Raman spectroscopic analysis to identify chemical changes associated with different subtypes of breast cancer tissue samples

Abdullah Chandra Sekhar Talari

Thesis submitted to the University of Sheffield for the degree of Doctor of Philosophy

Department of Material Science and Engineering,

The University of Sheffield

July 2015

Abstract

Breast cancer incidence rates are increasing in women worldwide with the highest rates reported in developed countries. A combination of screening approaches, Immuno-histopathology and gene profiling analyses are currently used for diagnosis and typing but have their own limitations in understanding disease and its subtypes. Raman spectroscopy (RS) has attained great attention from biomedical researchers due to its non-invasive and non-destructive detection approach. Chemometrics is one of the powerful tools used in spectroscopic research to enhance its sensitivity.

RS was used to characterise and differentiate two breast cancer and one normal breast cell lines (MDA-MB-436, MCF-7 and MCF-10A) and spectra of the cell lines have revealed basic differences in the concentration of biochemical compounds such as lipids, nucleic acids and proteins Raman peaks were found to differ in intensity and principal component analysis (PCA) was able to identify variations that lead to accurate and reliable separation of the three cell lines. Linear discriminant analysis (LDA) model of three cell lines was predicted with 100% sensitivity and 91% specificity.

RS studies were extended from single cells to multiple cell spheroids. Human breast cancer cell lines (T-47D) were grown as spheroids and a combination of RS and Cluster analysis were employed to understand biochemical fingerprint and differentiation of normal proliferating, hypoxic and necrotic regions of spheroids. These variations may be useful in identifying new spectral markers and further understanding of cancer metabolism.

Finally, Human breast biopsies on Tissue microarray (TMA) slide were analysed using RS and Chemometrics approaches. Biopsies were classified as luminal A, luminal B, HER2 and triple negative subtypes to understand chemical changes associated with breast cancer subtypes. Supervised and unsupervised algorithms were applied on biopsy data to explore intra and inter dataset biochemical changes associated with lipids, collagen and nucleic acid content.

In summary, RS has offered great potential understanding breast cancer from cell line level to multicellular spheroid to higher architecture of tissue. This study has explored new area to understand biochemical fingerprint of breast biopsies, which is complementary to current trends of molecular profiling and immuno histopathological approaches.

Acknowledgements

In the name of Allah (SWT), the most beneficent, the most merciful

First and foremost I would like to express my sincere gratitude to my supervisor Dr Ihtesham Ur Rehman for his extraordinary support, valuable guidance and diligent care towards my PhD project. His constant and consistent advice made my PhD period successful. I am also thankful for his contributions of time and ideas to make my PhD more fruitful.

I would like to thank my co-supervisor Professor Robert E Coleman for his valuable guidance and encouragement. I would like to thank Professor Ingunn Holen for her support and encouragement in my PhD program. I would like to thank Catherine Alyson, Mark Wagner and Robert Dickinson for their support in my lab work at Kroto Research Institute and Sheffield medical school.

I thank you to all my colleagues in Material Science and Engineering department and Kroto Research Institute for their constant support. I would like to thank my PhD colleagues Ceyla Yorucu for her extraordinary support, Salman Aziz and Ahtasham Raza for their exceptional support throughout my PhD program, Saad Bin Qasim, Peyman Zolgharnein, Muhammad Khawaja Hammad Uddin and Jasarat for their support during my doctoral program.

I would like to thank my family especially my parents, without their support I could not achieve this task. Finally I thank my wife and my beloved daughter (Leena) for their support and patience while I am finishing my writing up.

Abbreviations

^0C	degree Celsius
λ	wavelength
µ	micro
3D	Three dimensional
a.u.	Arbitrary units
A	Adenine
AS	Age-standardised
ATCC	American Type Culture Collection
BCC	Basal cell carcinoma
C	Cytosine
CA	Cluster Analysis
CARS	Coherent Anti-Stokes Raman Spectroscopy
CCD	Charged –Coupled device
cm^{-1}	per centimetre
CNAs	Copy-number alterations
CRM	Confocal Raman spectroscopy
CT	Computer Tomography
DCIS	Ductal Carcinoma in situ
DED	De-epidermised dermis
DMEM	Dulbecco's modified Eagle's Medium
DMSO	Dimethyl Sulphoxide
DNA	Deoxyribonucleic acid
DPSS	Diode-pumped solid state
ER	Estrogen receptor
EBV	Epistein-Barr virus
ECACC	European Collection of Cell Culture
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal Growth Factor

ESS	Elastic scattering spectroscopy
FAD	Flavin adenine nucleotide
FCS	Fetal Calf Serum
FISH	Florescence In Situ Hybridization
FS	Fluorescence Spectroscopy
FTIR	Fourier Transform Infrared Spectroscopy
FCA	fussy c-means cluster analysis
G	Guanine
gm	gram
GI	Gastrointestinal
h	Hours
H&E	Haematoxylin and Eosin
HER2	Human epidermal growth factor receptor 2
Ham's F12	Nutrient Mixture F12
HCA	Hierarchical clustering analysis
HCl	Hydrochloric acid
HPV	Human Papilloma Virus
IHC	Immuno histochemistry
IMS	Industrial methylated spirits
IDC	Invasive ductal carcinoma
K	Kilo
KMCA	K-means cluster analysis
L	litre
LDA	Linear Discriminant Analysis
m	metre
mM	millimolar
MRI	Magnetic Resonance Imaging
NIR	Near Infra-red
N	nano
NAD^+	Nicotinamide Adenine Nucleotide

NADH	Nicotinamide Adenine Nucleotide (reduced form)
NaOH	Sodium Hydroxide
OCT	Optical Coherence Tomography
PAH	Polycyclic aromatic hydrocarbon
PBS	Phosphate buffered saline
PCA	Principal Component Analysis
PET	Position Emission Tomography
pH	potentiometric hydrogen ion concentration
PLS	Partial Least Squares
PR	Progesterone receptor
RB1	Retinoblastoma gene
RNA	Ribonucleic acid
Rpm	revolutions per minute
RPMI	Roswell Park Memorial Institute
RS	Raman Spectroscopy
RSS	Resonance Raman Spectroscopy
SCC	Squamous cell carcinoma
SERS	Surface Enhance Raman Spectroscopy
SVM	Support Vector Machine
T	Thymine
TDLU	Terminal ductal lobular unit
TACS	Tumour-associated collagen signature
U	Uracil
UV	Ultra Violet
V	Volts
v	Volume
W	Watt
w	weight

Contents

Table of Figures

Table of Tables

Chapter 1

Introduction

1.1 Cancer – An Overview

Cancer is an uncontrolled growth of tissue caused by rapid cell proliferation. It is an accumulation of cells that leads to the development of a tumour, also known as a neoplasm. Neoplastic cells initially gather as a cluster in the body, which is referred to as the benign stage. These types of cancers are stable and recovery can easily be attained through surgical excision. Once neoplastic cells start to invade their surrounding tissues, this is referred to as the malignant stage, whereby cells have developed invasive behaviour and these are generally life threatening. When the tumour cells migrate away from primary locations and start forming remote colonies through lymphatic or blood circulation, they become metastatic. The classification of a cancer is dependent upon the type of cells, and the tissues where the disease originates. Carcinomas derive from epithelial tissues, sarcomas can develop from muscles cells and leukaemia derives from hemopoiteic cells (Alberts, 2015).

Most types of cancer cells form a mass which is called a tumour and is named after the organ of the body where it originated. An estimate across the world in 2008 revealed that 12.66 million people were diagnosed with cancer (Figure 1). It was also estimated that in the same year 7.56 million cancer deaths occurred out of a world estimated population of 6.75 billion people. Furthermore out of 27 cancers, lung cancer remains the most common cancer in both diagnosis (1.6millions (12.6)) and death rates (1.37 million (18.1)) during the same year. This is closely followed by breast cancer, which is the second most common cancer overall, with 1.38 million cases (10.9%) but ranks 5th in terms of deaths 61% (458000). This is followed by colorectal, stomach and prostate cancer both in terms of diagnosis and death rates i.e. 1.2 million cases and deaths

609900, 990000 cases and deaths 737000 913000 cases and 261000 deaths respectively (Ferlay et al., 2010, Reeves et al., 2007).

1.2 Breast Cancer

In the 18th century an eminent French Surgeon, Le Dran first reported the case of a breast cancer patient, whose previous two generations had experienced a history of the disease. In the mid-19th century, it was further reported by another French surgeon, Broca, who observed the disease in four previous generations. In l972, Lynch and his team reported hereditary breast cancer along with ovarian carcinoma in 34 families. A major breakthrough in breast cancer history was the identification of BRCA1 locus to 17q21 by Mary-Claire King and her colleagues in 1994 (Lindeman and Visvader, 2011).

1.3 Epidemiology

Breast cancer is the most common cancer in American women and it accounts for one in three cancers. Breast cancer figures of the year 2011 have revealed the following facts. More than one million women were diagnosed as invasive and nearly sixty thousand women were diagnosed as *in situ* breast cancer. Geographical differences also affected the occurrence of the disease. For example, those in the New Mexico and Delaware region were more prone to breast cancer compared to those in other regions. Incidence rates also varied relative to the different ethnic groups in the United States. The highest incidence rates were reported in non-hispanic white women in the area of California and Columbia whilst the highest mortality rates were reported in African American women (Torre et al., 2015).

Breast cancer is also the most commonly occurring cancer in Australian women, after skin cancer. According to studies by the Australian government, the incidence rates of breast cancer were increased in women aged between fifty to sixty nine years. The same studies found that nearly thirteen and a half thousand women were affected by the

disease in 2008. Breast cancer survival rates were lower in West, South and Northern regions of Australia compared to other areas. Breast cancer consequently has the second highest mortality rate amongst all cancers (Farley et al., 2010).

Western European countries appear to have the highest incidence rate for breast cancer whilst Eastern and Middle Africa account for the lowest values with almost five-fold variation in world age-standardised (AS) incidence rates (Figure 1). Amongst 27 different countries of Europe, Belgium had the highest rate of approximately 145 cases per 100,000 for European AS incidence rates in 2008 for female breast cancer whereas the lowest rates were noted in Greece with a value of 57 cases per 100,000. The incidence rate for females in the UK was found to be sixth amongst all European countries (EU-27) (Figure 2). Within United Kingdom, breast cancer is considered to be most common malignancy in England, Wales and Northern Ireland. While in Scotland breast cancer is second most common malignancy. The lifetime risk of this disease in women is 1 in 8, and nearly 50,000 women were diagnosed with breast cancer in UK during 2011. Since 1970, the incidence rates of lifetime risk increased by 72%. It was also noticed that the lifetime risk increases with age, nearly 8 in 10 women were diagnosed with the disease aged 50 or over. Breast cancer incidence AS rates in England, Wales, Scotland and Northern Ireland were 124.8, 123.3, 130.2 and 118.4 respectively. Northern Ireland showed lowest incident AS rates compared to other parts of the United Kingdom (Farley et al., 2010, Cancer Research UK., 2012).

The economically developed regions of the world have gained a significant increase in incidence of female breast cancer rates especially in Central, Eastern and Far Eastern parts of Europe and this has been greatly influenced by changes in lifestyle and/or reproductive behaviour, weight gain, alcohol consumption and the use of hormone replacement therapy (HRT). Similarly the risk has been found to be more pronounced in females who migrated from a low risk area to an area with a higher level of risk, thus

demonstrating the strong influence of lifestyle and environmental factors (Ferlay et al., 2010).

Approximately 10% of the population of developing countries is represented by Latin America and the Caribbean. Almost each year 10% of new cases of cancer are diagnosed in these countries. However, the total number of breast cancer cases is lower as compared to developed countries. Uruguay and Argentina have been shown to exhibit the highest number of cancer incidence rates in these parts of the world. According to general statistics, there are no regular screening programs available in these parts of the world (Cazap et al., 2008).

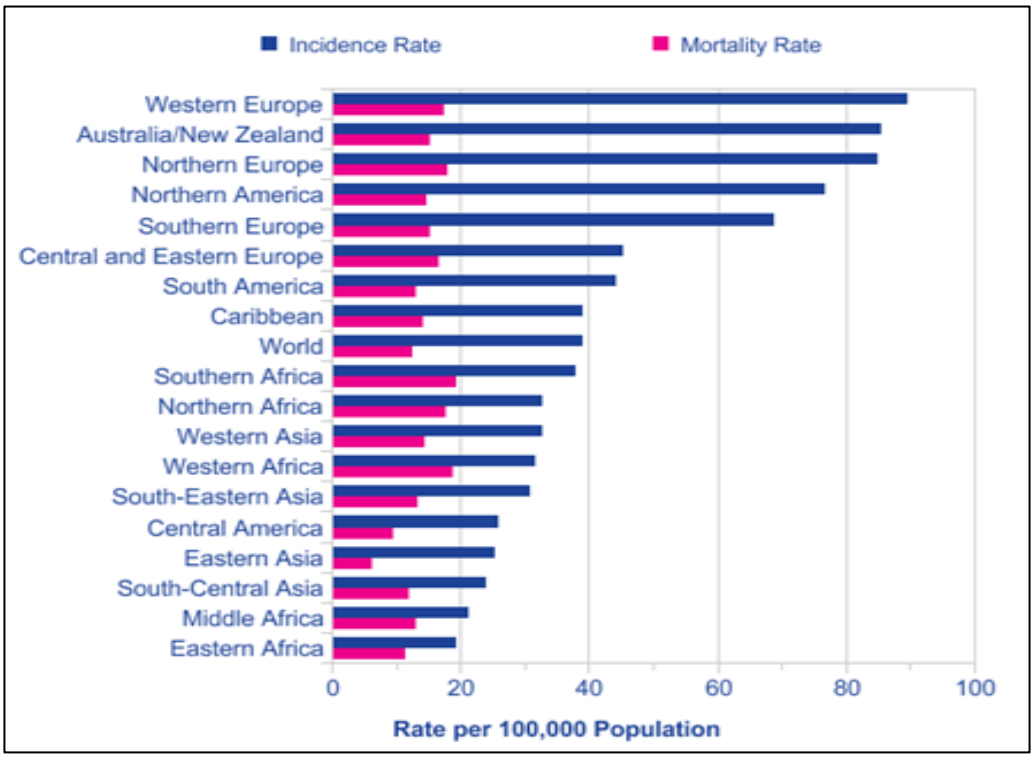

Figure 1: Worldwide statistics of breast cancer (Farley et al., 2010).

Figure 2: Breast cancer statistics of Europe (Farley et al., 2010).

1.4 Etiology

1.4.1 Geographic difference

Lifestyle based risk factors and environmental factors may affect the etiology of breast cancer. Breast cancer incidence rates were found to be considerably low in Asian and Mediterranean countries in comparison to the Western world. Cancer incidence rates were notably increasing in Asian women, whose ancestors migrated to United States a few generations previously. This may be due to lifestyle factors, which could have been adopted during this time period. In recent years, Japan has reported an increase in the incidence rates, perhaps due to the embracing of the Western lifestyle in their culture (Debruin and Josephy, 2002).

1.4.2 Alcohol consumption

Alcohol consumption may increase the risk of breast cancer occurrence. Studies have found that those women, who do not consume alcohol, show a significantly reduced likelihood of breast cancer occurrence, compared to women who consume alcohol.

Postmenopausal women, who are habituated to excessive drinking, appear to show an elevated risk of developing breast cancer. Alcohol intake may produce higher levels of estrogen and insulin like growth factor in the body (Scoccianti et al., 2014).

1.4.3 Diet

A number of studies have reported that dietary fat is one of the risk factors for developing breast cancer. The evidence is inconclusive in these studies due to a lack of definitive correlation between fat intake and the likelihood of cancer development. Estrogen exposure may also cause breast cancer risk through dietary factors, alcohol consumption, birth control pills and hormone replacement therapy. A Uruguay based case-control study has confirmed that consumption of red meat may elevate the risk of breast cancer. Recent studies have found carcinogenic materials such as polycyclic aromatic hydrocarbon (PAH) and heterocyclic aromatic amines while grilling meat. PAH associated DNA adducts were observed in breast cancer tissues and this may be, one of the causative agents in breast cancer development. But as of yet there is no clear evidence in this regard. Intake of vegetables and fruits may decrease the development of breast cancer (Alexander et al., 2010).

1.4.4 Tobacco Smoking

A weak relationship exists between smoking and breast cancer development. Recent studies have revealed that smoking could increase the risk of breast cancer incidence in women who started smoking at an early age. CYP1A1 (Cytochrome P450 family polypeptide A) genetic variants with nucleotide transition at specific gene locations were observed in these women (Ishibe et al., 1998). However, a very minute percentage of breast cancers have shown these risk factors. Scientists have established a correlation between smoking and N-Acetyltransferase (NAT 2) genotype. Postmenopausal women who have a three year smoking history had developed fast genotype. This will consequently result in a higher risk of developing breast cancer. Whereas

postmenopausal women, who had a non-smoking history had developed slow genotype, thus resulting in a lower risk of developing the disease (Reynolds, 2013).

1.4.5 Genetic risk factors

The major genes considered as risk factors for development of breast cancer are BRCA1 and BRCA2. The proteins, which are synthesized by these genes, could play an important role in cancer prevention. These genes, when mutated, contribute to 5 to 10% of hereditary breast cancers. The individuals who inherit the mutated copy of these genes have a high chance of developing the disease (Katsama et al., 2000). The other genes, which could be involved in developing cancer, are ATM, TP53, HER2/neu, CHEK2, PTEN, p53, BCL-2, CDH1 and STK 11. Over expression of BCL-2 gene has been observed in 10 to 40% of breast cancers. Mutations in p53 genes were highly associated with sporadic breast cancers. Mutated copies of these genes play an important role in cell cycle control, nucleic acid repair mechanisms and apoptosis pathways. Other risk factors implicated in breast cancer development are not having children, obesity, chemical pollutants and hormone therapy (Debruin and Josephy, 2002)

1.5 Breast Anatomy and Pathology

Breast or mammary gland is an important organ in the female body and provides nutrition to the infants. The major constituents of breast are adipose tissue, glandular tissue, lobes, lobules, alveoli, ducts and nipple. This structure is supported by connective tissue also known as cooper's ligaments (Figure 3). Breast is made up of mammary milk glands and it contains fifteen to twenty lobes. Each lobe is further made up of lobules, and each lobule made up of ten to hundred alveoli and they are interconnected with sequence of branched ducts. Approximately fifteen to twenty five ducts join together and form main duct. Histological studies have confirmed that more than seventeen main ducts will enter into the nipple. Recent investigations have demonstrated human breast structure using high-resolution ultrasound. Lactating breast

(milk production) is principally made up of adipose tissues and normal breast accumulates fat as increased with age, and it is glandular during lactation (Geddes, 2007, Ali and Coombes, 2002).

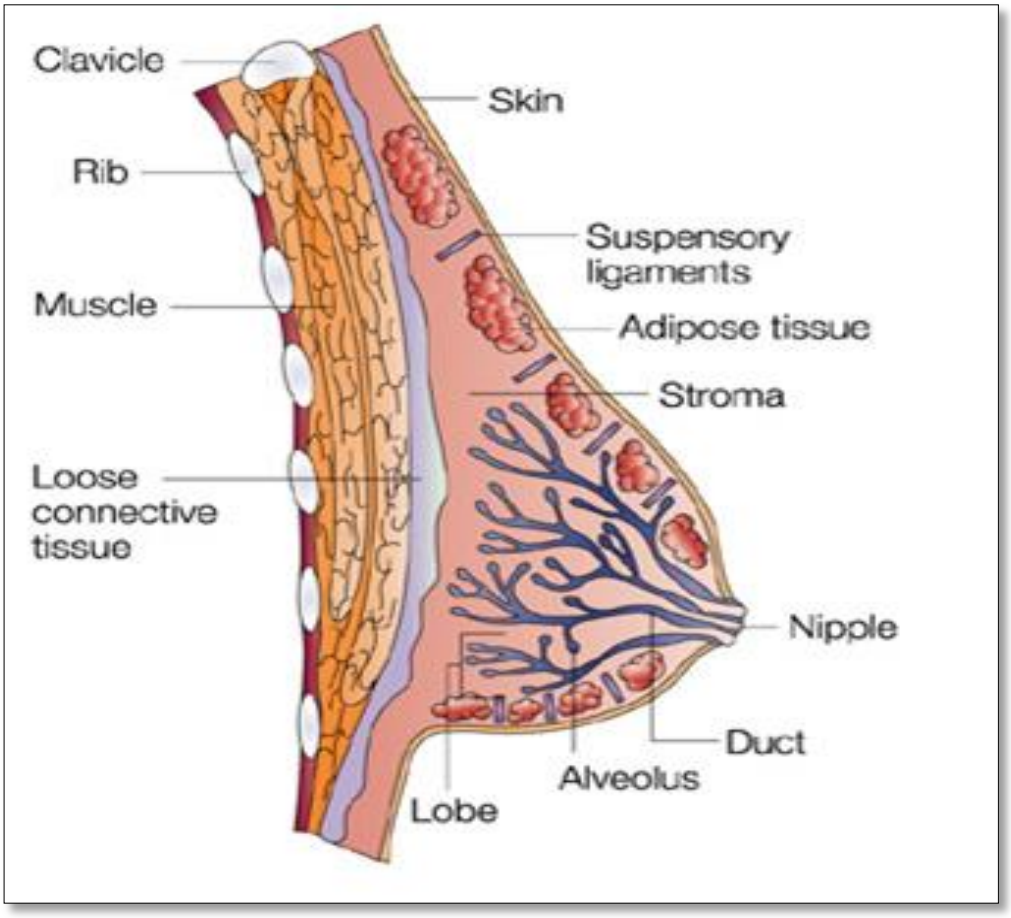

Figure 3: Schematic representation of Human breast.

The histological analysis of ducts and lobules show that these structures are surrounded by myoepithelial cells of transverse orientation. However they are lined by luminal epithelial cells of a single layer. The basement membrane serves to separate these structures from stroma. The occurrence of a breach in the basement membrane can be a distinguishing feature between carcinoma in situ and invasive carcinoma. Microenvironment of breast has components of stroma; the stromal components include some discrete cells such as immune cells, fibroblasts, adipocytes, extra cellular matrix and blood vessels (Bertos and Park, 2011).

1.6 Breast Cancer Subtypes and Stages

Based on morphology and structural organization, the tumours of breast have been divided into several categories. The invasive ductal carcinoma is the most commonly reported and observed one (approximately 75% cases; IDC NST) followed by invasive lobular carcinoma or ILC which is in about 10% of the cases. These two types of carcinomas together are responsible for 90% of all breast cancers. Some other types are histologically divided into categories such as tubular, medullary, metaplastic, neuroendocrine, mucinous (A and B), apocrine, adenoid cystic and micropapillary (The Royal College of Pathologists., 2015). The prognosis of cancer is dependant highly upon its histological type. It has been observed that the overall outcome of tubular, adenoid cystic, mucinous carcinomas is better than the ten year survival rate for ILC, medullary, IDC NST and apocrine variants. (Li et al., 2005, Weigelt et al., 2010)

Breast cancer is divided into four different clinical stages based on tumour, nodes and metastases stages (TNM staging)

TNM stage 1 can be divided into two main phases. In stage 1A the tumour is confined to breast area and has a diameter of $\leq$ 2cm. While in stage 2A the cancer spread to lymph nodes that are near to the breast area.

