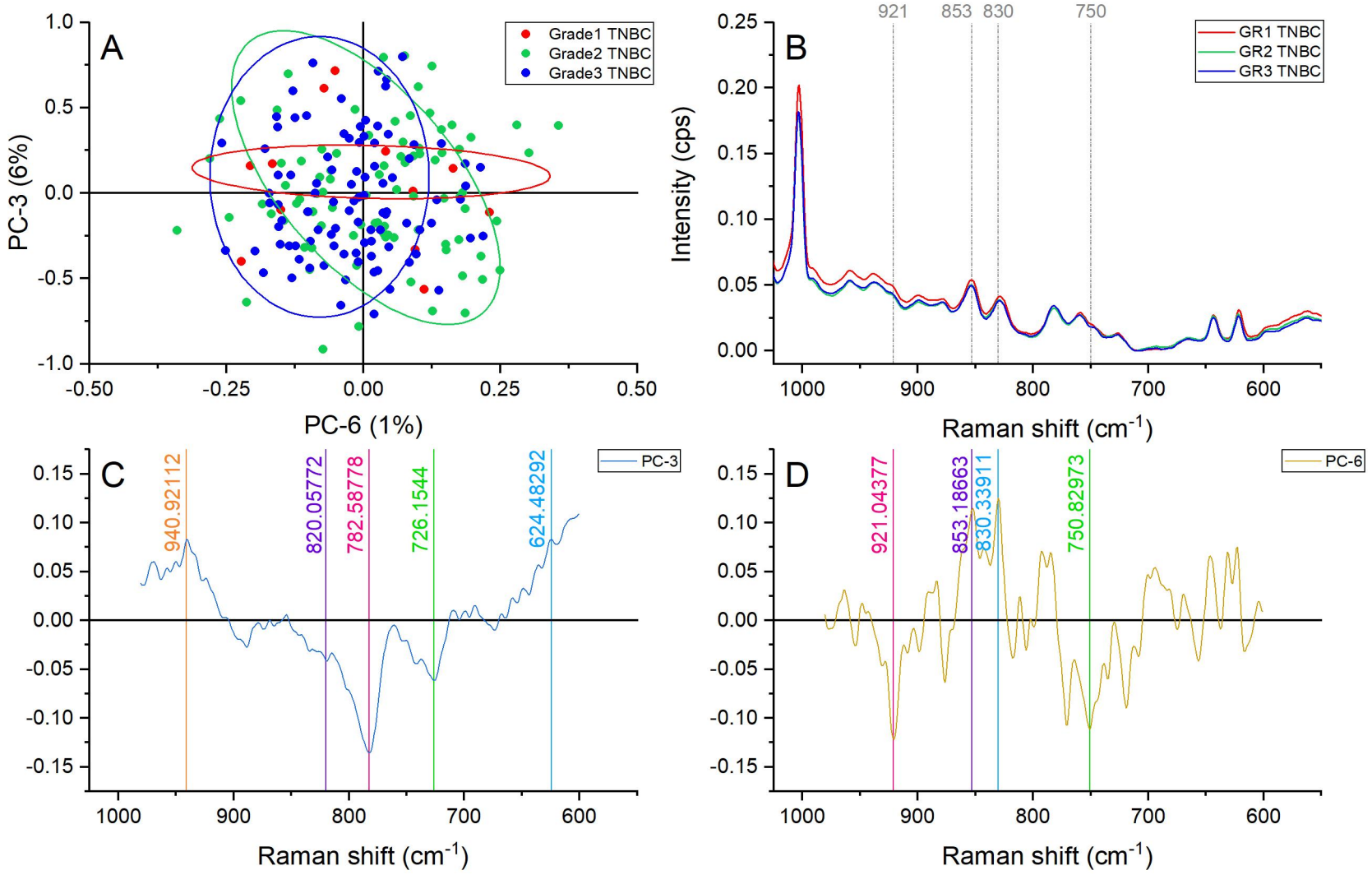

Figure 98. Principal component analysis of the amino acid and nucleic acid region (980-600 cm^{-1}). (A) Score plot using PC-3 and PC-6 accounting for 7% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the amino acid and nucleic acid region for TNBC samples grade 1, 2, and 3. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-3 Loading, (D) PC-6 loading.

7.2.3 HER2+ SAMPLES ALL GRADES

When compared, the average spectral profile of grade 1, 2, and 3 HER2+ displayed clear differences as presented in Figure 99 and Figure 100. In comparison with other subtype's analysis the HER2+ spectral profile showed more differences among all histological grades. The discussion of these specific areas is presented in the following sections.

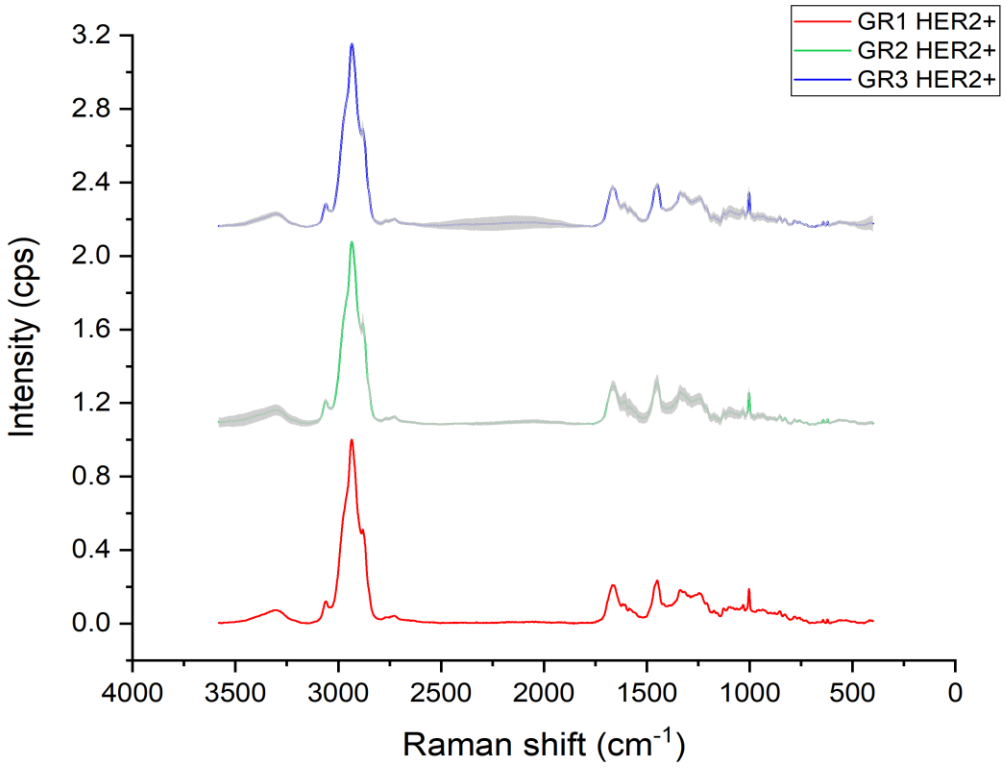

Figure 99. Raman average spectral profile of HER2+ samples representing grade 1, grade 2, and grade 3 cases, variance is represented with grey shadow.

7.2.3.1 LIPID REGION (3,100-2,680 CM^{-1}) ANALYSIS OF HER2+ SAMPLES ALL GRADES

Good grouping and separation was achieved in the lipid region using PC-1 and PC-4 as it can be seen in Figure 101. Six bands were identified as responsible for the separation and grouping of HER2+ samples of all grades. The only grade 1 sample with histological grade 1 presented a more intense peak at ≈2977 cm^{-1} which suggests a higher asymmetric vibration and content of CH_3 coming from lipids, specially fatty acids.[211] This fact correlates with our FTIR results, and can be explained by a less active lipid transformation, by peroxidation, degradation and active novo-lipogenesis creation, in lower grades.

PC-4 was responsible for the grade 2 and grade 3 separation. Higher contribution of CH vibrations in proteins (≈2916) was presented in grade 3 followed by grade 2 and 1 suggesting higher cell density and metabolism in higher grades. Additionally, lipid related bands located at ≈2993 cm^{-1} (cholesterol ester), and ≈2880 and 2847 cm^{-1} (CH_2 symmetric and asymmetric stretch in lipids) presented increased intensity in grade 2 samples.

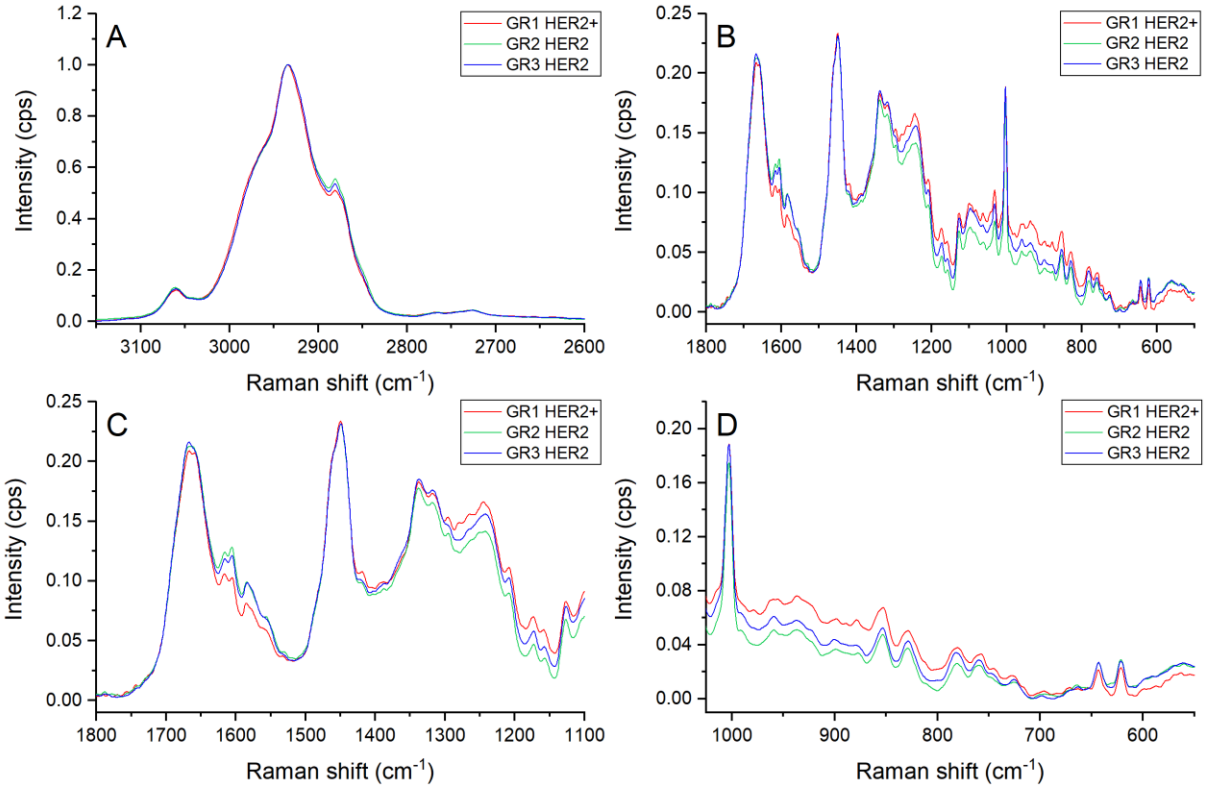

Figure 100. Raman average spectral profile of HER2+ samples representing grade 1, grade 2, and grade 3 cases. A) Lipid region (3100-2680 cm^{-1}), B) Fingerprint region (1800-500 cm^{-1}), C) Amides region (1800-1140 cm^{-1}) D) Amino acid and nucleic acids regions (980-600 cm^{-1}).

7.2.3.2 Fingerprint region (1,800-500 cm^{-1}) analysis of HER2+ samples all grades

The fingerprint principal component analysis of HER2+samples showed partial differentiation among grades and good grouping when PC-2 and PC-3 were used as presented in Figure 102. PC-2 identifies the following peaks as responsible for the clustering and separation of samples: ≈1601, 1455, 1361, 939, and 850 cm^{-1}.

The first three bands were assigned to phenylalanine, deoxyribose, and tryptophan vibrations hinting accelerated and active production of cell components.[74] Grade 3 showed higher contribution of these components followed by grade 2 and 1. The other two bands, ≈939 and 850 cm^{-1}, were identified as collagen characteristic peaks (proline and hydroxyproline)[203] and presented the opposite trend to the aforementioned bands with higher intensities in grade 1, follow by higher histological grades in order.

Moreover, the majority of peaks identified with PC-3 were associated also with collagen characteristic vibrations. As expected, grade 1 presented more intense band at those locations followed by grade 2 and 3. Both findings suggest that at early histological grades the collagen and ECM production remains the focus, while at higher and intermediate grades the proliferation of cellular material is priority.