Similarly, stage 2 is also divided into two phases. In stage 2A the tumour size may range from 0 to 2 cm and cancer cells can be found in up to 3 lymph nodes within axilla or breast bone. The tumour size greater than 2 cm but less than 5cm with no lymph nodes is also considered as stage 2A. While, in stage 2B the tumour is larger than 5cm but has not spread into the lymph nodes or breast bone.

Stage 3 is divided into three phases namely 3A, 3B and 3C. In stage 3A the tumour may be of any size or cancer cells can be found in 4 -9 glands under axilla or lymph nodes near breast glands. Or if a tumour size ranges from 2-5 cm and small clusters of breast

cancer cells are in lymph nodes. Or tumour size greater than 5cm and it has spread up to 3 lymph nodes. While, stage 3B refers that tumour has spread to the skin of breast or chest wall, or made the skin break down or caused swelling, or cancer may have spread to 9 lymph nodes in the axilla or in lymph glands near the breast bone. In stage 3C the tumour is spread into the 10 or more lymph nodes in the axilla, or to the nodes above or below of the collar bone, or near to the breast bone.

Stage 4 refers to metastatic state with tumour of any size, the lymph nodes may or may not be involved but the cancer has spread to other parts body such as bones, lungs, liver or brain and distant lymph nodes (Cancer Research UK., 2014).

While, the breast cancer subtype classification is depending upon immunoreactivity of progesterone receptor (PR), estrogen receptor (ER), and human epidermal Growth Factor 2 receptor (HER2) and can be divided into four subtypes;

- Luminal-A (ER+/PR+/HER2-)
- Luminal-B (ER+/PR+/HER2+) (or HER2- with high Ki67)
- HER2 over-expressing subtype (ER-/PR-/HER2+)
- Basal like/triple negative (ER-/PR-/HER2-)

1.7 Breast Tumours – Types and subtypes

Classifications of breast tumour types and subtypes are based on classical pathology of diseased tissues, histological variations in terms of structure and localization, immunopathology, genomics and transcriptomics.

Based on morphology and structural organization, breast tumours are classified into eleven histological types (Li et al., 2005, Weigelt et al., 2010). The different features of these tumour subtypes are mentioned in Table 1.

Table 1: Pathological classification based on morphology and structural organization of tissue biopsy samples

S. No	Histological type	Features
1	**Invasive ductal carcinoma**	Cancer starts spreading from wall of milk duct and invades breast tissue (≈80% cases)
2	**Invasive lobular carcinoma**	Cancer starts in the terminal ductal lobular unit (TDLU) and then spread in breast tissue (≈10% cases)
3	**Medullary carcinoma**	Cancer starts in the milk ducts and soft tumour resembles medulla of brain (≈3-5% cases)
4	**Neuroendocrine carcinoma**	Tumours derived from neuroendocrine cells of breast (≈0.1% cases)
5	**Tubular Carcinoma**	Tumour begins inside milk ducts and cells look tube-like under microscope (≈2% cases)
6	**Apocrine Carcinoma**	Tumour begins inside milk ducts and cells look like sweat gland cells under microscope (< 1%)
7	**Metaplastic Carcinoma**	Tumour is rapidly growing and consistently larger than typical IDC (< 1%)
8	**Mucinous Carcinoma**	Tumours begins in the milk ducts and produce mucus (< 2-3%)
9	**Adenoid cystic Carcinoma**	Resemble salivary gland counterpart and associated with ductal epithelium and basal or myoepithelial cells (≈0.1%)
10	**Micropapillary Carcinoma**	Clusters of tumour cells with predominant clear spaces and resemble angiolymphatic vessels (≈6%)

Based on immunopathological parameters, breast tumours are classified on four different subtypes. This classification is mainly based on principal markers such as progesterone receptor (PR), estrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2). Expression of ER and PR markers are characterised by immunohistochemistry (IHC) methods whereas HER2 marker is assessed by combination of IHC and FISH approaches (Parise et al., 2009, Wolff et al., 2007). The features of these subtypes are mentioned in Table 2.

Table 2: Subtype classification based on detection of antigens in cells of tissue section and tumour genetic characteristics.

S. No	Subtype Name	Features
1	**ER+**	ER+/HER2-
2	**HER2+**	ER-/HER2+
3	**Triple positive**	ER+/PR+/HER2+
4	**Triple Negative**	ER-/PR-/HER2-

Based on Micro-array based gene expression studies, breast tumours are classified into five subtypes (Table 3). This classification is one step advanced to immunopathological classification (Hu et al., 2006, Sorlie et al., 2001).

Table 3: Subtype classification based on genome-wide expression changes in diagnosis of disease.

S. No	Subtype name	Features
1	**Luminal A**	ER+/PR+/HER2-
2	**Luminal B**	ER+/PR+/HER2+
3	**Basal**	ER-/PR-/HER2-
4	**HER2**	ER-/PR-/HER2+
5	**Molecular apocrine**	ER-/HER2+

Genomic aberrations such as DNA copy-number alterations (CNAs) and gene expression profiles are useful to dissect disease complexity and stratify breast tumours into intrinsic gene expression subtypes (Natrajan et al., 2010, Jonsson et al., 2010).

Based on gene expression data and CNAs breast tumours are divided into 10 subtypes (Table 4). Carlos Caldas group has classified breast cancer based on CNAs and its associated gene expression changes. The ten integrative clusters were identified and these are distinct in its clinical features and outcomes (Dawson et al., 2013).

Table 4: Subtype classification based on alterations of the DNA of a genome of diseased biopsies

Integrative Cluster	CNAs Features
IntClust 1	Amplification of 17q23 and highest prevalence of GATA3 mutations (Luminal B intrinsic subtype)
IntClust 2	Amplification of 11q13/14 and this region possess CCND1, EMSY, PAK1(ER-positive tumours)
IntClust 3	Low prevalence of TP53 mutations and high frequency of PIK3CA, CDH1 and RUNX1 mutations (Luminal A tumours)
IntClust 4	Low levels of genomic instability, and CAN-devoid flat copy number landscape (both ER-positive and ER-negative tumours)
IntClust 5	ERBB2 amplification at 17q12, intermediate level of genomic instability and high proportion of TP53 mutations (mixture of ER-positive, ER-negative and triple negative tumours)
IntClust 6	Amplification of the 8p12 locus and high levels of genomic instability (ER-positive tumours)
IntClust 7	Intermediate levels of genomic instability including 16p gain and 16q loss and high frequency of 8q amplification (ER-positive luminal A tumours)
IntClust 8	1q gain/16q loss and high levels of PIK3CA, GATA3, and MAP2K4 mutations (ER-positive luminal A intrinsic subtype)
IntClust 9	8q cis-acting alterations and 20q amplification and deletions of PPP2R2A on 8p (Mixture of intrinsic subtypes)
IntClust 10	Highest rates of TP53 mutations and 5q loss and gains at 8q, 10p and 12p (mostly triple negative tumours)

1.8 Current Screening and Diagnosis and limitations

In UK, national screening programmes are open to every woman for breast cancer screening. In this process

1. **Palpation**

   This is a clinical examination but recommend to patients to physically examine any lumps in their own breasts including behind the nipple and axilla. If any benign lump is reported in this area, it is reported to doctor (Thistlethwaite and Stewart, 2007).

2. **Mammography**

   This imaging approach uses x-rays to identify breast cancers based on density changes. The limitations of this approach are generally 10-14 percent false negatives and it can be quite uncomfortable. Mammogram does not provide much information in younger women because of their high density of breast and less percentage of fat (Nyström et al., 1993, Bassett et al., 1991, Day, 1991)

3. **Ultrasound**

   This approach applies high frequency sound waves to detect breast cancers and it is usually helpful in younger women. This technique is widely used along with mammography and it is good in identifying breast tumours (Harper et al., 1981).

4. **Biopsy procedures**

   a. **Core needle biopsy**

      X-ray guided core needle is applied to anaesthetized area of breast tissue to collect a small portion of lump. Radiologist uses X-ray to guide the area of tumour and occasionally tissue sample needed is not visible on ultrasound. The limitations of this technique are slight bleeding, pain, and risk of infections, and sometimes missampling the area of suspected cancer (Weinfurtner et al., 2014).

   b. **Fine needle aspiration**

In this approach, fine needle is applied to collect suspected breast cancer cells but it can be painful and rarely needs local anaesthesia. This method has high sensitivity in discrimination between cysts and sold tumours but it cannot discriminate between in situ and IDC (Oyama et al., 2004).

5. **MRI (Magnetic resonance imaging)**

   This approach involves radio waves to identify breast lesions. It can reveal the size of tumours in lobular breast cancer patients more accurately which can guide surgical biopsy. This approach is more accurate than mammography and breast ultrasound. The limitations of this approach are high cost and lack of potential in visualization of both ductal carcinoma in situ and IDC (Orel, 2001).

1.9 Vibration Spectroscopy – novel diagnostic approach

Although vibrational spectroscopy has been used in the structural characterisation of biological molecules alongside X-ray diffraction studies since 1940's, it attained great resurge of attention in late 80's in disease diagnosis and screening approaches. RS is a vibrational approach and has allowed electromagnetic radiation such as UV, visible and near infrared regions to excite higher energy levels. The energy difference is used to vibrate molecules (Rehman, 2012).

The first vibrational approach towards biological studies was reported in 1930's in skin tissues. Later on this approach was employed to detect microorganisms such as bacteria and viruses. Initial studies were concentrated on exploring major biological components of tissues such as lipids, carbohydrates and proteins and these were helpful in understanding metabolisms. Infrared spectroscopy has attracted many scientists to work on biological tissues due to advancement in collection of spectra from aqueous solutions and subtraction of water peaks from raw spectra and explore molecular fingerprint of biological materials. Overall, the early era of vibrational research was dominated by infrared spectroscopy compared to RS. Raman analysis of biological tissue was

hampered with intense fluorescence and high cost of equipment. As time has progressed, advancements such as combination of microscopy and FTIR spectroscopy, powerful lasers, sensitive detector systems and fast computers have revolutionized infrared spectroscopy and it became powerful tool in non-invasive and label free approach (Banwell and McCash, 1983). The advent of FT-Raman spectroscopy and Nd:YAG lasers has eliminated fluorescence problems in early years and furthermore, invention of charge-coupled device (CCD) detectors, efficient filters has solved signal to noise ratio problems. Moreover, the advent of Confocal Raman spectroscopy (CRM), Surface Enhanced Raman Spectroscopy (SERS) and Coherent Anti-Stokes Raman Spectroscopy (CARS) has transformed RS related biological research. Novel diagnostic approach of RS in each disease is summarised in the following table.

Advantages of Spectroscopy

Optical spectroscopy has played vital role in cancer detection in recent years. The major advantages of this approach are mentioned below (Rehman, 2012)

- This is a non-invasive and label free approach and requires minimal sample preparation.
- Spectra can be acquired in a short space of time that leads to quick detection approach.
- It can provide chemical information.
- It can elucidate structural information.

Applications of spectroscopy in tissue imaging and disease diagnosis are pretty exciting. Raman has shown promising approach in diagnosis of various cancers on ex vivo biopsies. It may also helpful in margin assessment studies and that would reduce the need of reexcision surgeries resulting from positive margins. Implementation of these methods as tools in clinical field requires large scale of *ex vivo* and *in vivo* studies, and further development of diagnostic algorithms. Recent developments in Raman systems such as SERS, tip-enhanced Raman and

CARS have convinced the usage for *in vivo* human applications. SERS approach use nanoparticles to yield Raman information on the scale of 100nm. Biocompatible and non-toxic nanoparticles were used as SERS substrates in mouse models to diagnose cancer and inflammation. Tracking of functionalized nanoparticles in human body remains significant challenge and clinical usage of nanoparticles in human applications still in research phase.

1.10 Types of Spectroscopy

Based on regions of spectrum, molecular spectroscopy can be divided into five types namely nuclear magnetic resonance spectroscopy, electron spin resonance spectroscopy, rotational spectroscopy, vibrational spectroscopy and electronic spectroscopy. NMR and ESR spectroscopies were observed in radio frequency region. The charged particles like nucleus and electron and their dipole associated reversal spin causes an absorption or emission spectrum. Rotational spectroscopy was observed in microwave region and molecules, which have permanent dipole moment, exhibit changes in rotational levels and that energy transitions produce a spectrum. Vibrational spectroscopy lies in infrared region and this causes a dipole change. Different types of vibrations such as symmetric and anti-symmetric stretch, and bending vibrations make molecules infrared active. Electronic spectroscopy exhibit in the visible and UV region and this is caused by excitation of valence electrons leading to change in electric dipole.

Different types of spectroscopic techniques and their important features are summarised in Table 5 here below (Rehman, 2012)

Table 5: Different types of spectroscopy and features.

Spectroscopy	Feature	Region	Wave-length	Frequency (Hz)	Energy (joules/ mol)
NMR and ESR	Change of spin	Radio-frequency region	10m $^{-1}$ cm	3X10^{6} to 3X10^{10}	10^{-3} to 10^{-1}
Rotational spectroscopy	Change of orientation	Microwave region	1cm - 100μm	3X10^{10} to 3X10^{12}	10^{2}
Vibrational spectroscopy	Change of configuration	Infra-red region	100μm - 1μm	3X10^{12} to 3X10^{14}	10^{4}
Electronic Spectroscopy	Change of electronic distribution	Visible and UV region	1μm -10nm	3X10^{14} to 3X10^{16}	10^{7}

1.10.1 Elastic Scattering Spectroscopy (ESS)

It is a wavelength dependent scattering approach and it mainly relies on elastic scattering. The scattered light provides information regarding normal and cancerous tissues. Spectral signatures are unique to cellular and sub cellular organelles depending on size, chromatin intensity of nucleus and morphological features of cell organelles (Dhar et al., 2006).

1.10.2 Fluorescence Spectroscopy (FS)

The basic principle of fluorescence spectroscopy is to quantify intensity of emitted photons from sample molecules after collision with incident photons. In this approach, a fluorophore, a fluorescent chemical compound that binds to sample, is excited into a higher state from ground level after interaction with incident light at specific wavelength. Upon reaching higher state, the fluorescence signals are released at longer wavelengths than incident absorption. The emitted energy is dependent on both fluorophore and its chemical environment. This spectroscopy is applied to investigate structural changes and binding interactions between molecules (Lakowicz, 2013).

Tryptophan, an aromatic amino acid, is one of the main components of fluorophore and is helpful to understand biological interactions using their wavelength. In vivo FS was applied on skin biopsies to investigate tumour pathologies in basal and squamous cell carcinomas. Increase in Tryptophan and decrease in collagen emission signals related to skin malignancies were first reported (Brancaleon et al., 2001). A combination of diffuse reflectance and intrinsic fluorescence spectroscopes were used to diagnose breast cancer with highest sensitivity and specificity (Zhu et al., 2008, Volynskaya et al., 2008) .

1.10.3 Fourier Transform Infrared Spectroscopy (FTIR)

FTIR spectroscopy, an infra-red (IR) vibrational spectroscopic approach, can be used to explore biochemical changes at molecular level. The basic principle in IR spectroscopy is interaction of IR radiation with molecules which results in the transformation of signal time domain to frequency domain. Absorption spectrum is the result of FTIR spectroscopy whereas scattered spectrum is the result of RS. The energy absorbed by the molecules causes vibrations and that leads to net change in dipole moment. FTIR has been extensively used in non-aqueous samples due to high absorption of water.

FTIR applications have shown huge potential in the disease diagnosis such as breast cancer, brain cancer, bone cancer, cervical cancer, colon cancer, gastric cancer, lung cancer, oral cancer and skin cancer. FTIR applications were also helpful in studies of anti-cancer drugs, monitoring cell proliferation and growth, and in vitro glucose measurement. The development of fiber-optic probes has opened more avenues to investigate disease diagnosis in-vivo and in-vitro biopsies (Movasaghi et al., 2008).

1.10.4 Raman Spectroscopy

Raman spectroscopy had been extensively used to detect biochemistry of several biological compounds for many years. Raman has received much attention in past decade especially in biomedical applications such as diagnosis and monitoring disease

progress in various cancers. However, recent advancements in laser and detector technology have assisted Raman to become a potential approach in medical diagnostics (Mahadevan-Jansen et al., 1997).

Principle of Raman Spectroscopy

The basic principle involved in Raman spectroscopy is interaction between monochromatic light and matter. When light interacts with molecule, energy is transferred to molecule or vice versa. Most of the light will be scattered with same energy as the incident light and this phenomenon was first observed by Lord Rayleigh and is known as elastic or Rayleigh scattering (Rehman, 2012). A very minute fraction of light will scatter with different energy than the incident light and this was first discovered by Sir C. V. Raman in 1928. This phenomenon is commonly known as Raman effect or Raman scattering (Hanlon et al., 2000).

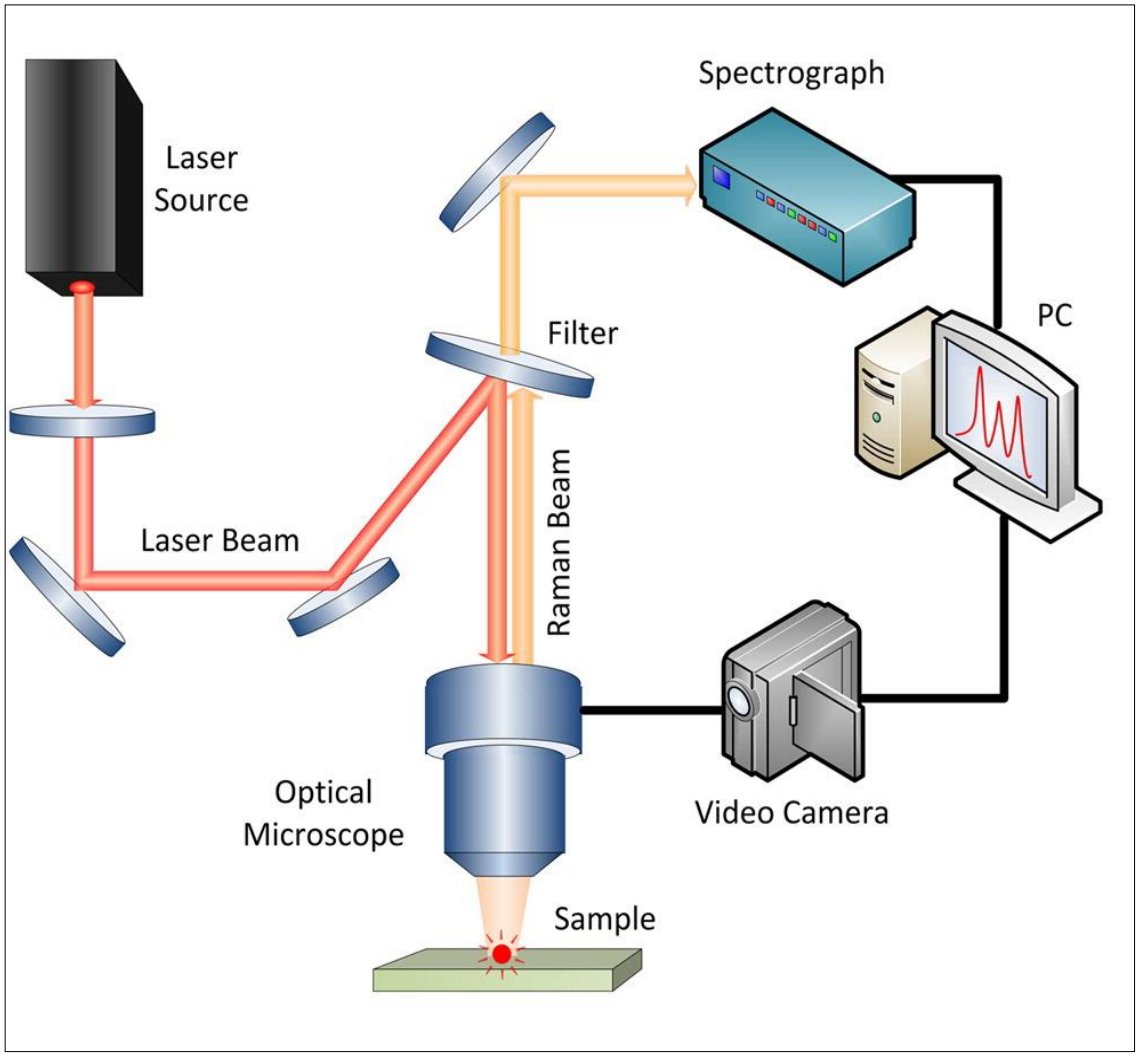

Figure 4: Schematic representation of Raman spectroscopy.

Raman scattering occurs when a molecule is excited to a vibrational state and it is associated with absorption and emission of a photon. When molecule absorbs energy during interaction, the scattered photon has lower energy than incident photon and this phenomenon is known as stokes Raman scattering (Rehman, 2012). During the interaction, if molecule loses energy, the scattered photon has higher energy than incident photon and it is known as anti-stokes Raman scattering (Figure 5). Raman spectrum is the representation of Raman scattered light. In graphical representation, x-axis denotes frequency shifts and y-axis denotes intensity absorption. The conventional units for frequency shifts and intensity absorption are wavenumber (cm^{-1}) and count per seconds (CPU) respectively (Hanlon et al., 2000). The first generation of Raman

instrumentation has used visible laser in their research. The major problems encountered here were fluorescence and instrumentation limitations. Tissue chromophores and complex biological compounds cause a lot of fluorescence in the spectra. Later on, invention of diode lasers and CCD cameras have diminished much florescence and improved spectral interpretations (Figure 4) (Mahadevan-Jansen et al., 1997).