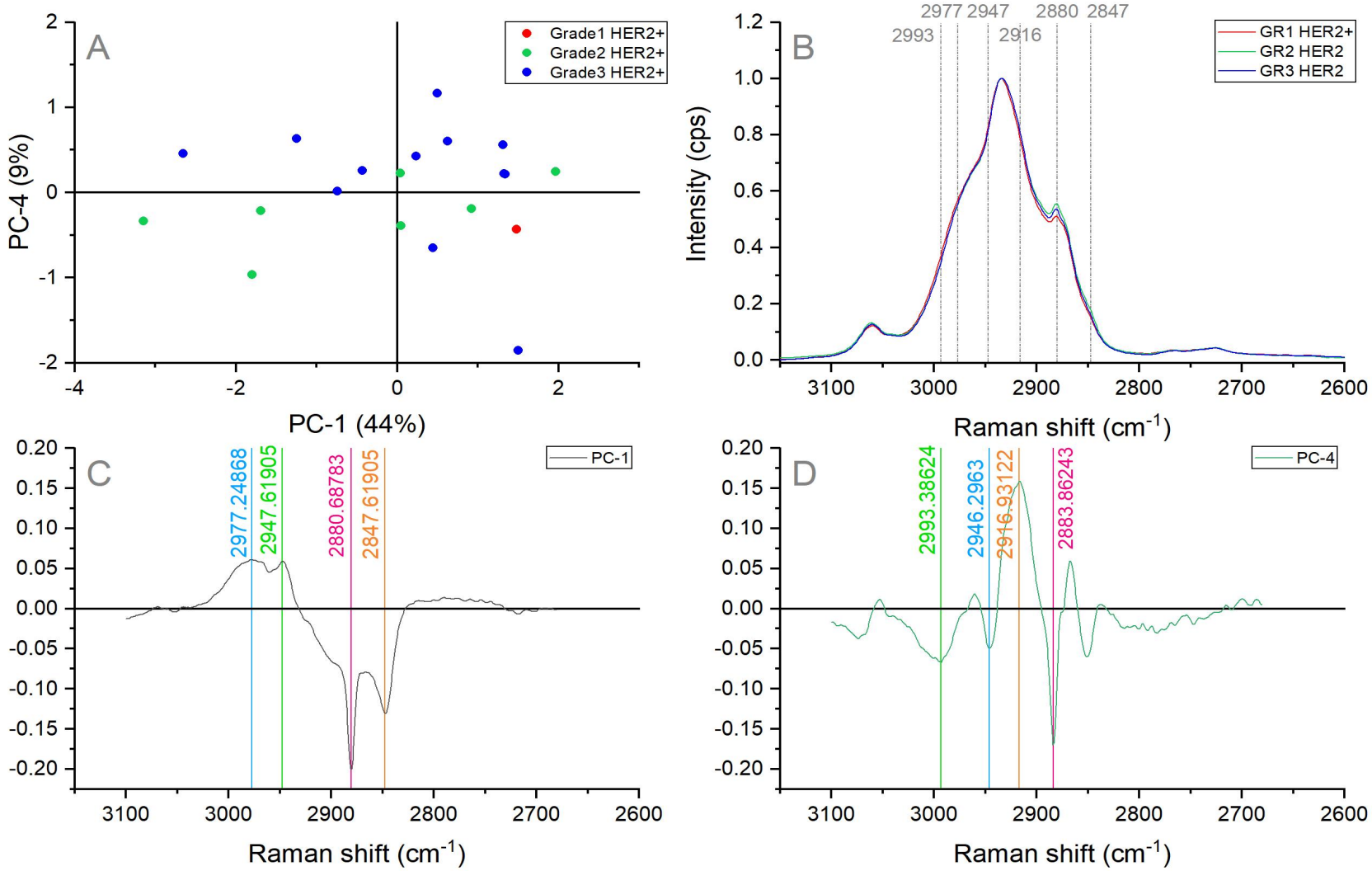

Figure 101. Principal component analysis of the lipid region (3,100-2,680 cm^{-1}). (A) Score plot using PC-1 and PC-4 accounting for 53% of the variance, (B) Average spectral profile of the lipid region for HER2+ samples grade 1, 2, and 3. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-1 Loading, (D) PC-4 loading.

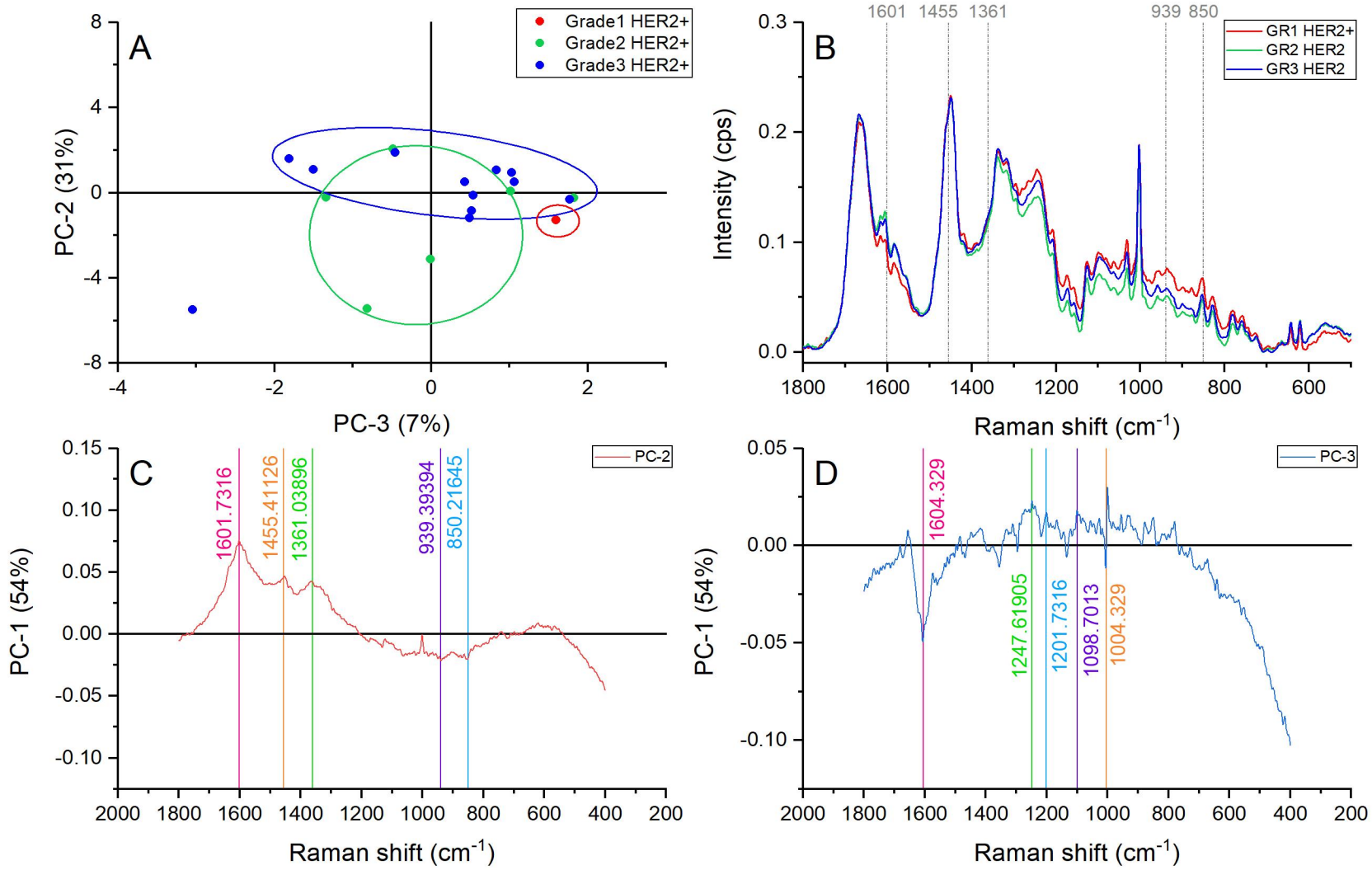

Figure 102. Principal component analysis of the fingerprint region (1,800-500 cm^{-1}). (A) Score plot using PC-2 and PC-3 accounting for 38% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the fingerprint region for HER2+ samples grade 1, 2, and 3. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-2 Loading, (D) PC-3 loading.

7.2.3.3 Amides region (1,800-1,140 cm^{-1}) analysis of HER2+ samples all grades

Important differences can be appreciated in the different amide regions of the average spectral profiles of all grades HER2+ breast cancer samples (Figure 103). These differences include intensity, sharpness and shape of the Raman peaks, which is particularly noticeable in the amide III region.

Principal component analysis of the amide region resulted in good grouping and identification of HER2+ samples with different grades as presented in Figure 104. Grade 1 and 3 HER2+ samples showed higher intensities in the peaks located at ≈1367 and 1179 cm^{-1}. The former band is assigned to the symmetric stretch of CH_3 groups in phospholipids, while the later corresponds to cytosine and guanine vibrations.[180] These contributions suggest that ongoing changes through grade 1 and 3 are occurring in the cellular membranes and the content of nucleic acids. However, only a single tumour sample was available with this subtype representing grade 1.

Grade 2 samples location in the score plot suggested more contribution of tyrosine and phenylalanine ring vibrations (≈1603 cm^{-1}), methylene deformation in biomolecules (≈1450 cm^{-1}), and ceramide deformation (≈1295 cm^{-1}) [180] in comparison with grade 1 and 3. These changes indicate adaptations in protein production, as well as methylation processes. The hyper-methylation in breast cancer has been investigated and has demonstrated affectation in the activity of genes in charge of steroid receptors, cell adhesion and matrix metalloproteinases inhibitors. [176]

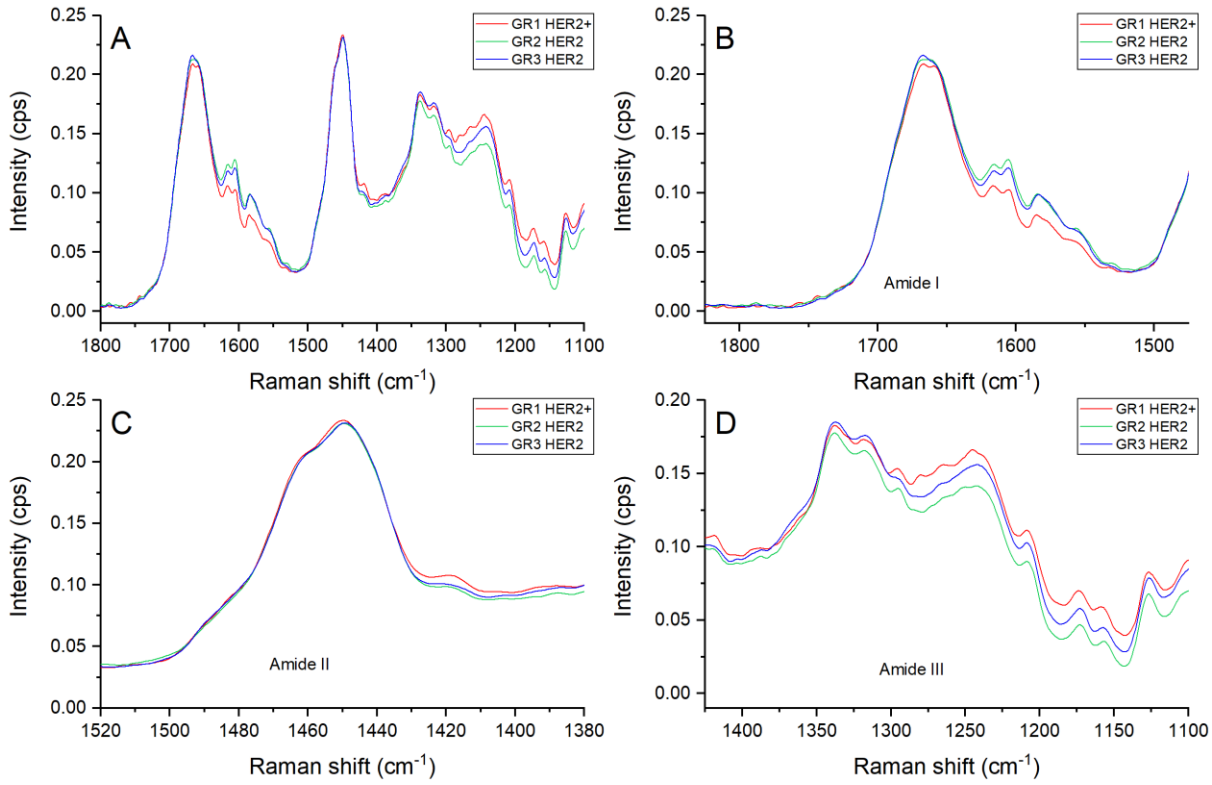

Figure 103. (A) Average spectral profile of the amide region (1,800-1,140 cm^{-1}) for HER2+ samples grade 1,2, and 3, (B) amide I region (1,800-1,510 cm^{-1}), (C) amide II region (1,510-1,390 cm^{-1}), (D) amide III region (1,390-1,140 cm^{-1}).

The only grade 1 sample presented important contribution in bands associated with DNA bases (≈1483 cm^{-1})[148], and CH_2 deformation and scissoring in lipids and acyl chains (≈1436 and 1419 cm^{-1})[211]. Therefore, active transformation in the lipid content can be assumed in early HER2+ grades.

7.2.3.4 Amino acid and nucleic acids regions (980-600 cm^{-1}) analysis of HER2+ samples all grades

The amino acid and nucleic acid region principal component analysis showed a good separation and clustering of the grade 1 ,2, and 3 HER2+ samples as can be appreciated in Figure 105. Grade 3 HER2+ samples seem to cluster on the positive PC-7 axis. This indicate a greater contribution of nucleic acids and proteins represented by the 780 and 647 cm^{-1} peaks assigned to uracil-based ring breathing mode and C-C twisting mode of tyrosine.[180]

Other protein related bands were identified with greater contributions in the grade 2 and 1. These peaks were located at ≈887, 856, 849, and 620 cm^{-1}. However, some of these bands (≈856 and 620 cm^{-1}) are characteristic of collagen vibrations which suggest changes in the tumour bed and its ECM to promote invasive processes and angiogenesis.[217] The rest of the listed peaks are associated with conformational changes favouring β-sheet conformations (≈887 cm^{-1}), and regular protein vibrations such as tyrosine (≈849 cm^{-1}).