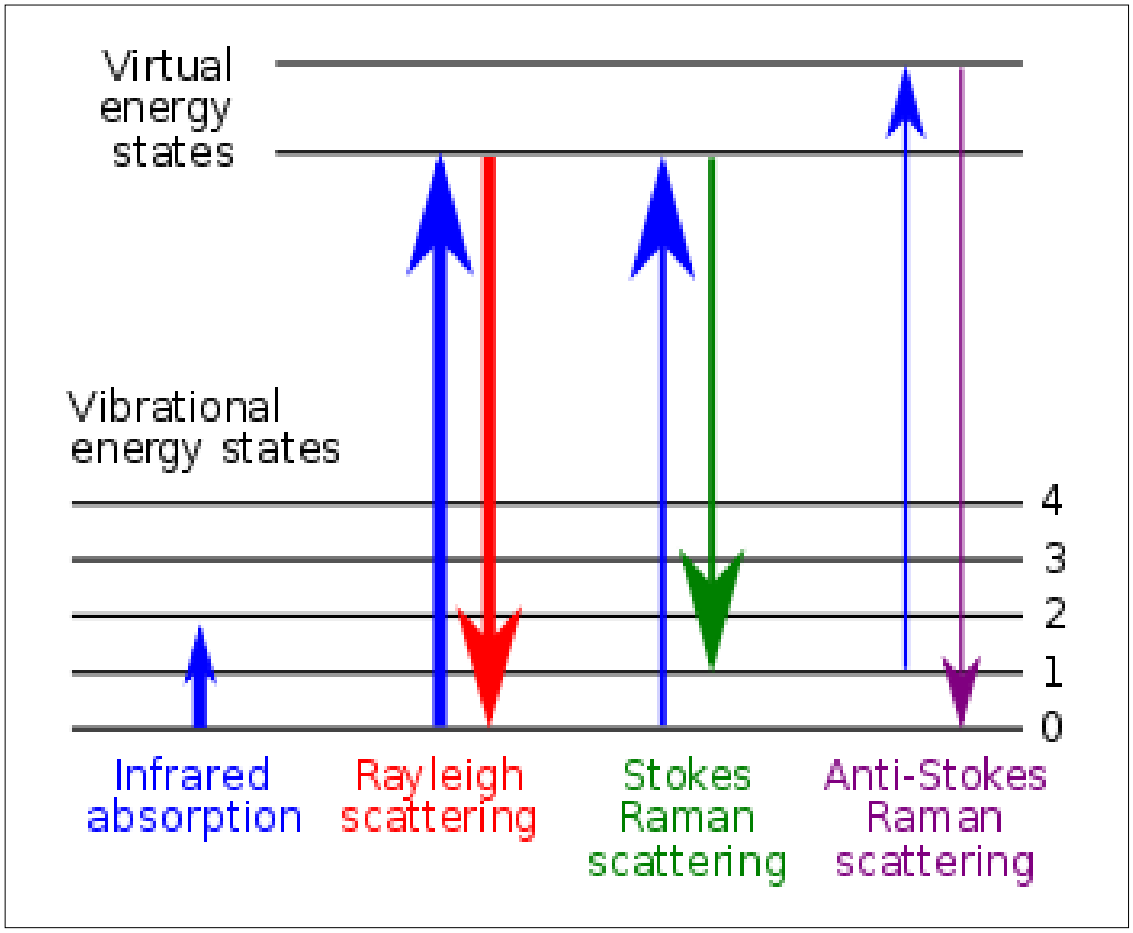

Figure 5: Raman scattering.

In Raman spectrum, each molecule express unique vibration and these molecular vibrations are commonly known as spectral peaks. These molecular vibrations are classified in two classes namely stretching and bending vibrations (Figure 6 & 7). In stretching vibrations, bond angles between the molecules remain same and bong lengths are changing during the vibration. There are two types of stretching vibrations namely asymmetric and symmetric. In bending vibrations, the bond length remains the same and bond angle changes during the vibration. There are four types of bending vibrations

namely rocking, twisting, wagging and scissoring (Lin-Vien et al., 1991).

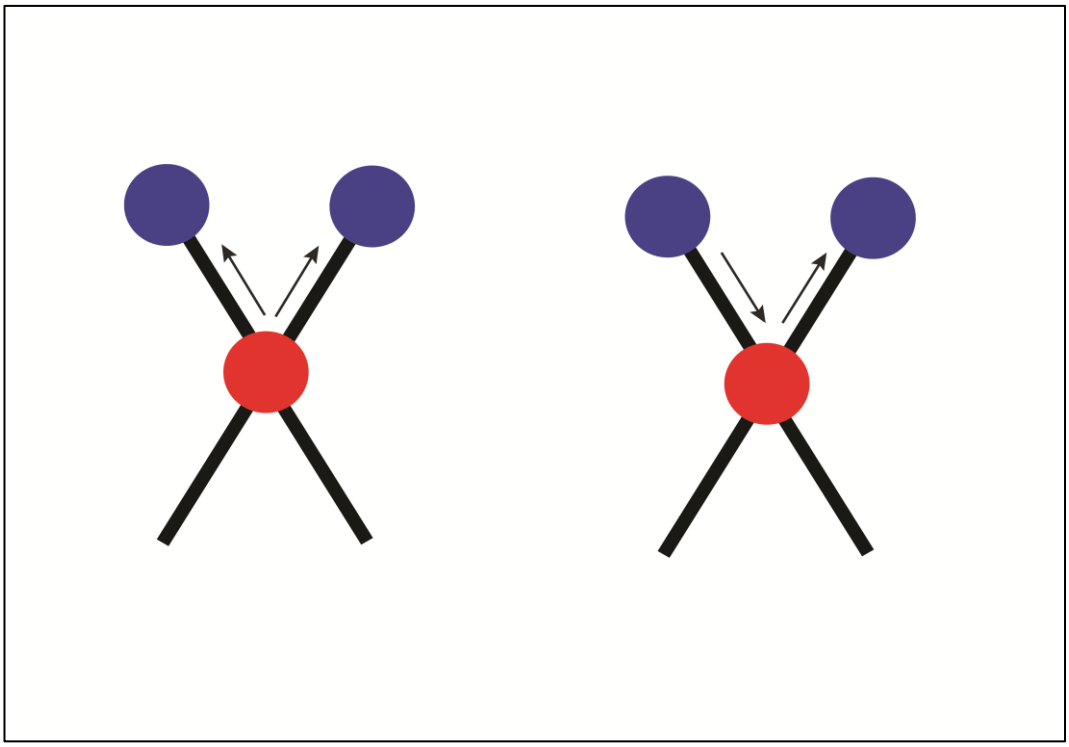

Figure 6: Schematic representation of stretching vibrations

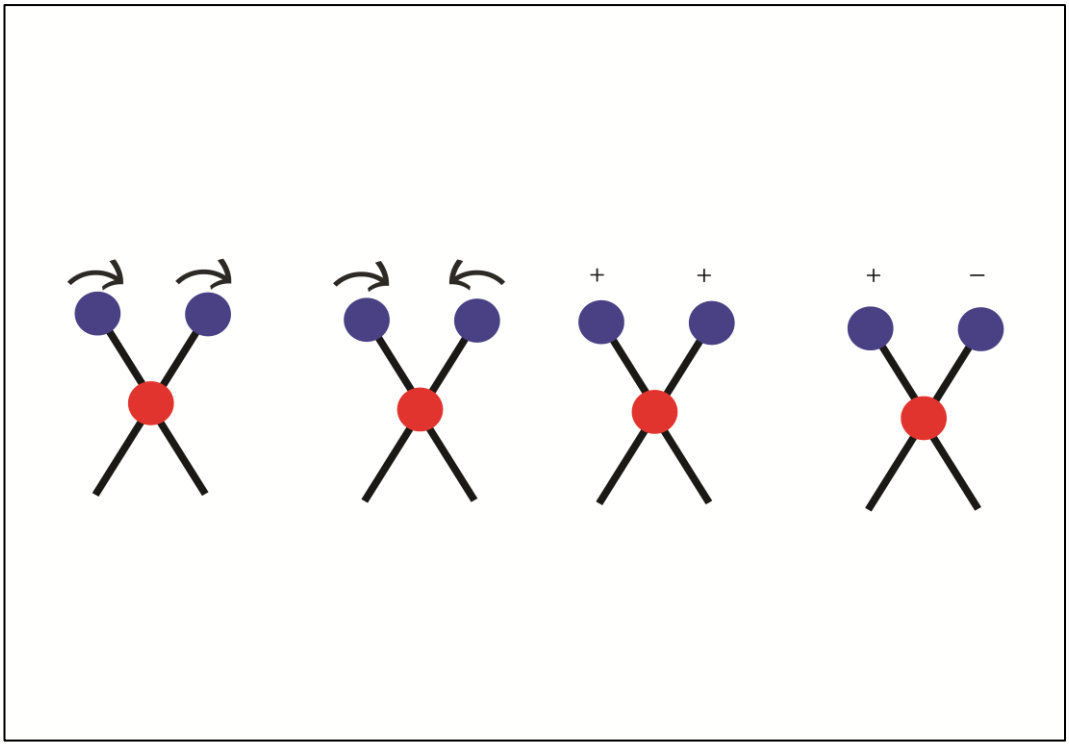

Figure 7: Schematic representation of bending vibrations

1.11 Types of Raman Systems

1.11.1 Surface Enhance Raman Spectroscopy

Surface enhanced Raman Spectroscopy (SERS) was first observed by Fleischmann and his colleagues in 1974 (Fleischmann et al., 1974). It involves substrates, which are generally rough metal surfaces and used to increase Raman signals by 10^3 to 10^6. The first use of gold and silver colloids as substrates was reported by Lee and Meisel and higher magnitudes up to 10^{14} were reported by Richard van Duyne (Lee and Meisel, 1982, Haynes et al., 2005). The frequently used SERS substrates are electrodes such as silver (Ag), gold (Au) and copper (Cu). Interactions between the laser and substrate results in the generation of plasmons, which creates an electromagnetic enhancement of the Raman signal. SERS was successfully applied on detection of breast cancer genes (Allain and Vo-Dinh, 2002), diagnosis of oral cancer (Kah et al., 2007), detection of CTCs in blood samples (Sha et al., 2008) and diagnosis of gastric and colorectal cancer (Feng et al., 2011a, Lin et al., 2011a).

1.11.2 Coherent Anti-stokes Raman Spectroscopy (CARS)

CARS is a type of non-linear spectroscopy, which employs multiple photons to generate coherent signals, several times the magnitude of spontaneous RS. Single continuous wave lasers are utilized in RS, in contrast to CARS, which employs two pulsed lasers (Tolles et al., 1977). Stokes scattering creates a photon that is similar to fluorescence emission. It hampers the detection of weaker Raman signals whereas anti-stokes scattering overcomes this problem. But it is weaker than stokes scattering. Two strong collinear lasers are used for irradiation. Frequency of the first laser is usually constant while second laser can be tuned create Raman active mode of interest. CARS signal is generated from vibrational motion of the molecules in the sample. A robust and directed signal is produced with CARS due to the coherent anti-stokes signal. The combination of CARS with laser scanning microscopes can provide high spectral images at video time rates (Cheng and Xie, 2004).

Recent studies using a combination of RS and CARS have gained attention in the

biomedical field especially in biomedical imaging of various biological tissues. CARS has successfully been employed to study pathological brain tissue (Evans et al., 2007) and lipid profiles of nerve cell membranes as well as myelin organization (Wang et al., 2005). Furthermore, based on myelin signals, CARS has been utilized to examine both the qualitative and quantitative distribution of grey matter in surrounding brain tissues (Evans et al., 2007). CARS was applied to colon tissue to investigate tumor features especially inflammatory margins using imaging (Krafft et al., 2009b).

Advantages of CARS:

1. CARS provides very high levels of spontaneous signals (~ 10^7) compared to RS.
2. It also generates spectra with very high resolution which is equivalent to resolution of monochromatic light.
3. It provides molecular structural characterization with higher collection efficiency and excellent spatial resolution compared to RS (Cheng and Xie, 2004).

1.11.3 Resonance Raman Spectroscopy (RRS)

Resonance Raman spectroscopy (RRS) utilizes specific excitation wavelengths which are close to the UV-visible absorption area. Overlapping of excitation wavelengths with the UV-visible range can produce higher scattering intensities, sometimes 6 times higher than normal RS. This phenomenon is known as Resonance Raman effect and optimum resonance depends on choosing the appropriate laser excitations corresponding to resonance conditions. Due to excitation coincidences, RRS shows high florescence compared to RS (Robert, 2009). RRS was applied to normal and breast cancer cells to investigate protein and nucleic acid changes (Yazdi et al., 1999)

and detection of different grades of brain cancer from normal meningeal tissues (Zhou et al., 2012).

Summary of Raman systems:

Type of Raman System	Raman features
Stimulated Raman system	• Non-linear phenomena in Raman signal • Pulsed lasers with 10^9 V/cm and improved signal-to-noise ratio • 4-5 orders of magnitude enhancement compared to the spontaneous Raman scattering
CARS system	• Two strong collinear lasers used for irradiation • Two coherent laser beams and resulting signal with Anti-stokes frequency • Frequency of the first laser is usually constant and second one can be tuned to create Raman active mode of interest
Resonance Raman system	• Excitation laser frequency crosses electronic excited states of samples and resulted into resonance • solved problems in usage in UV lasers in Raman spectroscopy • 3-6 orders of magnitude of intensity
SERS	• Certain metal surfaces (silver or gold colloids) with 5-6 orders of magnitude • Chemical enhancement followed by electromagnetic enhancement • SERS attached fibre optic probe (tip-enhanced Raman) applied in *in vivo* investigations

1.12 Chemometrics

Chemometrics, also known as multivariate data analysis, was first coined by Svante Wold in 1971. The use of chemometrics enables the extraction of complex information from chemical systems. It uses different disciplines such as applied mathematics, statistics and information science to extract crucial chemical information (Osawa et al., 2002). This can be achieved in two stages. In the first stage, chemical information is extracted as concentrations from spectral regions and in the second stage, these concentration profiles are compared using supervised and unsupervised algorithms to extract covariance factors that might help in understanding tissue pathology (Osawa et al., 2002). This form of multivariate data analysis is divided into two classes i.e. unsupervised and supervised algorithms.

1.13 Unsupervised algorithms

These algorithms are used to separate spectral data into groups based on systematic variation. Complex information is extracted from data sets and is commonly known as principal components (PCs) and each PC represents the linear combination of original spectral data components and these are orthogonal to each other. The most commonly used unsupervised algorithms in Raman based cancers studies are Principal component analysis (PCA) and Cluster analysis (CS) (Krafft et al., 2009c).

1.13.1 Principal Component Analysis (PCA)

As described earlier, this approach is performed on spectral data without any prior knowledge. Spectral data is converted into two matrices and are commonly known as PCs or scores. These scores are demonstrating the highest amount of variation. These are described as PC-1, PC-2 and PC-3 and so on. PC-1 shows highest amount of variation and PC-2 shows second highest amount of variation and so on. As the PC score increases the amount of variation decreases and latter PCs exhibits higher noise levels. PCA was used to separate normal and diseased tissues in breast (Chowdary et al.,

2006), lung (Magee et al., 2009), oral (Malini et al., 2006a), gastric (Bergholt et al., 2011b) and colon (Molckovsky et al., 2003) cancers.

1.13.2 Cluster Analysis (CA)

Cluster analysis is another example of an unsupervised approach and it divides data in to subsets based on variations. The aim of this approach is to enhance the degree of variation between clusters and reduce the degree of variation within each cluster. The amount of variation is defined as distance measure (Krafft et al., 2009c). There are two types of cluster analysis approaches, one is hierarchical cluster analysis (HCA) and the other is k-mean cluster analysis. HCA algorithm evaluates pairwise similarity from number of spectra (n) in order to calculate the symmetric distance matrix. This can be achieved in two steps. In the first step, it investigates the minimum distance from all spectra and extracts two similar spectra as a first cluster and then recalculates spectral differences between all remaining spectra and the first cluster. The second step involves searching for further similarity between remaining spectra or clusters and this process reiterates (n-1 times) to form a single cluster. This results in the formation of a 2D dendrogram, which corresponds to the lessening of clusters on one axis and respective spectral distances on the other axis. HCA can be divided into four groups based on the way in which the distance between clusters is measured. They are HCA-single linkage (nearest samples), HCA-complete linkage (farthest samples), HCA-average linkage (between nearest and farthest) and ward's method (measures homogeneity and heterogeneity within each cluster) (Krafft et al., 2009c, Beebe and Kowalski, 1987).

1.14 Supervised algorithms

These algorithms use training sets with reference data with prior knowledge of groups contained within to test new, unlabelled datasets. Initially few spectra are used to train a model with defined classes. The class membership of, unlabelled spectra are then predicted based on their proximity to each defined class in the training model. The most

commonly used supervised algorithms in Raman studies are linear discriminant analysis (LDA) and k-means cluster analysis (Krafft et al., 2009f).

1.14.1 Linear Discriminant Analysis (LDA)

LDA is arguably the most popular supervised chemometric tool used for spectroscopic studies. It utilizes information to estimate linear discriminant functions that optimally differentiate k classes. LDA uses this information to assign unknown spectra to predefined classes. Discriminant functions enhance the interclass variance as well as intraclass variance. The mahalanobis distance is a measurement of common covariance matrix used in LDA, whereas quadratic discriminant analysis is based on distance measured from class-specific covariance matrices. Raman associated LDA approach has been employed in several cancer studies to classify disease from normal tissues types (Kamemoto et al., 2010, Oshima et al., 2010, Li et al., 2012, Feng et al., 2011b, Yu et al., 2006, Talari et al., 2015a).

1.14.2 K-means cluster analysis

In K-means clustering, the whole data sets are divided into k number clusters and then indiscriminate selection of centroids. Each spectrum of cluster is allocated to the nearby centroid. New centroids are then calculated and new assignments are made. This process repeats until the cluster resolution achieved. K-means produces an averaged spectrum for each cluster together with a class membership matrix (Krafft et al., 2009f). Most Raman studies involve a k-mean cluster analysis approach based on PCA-transformed Raman data sets (Kneipp et al., 2003, Koljenović et al., 2002, Draux et al., 2009, Vogler et al., 2010).

1.15 Raman spectroscopy in Biological Tissues

1.15.1 Cervical cancer

In this study, near- infrared Raman spectroscopy was used as molecular marker detective to discriminate between normal and invasive cervical cancer tissue samples.

This study has concentrated on the final stage of invasive carcinoma and collected 14 tissue samples from four healthy individuals with no cervical cancer record and three patients with cervical squamous cell carcinoma. The wavelength of the laser used was 785nm to excite both normal and cancer cells. Several prominent bands were reported in the normal squamous cells such as δ (CCH) aliphatic of collagen (816 cm^{-1}), ring breathing in tyr & CCH deformation (854 cm^{-1}), C-C stretch (922 cm^{-1}), C-C skeletal stretch (938 cm^{-1}), phenylalanine & C-H in-plane bending (1003 cm^{-1}), O-P-O backbone stretch of DNA (1101 cm^{-1}), amide III (1247 cm^{-1}), δ (C=CH) (1273 cm^{-1}), CH_2 deformation of lipids (1321, 1450 cm^{-1}), DNA (1342 cm^{-1}) and amide I (1664 cm^{-1}). The outcome of this work has shown two specific differences between normal and malignant cells. Dominant Raman bands at 775 to 975 cm^{-1} and the amide III at 1248 cm^{-1} were present only in normal cell. This investigation has also revealed that C-H stretching mode at 2800 to 3100 cm^{-1} is six times lower in the factor in cancer cells when compared to normal cells (Kamemoto et al., 2010).

Gonzalez-Solis et al applied Raman spectroscopy along with multivariate approach to investigate cervical cancer using blood samples. In this study, serum was obtained by fresh blood samples from 20 patients and 10 healthy volunteers and in total 240 spectra were collected from these samples. 80 spectra were collected from 10 control patients, 160 spectra from 10 cervical cancer and 10 early cervical cancer patients. PCA and LDA were carried after removing the fluorescence contribution, smoothing and applying the baseline correction. Raman spectra have shown clear evidence in peak intensities between normal and cervical cancer states. The peaks of phenyl alanine and ß-carotenes of the cervical spectrum are higher than the peaks of the control spectrum. This study has also found that peaks of adenine, tryptophan, proline, glutathione, ß sheet and phospholipids, ß –carotene were present only in the cancer spectrum, whereas the peaks of amide III was present only in the control spectrum. This study has shown vivid

chemical variation in the serum samples of healthy and clinically diagnosed patients (Gonzalez-Solis et al., 2010).

1.15.2 Lung cancer

Oshima et al used Raman spectroscopy to demonstrate differences among cultures of normal and cancerous lung cell lines. Low to medium malignant Adenocarcinoma, squamous cell carcinoma, high malignant small cell carcinoma and large cell carcinoma, cell lines were cultured for Raman measurement. Single cell Raman spectra were obtained by using 532nm excitation wavelength instead of 785 or 633nm. High quality spectra were obtained at 60 sec/cell without any damage. Raman markers in lung cancer cell line and their malignancy were to establish in this study. The spectra has shown strong band for cyc-c and laser contributed much resonance in the spectra. PCA was successfully applied and 80% accuracy was achieved in discrimination between cancer cell lines. Step-wise LDA was performed on these cell lines to identify malignancy type and strain. The cross validation has provided 100% accuracy of discrimination. The current study has shown potentiality of 532nm laser in Raman diagnosis (Oshima et al., 2010).

Li et al has employed Surface Enhanced Raman Spectroscopy (SERS) to investigate peritoneal fluid, so that diagnosis of lung cancer can be possible at early stages. SERS was performed on saliva samples to detect vibrational features of the lung cancer pathology. Raman spectra of control and cancerous sample have shown peaks at 791 cm^{-1} (pyrimidine), 906 cm^{-1} (tyrosine), 1007 cm^{-1} (phenylalanine & carbamide), 1364 cm^{-1} (tryptophan) and 1720 cm^{-1} (lipid). The only minute differences such as peak intensities were found between the both spectra. This group has used multivariate analysis, such as, PCA to identify major differences and LDA to distinguish both groups. PCA and LDA were successfully employed and 82% total accuracy was

achieved. This study has shown promising approach of SERS in detection of lung cancer (Li et al., 2012).

Raman spectroscopy has been used by Huang et al to chemically characterize different organs of mice. This review has stated no spectral differences between blood pellet and whole blood. Oxygen saturation in hemoglobin has enhanced specific bands expression in the Raman spectra. Lipid and protein related CH_2 and CH_3 vibrations were expressed in the same regions of the spectra of serum, blood pellet and whole blood. Raman spectra of adipose tissue have shown much resemblance with spectra of palmitic acid. Stomach, colon and intestine has shared many spectral features, only two peaks were observed from protein and DNA and the remaining peaks were observed from lipid and protein. The spectra of teeth and bone has expressed calcium band at 950 cm^{-1} and it has been used as marker for detection of calcium phosphate in arthritis patients. The bands observed in this study are representing major biological molecules of those particular organs. Few Raman bands were appeared exclusively in those specific organs such as 1017, 2151 and 2747 cm^{-1} (spleen) 876 cm^{-1} (kidney) and 921 cm^{-1} (small intestine and colon) (Huang et al., 2011).