Furthermore, lipid characteristic peaks were also identified with higher intensities in grade 2 samples. These bands were ≈955 cm^{-1} representing carotenoids and cholesterol and ≈877 cm^{-1} representative of C-C-N^+ symmetric stretch in lipids.[180], [211]The importance of cholesterol and lipid adaptations in breast cancer has attracted attention in the past decades. Prior studies have proposed as hallmark of tumorigenic cells alterations in the cholesterol content of cellular plasma membranes. In similar studies, the regulation of the cholesterol-synthesis pathway has been indicated to also affect cellular proliferation in breast tumours. [219] On the other hand, increased novo lipogenesis has been associated with accelerated membrane formation caused by fast proliferation representing a hallmark of metastatic or invasive processes.[179], [220]

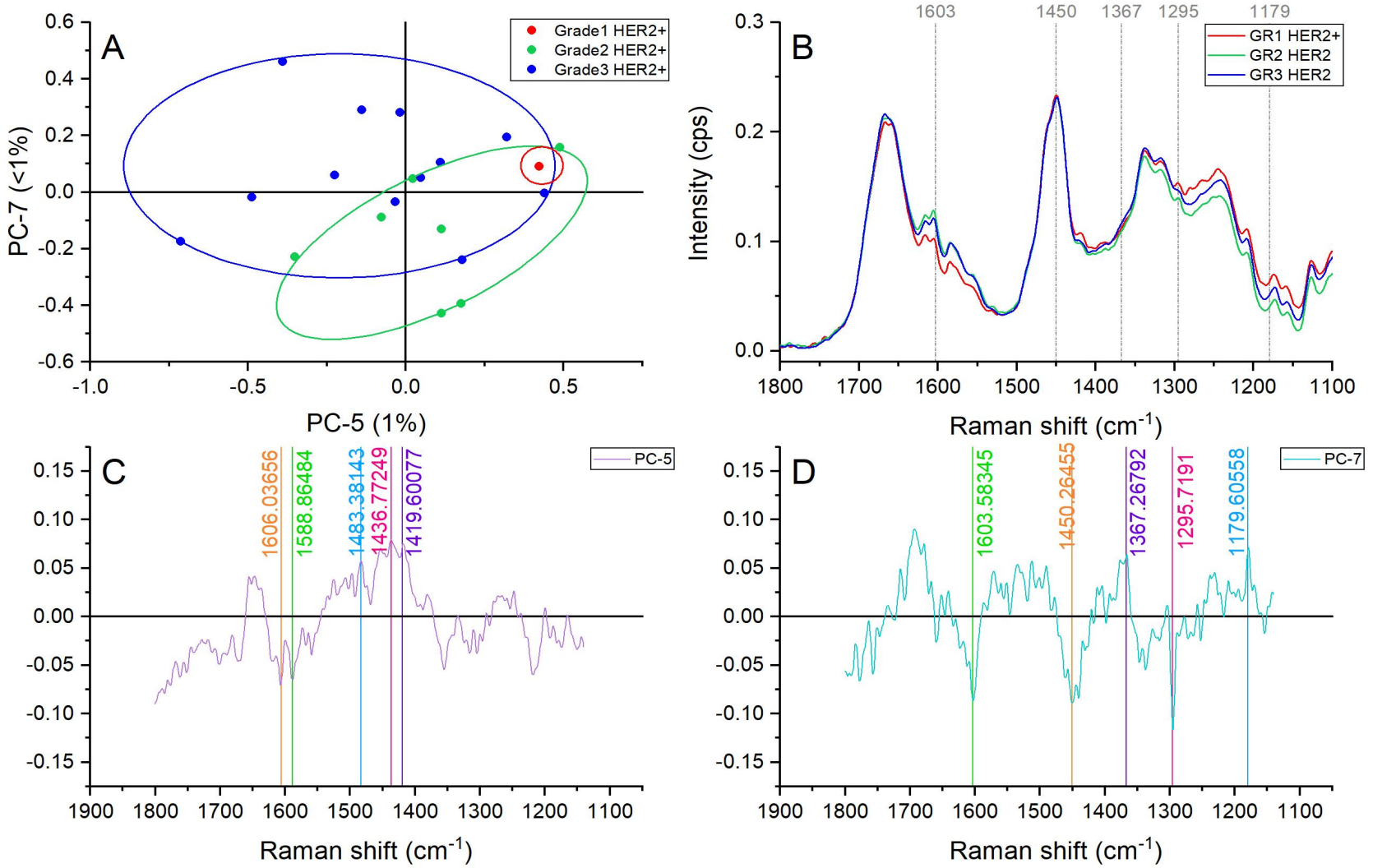

Figure 104. Principal component analysis of the amide region (1,800-1,140 cm^{-1}). (A) Score plot using PC-5 and PC-7 accounting for 2% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the amide region for HER2+ samples grade 1, 2, and 3. Dotted 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.

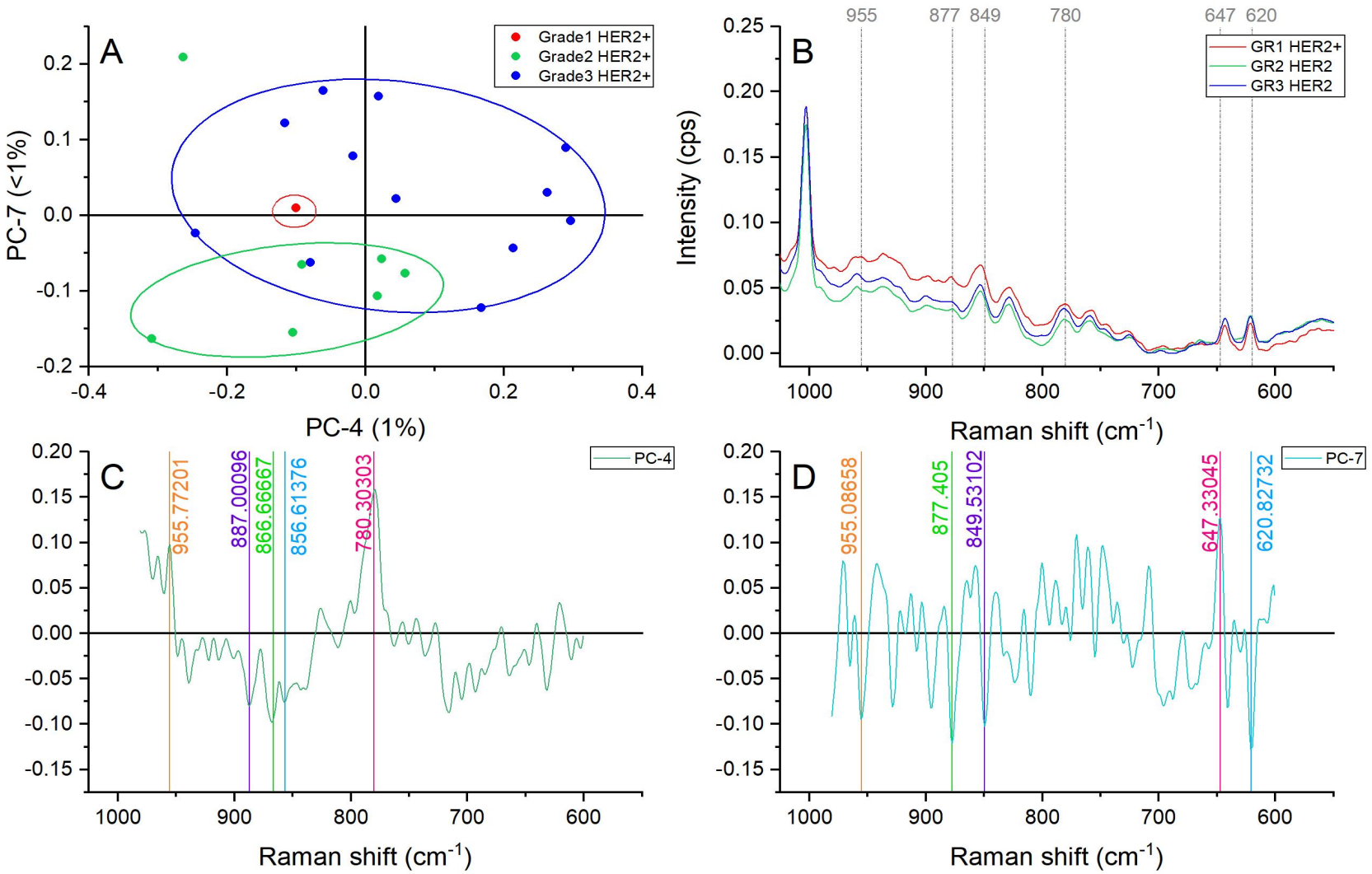

Figure 105. Principal component analysis of the amino acid and nucleic acid region (980-600 cm^{-1}). (A) Score plot using PC-4 and PC-7 accounting for 2% of the variance-ellipses were drawn subjectively based on visual trends, (B) Average spectral profile of the amino acid and nucleic acid region for HER2+ samples grade 1, 2, and 3. Dotted lines represent the peaks found as responsible for the separation based on the loadings presented in C&D, (C) PC-4 Loading, (D) PC7 loading.

Conclusions and future work

CHAPTER 8. CONCLUSIONS AND FUTURE WORK

Vibrational spectroscopy has proven to be an excellent technique for breast tissue analysis. The use of FTIR and Raman micro-spectroscopy on ex vivo breast samples provided excellent results containing important biochemical information. A combination of FTIR, Raman spectroscopy and chemometric methods was used for the first time in the study of African breast cancer in a relatively large cohort. This work contributes significantly to the field of cancer research by identifying key biochemical changes from normal breast to breast cancer as well as identifying chemical changes associated with different subtypes and grades of breast cancer.

The differentiation of normal breast and breast cancer was successfully achieved with both FTIR and Raman micro-spectroscopic techniques. The PCA model created with Raman spectroscopy presented a sensitivity of 90% and a specificity of 78%. The PCA model created with FTIR spectroscopy reached a sensitivity of 92% and specificity of 86%.

Although these numbers are promising and would provide good results in a screening test, in a definitive test at the time of surgery the presented prediction models will result in a 22% (Raman) and a 14% (FTIR) false positive rate. It is important to note that in a surgical context these false positives could result in women having unnecessary breast lumpectomies or mastectomies. As a consequence of removing healthy tissue there is the risk of tissue trauma and psychological impact. If these techniques are to be applied at the time of surgery, the protocols need to be optimised. For example, the incorporation of prediction protocols that consider analysing several points in odd numbers (n=2k+1) and the determination of the sample nature considering the majority of the predictions (k+1) could be used to improve the sensitivity of the presented models. This approach has been reported in protocols for glioma diagnosis in the UK. [221]

Progression of cancer comprises a wide variety of mechanisms where stroma, angiogenesis, inflammation, immune system, hormones, and xenobiotic play an important role. All these processes and mechanism will cause biochemical changes at tissue level. The importance of identifying the chemical pathway to the progression of the disease relies on the valuable predictive information which allows the clinician to take decisions to improve the treatment course and to give the patient a better chance against the disease. Hence, the biochemical changes identified in this research can facilitate the assessment of potential metastasis as well as the aggressiveness at different progression time-points. In clinical practice, tracking the progress of cancer supports several decisions, such as changes in therapy.

8. 1 COMPARISON OF NORMAL BREAST VS BREAST CANCER

This research established the lipid and protein contribution as main differentiators of normal breast and breast cancer. This study reported a dominance of lipids, specifically fatty acids, in the normal tissue Raman spectra. Whereas, protein dominated the malignant breast. The normal breast samples presented higher intensities of carotenoids, β-carotenoids, and cholesterol. In the breast cancer sections, the reduction of these components has been explained through lipid peroxidation. Complementarily, our FTIR results found lower content of lipids relative to the protein content for the cancer tissue, whereas the opposite trend was

true for the normal breast epithelia as confirmed by the Amide I/Lipids and Amide II/Lipids ratios.

Furthermore, FTIR identified transformations in the breast cancer lipids. These changes were proposed to be due to the increment of trans- lipid conformations in the breast cancer samples. In addition, a higher content of acyl chains in the cancerous area was identified, suggesting affectations on cell adhesion molecules in the cell membrane, as well as interactions between cell adhesion molecules and lipid components in the membrane.

Greater contributions of protein and nucleic acid bands were identified with Raman and FTIR spectroscopy in the cancerous samples in comparison with the normal breast samples. DNA methylation has been suggested as responsible for the reduction in the methylation ratio from normal to malignant tissue in the FTIR analysis.

The FTIR spectral profile of the breast cancer samples suggested a lower content of collagen relative to protein content. Both techniques indicated that the normal breast samples were dominated by the presence of collagen, while a higher content of proteins constituting the malignant epithelia were present in the cancerous samples.

When the FTIR results were compared with the Raman spectroscopy findings, the conformational differences in proteins were responsible for the grouping of normal breast samples. A lower intensity in α helix characteristic peaks confirmed conformational changes in the cancerous tissue. In addition, random coil vibrations presented higher intensities in the normal breast in comparison with the breast cancer samples. Amide I β-sheet conformations showed higher intensities in the Raman cancerous spectra. β-sheet cancer peaks in the FTIR breast cancer profile presented a higher intensity in comparison with the normal breast. The A_{1657}/A_{1635} ratio showed a higher relative quantity of α-helix conformation in the normal breast in comparison with cancerous tissue. β-sheet structural proteins were more present in cancer tissue. FTIR highlighted that structural changes formed an important part of breast cancer evolution.

Ceramide related peaks were only visible in the Raman breast cancer spectra. The ceramide peak served as an important biomarker due to its participation as a structural component of sphingolipids. Sphingolipids are highly relevant in carcinogenesis as they are linked to various cellular processes such as cancer development, progression, metastasis and resistance to therapy.

The heterogeneity of our breast cancer cohort was confirmed by the cluster analysis. This analysis was performed for the cancerous area and normal breast FTIR spectra. The infiltration of normal samples into clusters, formed mainly by cancer samples, confirmed the variability among cases. The cancerous samples included in this study represented different grades, subtypes, and as a result presented different characteristics that might have resembled the normal tissue.

8.2 Comparison of breast cancer grades and subtypes

The cancerous cohort represented a heterogeneous group of cases, therefore, a specific comparison based on grade and subtype provided a better insight of the biochemical changes associated with the cancer progression and evolution.

Raman and FTIR spectroscopy have been demonstrated as a functional evaluation and characterisation tool for breast cancer allowing qualitative analysis. Additionally, the understanding of positive (luminal) and negative (TNBC) hormonal receptors as well as HER2+ rich subtypes has been enriched using vibrational spectroscopy. These findings contribute in several ways to our understanding of breast cancer and provide a basis for analysis with vibrational spectroscopy. The work presented in this document supports future clinical implementations of these techniques.

8.2.1 Grade 1 breast cancer samples

The Raman analysis revealed a higher cholesterol ester contribution (≈2987 cm^{-1}) on the luminal and HER2+ samples than the TNBC samples. The intra-tumour accumulation of cholesteryl ester is highly relevant due to its close connection to cell proliferation. Luminal and HER2+ presented a higher order of lateral chain–chain interactions in lipids in comparison with TNBC, as indicated by the contribution of ≈2946 and ≈2883 cm^{-1}.

HER2+ Raman spectra presented a higher content of lipids and methylation processes, while the luminal grade 1 samples presented a split behaviour, and TNBC had the lowest contributions for these bands. Similar results were obtained with FTIR spectroscopy, however, the differences in intensity and shape of lipid peaks indicated varying content of fatty acids. These variations advocated the important role played by fatty acids in the early histological grades. A higher content of olefin =CH stretching was identified in HER2+ grade 1 than the TNBC and luminal grade 1 samples. In contrast with the Raman results, TNBC showed a higher overall contribution of lipids, and the grade 1 luminal and HER2+ samples presented a more dynamic trend.

Higher absorbance of proteins and nucleic acids vibrations was detected on the samples representing the luminal and TNBC subtypes using both techniques. Guanine and deformation of N-H in-plane in the FTIR spectra evidenced similarities between luminal and TNBC grade 1 samples. Whereas lipid peroxidation seemed to have an early active participation in TNBC in comparison with its positive counterpart (Luminal subtype). The similitude between the hormonal subtypes suggested a similar nucleic acid production at lower grades, and a slower overproduction of genetic material seemed to take place in the HER2+ grade 1 samples.

Structural changes in the proteins presented on grade 1 cancer cases were responsible for the discrimination of samples. The contribution of β-sheet structures of amide I in the FTIR data was similar on the luminal and TNBC. HER2+ showed less content of β-sheet structures. The spectral profiles of grade 1 luminal and TNBC showed wider amide I bands in comparison with the grade 1 HER2+ as a result of the low contribution of the β-sheet shoulder. In the same way, the Raman data revealed that luminal grade 1 spectra presented more intensity in the α-helical structure of proteins.

Ceramides play an important role in the breast cancer Raman profile of TNBC. The ceramide backbone vibrations provided higher contributions in the TNBC, than the luminal subtype and

HER2+ samples. In addition, TNBC grade 1 samples showed the highest contributions of nucleic acids and collagen.

8.2.2 Grade 2 breast cancer samples

TNBC grade 2 samples presented more intense Raman peaks at ≈1333, and 1313 cm^{-1} which were caused by nucleic acids and phosphates vibrations related to phospholipids and the twisting mode of collagen and lipids. Vibrations related to lipids, cholesterol and creatine were identified in the FTIR data presenting higher intensities in the luminal and TNBC subtype, whereas the HER2+ subtype samples were all located in the lower than average area. HER2+ data also presented lower contributions of lipids in the Raman analysis. Thus, more active lipid metabolism takes place in the hormonal subtypes in grade 2 samples in comparison with the HER2+ subtype.

The C-H deformation and C-O stretching associated with glycogen presented higher FTIR intensities in the TNBC and lower intensities in luminal and HER2+ grade 2 samples. The higher contribution of glycogen on the TNBC grade 2 served as an indicator of accelerated proliferation suggesting a more aggressive behaviour, which is characteristic of the TNBC subtype. In addition, the TNBC samples showed higher lipid methylation (CH_3) while HER2+ samples showed a more standard biochemical behaviour. The luminal and TNBC spectra intensities suggested a more active lipid metabolism in comparison with the HER2+ subtype.

Luminal and TNBC subtypes had higher protein production than HER2+ at grade 2 as indicated by the results of both techniques. The Raman analysis showed that HER2+ presented the lowest content of collagen among the grade 2 samples. Additionally, variations in the C-H stretching peak located at ≈2948 cm^{-1} were higher in the luminal and TNBC reflecting more biochemical activity on these subtypes in comparison with HER2+ cases.

Overproduction of genetic material in the luminal and HER2+ grade, and changes in the cell membrane fluidity, were suggested by the PO_2^- asymmetric stretching in nucleic acids and cholesterol ester FTIR vibrations. The significant participation of lipids and phospholipids in luminal cases was also confirmed with Raman spectroscopy. The protein-related Raman contributions suggested that TNBC grade 2 cases focus more on the overproduction of structural material to promote proliferation than the other subtypes.

Both spectroscopic analyses revealed conformational changes as part of the subtype differentiation. HER2+ grade 2 samples had greater contributions on the ≈1629 cm^{-1} Raman band suggesting dominant β-form structural conformations. HER2+ rich samples presented higher intensities of β-sheet FTIR bands indicating more of this secondary structure in comparison with the other subtypes.

TNBC grade 2 cases focused on the overproduction of genetic material to promote proliferation as confirmed by both techniques, whereas luminal grade 2 samples presented considerable structural changes in their proteins. This change translated to the rearrangement of hydrogen bonds within proteins resulting in structural changes in luminal grade 2 samples.

8.2.3 Grade 3 breast cancer samples

TNBC presented the highest Raman lipid contributions, followed by HER2+ and luminal respectively. Acyl chains and fatty acids also presented increased intensities in the TNBC grade 3 spectra. The same behaviour was confirmed when the grade 3 samples were analysed with FTIR, and lipid-related bands (≈3023, 2971, 2896, and 2896 cm^{-1}) presented a higher contribution in the TNBC than HER2+ and Luminal. Highly active lipid metabolism in TNBC grade 3 samples was proposed. The accumulation of fatty acids in TNBC suggested aggressive malignant activities that needed to be fuelled through these lipid components.

The Raman results proposed the accumulation of genetic and structural (proteins) material in the luminal subtype, while the TNBC subtype focused not only on this, but also on lipid transformations and collagen overproduction. The suggested lipid alterations included variations in phospholipids and carotenoids. The FTIR data suggested that TNBC had lower contents of nucleic acids in comparison with the positive hormonal receptor subtype. However, the Cytosine (NH_2), and C=N stretching presented in FTIR spectroscopy confirmed the overproduction of genetic content for replication and accelerated proliferation in TNBC.

Luminal grade 3 Raman spectra showed greater contributions of the C=C bending mode of phenylalanine. TNBC spectra presented greater contributions of amide III vibrations. The identified peaks suggested an accumulation of genetic material and ECM production, confirming a higher metabolism. HER2+ grade 3 Raman spectral data had higher intensities in the peaks associated with: aminoacid side chain vibrations in collagen, out-of-plane ring breathing in proteins, phosphodiester in nucleic acids, breathing of tryptophan and ring mode in DNA and RNA bases, and C-S stretching and C-C twisting in proteins. Active proliferation triggered by protein and nucleic acid content and ECM activation was proposed as part of the HER2+ cancer progression.

Luminal grade 3 samples presented structural changes suggesting adaptation towards β-sheet conformations according to the FTIR peak located at ≈1636 cm^{-1}. HER2+ grade 3 samples presented the lowest contribution of this secondary structure, while TNBC showed no particular preference. Thus, a reduction of the α-helix structure and increase in β-sheet structure was assumed in the luminal cases. Luminal grade 3 samples presented consistent important contributions in the C=N guanine FTIR vibration, followed by TNBC and finally HER2+ samples. These particular differences hinted towards guanine conformation changes, which disturb the reparative mechanisms of the cell and tissue.

8.2.4 Luminal breast cancer samples

Grade 1 and 2 luminal samples presented an increased intensity in the Raman bands representing CH_2 asymmetric stretching of lipids and fatty acids in comparison with grade 3 samples. Grade 2 spectra showed greater contributions of CH_2 deformation, scissoring in lipids, CH_2 deformation, and ceramide vibrations than grade 1 and grade 3 samples. These results were similar to those obtained with FTIR, where C=C cis conformation of fatty acid peaks had a greater contribution in grade 2 samples. Likewise, asymmetric stretching vibrations of CH_2 of acyl chains, CH_3 symmetric stretching, and stretching vibrations of CH_2 and CH_3 of phospholipids, cholesterol, and creatine, showed higher intensities and contributions in grade 2 cases.

By contrast, grade 3 luminal Raman spectra showed greater contribution of cholesterol ester and CH vibration in proteins. Breast tumours rich in cholesterol ester were associated with a higher histologic grade in addition to being linked to proliferation and aggressive potential. Grade 3 luminal cases also presented a higher contribution of the amide A, amide I, C=N, and NH_2 from adenine bands in the FTIR analysis than grade 2 and 1. These vibrations suggested that a higher content of proteins and nucleic acids were present in grade 3 samples, probably representing the aberration and poor differentiation associated with higher grades.

Grade 1 and 3 samples presented more intense Raman peaks than grade 2. However, grade 2 luminal samples showed higher and lower intensities than the average suggesting variable behaviour. The peak located at ≈1650 cm^{-1}assigned to amide I with random coils conformations suggested conformational changes in proteins as the histological grade increased.

Greater contributions of amide I with α-helix configuration, phenylalanine, hydroxyproline, and amide III vibrations were identified in grade 3 with Raman spectroscopy. This behaviour suggested that higher luminal breast cancer grades focused on cell proliferation by generating cellular components and stimulating the ECM. Additionally, grade 1 and grade 3 samples showed collagen related peaks. Thus, the active production of fibroblast products served as an indication of cancer progression through poor differentiation of the cancer cells.

FTIR spectroscopy detected the methylation of nucleotides in grade 1 luminal cases. Methylation modifies the genetic activity. Literature confirmed significant methylation in ER positive cell lines.

8.2.5 TNBC BREAST CANCER SAMPLES

Asymmetric stretching of CH_2 in lipids, CH_2 asymmetric stretching of lipids and proteins, and CH_3 symmetric stretching of lipids presented greater Raman contribution in grade 2 samples than grade 1 and 3 in the TNBC subtype. Therefore, a higher histological grade was suggested to have a lower content of lipids. FTIR results confirmed this, with intensities of C=C vibrations in fatty acids decreasing in higher grades.

Grade 1 samples kept intensities closer to the average in the FTIR spectra suggesting steady protein and fatty acid contents in early grades of TNBC. The analysis of the amide region revealed important overproduction of nucleic acids in grades 1 and 2 of the TNBC samples. The overproduction of this material was characterised by adenine vibrations in DNA.

Grade 3 samples showed higher FTIR intensities for CH_2 asymmetric bending and COO^- stretching presented in proteins and fatty acids than grade 2 and 1 TNBC samples, suggesting accumulation of structural cell contents (proteins) and cell membranes (fatty acids). The same trend was identified with the band located at ≈1236 cm^{-1} associated with phosphate I (PO_2^-) asymmetric stretching, mainly from nucleic acids. Greater contributions of the Raman bands representing CH_3 symmetric stretching (proteins), amide I α-helix conformation, amide III, and a combination of amide III and CH_2 wagging vibrations (nucleic acids), glycine backbones, and proline chains were detected in grade 3 samples, in comparison with grade 2 and grade 1. Furthermore, in the grade 3 samples a lower content of collagen was identified using the

Raman collagen characteristic peaks (phenylalanine and hydroxyproline vibrations) in comparison with the lower histological grades.

Grade 2 and 1 TNBC cases presented more β-sheet conformations of amide II in proteins (≈1534 cm^{-1}) in comparison with grade 3 samples. Thus, structural changes in proteins were proposed in initial grades of the TNBC cases to facilitate invasive processes.

8.2.6 HER2+ BREAST CANCER SAMPLES

HER2+ grade 1 presented higher asymmetric Raman vibrations and CH_3 content coming from lipids, especially fatty acids. This fact correlated with our FTIR results, and can be explained by a less active lipid transformation, peroxidation, degradation and active novo lipogenesis, in lower grades.