Nawaz et al reported that Confocal Raman Micro-spectroscopy (CRM) has the potential to analyse the efficacy of chemotherapeutic agent (cisplatin) on A549 adenocarcinoma cells. The purpose of this study is to discover chemical changes in cell membrane and cytoplasm of A549 cells after 96 hours exposure to cisplatin. Multiple spectra were obtained from different regions of cell (i.e. cell membrane, cytoplasm and nucleus), the spectral data was used to analyses the difference in cellular components and peak changes after Cisplatin treatment, multivariate analysis techniques like PCA (Principal component analyser), PLSR (Partial Least Square Regression) and PLS jack-knifing were used. They identified a band at 726 cm^{-1} that present in cytoplasm spectrum but not in nuclear spectrum, this band is assigned to tryptophan and CH_3 stretching to lipids.

Similarly, the Raman band at 1423 cm^{-1} (aromatic lipids) and band at 1510 cm^{-1} assigned to tryptophan and tyrosine were observed predominately in cell membrane and cytoplasm spectra rather in nuclear membrane spectra. The result of Cisplatin on the spectra of cell membrane illustrated the bands related to proteins at 671 $cm^{-1}$and 728 cm^{-1} (ring breathing of the tryptophan), 1030 cm^{-1} (C-H bending), 1094 cm^{-1} and 1126 cm^{-1} (C-N stretching) and 1655 cm^{-1} (amide I). The Raman peaks at 1371 cm^{-1} (CH_3 stretching) and 1448 cm^{-1} (CH deformation), related to cell membrane lipids, also undergo shifts 1371 to 1376 cm^{-1} and 1488 to 1450 cm^{-1}. They also extracted protein from control and treated Cisplatin cells, and analysed the spectral differences. They suggested major changes in the Raman bands include an intensification of the 904 cm^{-1} and 1131 cm^{-1} (C-C skeletal stretching), 1001 cm^{-1} (phenylalanine), 1061 cm^{-1} (C-N stretching), 1131 cm^{-1} and 639 cm^{-1} (tyrosine), 961 cm^{-1} and 1199 cm^{-1}, 1011 cm^{-1} (tryptophan), 1323 cm^{-1} and 1404 cm^{-1} (CH deformation), finally 1606 cm^{-1} (C=C bending). PLS Jack-kniffing analysis of cell membrane has demonstrated that cisplatin seems interacting with membrane lipids and causes some changes in their structure. Jack-kniffing results for cytoplasm have suggested that changes in cell physiology in term of ring breathing of tryptophan, RNA and CH bending were observed due to Cisplatin action (Nawaz et al., 2011).

1.15.3 Gastric Cancer

Feng et al reported combination of surface-enhanced Raman spectroscopy and multivariate approach to detect gastric cancer in on plasma samples. Raman spectra were collected from healthy and cancerous samples using green laser. Spectral differences between normal and cancerous samples were substantial and reproducible. Cancerous samples have shown higher intensities at 1330 (ν (C-H) of nucleic acids), 1445 ($\delta(CH_2)$ of collagen and phospholipids) and 1580 cm^{-1} (δ(C=C) of phenylalanine). Spectral data was further analysed by principal component analysis and Linear

discriminate analysis. Principal components 1, 2 and 7 have successfully differentiated cancer from normal groups. This diagnostic approach has revealed sensitivity and specificity of cancer detection were 79.5% and 91% respectively (Feng et al., 2011b).

1.15.4 Brain cancer

Sato et al applied Raman spectroscopy to identify changes at molecular level in live mice brain tissue. The idea behind this study is to detect structural and molecular differences in live animals in response to different physiological changes. This study has used near-infrared Raman probe which is made up of ball lens hollow fibres and Ti-sapphire laser. The purpose of this study is to detect protein and lipid alterations, and water cluster conformation under different conditions namely inhalation of sodium pentobarbital (SP), diethyl vapour (DE) and after euthanasia. The spectra of SP treated has shown peaks at 1664, 1446 and 1003 cm^{-1} are consigned phenylalanine respectively. The spectra of DE treated have shown peaks at 2846, 2881, and 2928 cm^{-1} are associated to the CH stretching vibrations of the protein and lipid groups. The peak at 3453 cm^{-1} associated to water cluster and it signify that the variation in conformational clusters due to the inhalation of DE. The intensity of the water band has decreased approximately 8% in the spectrum obtained after euthanasia. Fetal analysis of the animal has revealed high concentration of water in the olfactory lobes and low concentration of water in the frontal cortex. This study has provided new insight for Raman approach to carry further analysis in conscious animals (Sato et al., 2009).

Kirsch et al reported that Raman imaging could be used to investigate inter-cerebral tumours in brain metastasis. *Ex vivo* brain tissues were analysed to examine cerebral pathology through Raman mapping. Spectra of normal tissue has shown peaks at 2933 cm^{-1} is associated with (CH_2 and CH_3 vibrations) and 3245 cm^{-1} (O-H vibrations of water). Other prominent functional groups appeared in the spectra were phenylalanine, CH_2/CH_3 deformations, amide I and fatty acid vibrations. The fingerprint region also

possesses peaks related to oxygenated haemoglobin at 750, 1002, 1212, 1546, 1605 and 1619 cm^{-1}. Raman spectra of the tumour have demonstrated melanin peaks at 597, 976, 1404 and 1595 cm^{-1}. This study has signified the importance of melanin as molecular marker for future studies in brain metastases. Hence further development in the fibre probe research can increase the prospect of Raman spectroscopy as a diagnostic method to detect metastasis in various organs of mice (Kirsch et al., 2010).

1.15.5 Skin cancer

Larraona-puy et al has employed Raman microspectroscopy (RMS) in order to access its accuracy in detection and imaging basal cell carcinoma (BCC). This automated evaluation was carried out on skin tissues excised during normal and Mohs micrographic skin surgery. The approach has developed a database from various tissue sections from twenty different patients and it was used to construct the multivariate classification model. The spectra were categorized into BCC, dermis or epidermis and collagen type I spectral features has mainly contributed to differentiate between the dermis and BCC. On the other hand DNA peaks provided evidence in discriminating BCC from epidermis. Healthy dermis was characterised by proline and C-C backbone vibrations of proteins. This study has shown high sensitivity and specificity in the discrimination of normal healthy tissue from BCC. Moreover, this approach has developed 2 dimensional biochemical images by using supervised models, which further supported a strong correlation with histopathological detection (Larraona-Puy et al., 2009).

Cartaxo et al reported FT-Raman spectroscopy could be used to differentiate cutaneous melanoma from pigmented nevus. Spectral measurements were taken from normal, neoplastic and pigmented nevi tissues by using Nd:YAG 1064nm laser (resolution 4 cm^{-1}). Before classification, Raman spectra were first pre-processed by correcting the baseline and performing the vector normalization using the Minitab software. The

model of variance analysis with a constant factor was used at first phase, and this indicated that all variables showed an affiliation with their own group. In the next phase of the statistical study a discriminating analysis was used, the results of which demonstrated a significant differentiation between the three groups under study. Significant differences were found in the region of 800-980 cm^{-1}. The bands of 855 and 937 cm^{-1} are typical characteristics of the collagen spectrum and are mainly because of the proline vibration in the protein structure. The Raman peak shift were observed in spectra are correspondent to DNA's vibrational mode also increase in spectral intensity was observed in primary melanoma. This is because of the increased duplication of genetic material, responsible for the proliferation of malignant cells. This approach has shown efficiency of 75% in discrimination among normal, neoplastic and pigmented nevi groups (Cartaxo et al., 2010).

Larraona-puy et al further reported that RMS could be used to differentiate hair follicles from BCC in skin explants during MMS Spectral differences were observed between epidermis, BCC and hair follicle. Raman spectral data of hair follicles are very similar to epidermis, While BCC Raman spectral data was dissimilar and has shown high amount of DNA. Spectral differences can clearly differential between hair follicles and BCC. Linear discriminate analysis has detected BCC with specificity and sensitivity of around 90% and 85% respectively. This study has developed automated images future diagnosis purposes. Raman images have revealed that nucleic acids were abundant in hair follicles compared to epithelial cells and these results have shown greater correspondence with histological images (Larraona-Puy et al., 2011).

Konig et al carried out multiphoton autofluorescence/second harmonics generation and Coherent anti-Stokes Raman Scattering (CARS) tomography on skin biopsies to detect biochemical changes, tissue architecture, intra-tissue cell morphology and accumulation of various products of pharmaceutical and cosmetic industries. This hybrid tomograph

has provided valuable information about outer most skin layer. Intracellular lipids of stratum corneum have bright polygonal patterns and distribution of keratin was also revealed in this study. *In vivo* CARS was used to study diffusion pattern of water and oil on the skin. These studies have found accumulation of oil in peripheral skin and water content in lipid free areas. CARS was used to distinguish differences between normal healthy skin and psoriasis. This group has noticed that healthy corneocytes were enclosed by greater lipid intensities whereas pathological skin lacked such structures (Konig et al., 2011).

1.15.6 Oral cancer

A Fourier transform near- infrared (FT-NIR) Raman spectroscopy was used to differentiate between normal and different oral cell carcinomas. Normal oral mucosa was acquired from SCC patients and were characterised into groups by pathologist. Raman spectra were recorded by using Nd : YAG laser (1064 nm, resolution of 8 cm^{-1}). Raman spectra of SCC have shown various molecular vibrations such as amide I phenylalanine, tyrosine, CH_2 vibrations, C-C skeletal stretching, C-N stretching. Support Vector Machine was applied to develop new model for discrimination and this model was successfully separated SCC from normal mucosa with greater efficiency compared to normal and OLK. Proliferation rate and biochemistry were similar in both normal and OLK cells. Hence low efficiency discrimination was noted by multivariate approach. Therefore, combination of Raman and data mining approaches has demonstrated potentiality in oral cancer detection. Moreover, investigation offered real time and label free diagnostic approach for detection of different oral cancers(Li et al., 2010, Madhavi et al., 2012).

Su et al applied Raman confocal micro spectroscopy to detect biochemical variations between normal and malignant oral tissues and normal and oral squamous cell carcinoma. Spectral profiles of normal and malignant have shown many marked

differences and multiple Raman markers were identified in this study. These markers were mainly associated with DNA and protein vibrational features to distinguish between normal and cancerous one. The advantage of unsupervised algorithm is to extract spectral differences in the form of principal components and these new scores will represent highest variation in the data. The most pronounced differences in the Raman spectra were appeared in the region of 700 to 1800 cm^{-1}. Normal tissues has shown Raman bands at 756 and 1546 cm^{-1} corresponding to tryptophan and Phenylalanine at 1004 cm^{-1} was observed both in normal and malignant. The malignant tissue display Raman bands at 1341 and 1655 cm^{-1}, which corresponds to purines of DNA and amide I of protein respectively. The essential findings of this work have proven that Raman spectroscopy offers enormous potential to distinguish normal and malignant lesions and it can assist histopathologists in great deal (Su et al., 2012).

1.15.7 Leukaemia

Neugebauer et al reported that Raman imaging could be used to detection and differentiation of circulating tumour cells (CTC) from peripheral blood. CTCs play important role in tumour diagnosis and these can be used as molecular markers for tumour therapeutic studies. The purpose of this investigation is to detect leukocytes, myeloid leukaemia cells and solid tumour cells by using combination of Raman spectroscopy and statistical approaches. Spectra were taken from dry cell pellet using 785 nm laser (resolution 4 cm^{-1}). General features of CTCs Raman spectra looked alike but leukocytes have shown higher intensities. Negative differences were observed due to high amount of DNA, lipids and proteins of cancer cells. Unsupervised statistical method such as HCA was employed on fingerprint region of the spectra and the resultant dendrogram has well separated leukocytes from other cell types. PCA studies were further supported HCA clustering results. Supervised approach such as SVM model was classified the cells with greater accuracy of approximately 98%. This

pioneering research could assist future studies in cancer detection as well as progression (Neugebauer et al., 2010).

1.15.8 Bladder cancer

Shapiro et al has employed Raman microscope to detect urothelial carcinoma commonly known as bladder cancer (BCa) from normal epithelial cells of the voided urine. Previous studies have attempted to discriminate BCa from normal bladder tissue by using Raman spectroscopy. Raman spectra were obtained normal and cancerous tissues by using 532nm visible laser. A distinct peak at 1584 cm^{-1} was observed in malignant tissue and it was absent in normal tissues. This group has developed a model based on peak height (1584 cm^{-1}) of normalized spectrum and used set of thresholds of the height to classify healthy, low and high grade cancers. Raman molecular imaging (RMI) yields a 92% of sensitivity and 91% of specificity for detecting BCa. This study has shown high accuracy in differentiating low-grade tumours from high-grade tumours (Shapiro et al., 2011).

A summary of important Raman spectroscopic investigations in biological tisses were summarised in Table 6.

Table 6: Raman spectroscopic diagnostic applications in various cancers.

Disease	Raman Approach	Biological tissue/ sample	Results	Ref
Brain Cancer	Raman Spectroscopy	Healthy and cancerous porcine tissues	Lipid content and lipid composition has differentiated from normal to cancerous	(Koehler et al., 2009)
Cervical Cancer	Raman micro-spectroscopy	Epithelial layers of normal and cervical biopsies	Collagen , glycogen and DNA peaks were differed and diagnostic accuracy	(Keller et al., 2008)

			was nearly 100%	
Lung cancer	Near infrared Raman spectroscopy	Normal, malignant tissues including squamous cell carcinoma (SCC) and adenocarcinoma	Disease classification is based on lipids, nucleic acids and aromatic amino acids. Malignant tissue have shown high nucleic acid and aromatic amino acid content and low in phospholipid and collagen based amino acid content compared to normal tissue	(Huang et al., 2005)
Lung Cancer	Surface Enhanced Raman Spectroscopy and chemometrics approach such as PCA and LDA	Saliva samples from normal and cancerous patient	Normal and cancerous samples have shown band intensity differences in pyrimidine bases, tyrosine, tryptophan and lipids. Multivariate approach has shown disease accuracy of 82%	(Li et al., 2011)
Gastric Cancer	Near infrared multichannel Raman spectroscopy and chemometrics	Normal and surgically resected samples from cancerous lesions	Sensitivity, specificity and overall accuracy of diagnostic algorithm is 95%, 100% and 98%, respectively.	(Kawabata et al., 2011)

			Approach will be helpful to clinicians in endoscopic submucosa surgery	
Oesophageal Cancer	Raman Spectroscopy	Frozen normal and oesophageal biopsies	Biochemical components such as lipids, glycogen and nucleic acid content were differed from normal to cancerous Potential for real-time diagnosis	(Shetty et al., 2006)
Nasopharyngeal cancer	SERS and PCA & LDA	Normal and tumour blood plasma	Collagen, phenylalanine and phospholipids were higher in cancerous than normal plasma Diagnostic accuracy in terms of sensitivity and specificity is 79.5 and 91%, respectively	(Feng et al., 2011b)

1.16 Raman Spectroscopy in Breast Cancer Research

Raman spectroscopy has been a successfully applied in breast cancer research. Initially Raman was employed to identify biochemical fingerprint of human breast biopsies. Near Infrared (IR) lasers successfully recognized major molecular components such as lipids including oleic acid and its derivative fatty acids. Visible lasers provided information on carotenoids (Frank et al., 1994). Raman was subsequently applied to differentiate normal breast from infiltrating ductal carcinoma. Significant and

reproducible Raman shifts were observed between normal and cancerous samples, associated with both fatty acids and proteins (Frank et al., 1995e). Furthermore, near-IR Raman studies identified biochemical components such as amide I, amide III, phospholipids, proteins and phenylalanine and consistent peak shifts were observed between the normal and cancerous tissues (Manoharan et al., 1998).

Haka et al has reported Raman to detect different grades of breast cancer by using qualitative and quantitative chemical information. This approach has used microspectroscopic model, which is built upon nine basic components of breast tissue. Modelling method has exactly mimicked the breast tissue and these models were used to predict and characterize the disease status of the breast biopsies. This study has established correlation between biochemical composition and disease aggressiveness. They developed a diagnostic algorithm and it is specific for pathological diagnosis with 94% sensitivity and 96% specificity (Haka et al., 2005).

Rehman et al employed Raman spectroscopy to identify chemical changes in ductal carcinoma in situ and invasive carcinoma. This study has aimed to differentiate different cancer stages in these two carcinomas. Tissue biopsies were analysed using 786 nm laser with spectral resolution of 4 cm^{-1}. Spectral differences were observed in the fingerprint region of normal and the two carcinomas. Spectral features of carcinoma biopsies have shown clear shift towards proteins and backwards to lipids and triacylglycerides. This is typical picture of cancer progression in malignant tissues. Different stages of IDC have shown spectral differences in terms of lipids, protein and nucleic acid composition. Different grades of DCIS have shown spectral variations in asymmetric and symmetric vibrations of lipids at high frequency region and phosphodiester vibrations in fingerprint region. This study has provided unique snapshot of chemical fingerprint of different grades of carcinoma (Rehman et al., 2007)

Haka et al has investigated in-vitro fresh-frozen tissues of different types of breast carcinoma by using Raman spectroscopy. These tissue samples were collected from different sites of different patients. Raman spectra were recorded using 830nm laser with resolution of 8 cm^{-1}. In order to extract information from the Raman spectra a spectroscopic model was employed. The modelling approach has involved fit coefficients, which resemble chemical composition of the disease. The results of this approach have shown much similarity with pathological diagnosis. The resulting diagnostic algorithm which classifies tissues not even as benign or malignant, but also according to pathological diagnosis has attained sensitivity of 94%, a specificity of 96%, and a total test efficiency of 95% for the diagnosis of cancer. Furthermore, their results seemed similar to traditional histopathology (Haka et al., 2009).

Saha et al applied Raman spectroscopy to identify microcalcifications based on calcium deposits present in the breast cell. Raman spectra were acquired from tissue samples collected through core biopsy method. Cellular components such as calcium, fat and collagen were used to develop new algorithm to identify calcifications. The spectra of breast biopsies have shown Raman peaks at 912 cm^{-1} and 1477 cm^{-1} (calcium oxalate) and 960 cm^{-1} (calcium hydroxyapatite). This group has detected calcium deposits at the depth of up to 2.15 mm depth in the tissue biopsies. This algorithm has shown positive and negative predicted value of 85 % and sensitivity of 86% for detection of microcalcifications. This study has identified microcalcifications and discriminated type I microcalcifications from type II (Saha et al., 2011).

Abramczyk et al has used Raman imaging to discriminate normal from cancerous breast tissue. The main aim of this study is not only to avoid histological studies based on biopsies, which are often prone to subjective interpretations, but also to understand chemical properties of the tissue. Optical fibres coupled catheters to micro-Raman spectrometer has been used in this study. This Raman images have shown clear

differences in carotenoids and fatty acid composition. These images have also revealed the differences between metabolic products of cancerous tissue and its surroundings. Spectral data has confirmed that non-cancerous tissue is almost identical to monosaturated oleic acid, which is common component of triglycerides and adipose tissue, whereas cancerous tissue is likely to be dominated by arachidonic acid derivatives mainly cyclic eicosanoids catalyzed by cyclooxygenase. This study has suggested potential role of carotenoids and lipids in breast cancer detection (Abramczyk et al., 2011, Abramczyk et al., 2012).

A summary of important Raman spectroscopic investigations in breast cancer research were summarised in Table 7.

Table 7: Brief summary of Raman spectroscopy application in breast cancer diagnosis

S.No	Raman Approach	Biological tissue/ sample	Results	Ref
1	Near-IR Raman spectroscopy	Human normal breast tissue specimens	Major molecular components of breast such as lipids including oleic acid and its derivative fatty acids identified. Visible lasers provided information on carotenoids	(Frank et al., 1994)

2	Raman Spectroscopy & multivariate approach such as logistic regression	Excisional breast biopsies including normal, fibrocystic , fibroadenoma and IDC breast tissues	Diagnostic algorithm has successfully differentiated normal and benign tissues from cancerous tissues with sensitivity and specificity of 94 % and 96%, respectively	(Haka et al., 2009)
3	Near-IR Raman spectroscopy	Breast tissue samples of DCIS and IDC	Different stages of IDC have shown spectral differences in terms of major molecular components Different grades of DCIS have shown s9pectral variations in asymmetric and symmetric vibrations of lipids at high frequency region and phosphodiester vibrations in fingerprint region	(Rehman et al., 2007)
4	Portable optical fibre probe with	Fresh stereotactic breast needle	Detection of microcalcifications	(Saha et al.,

	830nm diode laser	biopsies	(type 1 and 2) in breast tissue based on their chemical composition	2011)
5	Raman spectrometer with Nd:YAG 532 nm and 488, 514 nm lasers	Ductal and lobular carcinoma biopsies	Raman images have provided difference between metabolic products of cancerous tissue and its surroundings	(Abramczyk et al., 2011)

Aim and Objectives

1. Identification of chemical differences associated with different breast cancer cell lines using Raman spectroscopy

Human normal and breast cancer cell lines (MDA-MB-436 (ER^-PR^-), MCF 10A (non-malignant) and MCF-7 (ER^+)) were cultured, trypsinized and Agar cell plugs were made and fixed in formalin followed by paraffin wax embedding.

Cell sections were analysed using Thermo Nicolet DXR Raman system equipped with a 532nm laser at $4cm^{-1}$ resolution and further analysed by both supervised and unsupervised algorithms

2. Analysis of 2D and 3D models of breast cancer cell line (T47D) using combination of Raman Spectroscopy and Chemometrics

Human breast cancer cells (T47 D) were grown on 1.5% agarose coated 96 well plates and change media once in 3-4 days until 21 days.

Spheroids were analysed using Thermo Nicolet DXR Raman system equipped with a 532nm laser at $4cm^{-1}$ resolutions and further analysed by using multivariate approach.

3. Raman spectroscopic analysis of human different breast cancer subtype biopsies on TMA slide

Human different breast cancer biopsies (0.6 mm in diameter in size) were fixed in formalin followed by paraffin wax embedding) & were de-waxed and sections of each biopsy were cut using microtome

Cell sections were analysed using Thermo Nicolet DXR Raman system equipped with a 532nm laser at $4cm^{-1}$ resolution

Spectra were analysed using 'The Unscrambler (Multivariate Data Analysis Software)'

Chapter 3

Materials and Methods

3.1 Raman Analysis of Breast Cancer Cell lines

3.1.1 Materials

1. Dulbecco's modified Eagle Medium (DMEM) -

It was purchased from Gibco, Invitrogen, Paisley, UK and it contains 4500 mg/ml glucose GlutaMAX TM and sodium pyruvate

2. HAM'S nutrient mix F12

It was purchased from Gibco, Invitrogen, Paisley, UK and supplemented with L-glutamine and sodium bicarbonate.