Higher contributions of CH Raman vibrations in proteins (≈2916) and FTIR vibrations (N-H symmetric stretching, N-H amide B stretching, and CH_3 symmetric stretching mainly from proteins and lipids) were present in grade 3 in comparison to grade 2 and 1, suggesting higher cell density and metabolism in higher grades. Grade 3 HER2+ samples had more significant contributions from vibrations associated with nucleic acids in both techniques, than grade 2 and grade 1 samples. Additionally, lipid related bands located at ≈2993 cm^{-1} (cholesterol ester), and ≈2880 and 2847 cm^{-1} (CH_2 symmetric and asymmetric stretch in lipids) presented increased intensity in grade 2 samples.

Grade 1 and 2 showed high FTIR contributions of phosphate and important lipid vibrations (wagging of phospholipids, fatty acids, triglycerides). Moreover, Grade 2 samples showed more ceramide deformation (≈1295 cm^{-1}) in comparison with grade 1 and 3.

As the histological grade increased, the contribution of the Raman bands located at ≈939 and 850 cm^{-1}, identified as collagen characteristic peaks, increased. Therefore, early histological grades involved collagen and ECM production, while at higher and intermediate grades the proliferation of cellular material played the main role.

8.3 FUTURE WORK

This research listed and revised substantial changes caused by breast cancer and their different grades and subtypes using vibrational spectroscopy. The presented data highlights the importance of FTIR and Raman spectroscopy and their outstanding clinical potential. It is clear that more work is required to improve the technology and facilitate its clinical use and data processing. To accomplish this, multidisciplinary collaboration between researchers, companies, and clinicians is crucial.

This research will be of significant value to new researchers exploring the use of vibrational spectroscopy to study breast cancer and help in understanding the chemical pathway to the progression of disease. It would be interesting to carry these findings further and focus some research on improving the sensitivity and specificity of the prediction model. This could be achieved using machine learning and taking advantage of the big cohort data. The use of artificial intelligence and machine learning can represent an import tool in the spectroscopic

analysis of vibrational data. The combination of vibrational spectroscopy and these strategies and algorithms can offer an effective and low-cost solution to healthcare and diagnostics.

Machine learning approaches consist of a combination of several classification systems. The use of several classifiers allows improving the accuracy of the classification. These algorithms are designed to maximise the number of accurate classification. Ideally, artificial intelligence (AI) and machine learning algorithms should be able to adapt to new data without retraining. AI and machine learning have been widely used in different fields using vibrational spectroscopy data, for example, the analysis of ovarian and breast cancer tissue, cervical cytology, food analysis, drugs concentration, and soil analysis. [125], [222]–[225]

The use of self-learning classifier would be an ideal next step for this research. In general terms, in the training stage, a model is created containing several rules. These rules encompass relevant information that is characteristic of specific classes, for example, the relevant database of the normal and malignant breast tissue biochemical behaviour identified in this work. When new data is included in this model, it should evolve and produce more accurate classification rates, adapting and improving the robustness of the final model.

Additionally, once the new and improved algorithm has been improved, the new classification algorithm could be tested in different fixated samples. These systems will then provide the opportunity of immediate feedback to clinicians and pathologists, significantly improving patient care. This will help in increasing confidence among the clinicians and rapidly bringing these techniques to the armoury of surgeons, oncologists, and pathologists.

Vibrational spectroscopy has the potential to offer a real-time analysis in preserved breast samples with excellent accuracy and is capable of maintaining this accuracy after several years. Moreover, FTIR and Raman spectroscopy can provide practitioners and researchers with chemical information reflecting changes at tissue level. In clinical practice, these changes influence the treatment strategies. In practical research, these insights help to understand carcinogenesis and invasive progressions, accelerating the development of new treatments.

References

CHAPTER 9. REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA. Cancer J. Clin.*, 2018.

[2] M. A. Knowles, *Introduction to the cellular and molecular biology of cancer*, 4th ed. New York: New York : Oxford University Press, 2005, 2005.

[3] Y. Wen, D. Zhang, H. Liu, F. Wang, and Y. Zhang, "Heterogeneity in Breast cancer," *Cancer Genet. Epigenetics*, vol. 3, 2015.

[4] A. Marusyk and K. Polyak, "Tumor heterogeneity: causes and consequences," *Biochim. Biophys. Acta (BBA)-Reviews Cancer*, vol. 1805, no. 1, pp. 105–117, 2010.

[5] K. Polyak, "Heterogeneity in breast cancer," *J. Clin. Invest.*, vol. 121, no. 10, p. 3786, 2011.

[6] N. R. Bertos and M. Park, "Breast cancer—one term, many entities?," *J. Clin. Invest.*, vol. 121, no. 10, pp. 3789–3796, 2011.

[7] J. S. Ross and G. N. Hortobagyi, *Molecular oncology of breast cancer*. Sudbury, Mass.: Sudbury, Mass. : Jones and Bartlett Publishers, 2005, 2005.

[8] P. P. Rosen, *Rosen's breast pathology*. Lippincott Williams & Wilkins, 2001.

[9] B. Tiede and Y. Kang, "From milk to malignancy: the role of mammary stem cells in development, pregnancy and breast cancer," *Cell Res.*, vol. 21, no. 2, pp. 245–257, 2011.

[10] D. J. Dabbs, *Breast pathology*. Elsevier Health Sciences, 2012.

[11] J. Russo and I. H. Russo, "Histological Evaluation of the Normal Breast," in *Techniques and Methodological Approaches in Breast Cancer Research*, New York, NY: Springer New York, 2014, pp. 45–73.

[12] K. Saladin, *Human Anatomy*. New York: McGraw-Hill, 2012.

[13] D. Forman, "Cancer incidence and survival by major ethnic group, England, 2002–2006," *London Natl. Cancer Intell. Netw.*, 2009.

[14] A. Januszewski, N. Tanna, and J. Stebbing, "Ethnic variation in breast cancer incidence and outcomes—the debate continues," *Br. J. Cancer*, vol. 110, no. 1, p. 4, 2014.

[15] S. Farooq and M. P. Coleman, "Breast cancer survival in South Asian women in England and Wales," *J Epidemiol Community Heal.*, vol. 59, no. 5, p. 402, 2005.

[16] C. A. Clarke, D. W. West, B. K. Edwards, L. W. Figgs, J. Kerner, and A. G. Schwartz, "Existing data on breast cancer in African-American women," *Cancer*, vol. 97, no. S1, pp. 211–221, 2003.

[17] C. I. Li, K. E. Malone, and J. R. Daling, "Differences in breast cancer stage, treatment, and survival by race and ethnicity," *Arch. Intern. Med.*, vol. 163, no. 1, p. 49, 2003.

[18] A. Daramola, A. Banjo, F. Abdulkareem, and A. Shaaban, "Breast cancer reporting in Lagos, implications for training and education in Africa," *Cancer Res.*, vol. 73, 2013.

[19] F. Dewaard and D. Trichopoulos, “A UNIFYING CONCEPT OF THE ETIOLOGY OF BREAST-CANCER,” *Int. J. Cancer*, vol. 41, no. 5, pp. 666–669, 1988.

[20] B. Macmahon, P. Cole, and J. Brown, “ETIOLOGY OF HUMAN BREAST-CANCER - REVIEW,” *J. Natl. Cancer Inst.*, vol. 50, no. 1, pp. 21–42, 1973.

[21] M. D. Althuis, J. H. Fergenbaum, M. Garcia-Closas, L. A. Brinton, M. P. Madigan, and M. E. Sherman, “Etiology of hormone receptor-defined breast cancer: A systematic review of’the literature,” *Cancer Epidemiol. Biomarkers Prev.*, vol. 13, no. 10, pp. 1558–1568, 2004.

[22] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, “Estimating the global incidence and mortality in 2018: GLOBOCAN sources and methods.,” 2018.

[23] P. K. Ravert and C. Huffaker, “Breast cancer screening in women: An integrative literature review,” *J. Am. Acad. Nurse Pract.*, vol. 22, no. 12, pp. 668–673, 2010.

[24] L. Witek-Janusek, S. Gabram, and H. L. Mathews, “Psychologic stress, reduced NK cell activity, and cytokine dysregulation in women experiencing diagnostic breast biopsy,” *Psychoneuroendocrinology*, vol. 32, no. 1, pp. 22–35, Jan. 2007.

[25] G. Thomas *et al.*, “Evaluating feasibility of an automated 3-dimensional scanner using Raman spectroscopy for intraoperative breast margin assessment,” *Sci. Rep.*, vol. 7, no. 1, p. 13548, 2017.

[26] S. R. Lakhani *et al.*, “WHO Classification of Tumours of the Breast. 2012,” *Lyon IARC*, vol. 4, 2012.

[27] I. Ellis *et al.*, “Pathology Reporting of Breast Disease in Surgical Excision Specimens Incorporating the Dataset for Histological Reporting of Breast Cancer, The Royal Collage of Pathologists, London (2016),” 2016.

[28] J. P. Thiery, “Epithelial- mesenchymal transitions in tumour progression,” *Nat. Rev. Cancer*, vol. 2, no. 6, p. 442, 2002.

[29] M. Alizart, J. Saunus, M. Cummings, and S. R. Lakhani, “Molecular classification of breast carcinoma,” *Diagnostic Histopathol.*, vol. 18, no. 3, pp. 97–103, 2012.

[30] H. Kennecke *et al.*, “Metastatic behavior of breast cancer subtypes,” *J. Clin. Oncol.*, vol. 28, no. 20, pp. 3271–3277, 2010.

[31] C. M. Perou *et al.*, “Molecular portraits of human breast tumours,” *Nature*, vol. 406, no. 6797, pp. 747–752, 2000.

[32] J. D. Brenton, L. A. Carey, A. A. Ahmed, and C. Caldas, “Molecular classification and molecular forecasting of breast cancer: Ready for clinical application?,” *J. Clin. Oncol.*, vol. 23, no. 29, pp. 7350–7360, 2005.

[33] L. Pusztai, C. Mazouni, K. Anderson, Y. Wu, and W. F. Symmans, “Molecular classification of breast cancer: Limitations and potential,” *Oncologist*, vol. 11, no. 8, pp. 868–877, 2006.

[34] M. Guedj *et al.*, “A refined molecular taxonomy of breast cancer,” *Oncogene*, vol. 31, no. 9, pp. 1196–1206, 2012.

[35] C. Curtis *et al.*, “The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups,” *Nature*, vol. 486, no. 7403, pp. 346–352, Jun. 2012.

[36] N. National institute for Health and Care Excellence, “Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast,” 2018.

[37] M. C. U. Cheang *et al.*, “Ki67 Index, HER2 Status, and Prognosis of Patients With Luminal B Breast Cancer,” *JNCI J. Natl. Cancer Inst.*, vol. 101, no. 10, pp. 736–750, May 2009.

[38] EARLY BREAST CANCER TRIALISTS’ COLLABORATIVE GROUP, “Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy: 133 randomised trials involving 31 000 recurrences and 24 000 deaths among 75 000 women,” *Lancet*, vol. 339, no. 8784, pp. 1–15, Jan. 1992.

[39] C. Liedtke *et al.*, “Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer,” *J. Clin. Oncol.*, vol. 26, no. 8, pp. 1275–1281, 2008.

[40] E. A. Rakha, M. E. El-Sayed, A. R. Green, A. H. S. Lee, J. F. Robertson, and I. O. Ellis, “Prognostic markers in triple-negative breast cancer,” *Cancer*, vol. 109, no. 1, pp. 25–32, Jan. 2007.

[41] R. D. Chacón and M. V Costanzo, “Triple-negative breast cancer,” *Breast Cancer Res.*, vol. 12, no. S2, p. S3, Oct. 2010.

[42] B. D. Lehmann *et al.*, “Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies,” *J. Clin. Invest.*, vol. 121, no. 7, p. 2750, 2011.

[43] W. J. Irvin and L. A. Carey, “What is triple-negative breast cancer?,” *Eur. J. Cancer*, vol. 44, no. 18, pp. 2799–2805, 2008.

[44] W. D. Foulkes, I. E. Smith, and J. S. Reis-Filho, “Triple-Negative Breast Cancer,” *N. Engl. J. Med.*, vol. 363, no. 20, pp. 1938–1948, 2010.

[45] E. Rakha, I. Ellis, and J. Reis-Filho, “Are triple- negative and basal- like breast cancer synonymous? [1],” *Clin. Cancer Res.*, vol. 14, no. 2, p. 618, 2008.

[46] S. Badve *et al.*, “Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists,” *Mod. Pathol.*, vol. 24, no. 2, pp. 157–167, 2011.

[47] R. Schmadeka, B. E. Harmon, and M. Singh, “Triple- negative breast carcinoma: current and emerging concepts,” *Am. J. Clin. Pathol.*, vol. 141, no. 4, p. 462, 2014.

[48] J. Pollard *et al.*, “Differential expression of microRNAs in breast cancers from four different ethnicities,” *Pathobiology*, vol. 85, no. 4, pp. 220–226, 2018.

[49] A. Sharma-Oates, A. M. Shaaban, I. Tomlinson, L. Wynne, J.-B. Cazier, and S. Sundar, “Heterogeneity of germline variants in high risk breast and ovarian cancer susceptibility genes in India,” *Precis. Clin. Med.*, vol. 1, no. 2, pp. 75–87, 2018.

[50] N. A. Titloye *et al.*, “Histological Features and Tissue Microarray Taxonomy of Nigerian Breast Cancer Reveal Predominance of the High-Grade Triple-Negative Phenotype,” *Pathobiology*, vol. 83, no. 1, pp. 24–32, 2016.