3. RPMI 1640 media
4. Horse serum (Gibco, Invitrogen, Paisley, UK),
5. Fetal bovine serum (FBS) (Gibco, Invitrogen, Paisley, UK)
6. Penicillin and Streptomycin (Gibco, Invitrogen, Paisley, UK),
7. Cholera toxin (Calbiochem),
8. Insulin (Sigma),
9. Epidermal Growth Factor (EGF) (Sigma),
10. Fungizone
11. Trypsin (Invitrogen life technologies)
12. Hydrocortisol (Sigma)
13. Dimethyl sulfoxide (DMSO) (Gibco, BRL. Pasiley, UK)
14. T75 flasks & plastic ware

All plastic ware were supplied by Nunc, Thermo Scientific, Loughborough, Leicestershire, UK.

DMEM provides nutrients for cells and it constitute basal medium, four folds of amino acids (Serine and Glycine), ferric nitrate and vitamins. Similarly, HAM's nutrient mix

provides nutrients for cells. It was initially developed to support growth of several clones of CHO (hamster cell lines) and HeLa cell lines. RPMI 1640 media was developed by Moore (Rosewell Park Memorial Institute). It is a combination of bicarbonate buffering system and amino acids. Pencillin/Streptomycin prevent bacterial contamination from both gram negative and positive bacteria. Cholera toxin increases cAMP levels in cultured cells, which in turn promotes growth. Insulin will bind insulin receptor on cells and increases cell proliferation. Fungizone is used to prevent contamination of cell cultures from yeast and multicellular fungi. Trypsin/EDTA provides Mg2+ and Ca2+ ions in media and these ions act as chelators. Hydrocortisol is steroid and it promotes cell growth and proliferation. Glutamine is used in protein metabolism, while DMSO is used in cryopreservation.

3.1.2 Breast Cancer Cell Lines

There are three cell lines were used in this experiment. Two breast cancer cell lines (MCF-7, and MDA-MB-436) and one normal breast cell line (MCF-10A) were used and the cell lines were purchased from American Type Culture Collection (ATCC, Manassas, VA, USA). The brief description of these cell lines was mentioned in Table 8.

Table 8: Breast cancer cell lines.

Cell line	Tissue origin	Receptors expressed	Source
MCF-10A (Catalog No: CRL-10317™)	Mammary gland of human breast	Non tumorigenic and EGF positive	American Type Culture Collection (ATCC) VA, USA
MCF-7 (Catalog No: HTB-22™)	Breast adenocarcinoma (from metastatic site)	Estrogen receptor positive	American Type Culture Collection (ATCC) VA, USA
MDA-MB-436 (Catalog No: HTB-132™)	Breast adenocarcinoma	Tumorigenic and triple negative (ER^- PR^-) subtype	American Type Culture Collection (ATCC) VA, USA

3.1.3 MCF-10A

The MCF-10A line was isolated from the breast tissue of a 36-year- old female with fibrocystic disease. It was established by culturing in serum free media with low calcium concentration. It is of non- tumorigenic epithelial origin and expresses receptors for epidermal growth factor (EGF). MCF-10A are described as normal breast epithelial cell line because of their diploid karyotype and dependent on growth factors for their proliferation. This cell line expresses breast specific antigens such as cytokeratins, sialomucins and milk fat globules and they were detected by MC-5 and MFA breast monoclonal antibodies. MCF-10A has shown 3D growth in collagen and laminins mixture and display domes in confluent cultures, which are similar to acini of human breast.

There are three types of media used for MCF-10A cell line, namely;

1. Growth media

2. Resuspension media
3. Freezing media

Table 9: Growth media of MCF-10A.

Media Component	Final concentration	Storage
DMEM/F12	89%	4º C
Horse serum	10%	-20º C
EGF (100 mg/ml)	5 ng/ml	-20º C
Penicillin/Streptomycin	100 i.u/ml of each	-20º C
Hydrocortisol	4 ng/ml	4º C
Insulin (10mg/ml)	5 µg/ml	4º C
Cholera toxin (1mg/ml)	8.47 ng/ml	4º C

Table 10: Resuspension media of MCF-10A.

Media Component	Final concentration	Storage
DMEM/F12	89%	4º C
Horse serum	10%	-20º C
Penicillin/Streptomycin	100 i.u/ml of each	-20º C

Table 11: Freezing media of MCF-10A.

Media Component	Final concentration	Storage
DMEM/F12	89%	4º C
Horse serum	10%	-20º C
Fetal bovine serum	10%	-20º C
EGF (100 mg/ml)	5 ng/ml	-20º C
Penicillin/Streptomycin	100 i.u/ml of each	-20º C
Hydrocortisol	4 ng/ml	4º C
Insulin (10mg/ml)	5 µg/ml	4º C
Cholera toxin (1mg/ml)	8.47 ng/ml	4º C
DMSO		

3.1.4 MCF-7

The MCF-7 cell line was isolated from the metastatic site of a 69-year-old female with breast adenocarcinoma. It expresses estrogen receptors.

Table 12: Media components of MCF-7 cell line.

Media Component	Final concentration	Storage
RPMI	89%	4º C
Fetal serum	10%	-20º C
Penicillin	100 i.u/ml	-20º C
Streptomycin	100 i.u/ml	-20º C
L-Glutamine	2 mM	-20º C
Amphotericin	0.625 µg/ml	-20º C

3.1.5 MDA-MB-436

The MDA-MB-436 cell line was isolated from the breast tissue of a 51-year-old female with metastatic adenocarcinoma. It expresses EGF and transforming growth factor alpha (TGF alpha) receptors and is considered to reflect the triple negative (ER^- PR^-) subtype of breast cancer.

3.2 Methods

3.2.1 Culturing of Breast Cancer Cell Lines

The total cell culture work of this section was carried in laminar flow hoods of Sheffield medical school under appropriate aseptic conditions. MCF-10A cell was grown in a 1:1 mixture of Dulbecco's modified Eagle Medium (DMEM) and HAM'S nutrient mix F12 (Invitrogen), 5% horse serum (Invitrogen), 100 IU/ml penicillin, 100 IU/ml streptomycin (Invitrogen), 100n/ml cholera toxin (Sigma), 10µg/ml insulin (Sigma), 20ng/ml EGF (Sigma) and 0.5mg/ml hydrocortisol. MCF 7 and MDA-MB-436 cell lines were grown in RPMI media, 5% fetal calf serum, 2mM L-glutamine, 100 IU/ml penicillin, 100 IU/ml streptomycin (Invitrogen) and 0.62510µg/ml amphotericin (Talari et al., 2015a). The three cell lines were grown as monolayers in T75 tissue culture flasks

and kept in incubators at 37^{o}C in a humidified atmosphere with 5% CO2. Until 70% confluence, Cell culture medium was changed every 2-3 days and subculturing required for these cell lines. The three cell lines, which were used in the experiments, were between cell passage 5 and 19 cell passage describes the number of times the cell were split to grow under cell culture conditions for extended time periods.

3.2.2 Subculturing of MCF-10A, MCF-7 and MDA-MB-436 cells

Subculturing by definition is transfer of some cells from a previous culture to fresh growth media for the aim of prolonged life and the cell number expansion. This procedure involves discarding the culture media from T-75 tissue culture flasks, which has 70-80% confluence and briefly rinse the cell layers with PBS. 1ml of 1X trypsin EDTA solution was added to flasks and kept flasks at 37^{o}C in incubators for 5 minutes. Flasks were observed under inverted light microscope to observe detachment of cells from flask. 9 ml of growth media was added to each flask and transferred cell suspension to centrifuge tube and spin at 100Xg for 5 minutes. Supernatant was discarded and resuspend the cells with fresh growth medium and disperse cell suspension aliquots to new tissue culture flasks with the ratio of 1:3 to 1:4.

3.2.3 Preparation of Agar Plugs

Once cells reached 70-80% confluence, cells were washed with phosphate buffer saline (PBS) and the cells were trypsinized to make viable, large pellets with a good number of cells for spectroscopic studies that would be representative of large cell clusters. Cell pellets were re-suspended with PBS followed by the addition of liquid agarose. Agar plugs were fixed in 3.7% formaldehyde for room temperature for at least 24 hours before histological processing (Talari et al., 2015a).

3.2.4 Histological Processing of Agar Plugs

Histological processing including paraffin wax embedding was carried out using a Leica TP1020 benchtop processor (Leica Microsystems, Milton Keynes, UK). The steps involved in histological processing were mentioned in the below Table 13

Table 13: Histological processing of agar plugs and steps involved.

Station	Solution	Time	Vacuum
I	10% formalin	120	No
II	70% ethanol	120	No
III	70% ethanol	120	No
IV	95% ethanol	120	No
V	95% ethanol	120	No
VI	100% ethanol	120	No
VII	100% ethanol	120	No
VIII	Xylene	120	No
IX	Xylene	120	No
X	Paraffin wax at 57°C	120	Yes
XI	Paraffin wax at 57°C	120	Yes
		22 hours +15 min drain time	

Once processing has done, Agar plugs were embedded using Leica EG1160 embedding unit, dispenser and hot plate (Leica microsystems, Milton Keynes UK) (Talari et al., 2015a).

3.2.5 Raman sample preparation

Histological sections of each cell line, 20µm thick, were cut using a microtome (Leica Microsystems, Milton Keynes, UK). Each section was mounted onto a glass slide and firmly fixed. De-waxing was achieved using Xylene treatment for 30 minutes, 50% alcohol for 5 minutes, 70 % alcohol for 5 minutes and 100% alcohol for 5 minutes. The aim of Xylene treatment is to remove the wax from the sample and alcohol treatment rehydrates the sample (Talari et al., 2015a).

3.2.6 Raman Measurements

High quality Raman spectra were collected from cell lines using a non-invasive dispersive micro-Raman system dxr™ (Thermo Nicolet) equipped with a 532nm laser. Before collecting spectral data from cell lines, the system was fully calibrated using polystyrene as standard. The sensitivity in terms of signal to noise ratio of the laser was 1000:1. Laser power of 10mw was focussed on the sample. A 50x long working distance objective was used and a spectrograph aperture was set to 50 micron pinhole. Spectral collection exposure time was set to 50 seconds with 5 exposures. Twenty spectra were collected from each cell line over the spectral range 400-3400 cm^{-1}. Thermo Nicolet OMNIC™ software was employed for data acquisition(Talari et al., 2015a).

3.2.7 Data Processing and analysis

Spectral range were analyze between 600-3400 cm^{-1}. Three cell lines were examined using DXR Raman confocal microscope. A total of 270 spectra were collected for this study (MCF-10 A = 90, MDA-MB-436 = 90 and MCF-7 = 90). A mean spectrum of each cell line was collected for comparison studies. Base line corrections and smoothing

was accomplished using OMNIC Atlμs ™ software of Thermo Scientific, Madison, WI, USA)

Data analysis was performed using Unscrambler X 10.2software (Camo software, Oslo, Norway). Multivariate approach was used to quantitative and qualitative analysis of different spectral regions. Quantitative analysis includes peak intensity of lipids, amide I and amide III. Both unsupervised and supervised approaches were used in data analysis. Principal component analysis (PCA) was performed on full spectral range (3300-200 cm^{-1}), fingerprint range (1800-500 cm^{-1}), lipids region (3100-2680 cm^{-1}), amide I (1800-1530 cm^{-1}) and amide III (1380-1190 cm^{-1}).

Linear discrimination analysis (LDA) was applied on classification and prediction accuracy. LDA model was setup over full spectral range. Spectral processing for all LDA models were baseline correction and unit vector normalization. Five samples from each group were left out at each pass until total number of 20 spectra of each cell line predicted (Table 14).

Table 14: LDA model set up for three breast cell line models.

Cell line	Number of spectra used in LDA model at each pass	Number of spectra unknown at each pass
MCF-7	35	5
MCF-10 A	35	5
MDA-MB-436	20	5

Diagnostic test evaluation is represented in sensitivity and specificity. Sensitivity of this model is calculated using correctly classified false positive, false negative and true negative. Formulas used for calculation of sensitivity and specificity are

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

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

3.3 Raman Analysis of 3D Spheroids

Initially MCF-7, MDA-MB-436 and MCF7 cell lines were used for spheroid cultures. None of these cell lines have grown to spheroids until 14 days. T-47D cell line was chosen for spheroid culture and spheroids were cultured until 14 days successfully.

3.3.1 Cell Line

T-47D epithelial cell line was purchased from American Type Culture Collection (ATCC, Manassas, VA, USA). This cell line was isolated from the breast tissue of a 54-year- old female with infiltrating ductal carcinoma disease. It expresses both receptors for estrogen and progesterone. T-47D cell line was cultured in RPMI media (Invitrogen), 5% fetal calf serum (Invitrogen), 2mM L-glutamine (Invitrogen), 100 IU/ml penicillin (Invitrogen), 100 IU/ml streptomycin (Invitrogen) and 0.62510µg/ml amphotericin (Invitrogen). This cell line was grown as monolayers in T75 tissue culture flasks and kept in incubators at 37^{o}C in a humidified atmosphere with 5% CO_2. Until 70% confluence, Cell culture medium was changed every 2-3 days and subculturing was done using 1X Trypsin EDTA solution.

3.3.2 Culture of Spheroids

T-47D tumour cell spheroids were made using liquid overlay method (Offner et al 1993). 1.5 % Agarose (Electrophoresis grade Invitrogen) solution was prepared using RPMI media (without FCS) and heated in a microwave until it dissolved. Using a Varipipette (Eppendorf), agarose was aliquot into a 96 well plate with 100µl/well and left 30 minutes to cool and set. Cells were trypsinized from 70-80% confluence T75 flasks and seeded into wells at concentration of 2000 cells /200µl / well. Cells were fed with fresh medium using multi-channel pipette and carefully removed old medium until 14 days.

3.3.3 Histology of Spheroids

After 14 days, T-47D spheroids were washed with PBS and fixed in 3.7% formaldehyde at room temperature for a minimum of 24 hours before histological processing. Histological processing, sample preparation, Raman measurements and data processing were kept constant throughout this project as mentioned in section 3.1 to 3.4, unless otherwise stated.

3.3.4 Chemometric Analysis

Chemometric methods such as PCA and Cluster Analysis (CA) were used in this study. PCA was performed on high-wavenumber region (3200-2600 cm^{-1}), finger print region (1800-600 cm^{-1}), amide I region (1800-1510 cm^{-1}), amide II region (1510-1390 cm^{-1}), amide III region (1390-1140 cm^{-1}) and nucleic acid region (980-600 cm^{-1}). CA was performed on high-wavenumber, amide I and nucleic acid regions using Wards's method squared Euclidean distance.

Linear discrimination analysis (LDA) was employed on classification and prediction accuracy. LDA model was setup over full spectral range. Spectral processing for all LDA models were baseline correction and unit vector normalization. Four or Five

samples from each group were left out at each pass until total number of 20 spectra of each spheroid region predicted (Table 15).

Table 15: LDA model set up for three regions of T-47D models.

Cell line	Number of spectra used in LDA model at each pass	Number of spectra unknown at each pass
Normal proliferating region	20	4 or 5
Hypoxic region	19	4 or 5
Necrotic region	21	4 or 5

3.4 Raman Analysis of breast cancer biopsies

3.4.1 Tissue Micro Array (TMA)

Section three involves Raman study of breast cancer biopsy samples from tissue micro array (TMA) slide. TMA approach has gained much attention in pathological research analysis. H. Battifora initially developed this approach in 1986 and later on J.Kononen and his colleagues developed it as proper technique in 1998. This approach has revolutionized in multiplex histological studies. Results of this approach are reliable and provided new platform correlated gene expression studies. The main advantages of this approach is as follows

1. High-throughput analysis of number of tissue biopsies in a single slide within a short time.
2. It will be helpful in comparative analysis of gene and protein expression studies of same tissue under same conditions

3. It increases research effectiveness with minimal quantities of tissues.
4. This approach has been useful in semi-quantitative scoring immunohistochemistry, FISH and in situ PCR.

The major limitations of this approach are as follows

1. Pathological cores may not representative of whole tumour because due to its heterogeneity.
2. It has time consuming steps including preparation of individual paraffin blocks and
3. This approach entails qualified staff and costly equipment
4. Restricted sample quantity has offered difficulties in interpretation in terms of spatio-temporal data analysis of each section

TMA slide preparation:

The first step in TMA slide preparation is study of clinical or tumour biopsy before making decision on area of investigation. TMA approach generally allows tissue cores as small as 0.6 mm in diameter.

The major steps involved in the TMA slide preparation are

1. Selection of donor tissue and identification of areas of interest for arraying
2. Preparation of paraffin blocks and making array pores in the blocks
3. Punching tissue cylinders (core biopsies) from marked representative areas of donor paraffin block.
4. Inserting the tissue cylinders into the recipient paraffin block with low or high densities.
5. Embedding recipient block to get TMA paraffin blocks
6. Cutting TMA sections using a microtome

3.4.2 Sample quantity

In this study we have examined different breast cancer patient biopsies on TMA slide. TMA and ethical approval was provided by Clinical Trials Research Unit (CTRU), University of Leeds. Clinical trial name was Azure and trial Data Transfer Agreement (DTA) ID is Azure_07. TMA slide has provided information about patient trial number, ER status, PR status, HER2 status, histological grade and tumour pathology.

ER and PR status has described as 0-8, where 0-2 is regarded as negative, sometimes 3 is also considered as negative. HER status has descried as 0-3, where 0 and 1 is negative, and 3 is positive. Score 2 are equivocal and need to perform an in situ hybridisation method to ascertain true status. Approximately 30% or slightly less of 2+ are amplified and therefore considered as positive. Histological grade is represented as low grade, moderate and high grade. Tumour pathology of breast biopsies were ductal No Special Type (NST), lobular, mixed (ductal /lobular), mucinous, tubular and other types. Majority of breast biopsies belong to ductal NST. A total 132 biopsy samples were present on TMA slide and known ER, PR and HER2 status biopsies were used for spectral collection. 30 spectra were collected from each biopsy and in total 3,960 spectra were collected from patient biopsies.

Sample preparation and Raman spectroscopic measurements were kept constant throughout this study as mentioned in section 3.1 to 3.4, unless otherwise stated.

Chemometric methods such as PCA and Cluster Analysis (CA) were used in this study. PCA was performed on high-wavenumber region (3200-2600 cm^{-1}), finger print region (1800-600 cm^{-1}), amide I region (1800-1510 cm^{-1}), amide II region (1510-1390 cm^{-1}), amide III region (1390-1140 cm^{-1}) and nucleic acid region (980-600 cm^{-1}). CA was performed on high-wavenumber, amide I and nucleic acid regions using Wards's method squared Euclidean distance. Linear discrimination analysis (LDA) was applied on classification and prediction accuracy. LDA model was setup over full spectral

range. Spectral processing for all LDA models were baseline correction and unit vector normalization. Five samples from each group were left out at each pass until total number of 30 spectra of each subtype predicted (Table 16).

Table 16: LDA model set up for four subtypes of TMA biopsies

Cell line	Number of spectra used in LDA model at each pass	Number of spectra unknown at each pass
Luminal A	30	5
Luminal B	30	5
HER2 positive	30	5
Triple negative	30	5

Chapter 4

Results and Discussion

4.1 Breast Cancer Cell lines - Results

In this section, Spectral analysis of normal (MCF-10A) and two breast cancer (MCF-7 and MDA-MB-436) cell lines was conducted. Raman spectra of the cell lines have revealed basic differences in the concentration of biochemical compounds, such as proteins, lipids and nucleic acids. Raman peaks differ in peak intensities and shifts in normal and cancer cell lines and Chemometric approach may assist in accurate and reliable classification with improved sensitivity and specificity. Vibrational spectroscopy has gained much attention in recent years in cancer research and especially used to explore the chemical fingerprints different biological tissues including normal and malignant types. A combination of Raman spectroscopy and multivariate approach was used for the first time to identify subtypes at cellular level.

4.1.1 Raman Peak analysis

A total of 120 spectra were collected from MCF-10A, MCF-7 and MDA-MB-436 cell lines. Mean spectra of three cell lines were extracted and identified peaks assignments (Figure 8). The spectra obtained from each subtype showed significant differences in terms of peak heights and peak shifts. All peaks were carefully examined and assigned to their respective biological groups. Peak assignments were identified in the entire spectral range (400 – 3200 cm^{-1}) of each cell line and summarised in table (appendix). To understand the chemical structural properties of cell lines and tissues, it is important to characterise these as precisely and accurately as possible. Therefore, all possible chemical bonds and functional groups were identified and their details are given below:

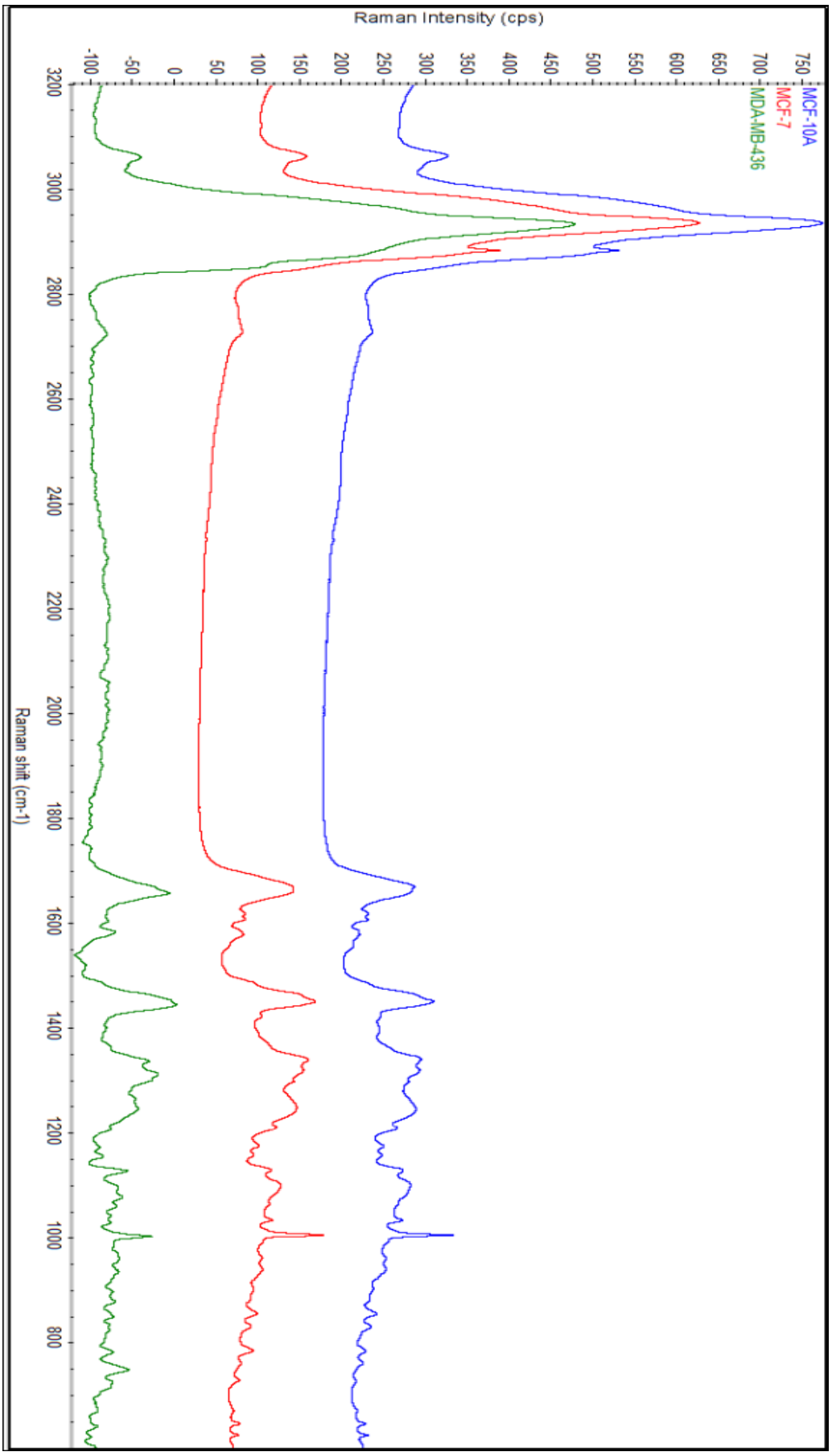

Figure 8: Mean spectra of breast cancer cell lines (MCF-10A, MCF-7 and MDA-MB-436).

4.1.2 C-H region (3200 - 2700 cm^{-1})

Spectral peak observed at 3060 cm^{-1} in MDA-MB-436, 3061 cm^{-1} in MCF-7 and 3062 cm^{-1} MCF-10A was assigned C-H stretching of lipids. A small peak, present as shoulder was observed at 2881 cm^{-1} in MCF-7 and MCF-10A and attributed to CH_2 asymmetric stretching vibrations of lipids and proteins, whereas this shoulder peak was absent in MDA-MB-436 (Figure 10). A strong Raman peak observed in high-wavenumber region at 2934 cm^{-1} in MCF-10A and MCF-7 assigned to CH_2 asymmetric stretching vibrations of lipids (Figure 9). A strong Raman band was observed at 2932 cm^{-1} at MDA-MB-436, which corresponds to CH_2 asymmetric stretch of lipids and fatty acids. A weak band was observed in MCF-10A and MCF-7 at 2726 cm^{-1}, which represents stretching vibrations of CH, NH and OH groups of lipids. A similar weak band was observed in MDA-MB-436 at 2722 cm^{-1}, which represents stretching vibrations of CH, NH and OH groups of lipids (Talari et al., 2015k).

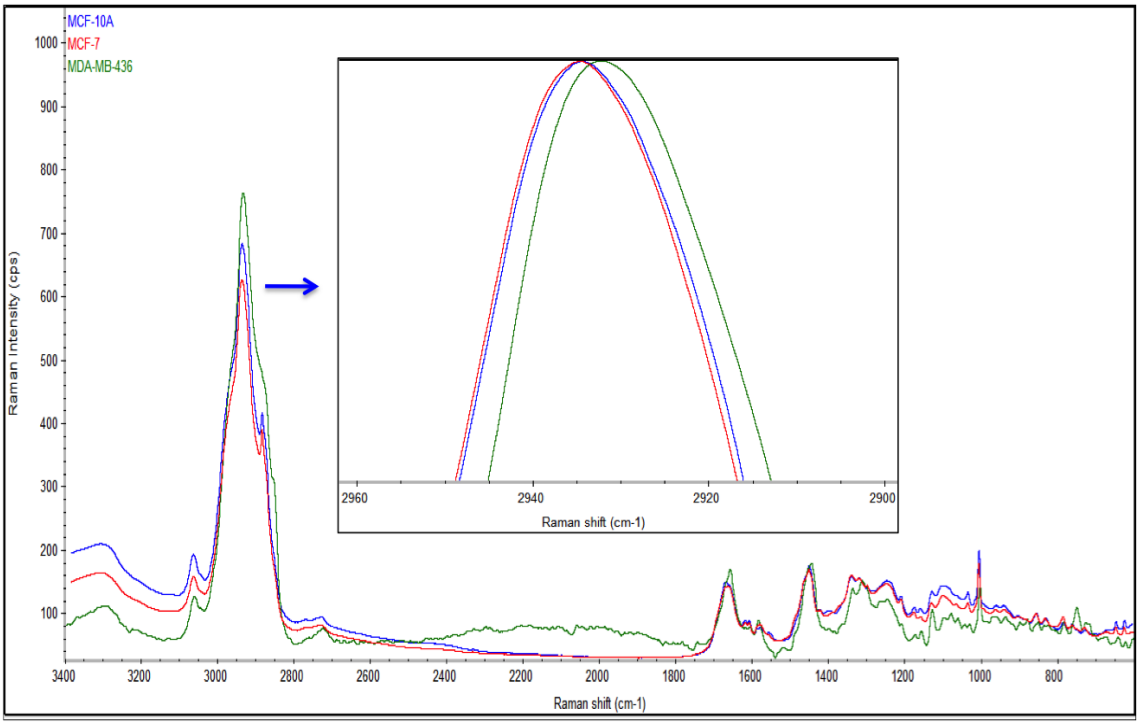

Figure 9: Mean Raman spectra of MCF-10A (blue), MCF-7 (red) and MDA-MB-436 (green) with major lipid peak shift.

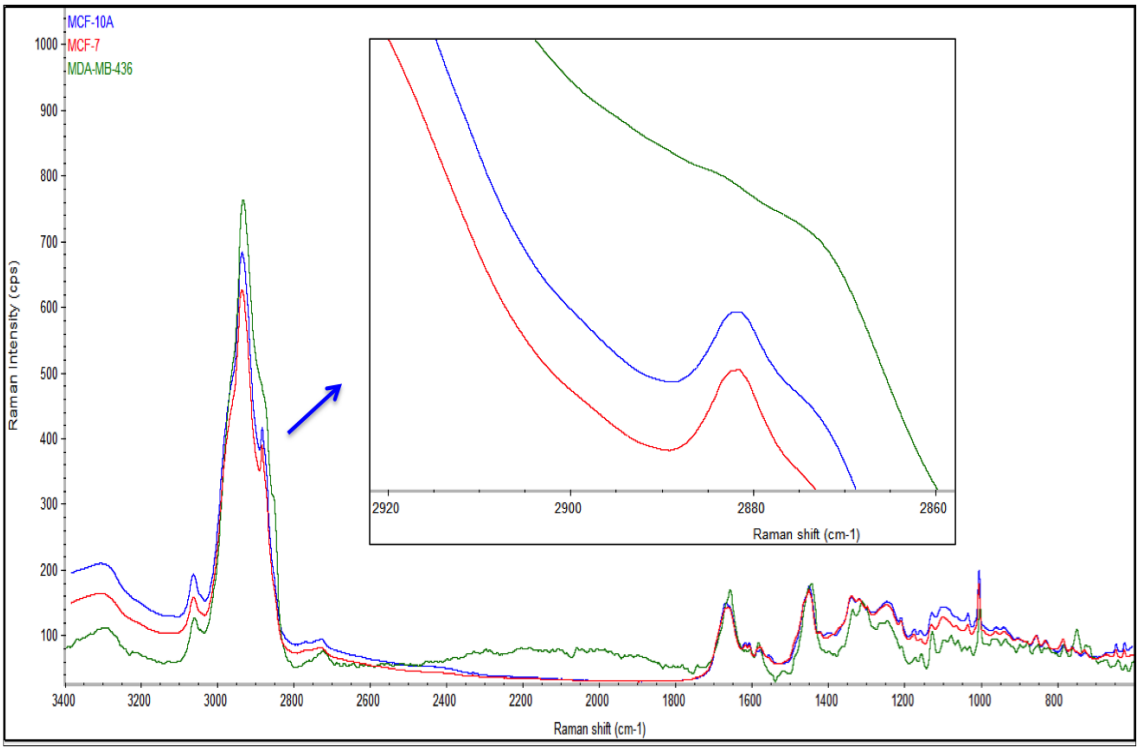

Figure 10: Mean Raman spectra of MCF-10A (blue), MCF-7 (red) and MDA-MB-436 (green) with major lipid shoulder peak shift.

4.1.3 Amide I region (1670 - 1655 cm^{-1})

Raman peaks at 1668 cm^{-1} (MCF-10A), 1659 cm^{-1} (MCF-7) and 1655 cm^{-1} (MDA-MB-436) are characteristic assignments of C=O stretching vibrations of proteins (Figure 11-12). These vibrations are mainly due to α-helical confirmation of proteins. A significant shift was observed in Amide I band in these three cell lines (Talari et al., 2015k).

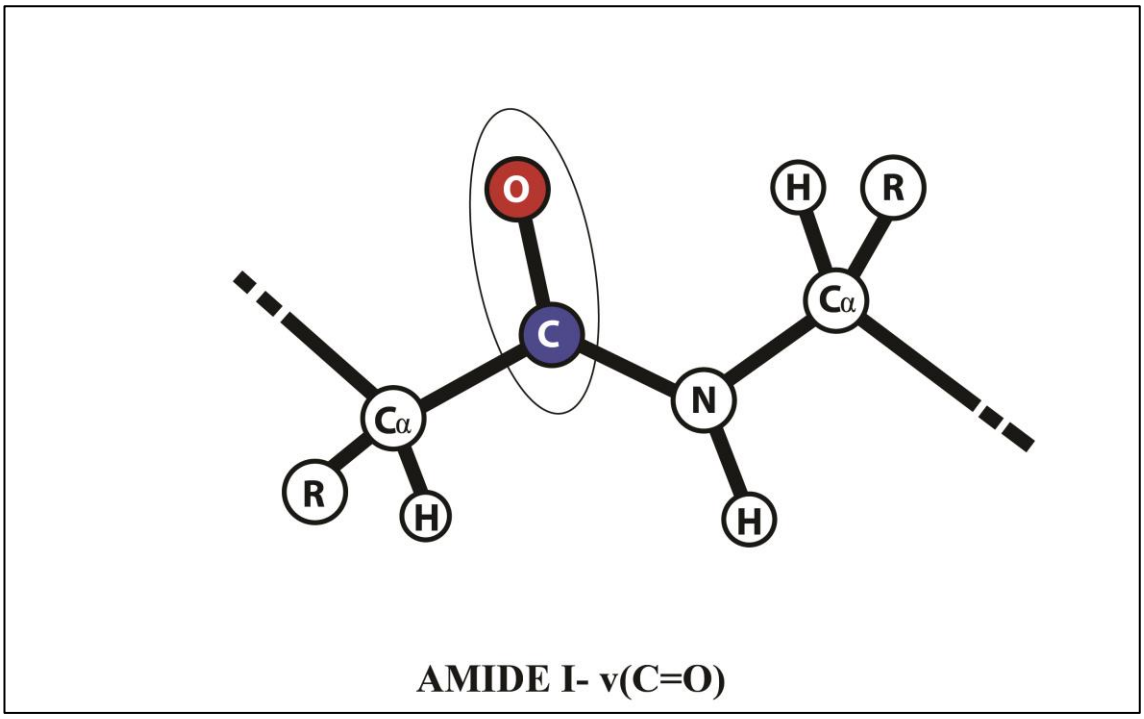

Figure 11: The structure of amide I.

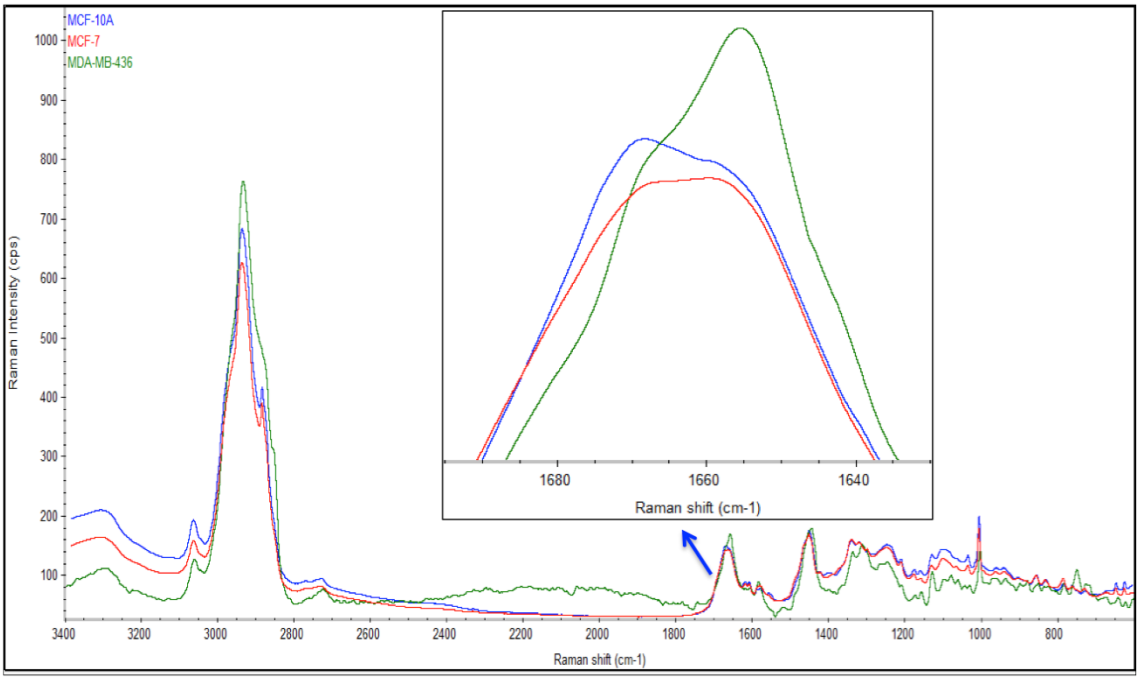

Figure 12: Mean Raman spectra of MCF-10A (blue), MCF-7 (red) and MDA-MB-436 (green) with major amide I peak shift.

4.1.4 Amide II region (1650 - 1290 cm^{-1})

A Raman peak arising at 1616 cm^{-1} was observed only in MCF-10A and MCF-7 that corresponds to C=C stretching vibrations of aromatic amino acids such as tyrosine and tryptophan (Figure 13). Peak present at 1605 cm^{-1} was appeared in MCF-10A and

MCF-7 which correspond to ring vibration of phenylalanine or tyrosine or C=C stretching mode of proteins or NH_2 vibrations of cytosine. A Raman band was observed inMDA-MB-436 at 1582 cm^{-1} and 1517 cm^{-1}, which correspond to δ (C=C) vibrations of phenylalanine and C-C stretching vibration of β carotene respectively. Raman peaks were observed at 1576 cm^{-1} (MCF-10A) and 1577 cm^{-1} (MCF-7) representing nucleic acid contents of cells or purine base such as guanine. CH_2 and CH_2, CH_3 deformations were observed in MCF-7 at 1448 cm^{-1} and CH vibrations of proteins and lipids were observed in MCF-10A at 1449 cm^{-1}. Raman band was observed in MDA-MB-436 at 1442 cm^{-1} that represents either fatty acids or CH_2 bending mode or CH_3, CH_2 deformations of collagen. Raman peaks were observed at 1337 cm^{-1} and 1338 cm^{-1} in MCF-7 and MCF-10A which corresponds to amide III or CH deformation of proteins or CH_2 wagging vibrations of amino acids such as glycine and proline or ring breathing modes of purine bases such as adenine and guanine. Raman bands at 1334 cm^{-1} and 1310 cm^{-1} were appeared in MDA-MB-436 that represents tryptophan and lipid specific vibrations respectively (Movasaghi et al., 2007a).

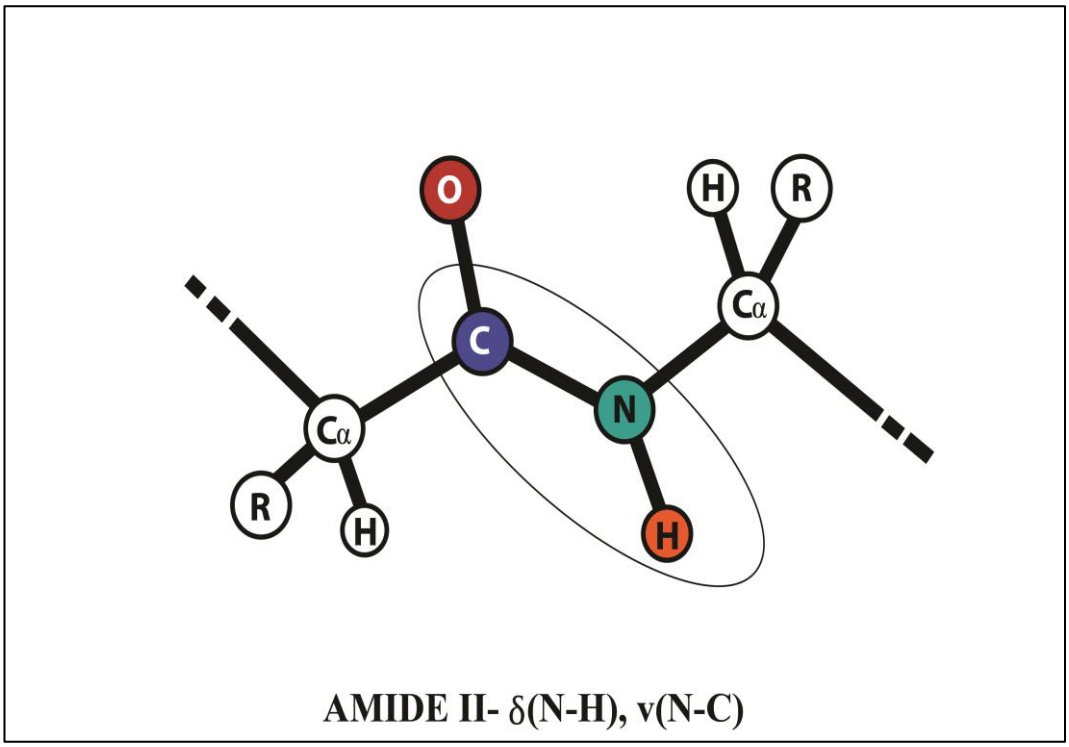

Figure 13: The structure of amide II.

4.1.5 Amide III region (1244- 1208 cm^{-1})

Raman bands were appeared between 1243-1245 cm^{-1} in three cell lines representing amide III or phosphate backbone of nucleic acids (Figure 14). CH_2 wagging and C-N stretching vibrations attributed amide III and asymmetric stretching vibrations of phosphodiester groups were attributed phosphate backbone. Pyrimidine bases such as cytosine and thymine were also contributed towards this peak. Raman peak at 1208 cm^{-1} in MCF-10A, MCF-7 and MDA-MB-436 were representing $\nu(C\text{-}C_6H_5)$ vibrations of aromatic amino acids such as phenylalanine, tyrosine and tryptophan of proteins (Figure 15) (Movasaghi et al., 2007a).

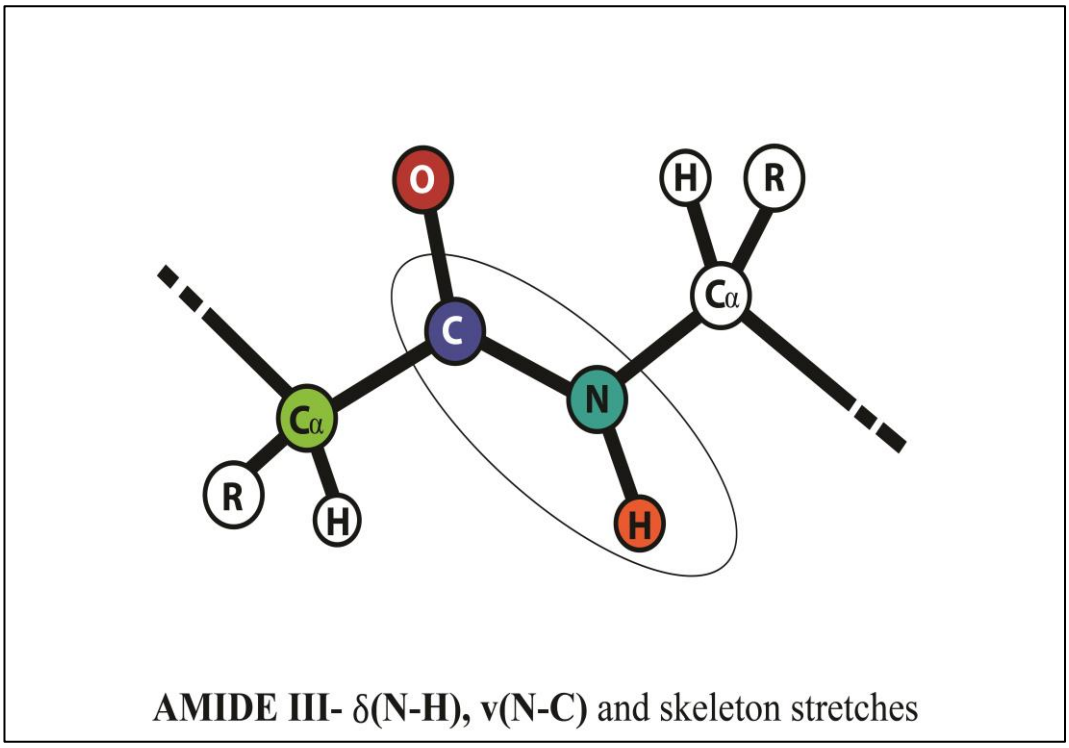

Figure 14: The structure of amide III

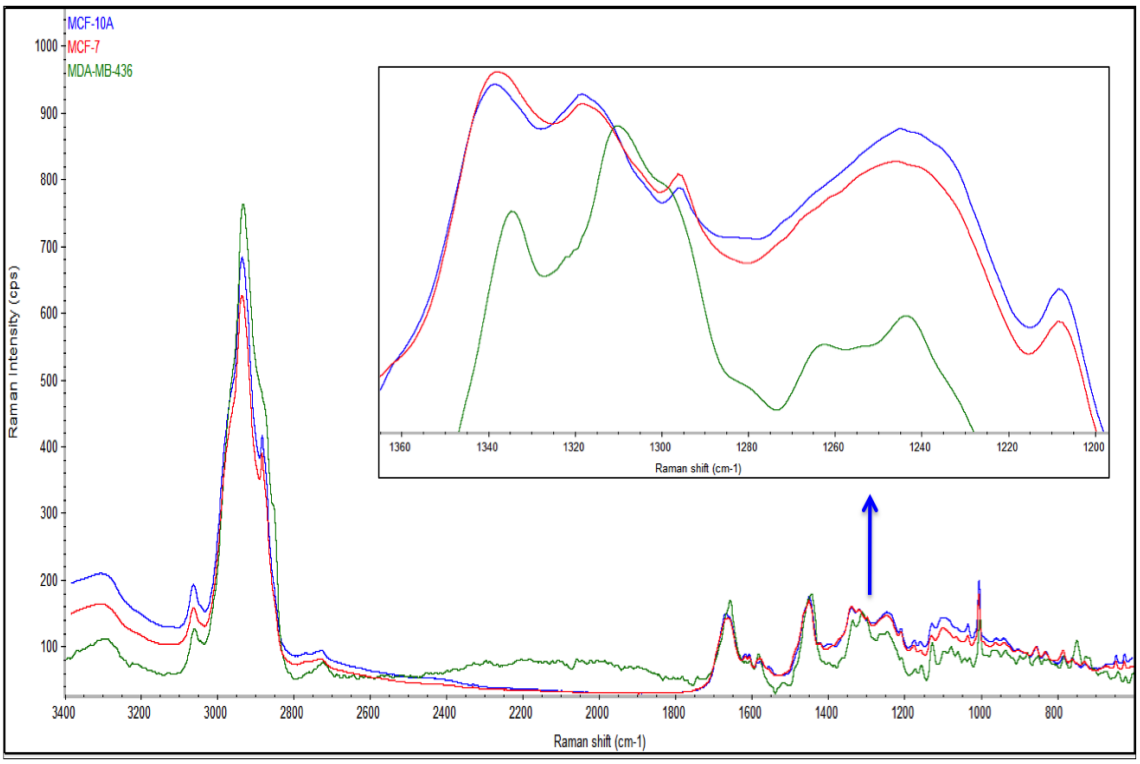

Figure 15: Mean Raman spectra of MCF-10A (blue), MCF-7 (red) and MDA-MB-436 (green) with major amide III peak shift.