[51] D. Huo *et al.*, “Population differences in breast cancer: survey in indigenous African women reveals over-representation of triple-negative breast cancer,” *J. Clin. Oncol.*, vol. 27, no. 27, pp. 4515–4521, 2009.

[52] A. J. Agboola *et al.*, “Molecular characteristics and prognostic features of breast cancer in

Nigerian compared with UK women," *Breast Cancer Res. Treat.*, vol. 135, no. 2, pp. 555–569, 2012.

[53] A. Stark *et al.*, "African ancestry and higher prevalence of triple-negative breast cancer," *Cancer*, vol. 116, no. 21, pp. 4926–4932, 2010.

[54] R. C. Millikan *et al.*, "Epidemiology of basal-like breast cancer," *Breast Cancer Res. Treat.*, vol. 109, no. 1, pp. 123–139, 2008.

[55] H. Y. Chang *et al.*, "Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 10, pp. 3738–3743, 2005.

[56] S. Cleator, W. Heller, and R. C. Coombes, "Triple-negative breast cancer: therapeutic options," *Lancet Oncol.*, vol. 8, no. 3, pp. 235–244, Mar. 2007.

[57] D. P. Atchley *et al.*, "Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer.," *J. Clin. Oncol.*, vol. 26, no. 26, pp. 4282–8, Sep. 2008.

[58] N. MCCARTHY, G. MITCHELL, M. BILOUS, N. WILCKEN, and G. J. LINDEMAN, "Triple-negative breast cancer: making the most of a misnomer," *Asia. Pac. J. Clin. Oncol.*, vol. 8, no. 2, pp. 145–155, Jun. 2012.

[59] D. Zardavas, A. Irrthum, C. Swanton, and M. Piccart, "Clinical management of breast cancer heterogeneity," *Nat. Rev. Clin. Oncol.*, vol. 12, no. 7, p. 381, 2015.

[60] G. R. Oxnard *et al.*, "When progressive disease does not mean treatment failure: reconsidering the criteria for progression," *J. Natl. Cancer Inst.*, vol. 104, no. 20, pp. 1534–1541, 2012.

[61] C. J. J. Conti, "Mechanisms of Tumor Progression," in *Comprehensive Toxicology*, Elsevier, 2010, pp. 335–347.

[62] M. A. Cortez, C. Ivan, P. Zhou, X. Wu, M. Ivan, and G. A. Calin, "microRNAs in cancer: from bench to bedside," in *Advances in cancer research*, vol. 108, Elsevier, 2010, pp. 113–157.

[63] K. H. Brettingham-Moore and P. C. Taberlay, "Cancer Epigenetics," in *Drug Discovery in Cancer Epigenetics*, Elsevier, 2016, pp. 41–59.

[64] R. Y. Huang and D. A. Reardon, "Imaging Studies in Immunotherapy," in *Translational Immunotherapy of Brain Tumors*, Elsevier, 2017, pp. 149–179.

[65] H. Korkaya, A. Davis, and M. S. Wicha, "10 – Cancer Stem Cells and the Microenvironment," *Mol. Basis Cancer*, pp. 157-164.e3, 2015.

[66] J. Yokota, "Tumor progression and metastasis," *Carcinogenesis*, vol. 21, no. 3, pp. 497–503, 2000.

[67] K. Strimbu and J. A. Tavel, "What are biomarkers?," *Curr. Opin. HIV AIDS*, vol. 5, no. 6, pp. 463–6, Nov. 2010.

[68] J. S. Ross *et al.*, "Breast cancer biomarkers and molecular medicine," *Expert Rev. Mol. Diagn.*, vol. 3, no. 5, pp. 573–585, Sep. 2003.

[69] S. M. Hanash, S. J. Pitteri, and V. M. Faca, "Mining the plasma proteome for cancer biomarkers," *Nature*, vol. 452, no. 7187, pp. 571–579, Apr. 2008.

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

[71] G. Gauglitz and D. S. Moore, *Handbook of spectroscopy [electronic resource]*, Second, Co. Weinheim, Germany : Wiley-VCH, 2014, 2014.

[72] C. Kendall *et al.*, "Vibrational spectroscopy: a clinical tool for cancer diagnostics," *Analyst*, vol. 134, no. 6, pp. 1029–1045, 2009.

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

[74] A. C. S. Talari, C. A. Evans, I. Holen, R. E. Coleman, and I. U. Rehman, "Raman spectroscopic analysis differentiates between breast cancer cell lines," *J. Raman Spectrosc.*, 2015.

[75] B. Stuart, *Infrared spectroscopy : fundamentals and applications*. J. Wiley, 2004.

[76] D. I. Ellis and R. Goodacre, "Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy," *Analyst*, vol. 131, no. 8, pp. 875–885, 2006.

[77] R. Manoharan, Y. Wang, and M. S. Feld, "Histochemical analysis of biological tissues using Raman spectroscopy," *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 52, no. 2, pp. 215–249, 1996.

[78] C. J. Frank, D. C. B. Redd, T. S. Gansler, and R. L. McCreery, "Characterization of human breast biopsy specimens with near-IR Raman spectroscopy," *Anal. Chem.*, vol. 66, no. 3, pp. 319–326, Feb. 1994.

[79] M. Marro, C. Nieva, R. Sanz-Pamplona, and A. Sierra, "Molecular monitoring of epithelial-to-mesenchymal transition in breast cancer cells by means of Raman spectroscopy," *Biochim. Biophys. Acta - Mol. Cell Res.*, vol. 1843, no. 9, pp. 1785–1795, 2014.

[80] R. Manoharan *et al.*, "Raman spectroscopy and fluorescence photon migration for breast cancer diagnosis and imaging," *Photochem. Photobiol.*, vol. 67, no. 1, pp. 15–22, 1998.

[81] M. E. Monaco, "Fatty acid metabolism in breast cancer subtypes," *Oncotarget*, vol. 8, no. 17, p. 29487, 2017.

[82] W. B. Kinlaw, P. W. Baures, L. E. Lupien, W. L. Davis, and N. B. Kuemmerle, "Fatty acids and breast cancer: make them on site or have them delivered," *J. Cell. Physiol.*, vol. 231, no. 10, pp. 2128–2141, 2016.

[83] H. Abramczyk, J. Surmacki, B. Brożek-Płuska, Z. Morawiec, and M. Tazbir, "The hallmarks of breast cancer by Raman spectroscopy," *J. Mol. Struct.*, vol. 924, pp. 175–182, 2009.

[84] R. E. Kast *et al.*, "Raman spectroscopy can differentiate malignant tumors from normal breast tissue and detect early neoplastic changes in a mouse model," *Biopolymers*, vol. 89, no. 3, pp. 235–241, 2008.

[85] C. J. Frank, R. L. McCreery, and D. C. B. Redd, "Raman spectroscopy of normal and diseased human breast tissues," *Anal. Chem.*, vol. 67, no. 5, pp. 777–783, 1995.

[86] A. S. Haka *et al.*, "Diagnosing breast cancer using Raman spectroscopy: prospective analysis," *J. Biomed. Opt.*, vol. 14, no. 5, p. 54023, 2009.

[87] H. Abramczyk, B. Brozek-Pluska, J. Surmacki, J. Jablonska, and R. Kordek, "The label-free Raman

imaging of human breast cancer," *J. Mol. Liq.*, vol. 164, no. 1, pp. 123–131, 2011.

[88] B. Brożek-Płuska, I. Placek, K. Kurczewski, Z. Morawiec, M. Tazbir, and H. Abramczyk, "Breast cancer diagnostics by Raman spectroscopy," *J. Mol. Liq.*, vol. 141, no. 3, pp. 145–148, 2008.

[89] S. Rehman *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.*, vol. 38, no. 10, pp. 1345–1351, 2007.

[90] M. V. P. Chowdary, K. K. Kumar, J. Kurien, S. Mathew, and C. M. Krishna, "Discrimination of normal, benign, and malignant breast tissues by Raman spectroscopy," *Biopolymers*, vol. 83, no. 5, pp. 556–569, 2006.

[91] Y. Yamashita, S. Nishiumi, S. Kono, S. Takao, T. Azuma, and M. Yoshida, "Differences in elongation of very long chain fatty acids and fatty acid metabolism between triple-negative and hormone receptor-positive breast cancer," *BMC Cancer*, vol. 17, no. 1, p. 589, 2017.

[92] M. P. Ogrodzinski, J. J. Bernard, and S. Y. Lunt, "Deciphering metabolic rewiring in breast cancer subtypes," *Transl. Res.*, 2017.

[93] S. Beloribi-Djefaflia, S. Vasseur, and F. Guillaumond, "Lipid metabolic reprogramming in cancer cells," *Oncogenesis*, vol. 5, no. 1, p. e189, 2016.

[94] J. A. Menendez and R. Lupu, "Fatty acid synthase regulates estrogen receptor-α signaling in breast cancer cells," *Oncogenesis*, vol. 6, no. 2, p. e299, 2017.

[95] H. Abramczyk and B. Brozek-Pluska, "New look inside human breast ducts with Raman imaging. Raman candidates as diagnostic markers for breast cancer prognosis: Mammaglobin, palmitic acid and sphingomyelin," *Anal. Chim. Acta*, vol. 909, pp. 91–100, 2016.

[96] S. K. Paidi, A. Rizwan, C. Zheng, M. Cheng, K. Glunde, and I. Barman, "Label-Free Raman Spectroscopy Detects Stromal Adaptations in Premetastatic Lungs Primed by Breast Cancer," *Cancer Res.*, vol. 77, no. 2, pp. 247–256, 2017.

[97] K. E. Shafer-Peltier, A. S. Haka, J. T. Motz, M. Fitzmaurice, R. R. Dasari, and M. S. Feld, "Model-based biological Raman spectral imaging," *J. Cell. Biochem.*, vol. 87, no. S39, pp. 125–137, 2002.

[98] A. S. Haka *et al.*, "Diagnosing Breast Cancer by Using Raman Spectroscopy," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 35, pp. 12371–12376, 2005.

[99] M. M. Mariani, L. Maccoux, C. Matthaus, M. Diem, J. Hengstler, and V. Deckert, "Micro-Raman Detection of Nuclear Membrane Lipid Fluctuations in Senescent Epithelial Breast Cancer Cells," *Anal. Chem.*, vol. 82, no. 10, pp. 4259–4263, 2010.

[100] X. F. Ling *et al.*, "FTIR spectroscopic explorations of clinical practice of breast cancer," *Spectrosc. Spectr. Anal.*, vol. 25, no. 2, pp. 198–200, 2005.

[101] S. Zhou *et al.*, "FTIR spectroscopic characterization of freshly removed breast cancer tissues," *Zhonghua Zhong Liu Za Zhi*, vol. 28, no. 7, pp. 512–4, Jul. 2006.

[102] R. Eckel, H. Huo, H.-W. Guan, X. Hu, X. Che, and W.-D. Huang, "Characteristic infrared spectroscopic patterns in the protein bands of human breast cancer tissue," *Vib. Spectrosc.*, vol. 27, no. 2, pp. 165–173, 2001.

[103] Y. Ci, T. Gao, J. Dong, X. Kan, and Z. Guo, "FTIR assessment of the secondary structure of proteins in human breast benign and malignant tissues," *Chinese Sci. Bull.*, vol. 44, no. 24, pp.

2215–2221, Dec. 1999.

[104] W. M. Elshemey, A. M. Ismail, and N. S. Elbialy, “Molecular-Level Characterization of Normal, Benign, and Malignant Breast Tissues Using FTIR Spectroscopy,” *J. Med. Biol. Eng.*, vol. 36, no. 3, pp. 369–378, Jun. 2016.

[105] T. Gao, J. Feng, and Y. Ci, “Human breast carcinomal tissues display distinctive FTIR spectra: implication for the histological characterization of carcinomas,” 1999.

[106] S. Rehman, Z. Movasaghi, J. A. Darr, and I. U. Rehman, “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.*, vol. 45, no. 5, pp. 355–368, 2010.

[107] S. Kumar, C. Desmedt, D. Larsimont, C. Sotiriou, and E. Goormaghtigh, “Change in the microenvironment of breast cancer studied by FTIR imaging,” *Analyst*, vol. 138, no. 14, pp. 4058–4065, 2013.

[108] G. Devi, T. S. Renuga Devi, and S. Gunasekaran, “FTIR spectroscopic study on benign and cancerous human breast tissues - A Run Chart analysis,” *Int. J. Pharm. Sci. Rev. Res.*, vol. 2, no. 2, pp. 73–77, 2010.

[109] H. Fabian, P. Lasch, M. Boese, and W. Haensch, “Infrared microspectroscopic imaging of benign breast tumor tissue sections,” *J. Mol. Struct.*, vol. 661, pp. 411–417, 2003.

[110] R. R. Alfano *et al.*, “Human breast tissues studied by IR Fourier transform Raman spectroscopy,” *Lasers life Sci.*, vol. 4, no. 1, pp. 23–28, 1991.

[111] A. S. Haka, K. E. Shafer-Peltier, M. Fitzmaurice, J. Crowe, R. R. Dasari, and M. S. Feld, “Identifying microcalcifications in benign and malignant breast lesions by probing differences in their chemical composition using Raman spectroscopy,” *Cancer Res.*, vol. 62, no. 18, pp. 5375–5380, 2002.

[112] H. Abramczyk, B. Brozek-Pluska, J. Surmacki, J. Jablonska-Gajewicz, and R. Kordek, “Raman ‘optical biopsy’of human breast cancer,” *Prog. Biophys. Mol. Biol.*, vol. 108, no. 1, pp. 74–81, 2012.