4.1.6 Phenylalanine region (1000 - 1207 cm^{-1})

A Raman peak was observed in 1170 , 1153, 1124, 1075, 1058 and 1039 cm^{-1} in MDA-MB-436 and representing C-H in-plane bending mode of tyrosine, carbohydrates, ν(C-C) skeletal of acyl vibrations of lipids, phosphate vibrations (charwat-pessler et al., 2013) respectively. Raman peak at 1003 cm^{-1} in MCF-10A and MCF-7 and 1000 cm^{-1} in MDA-MB-436 corresponds C-C aromatic stretching of phenylalanine (Talari et al., 2015k).

4.1.7 Nucleic acid and Amino acid region (960 - 750 cm^{-1})

Spectral region of 960- 750cm^{-1} has mainly represented vibration modes of either aromatic amino acids or nucleic acid components of each cell line. Raman peaks at 936 cm^{-1} in MCF-7 have represented bio molecular assignment of ν(C-C) of α-helix conformation for proteins whereas MCF-10A and MDA-MB-436 have shown Raman peak at 958 cm^{-1}, which represents ν_1 (PO_3^{-4}) of phosphate peak. Raman bands of 853, 828, 782 and 758 cm^{-1} in were observed in both MCF-7 and MCF-10A. Raman peak at 853 cm^{-1} have represented ring-breathing mode of tyrosine and C-C stretching of proline or glycogen. The peak that appeared at 828 cm^{-1}, has represented either out-of-plane ring breathing mode or tyrosine or O-P-O stretching vibrations of DNA and RNA. O-P-O stretching vibrations are raised from phosphodiester groups of phosphate backbone of nucleotides. The peak that appeared at 782 cm^{-1} has represented pyrimidine bases of DNA and RNA such as thymine, cytosine and uracil. Indole ring of tryptophan has contributed Raman peak at 758 cm^{-1} in both normal and estrogen receptor expressed cell lines. Raman peak of 932 cm^{-1} in MDA-MB-436 has represented skeletal vibrations of C-C or α-helix of proteins. Phospholipids have contributed Raman band at 875 in MDA-MB-436, which is mainly raised from antisymmetric stretch vibration of choline group. C-O-C skeletal mode vibrations, monosaccharide such as α-glucose and disaccharides such as maltose has expressed Raman band at 847 in MDA-MB-436 cell line.

4.1.8 C-C region (750 to 600 cm^{-1})

MCF-7 and MCF-10A have shown Raman peaks at 643 and 621 attributed to C-C twisting mode of tyrosine and phenylalanine respectively. C-C stretching of proline peak was observed in MCF-7 at 726 and in MCF-10A at 725.

4.1.9 Peak Height Analysis

Peak height analysis was performed on three spectral regions of among breast cancer cell lines. Peak intensities have shown clear differences and quantification of peak heights were performed on lipids (2934 cm-1), amide I (1658 cm-1) and amide III (1244 cm-1) regions. The absolute intensity of each region is increased from MCF-10A to MCF-7 and MDA-MB-436 as shown in Figure 16-18.

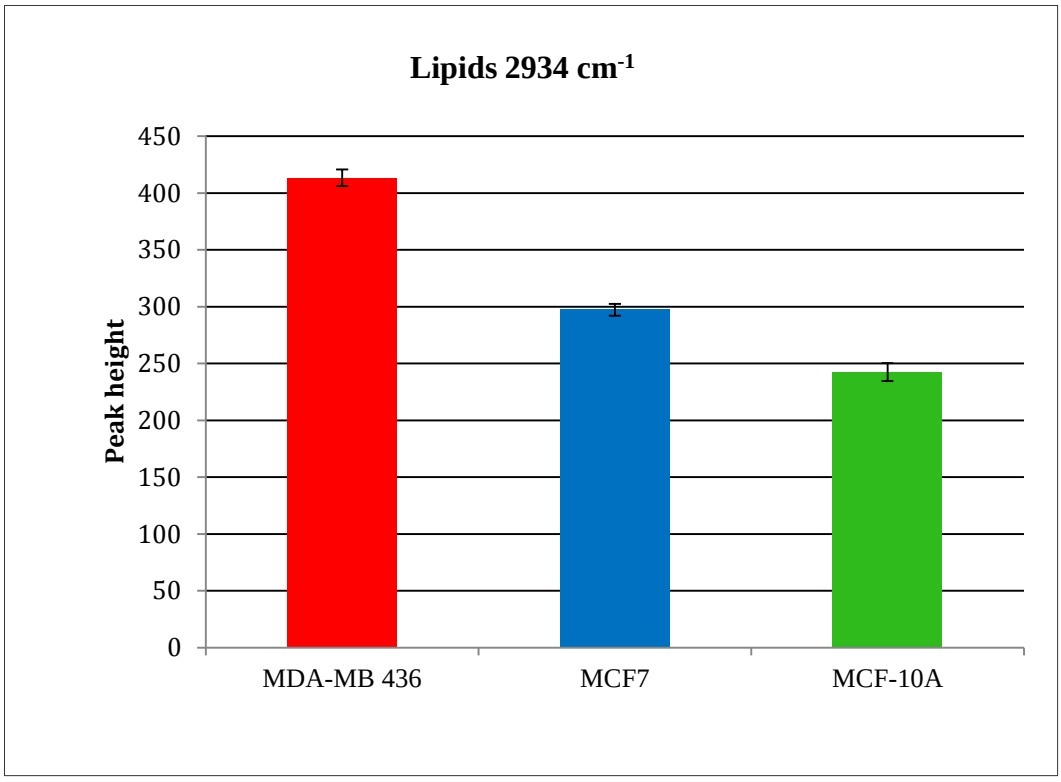

Figure 16: Peak height analysis of lipid region at 2934 cm-1 for MDA-MB-436 (red), MCF-7 (blue) and MCF-10A (green) cell lines

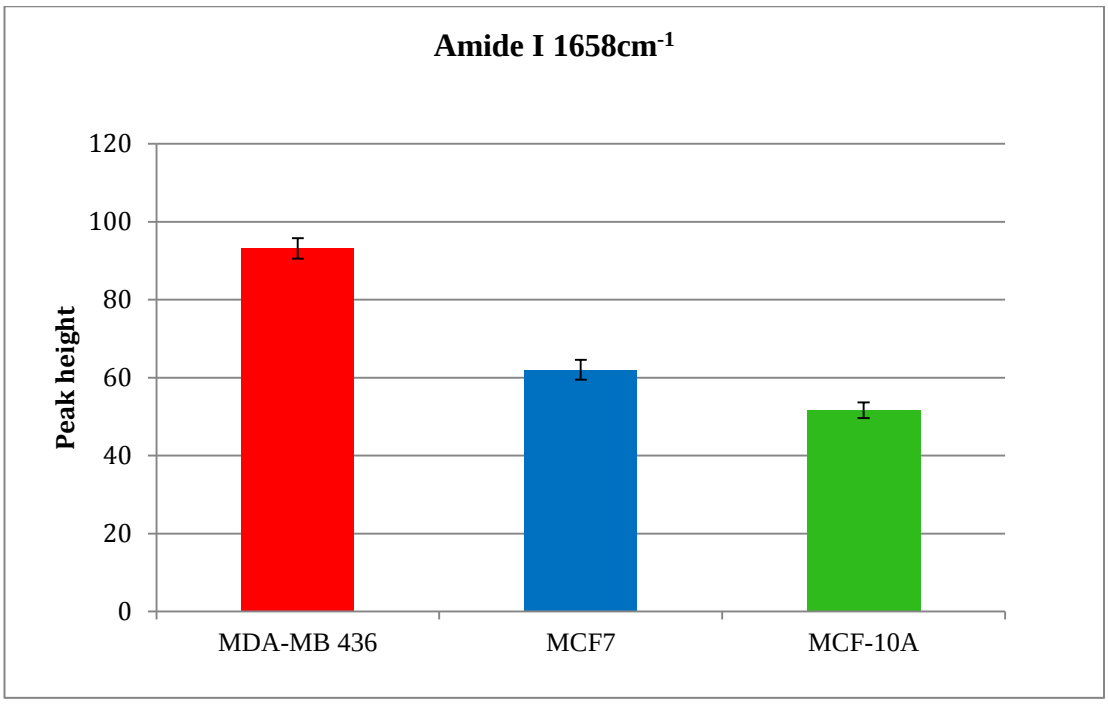

Figure 17: Peak height analysis of Amide I region at 1658 cm^{-1} for MDA-MB-436 (red), MCF-7 (blue) and MCF-10A (green) cell lines.

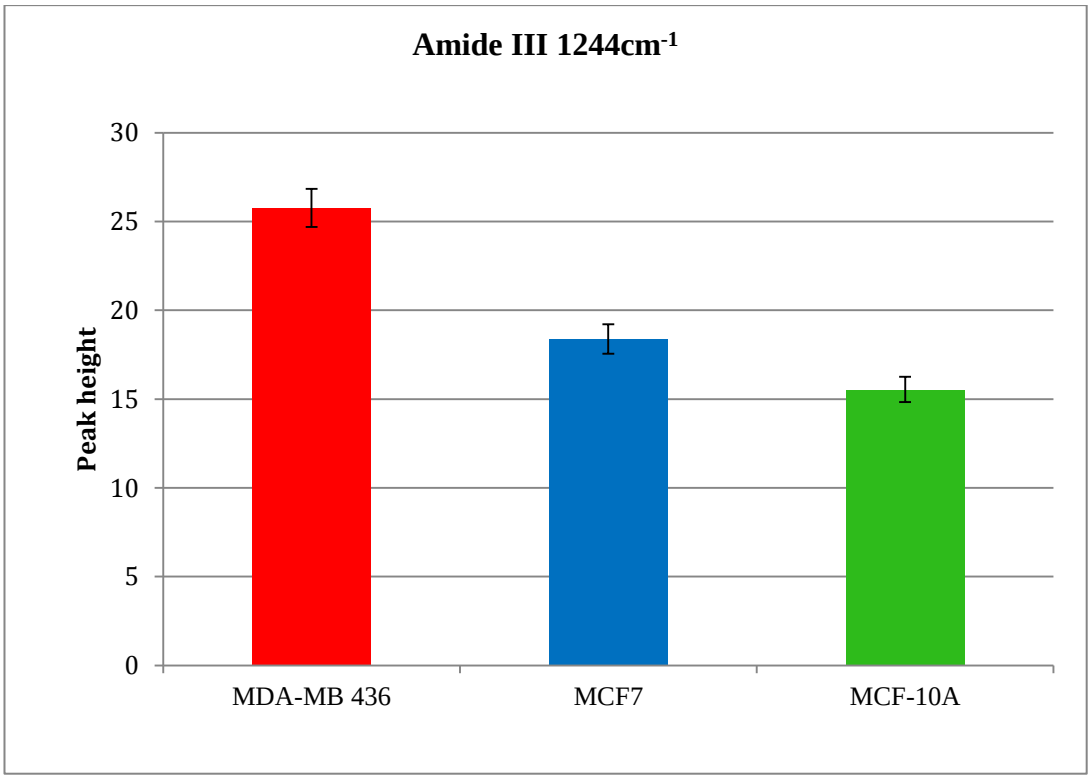

Figure 18: Peak height analysis of lipid region at 2934 cm-1 for MDA-MB-436 (red), MCF-7 (blue) and MCF-10A (green) cell lines.

4.2. Chemometrics analysis

A combination of supervised and unsupervised algorithms was used to analyze different spectral regions of normal and breast cancer cell lines. Raman peaks were found to differ in intensities and peak shifts but chemometric approaches able to identify variations that lead to accurate and reliable separation of three cell lines.

4.2.1 Principal component Analysis (PCA)

PCA was performed over whole spectral range (3200-600 cm-1), fingerprint range (1800-500cm^{-1}), lipids (3100-2680 cm^{-1}), amide I (1800-1530) and amide III regions (1380-1190 cm-1). PCA was conducted among three cell lines to identify chemical variations, which can contribute towards cell line separation.

4.2.1.1 Whole spectral range

PCA with the whole spectral range have shown good overall separation between the three cell lines (Figure 19), In this PCA plot of whole spectral region, the first principal component (PC-1) has separated the two cancer cell lines. Whilst the second principal component (PC-2) has separated MCF-10A (normal) from MCF-7 and MDA-MB-436. In this PCA plot, the first principal component (PC-1) has separated MCF-7A from MDA-MB-436 with higher extent.

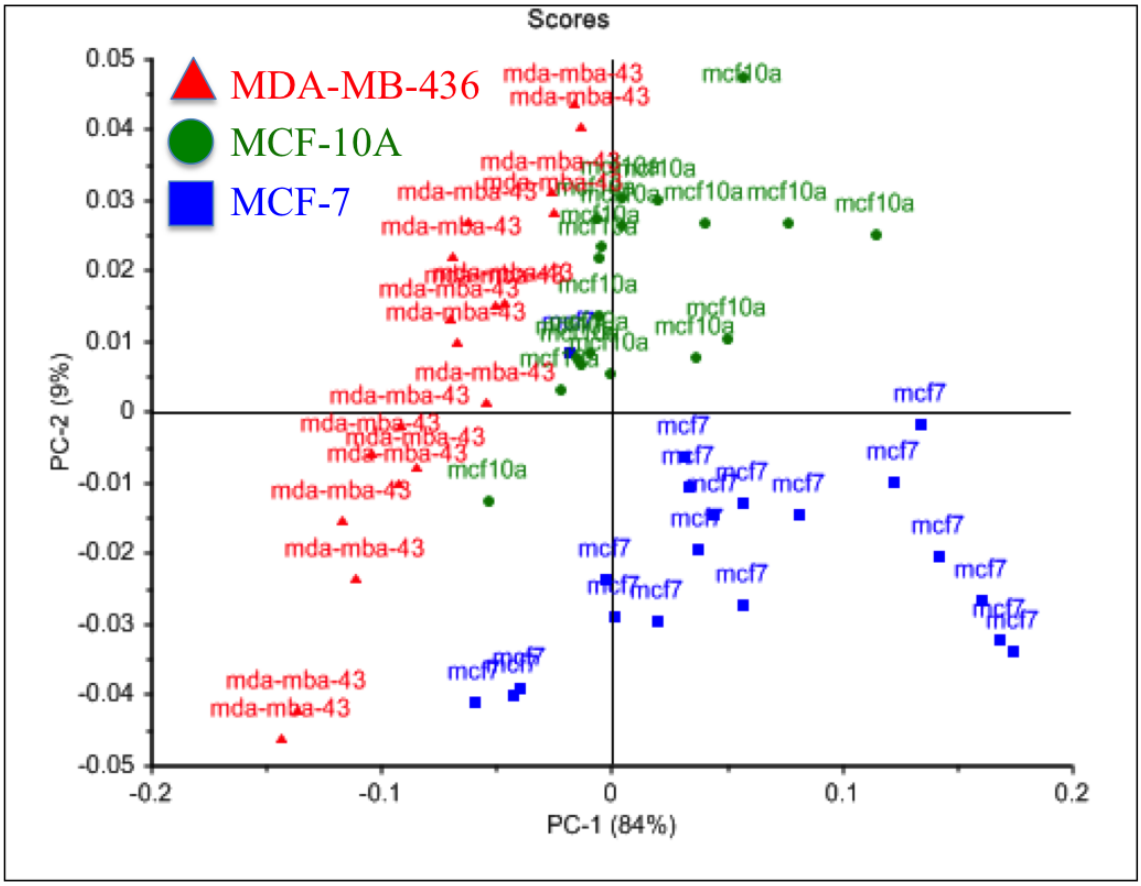

Figure 19: A two-dimensional principal component analysis (PCA) score plot of whole spectral range for MDA-MB-436 (red), MCF-10A (green) and MCF-7 (blue) cell lines based on first (84%) and second (9%) principal components (PC).

4.2.1.2 Fingerprint region

The PCA of the fingerprint region shows overall good separation among three cell lines. In this PCA plot, first PC (PC-1) separates MCF-7 from less extent with MDA-MB-436 and MCF-10A whereas the second PC (PC-2) has pretty much separated triple negative subtype from the normal cell line and estrogen receptor subtype (Figure 20).

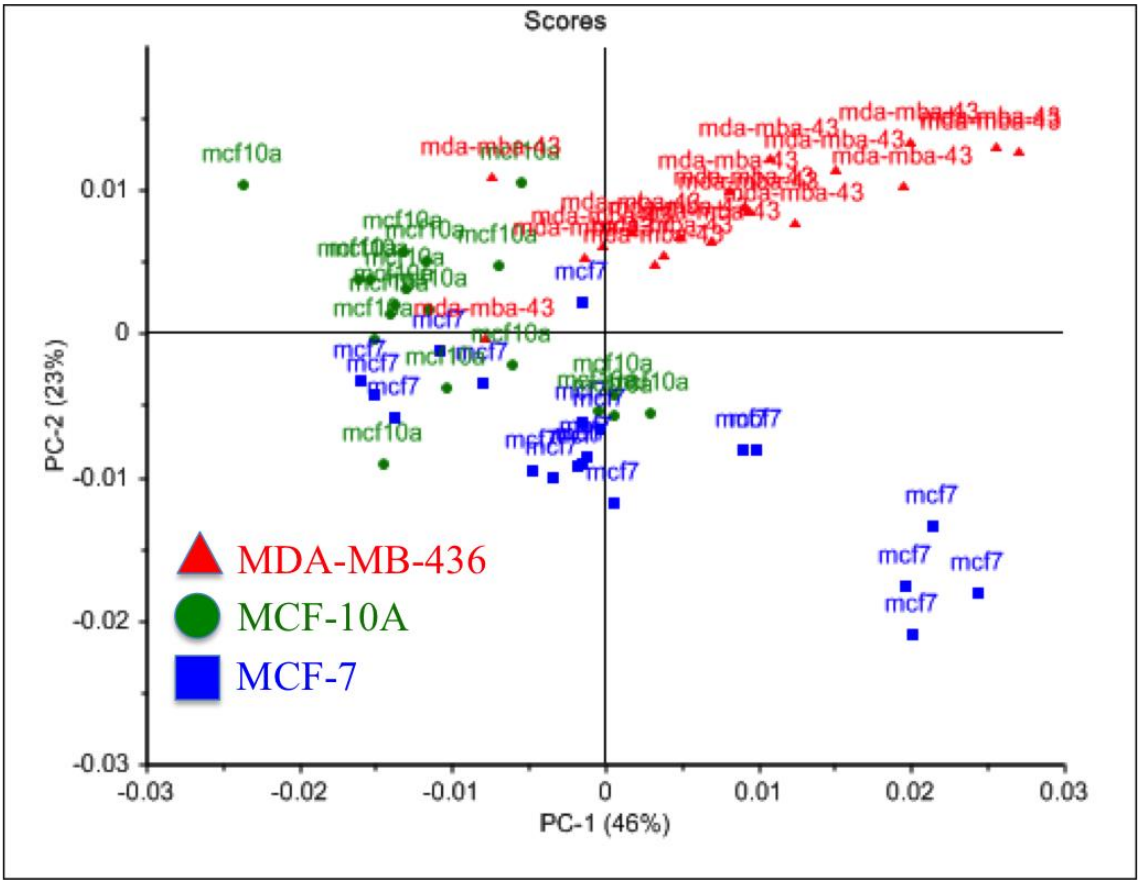

Figure 20: A two-dimensional principal component analysis (PCA) score plot of Fingerprint range for MDA-MB-436 (red), MCF-10A (green) and MCF-7 (blue) cell lines based on first (46%) and second (23%) principal components (PC).

4.2.1.3 Lipid region

PCA was performed over the lipid region between 3100 – 2680 cm^{-1}. In the first PCA plot, the PC-1 (88%) separates MDA-MB 436 from MCF-7. PC-4 separates breast cancer subtypes from the normal breast cell line with lesser extent. In the second PCA plot, PC-1 has clearly separated triple negative subtype from estrogen receptor subtype whereas PC-2 has separated MCF-7 from MDA-MB-436 and MCF-10A cell lines. The 3D plot has shown 3 clear clusters of cell lines using PC-1, PC-2 and PC-4 (Figures 21-23).