[113] A. S. Haka *et al.*, “In vivo margin assessment during partial mastectomy breast surgery using Raman spectroscopy,” *Cancer Res.*, vol. 66, no. 6, pp. 3317–3322, 2006.

[114] J. L. Pichardo-Molina *et al.*, “Raman spectroscopy and multivariate analysis of serum samples from breast cancer patients,” *Lasers Med. Sci.*, vol. 22, no. 4, pp. 229–236, 2007.

[115] N. Stone, R. Baker, K. Rogers, A. W. Parker, and P. Matousek, “Subsurface probing of calcifications with spatially offset Raman spectroscopy (SORS): future possibilities for the diagnosis of breast cancer,” *Analyst*, vol. 132, no. 9, pp. 899–905, 2007.

[116] J. Smith, C. Kendall, A. Sammon, J. Christie-Brown, and N. Stone, “Raman spectral mapping in the assessment of axillary lymph nodes in breast cancer,” *Technol. Cancer Res. Treat.*, vol. 2, no. 4, pp. 327–331, 2003.

[117] C. Yu, E. Gestl, K. Eckert, D. Allara, and J. Irudayaraj, “Characterization of human breast epithelial cells by confocal Raman micro spectroscopy,” *Cancer Detect. Prev.*, vol. 30, no. 6, pp. 515–522, 2006.

[118] K. W. Short, S. Carpenter, J. P. Freyer, and J. R. Mourant, “Raman spectroscopy detects biochemical changes due to proliferation in mammalian cell cultures,” *Biophys. J.*, vol. 88, no. 6, pp. 4274–4288, 2005.

[119] J. R. Mourant, Y. R. Yamada, S. Carpenter, L. R. Dominique, and J. P. Freyer, “FTIR spectroscopy demonstrates biochemical differences in mammalian cell cultures at different growth stages,” *Biophys. J.*, vol. 85, no. 3, pp. 1938–1947, 2003.

[120] W. Yang, X. Xiao, J. Tan, and Q. Cai, “In situ evaluation of breast cancer cell growth with 3D ATR-FTIR spectroscopy,” *Vib. Spectrosc.*, vol. 49, no. 1, pp. 64–67, 2009.

[121] N. S. Ozek, S. Tuna, A. E. Erson-Bensan, and F. Severcan, “Characterization of microRNA-125b expression in MCF7 breast cancer cells by ATR-FTIR spectroscopy,” *Analyst*, vol. 135, no. 12, pp. 3094–3102, 2010.

[122] S. Kumar, E. Goormaghtigh, and T. S. Shabi, “A FTIR imaging characterization of fibroblasts stimulated by various breast cancer cell lines,” *PLoS One*, vol. 9, no. 11, 2014.

[123] M. Verdonck *et al.*, “Breast cancer and melanoma cell line identification by FTIR imaging after formalin-fixation and paraffin-embedding,” *Analyst*, vol. 138, no. 14, pp. 4083–4091, 2013.

[124] N. I. Afanasyeva, S. F. Kolyakov, S. G. Artjushenko, V. V Sokolov, and G. A. Frank, “Minimally invasive and ex vivo diagnostics of breast cancer tissues by fiber optic evanescent wave Fourier transform IR (FEW-FTIR) spectroscopy,” in *Optical Biopsy II Conference*, 1998, vol. 3250, pp. 140–147.

[125] M. Sattlecker, R. Baker, N. Stone, and C. Bessant, “Support vector machine ensembles for breast cancer type prediction from mid-FTIR micro-calcification spectra,” *Chemom. Intell. Lab. Syst.*, vol. 107, no. 2, pp. 363–370, 2011.

[126] P. Tian *et al.*, “Intraoperative detection of sentinel lymph node metastases in breast carcinoma by Fourier transform infrared spectroscopy,” *Br. J. Surg.*, vol. 102, no. 11, p. 1372, 2015.

[127] C. Liu *et al.*, “Infrared absorption of human breast tissues in vitro,” *J. Lumin.*, vol. 119, pp. 132–136, 2006.

[128] M. Smolina and E. Goormaghtigh, “FTIR imaging of the 3D extracellular matrix used to grow colonies of breast cancer cell lines,” *Analyst*, vol. 141, no. 2, pp. 620–629, 2016.

[129] P. K. Hopke, “The evolution of chemometrics,” *Anal. Chim. Acta*, vol. 500, no. 1, pp. 365–377, 2003.

[130] S. Wold, “Chemometrics, why, what and where to next?,” *J. Pharm. Biomed. Anal.*, vol. 9, no. 8, pp. 589–596, 1991.

[131] J. N. Miller, *Statistics and chemometrics for analytical chemistry [electronic resource]*, 5th ed. Harlow, England ; New York: Harlow, England ; New York : Pearson/Prentice Hall, 2005, 2005.

[132] P. Geladi, “Chemometrics in spectroscopy. Part 1. Classical chemometrics,” *Spectrochim. Acta Part B-Atomic Spectrosc.*, vol. 58, no. 5, pp. 767–782, 2003.

[133] J. E. Jackson, *A user's guide to principal components*. New York ; Chichester: New York ; Chichester : Wiley, c1991, 1991.

[134] A. J. Izenman, “Linear discriminant analysis,” in *Modern Multivariate Statistical Techniques*, Springer, 2013, pp. 237–280.

[135] S. Wang, J. F. Lu, X. J. Gu, H. S. Du, and J. Y. Yang, “Semi-supervised linear discriminant analysis for dimension reduction and classification,” *Pattern Recognit.*, vol. 57, pp. 179–189, 2016.

[136] Y. Q. Guo, T. Hastie, and R. Tibshirani, “Regularized linear discriminant analysis and its application in microarrays,” *Biostatistics*, vol. 8, no. 1, pp. 86–100, 2007.

[137] H. P. Chan *et al.*, “COMPUTER-AIDED CLASSIFICATION OF MAMMOGRAPHIC MASSES AND NORMAL TISSUE - LINEAR DISCRIMINANT-ANALYSIS IN TEXTURE FEATURE SPACE,” *Phys. Med. Biol.*, vol. 40, no. 5, pp. 857–876, 1995.

[138] H. Ding *et al.*, “Discrimination of inflammatory bowel disease using Raman spectroscopy and linear discriminant analysis methods,” in *Biomedical Vibrational Spectroscopy 2016: Advances in Research and Industry*, 2016, vol. 9704.

[139] S. S. Nazeer, A. Saraswathy, A. K. Gupta, and R. S. Jayasree, “Fluorescence spectroscopy to discriminate neoplastic human brain lesions: a study using the spectral intensity ratio and multivariate linear discriminant analysis,” *Laser Phys.*, vol. 24, no. 2, 2014.

[140] S. E. Pinder *et al.*, “The manufacture and assessment of tissue microarrays: suggestions and criteria for analysis, with breast cancer as an example,” *J Clin Pathol*, vol. 66, no. 3, p. 169, 2013.

[141] G. Callagy *et al.*, “Molecular classification of breast carcinomas using tissue microarrays,” *Diagnostic Mol. Pathol.*, vol. 12, no. 1, pp. 27–34, 2003.

[142] R. Simon, M. Mirlacher, and G. Sauter, “Tissue microarrays in cancer diagnosis,” *Expert Rev. Mol. Diagn.*, vol. 3, no. 4, pp. 421–430, 2003.

[143] R. L. Camp, L. A. Charette, and D. L. Rimm, “Validation of tissue microarray technology in breast carcinoma,” *Lab. Investig.*, vol. 80, no. 12, pp. 1943–1949, 2000.

[144] R. Simon, M. Mirlacher, and G. Sauter, “Tissue microarrays,” *Mol. Diagnosis Cancer Methods Protoc.*, pp. 377–389, 2004.

[145] C. Kim and S. Paik, “Tissue microarrays,” in *Biomarkers in Breast Cancer*, Springer, 2006, pp. 31–44.

[146] A. M. Quinn, P. Tang, Q. Yang, and P. A. Bourne, “Tissue array construction: pitfalls, problems, and progress,” *Appl. Immunohistochem. Mol. Morphol.*, vol. 16, no. 3, pp. 287–290, 2008.

[147] S. A. Mian, H. E. Colley, M. H. Thornhill, and I. u Rehman, “Development of a dewaxing protocol for tissue-engineered models of the oral mucosa used for Raman spectroscopic analysis,” *Appl. Spectrosc. Rev.*, vol. 49, no. 8, pp. 614–617, 2014.

[148] I. ur Rehman, *Vibrational spectroscopy for tissue analysis*. Boca Raton: Boca Raton : CRC Press, c2013, 2013.

[149] Z. Movasaghi, S. Rehman, and I. U. Rehman, “Fourier transform infrared (FTIR) spectroscopy of biological tissues,” *Appl. Spectrosc. Rev.*, vol. 43, no. 2, pp. 134–179, 2008.

[150] A. A. Spector and M. A. Yorek, “Membrane lipid composition and cellular function.,” *J. Lipid Res.*, vol. 26, no. 9, pp. 1015–35, Sep. 1985.

[151] D. Sheng *et al.*, “A study of structural differences between liver cancer cells and normal liver cells using FTIR spectroscopy,” *J. Mol. Struct.*, vol. 1099, pp. 18–23, Nov. 2015.

[152] G. L. Johanning, “Modulation of breast cancer cell adhesion by unsaturated fatty acids,”

Nutrition, vol. 12, no. 11–12, pp. 810–816, Nov. 1996.

[153] M. He, S. Guo, and Z. Li, "In situ characterizing membrane lipid phenotype of breast cancer cells using mass spectrometry profiling," *Sci. Rep.*, vol. 5, no. 1, p. 11298, Sep. 2015.

[154] E. Bogomolny, S. Argov, S. Mordechai, and M. Huleihel, "Monitoring of viral cancer progression using FTIR microscopy: A comparative study of intact cells and tissues," *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1780, no. 9, pp. 1038–1046, Sep. 2008.

[155] G. M. S. El-Bahy, "FTIR AND RAMAN SPECTROSCOPIC STUDY OF FENUGREEK (Trigonella foenum graecum L.) SEEDS," 2005.

[156] P. Garidel, "Mid-FTIR-Microspectroscopy of stratum corneum single cells and stratum corneum tissue," *Phys. Chem. Chem. Phys.*, vol. 4, no. 22, pp. 5671–5677, Nov. 2002.

[157] A. Derenne, O. Vandersleyen, and E. Goormaghtigh, "Lipid quantification method using FTIR spectroscopy applied on cancer cell extracts," *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids*, vol. 1841, no. 8, pp. 1200–1209, Aug. 2014.

[158] R. Clarke, H. W. van den Berg, and R. F. Murphy, "Reduction of the membrane fluidity of human breast cancer cells by tamoxifen and 17β-estradiol," *JNCI J. Natl. Cancer Inst.*, vol. 82, no. 21, pp. 1702–1705, 1990.

[159] M. Hilvo *et al.*, "Novel Theranostic Opportunities Offered by Characterization of Altered Membrane Lipid Metabolism in Breast Cancer Progression," *Cancer Res.*, vol. 71, no. 9, pp. 3236–3245, May 2011.

[160] C. R. Santos and A. Schulze, "Lipid metabolism in cancer," *FEBS J.*, vol. 279, no. 15, pp. 2610–2623, Aug. 2012.

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

[162] N. P. Camacho, P. West, P. A. Torzilli, and R. Mendelsohn, "FTIR microscopic imaging of collagen and proteoglycan in bovine cartilage," *Biopolymers*, vol. 62, no. 1, pp. 1–8, Jan. 2001.

[163] M. C. Chang and J. Tanaka, "FT-IR study for hydroxyapatite/collagen nanocomposite cross-linked by glutaraldehyde," *Biomaterials*, vol. 23, no. 24, pp. 4811–4818, Dec. 2002.

[164] G. I. Dovbeshko *et al.*, "Surface enhanced IR absorption of nucleic acids from tumor cells: FTIR reflectance study," *Biopolymers*, vol. 67, no. 6, pp. 470–486, Jan. 2002.

[165] A. L. M. Batista de Carvalho *et al.*, "Chemotherapeutic response to cisplatin-like drugs in human breast cancer cells probed by vibrational microspectroscopy," *Faraday Discuss.*, vol. 187, no. 0, pp. 273–298, Jun. 2016.

[166] G. Zandomeneghi, M. R. H. Krebs, M. G. McCammon, and M. Fändrich, "FTIR reveals structural differences between native β-sheet proteins and amyloid fibrils," *Protein Sci.*, vol. 13, no. 12, pp. 3314–3321, Jan. 2009.

[167] G. I. Dovbeshko, N. Y. Gridina, E. B. Kruglova, and O. P. Pashchuk, "FTIR spectroscopy studies of nucleic acid damage," *Talanta*, vol. 53, no. 1, pp. 233–246, Oct. 2000.

[168] C. E. Zois and A. L. Harris, "Glycogen metabolism has a key role in the cancer microenvironment and provides new targets for cancer therapy," *J. Mol. Med.*, vol. 94, no. 2, pp. 137–154, Feb.

2016.

[169] B. J. Grant, A. A. Gorfe, and J. A. McCammon, “Large conformational changes in proteins: signaling and other functions,” *Curr. Opin. Struct. Biol.*, vol. 20, no. 2, pp. 142–147, Apr. 2010.

[170] L. B. Mostaço-Guidolin, L. S. Murakami, M. R. Batistuti, A. Nomizo, and L. Bachmann, “Molecular and chemical characterization by Fourier transform infrared spectroscopy of human breast cancer cells with estrogen receptor expressed and not expressed,” *J. Spectrosc.*, vol. 24, no. 5, pp. 501–510, 2010.

[171] P. R. Palaniappan and V. Vijayasundaram, “FTIR study of arsenic induced biochemical changes on the liver tissues of fresh water fingerlings Labeo rohita,” *Rom J Biophys*, vol. 18, pp. 135–144, 2008.

[172] J. Anastassopoulou *et al.*, “Microimaging FT-IR spectroscopy on pathological breast tissues,” *Vib. Spectrosc.*, vol. 51, no. 2, pp. 270–275, Nov. 2009.

[173] E. Currie, A. Schulze, R. Zechner, T. C. Walther, and R. V. Farese, “Cellular Fatty Acid Metabolism and Cancer,” *Cell Metab.*, vol. 18, no. 2, pp. 153–161, Aug. 2013.

[174] C. P. Rajneesh, A. Manimaran, K. R. Sasikala, and P. Adaikappan, “Lipid peroxidation and antioxidant status in patients with breast cancer,” *Singapore Med. J.*, vol. 49, no. 8, p. 640, 2008.

[175] F. Tas *et al.*, “Oxidative Stress in Breast Cancer,” *Med. Oncol.*, vol. 22, no. 1, pp. 011–016, 2005.

[176] P. M. Das and R. Singal, “DNA Methylation and Cancer,” *J Clin Oncol*, vol. 22, pp. 4632–4642, 2004.

[177] X. Yang, L. Yan, and N. E. Davidson, “DNA methylation in breast cancer.,” *Endocr. Relat. Cancer*, vol. 8, no. 2, pp. 115–127, 2001.

[178] M. Widschwendter *et al.*, “Association of Breast Cancer DNA Methylation Profiles with Hormone Receptor Status and Response to Tamoxifen,” 2004.

[179] E. Rysman *et al.*, “De novo Lipogenesis Protects Cancer Cells from Free Radicals and Chemotherapeutics by Promoting Membrane Lipid Saturation,” *Cancer Res.*, vol. 70, no. 20, pp. 8117 LP – 8126, Oct. 2010.

[180] A. Talari *et al.*, “Raman Spectroscopy of Biological Tissues,” *Appl. Spectrosc. Rev.*, vol. 50, no. 1, pp. 46–111, Jan. 2015.

[181] D. Lazaro-Pacheco, A. M. Shaaban, S. Rehman, and I. Rehman, “Raman spectroscopy of breast cancer,” *Appl. Spectrosc. Rev.*, pp. 1–37, Apr. 2019.

[182] D. Hartmann *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.*, vol. 44, no. 4, pp. 620–628, Apr. 2012.

[183] S. Schiffmann *et al.*, “Ceramide synthases and ceramide levels are increased in breast cancer tissue,” *Carcinogenesis*, vol. 30, no. 5, pp. 745–752, 2009.

[184] K. Sakai *et al.*, “Composition and turnover of phospholipids and neutral lipids in human breast cancer and reference tissues,” *Carcinogenesis*, vol. 13, no. 4, pp. 579–584, 1992.

[185] D. P. Lau *et al.*, “Raman spectroscopy for optical diagnosis in normal and cancerous tissue of

the nasopharynx?preliminary findings," *Lasers Surg. Med.*, vol. 32, no. 3, pp. 210–214, Mar. 2003.

[186] S. Zhang *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.*, vol. 66, no. 3, pp. 626–632, 1997.

[187] B. Brozek-Pluska, J. Musial, R. Kordek, E. Bailo, T. Dieing, and H. Abramczyk, "Raman spectroscopy and imaging: applications in human breast cancer diagnosis," *Analyst*, vol. 137, no. 16, pp. 3773–3780, 2012.

[188] B. Brozek-Pluska, J. Jablonska-Gajewicz, R. Kordek, and H. Abramczyk, "Phase Transitions in Oleic Acid and in Human Breast Tissue As Studied by Raman Spectroscopy and Raman Imaging," *J. Med. Chem.*, vol. 54, no. 9, pp. 3386–3392, 2011.

[189] M. W. Conklin *et al.*, "Aligned collagen is a prognostic signature for survival in human breast carcinoma," *Am. J. Pathol.*, vol. 178, no. 3, pp. 1221–1232, 2011.

[190] A. C. S. Talari, A. Raza, S. Rehman, and I. U. Rehman, "Analyzing normal proliferating, hypoxic and necrotic regions of T-47D human breast cancer spheroids using Raman spectroscopy," *Appl. Spectrosc. Rev.*, vol. 52, no. 10, pp. 909–924, 2017.

[191] C. Yu, E. Gestl, K. Eckert, D. Allara, and J. Irudayaraj, "Characterization of human breast epithelial cells by confocal Raman microspectroscopy," *Cancer Detect. Prev.*, vol. 30, no. 6, pp. 515–522, Jan. 2006.

[192]) R C ; J L Lord *et al.*, "Laser-Excited Raman Spectroscopy of Biomolecules. VI. Some Polypeptides as Conformational Models," 1971.

[193] M. R. Wrensch *et al.*, "Breast fluid cholesterol and cholesterol beta-epoxide concentrations in women with benign breast disease.," *Cancer Res.*, vol. 49, no. 8, pp. 2168–74, Apr. 1989.

[194] J. Kneipp, T. B. Schut, M. Kliffen, M. Menke-Pluijmers, and G. Puppels, "Characterization of breast duct epithelia: a Raman spectroscopic study," *Vib. Spectrosc.*, vol. 32, no. 1, pp. 67–74, Aug. 2003.

[195] G. J. Thomas, B. Prescott, and D. W. Urry, "Raman amide bands of type-II ?-turns in cyclo-(VPGVG)3 and poly-(VPGVG), and implications for protein secondary-structure analysis," *Biopolymers*, vol. 26, no. 6, pp. 921–934, Jun. 1987.

[196] A. Rygula, K. Majzner, K. M. Marzec, A. Kaczor, M. Pilarczyk, and M. Baranska, "Raman spectroscopy of proteins: a review," *J. Raman Spectrosc.*, vol. 44, no. 8, pp. 1061–1076, Aug. 2013.

[197] A. Orimo *et al.*, "Stromal Fibroblasts Present in Invasive Human Breast Carcinomas Promote Tumor Growth and Angiogenesis through Elevated SDF-1/CXCL12 Secretion," *Cell*, vol. 121, no. 3, pp. 335–348, May 2005.

[198] K. Glunde, C. Jie, and Z. M. Bhujwalla, "Molecular causes of the aberrant choline phospholipid metabolism in breast cancer," *Cancer Res.*, vol. 64, no. 12, pp. 4270–4276, 2004.

[199] Y. L. Ting, D. Sherr, and H. Degani, "Variations in energy and phospholipid metabolism in normal and cancer human mammary epithelial cells.," *Anticancer Res.*, vol. 16, no. 3B, pp. 1381–8, 1996.

[200] J. Ruiz-Cabello and J. S. Cohen, “Phospholipid metabolites as indicators of cancer cell function,” *NMR Biomed.*, vol. 5, no. 5, pp. 226–233, Sep. 1992.

[201] H. Abramczyk and B. Brozek-Pluska, “Raman Imaging in Biochemical and Biomedical Applications. Diagnosis and Treatment of Breast Cancer,” *Chem. Rev.*, vol. 113, no. 8, pp. 5766–5781, Aug. 2013.

[202] D. C. Allred *et al.*, “Association of p53 protein expression with tumor cell proliferation rate and clinical outcome in node-negative breast cancer,” *JNCI J. Natl. Cancer Inst.*, vol. 85, no. 3, pp. 200–206, 1993.

[203] M. G. Martinez, A. J. Bullock, S. MacNeil, and I. U. Rehman, “Characterisation of structural changes in collagen with Raman spectroscopy,” *Appl. Spectrosc. Rev.*, pp. 1–34, Jan. 2019.

[204] D. de Gonzalo-Calvo *et al.*, “Intratumor cholesteryl ester accumulation is associated with human breast cancer proliferation and aggressive potential: a molecular and clinicopathological study,” *BMC Cancer*, vol. 15, no. 1, p. 460, Dec. 2015.

[205] A. A. Dmitriev and N. V. Surovtsev, “Temperature-Dependent Hydrocarbon Chain Disorder in Phosphatidylcholine Bilayers Studied by Raman Spectroscopy,” *J. Phys. Chem. B*, vol. 119, no. 51, pp. 15613–15622, Dec. 2015.

[206] M. Procházka, J. Štěpánek, and P.-Y. Turpin, “Interaction of phospholipid dispersions with water-soluble porphyrins as monitored by their Raman temperature profiles,” *Chem. Phys. Lipids*, vol. 132, no. 2, pp. 145–156, Dec. 2004.

[207] C. Krafft, L. Neudert, T. Simat, and R. Salzer, “Near infrared Raman spectra of human brain lipids,” *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 61, no. 7, pp. 1529–1535, May 2005.

[208] L. Seballos, J. Z. Zhang, and R. Sutphen, “Surface-enhanced Raman scattering detection of lysophosphatidic acid,” *Anal. Bioanal. Chem.*, vol. 383, no. 5, pp. 763–767, Nov. 2005.

[209] E. Hietanen, K. Punnonen, R. Punnonen, and O. Auvinen, “Fatty acid composition of phospholipids and neutral lipids and lipid peroxidation in human breast cancer and lipoma tissue,” *Carcinogenesis*, vol. 7, no. 12, pp. 1965–1969, Dec. 1986.

[210] M. Nonaka, E. Li-Chan, and S. Nakai, “Raman spectroscopic study of thermally induced gelation of whey proteins,” *J. Agric. Food Chem.*, vol. 41, no. 8, pp. 1176–1181, 1993.

[211] K. Czamara, K. Majzner, M. Z. Pacia, K. Kochan, A. Kaczor, and M. Baranska, “Raman spectroscopy of lipids: a review,” *J. Raman Spectrosc.*, vol. 46, no. 1, pp. 4–20, Jan. 2015.

[212] M. E. Darvin *et al.*, “Noninvasive detection of beta-carotene and lycopene in human skin using Raman spectroscopy,” *LASER PHYSICS-LAWRENCE-*, vol. 14, no. 2, pp. 231–233, 2004.

[213] W. Cheng, M. Liu, H. Liu, and S. Lin, “Micro-Raman spectroscopy used to identify and grade human skin pilomatrixoma,” *Microsc. Res. Tech.*, vol. 68, no. 2, pp. 75–79, 2005.

[214] W. Akhtar and H. G. M. Edwards, “Fourier-transform Raman spectroscopy of mammalian and avian keratotic biopolymers,” *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 53, no. 1, pp. 81–90, Jan. 1997.

[215] E. R. Nelson, C. Chang, and D. P. McDonnell, “Cholesterol and breast cancer pathophysiology,” *Trends Endocrinol. Metab.*, vol. 25, no. 12, pp. 649–655, Dec. 2014.

[216] R. Kalluri and M. Zeisberg, “Fibroblasts in cancer,” *Nat. Rev. Cancer*, vol. 6, no. 5, pp. 392–401, May 2006.

[217] P. P. Provenzano *et al.*, “Collagen density promotes mammary tumor initiation and progression,” *BMC Med.*, vol. 6, no. 1, p. 11, Dec. 2008.

[218] H. Abramczyk and B. Brozek-Pluska, “New look inside human breast ducts with Raman imaging. Raman candidates as diagnostic markers for breast cancer prognosis: Mammaglobin, palmitic acid and sphingomyelin,” *Anal. Chim. Acta*, vol. 909, pp. 91–100, Feb. 2016.

[219] G. Llaverias *et al.*, “Role of Cholesterol in the Development and Progression of Breast Cancer,” *Am. J. Pathol.*, vol. 178, no. 1, pp. 402–412, Jan. 2011.

[220] D. Chaturvedi, S. A. Balaji, V. K. Bn, F. Ariese, S. Umapathy, and A. Rangarajan, “Different Phases of Breast Cancer Cells: Raman Study of Immortalized, Transformed, and Invasive Cells.,” *Biosensors*, vol. 6, no. 4, Nov. 2016.

[221] J. R. Hands *et al.*, “Attenuated total reflection Fourier transform infrared (ATR-FTIR) spectral discrimination of brain tumour severity from serum samples,” *J. Biophotonics*, vol. 7, no. 3-4, pp. 189–199, 2014.

[222] K. Gajjar *et al.*, “Fourier-transform infrared spectroscopy coupled with a classification machine for the analysis of blood plasma or serum: a novel diagnostic approach for ovarian cancer,” *Analyst*, vol. 138, no. 14, p. 3917, Jun. 2013.

[223] D. I. Ellis, D. Broadhurst, S. J. Clarke, and R. Goodacre, “Rapid identification of closely related muscle foods by vibrational spectroscopy and machine learning,” *Analyst*, vol. 130, no. 12, p. 1648, Nov. 2005.

[224] A. Morellos *et al.*, “Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy,” *Biosyst. Eng.*, vol. 152, pp. 104–116, Dec. 2016.

[225] C. Cheng, J. Liu, C. Zhang, M. Cai, H. Wang, and W. Xiong, “An overview of infrared spectroscopy based on continuous wavelet transform combined with machine learning algorithms: application to chinese medicines, plant classification, and cancer diagnosis,” *Appl. Spectrosc. Rev.*, vol. 45, no. 2, pp. 148–164, 2010.