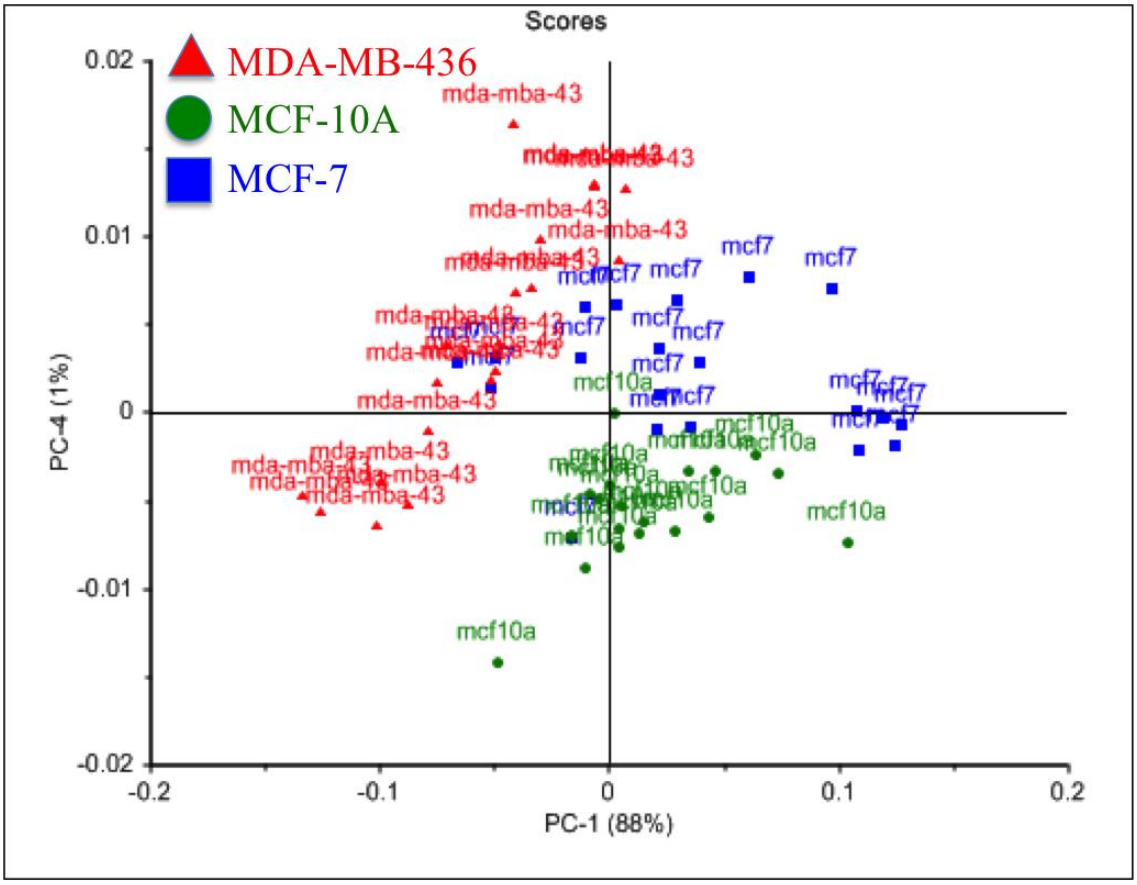

Figure 21: A two-dimensional principal component analysis (PCA) score plot of lipid region for MDA-MB-436 (red), MCF-10A (green) and MCF-7 (blue) cell lines based on first (88%) and fourth (1%) principal components (PC).

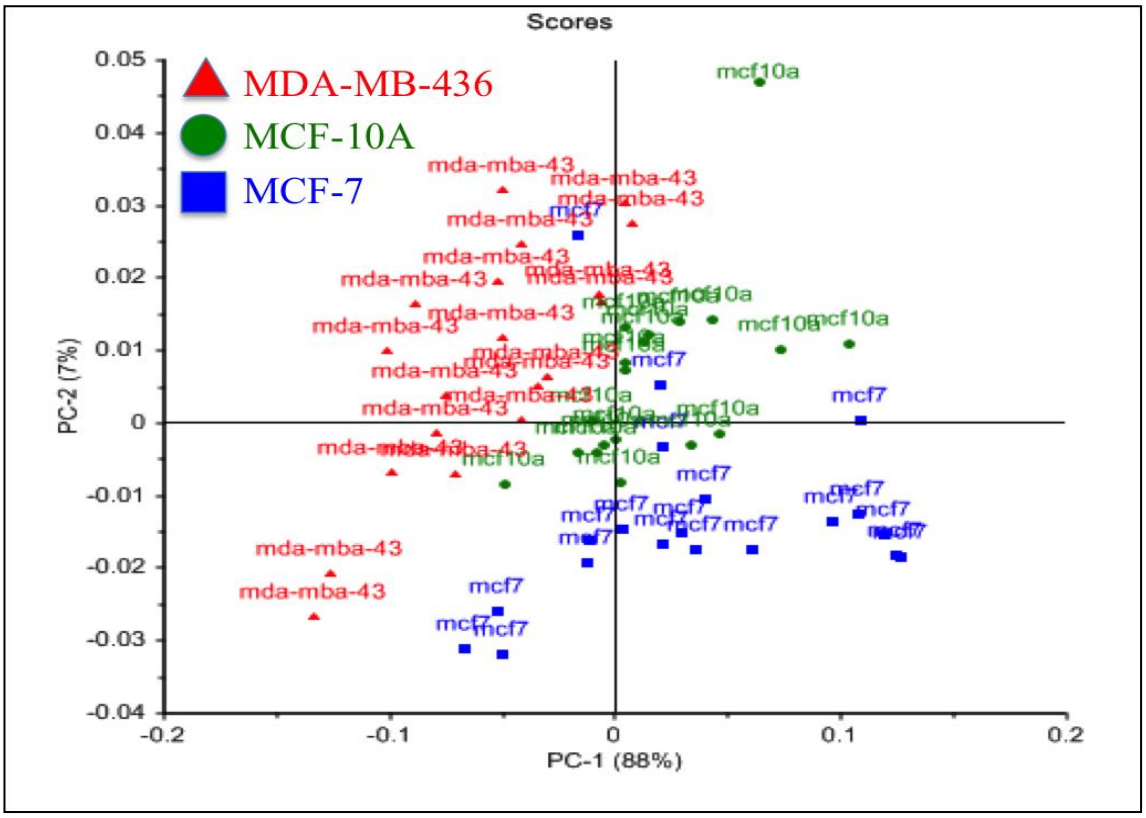

Figure 22: A two-dimensional principal component analysis (PCA) score plot of lipid region for MDA-MB-436 (red), MCF-10A (green) and MCF-7 (blue) cell lines based on first (88%) and second (7%) principal components (PC).

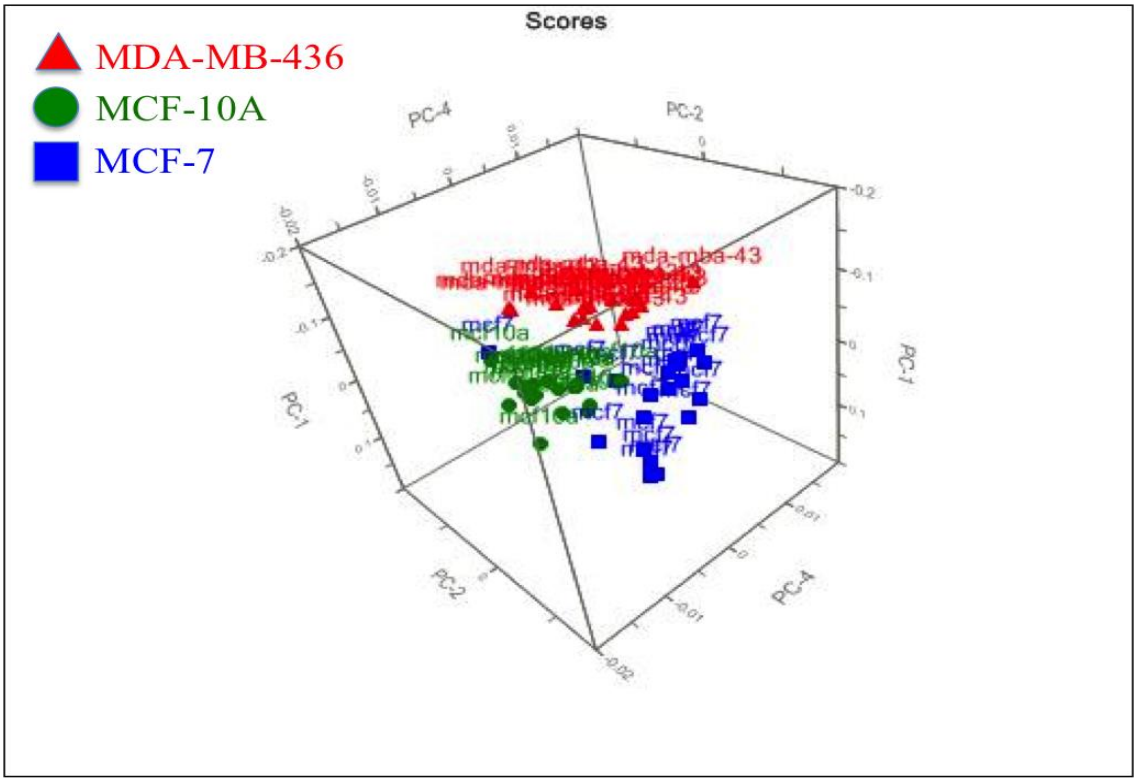

Figure 23: A three-dimensional principal component analysis (PCA) plot of lipid region for MDA-MB-436 (red), MCF-10A (green) and MCF-7 (blue) cell lines based on first (88%), second (2%) and fourth (1%) principal components (PC).

4.2.1.4 Amide I

PCA was used to analyse amide I region between 1800-1530 cm^{-1}. The PCA plot of amide I indicated that principal component PC-2 has clearly separated MCF-7 cell line from IMDA-MB 436 cell line. PC-1 has pretty much separated MCF-7 from MDA-MB-436 and MCF-10A. In the first PCA plot PC-1 separated normal cell line from estrogen receptor expressed cell line. Whereas PC-3 pretty much separated MCF-7 from MDA-MB-436. The 3D plot of amide I has shown clear cluster formation of three cell lines (Figures 24-26).

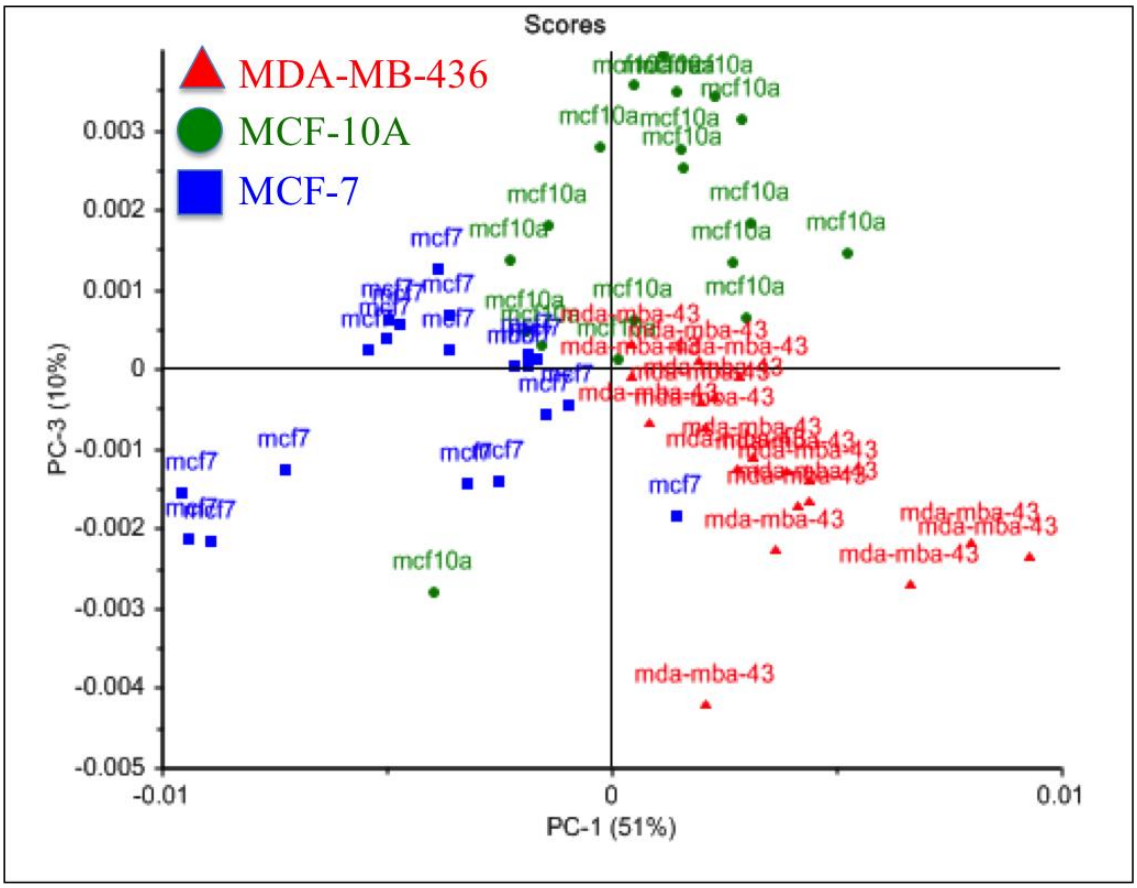

Figure 24: A two-dimensional principal component analysis (PCA) score plot of amide I region for MDA-MB-436 (red), MCF-10A (green) and MCF-7 (blue) cell lines based on first (51%) and third (10%) principal components (PC).

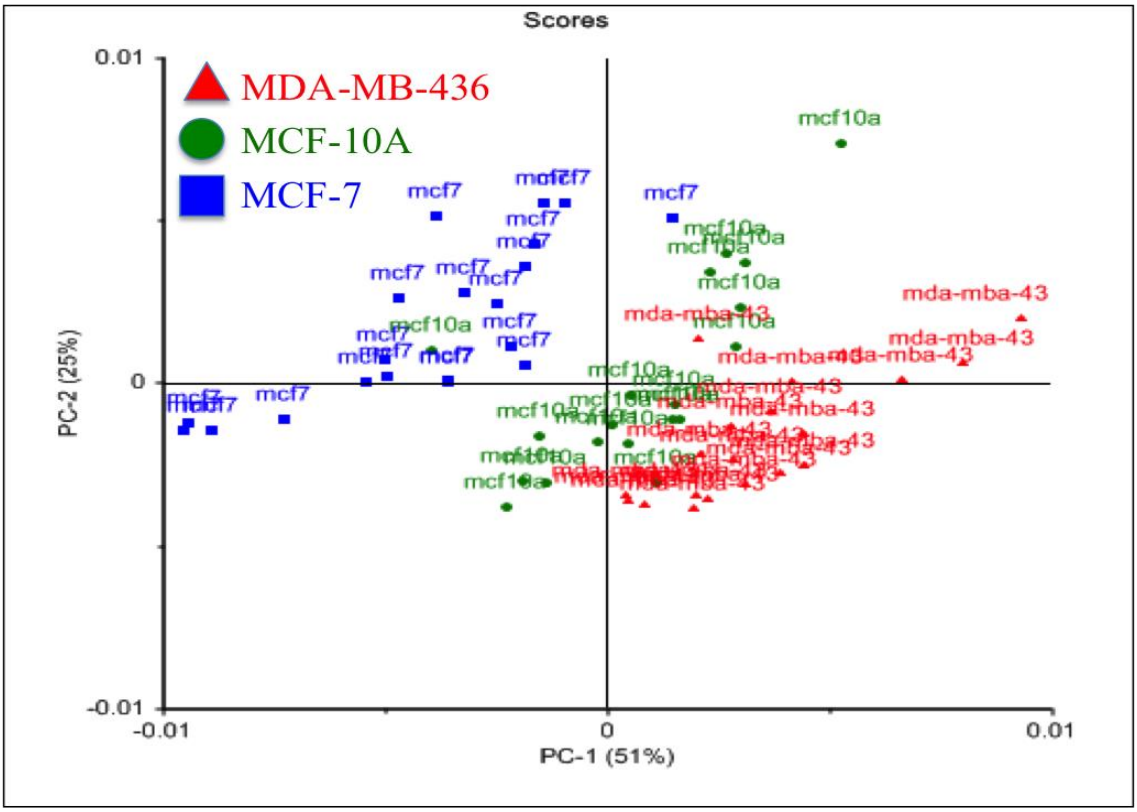

Figure 25: A two-dimensional principal component analysis (PCA) score plot of amide I region for MDA-MB-436 (red), MCF-10A (green) and MCF-7 (blue) cell lines based on first (51%) and second (25%) principal components (PC).

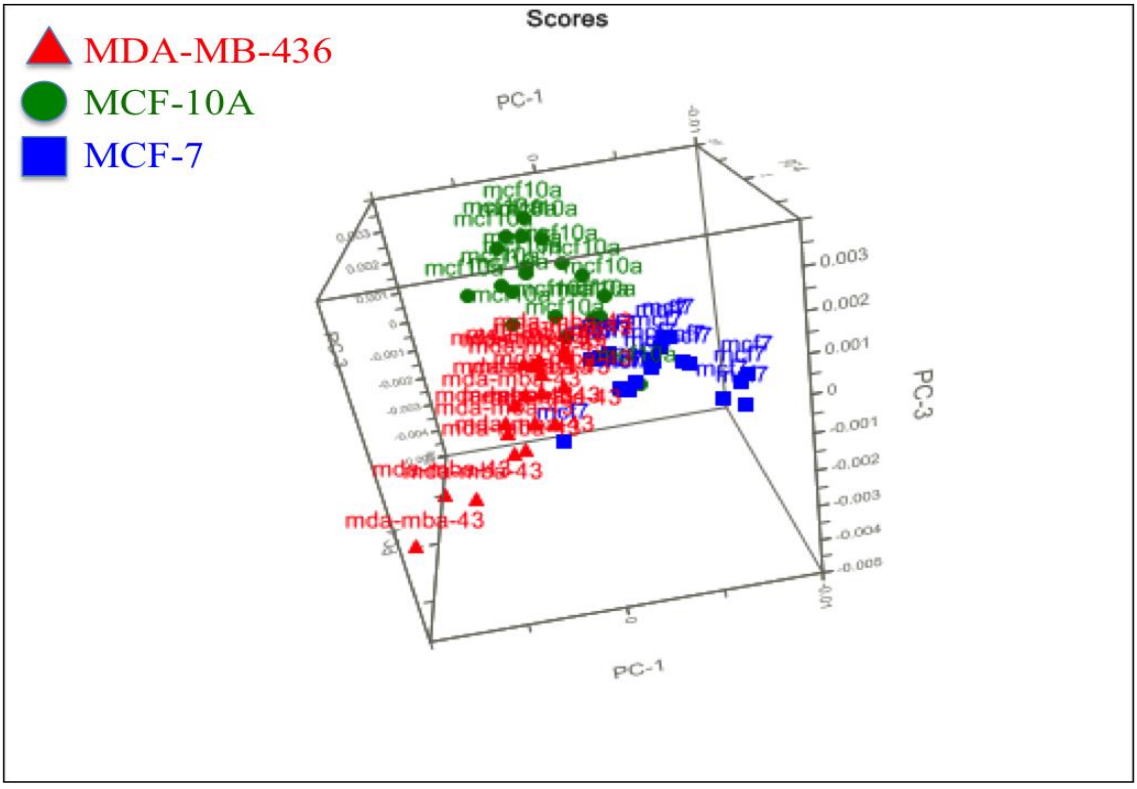

Figure 26: A three-dimensional principal component analysis (PCA) plot of Amide I region for MDA-MB-436 (red), MCF-10A (green) and MCF-7 (blue) cell lines based on first (51%), second (25%) and third (10%) principal components (PC).

4.2.1.5 Amide III

PCA was performed on amide III region (1380-1190 cm^{-1}) to identify chemical differences between three cell lines. The amide III region provides information on lipids, proteins and nucleic acid contents. PCA loading plot of amide III has indicated that PC-1 has pretty much separated MCF-7 cell line from MCF-10-A and MDA-MB436 whereas PC-2 has differentiated MCF-10-A and MCF-7 from MDA-MB436. The second PC-A loading plot of amide III has indicated that PC-3 has well separated MCF-7A and MDA-MB436 from MCF-10A (Figures 27-29).

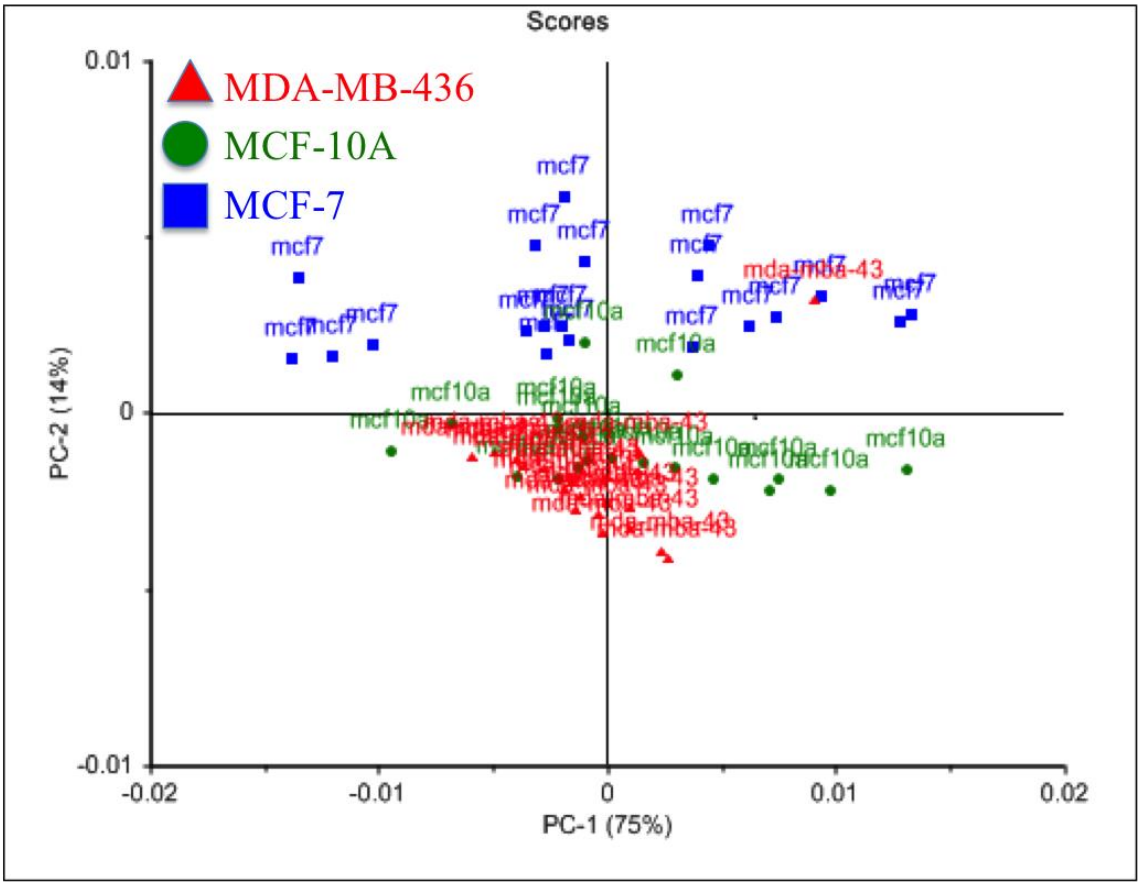

Figure 27: A two-dimensional principal component analysis (PCA) score plot of amide III region for MDA-MB-436 (red), MCF-10A (green) and MCF-7 (blue) cell lines based on first (75%) and second (14%) principal components (PC).

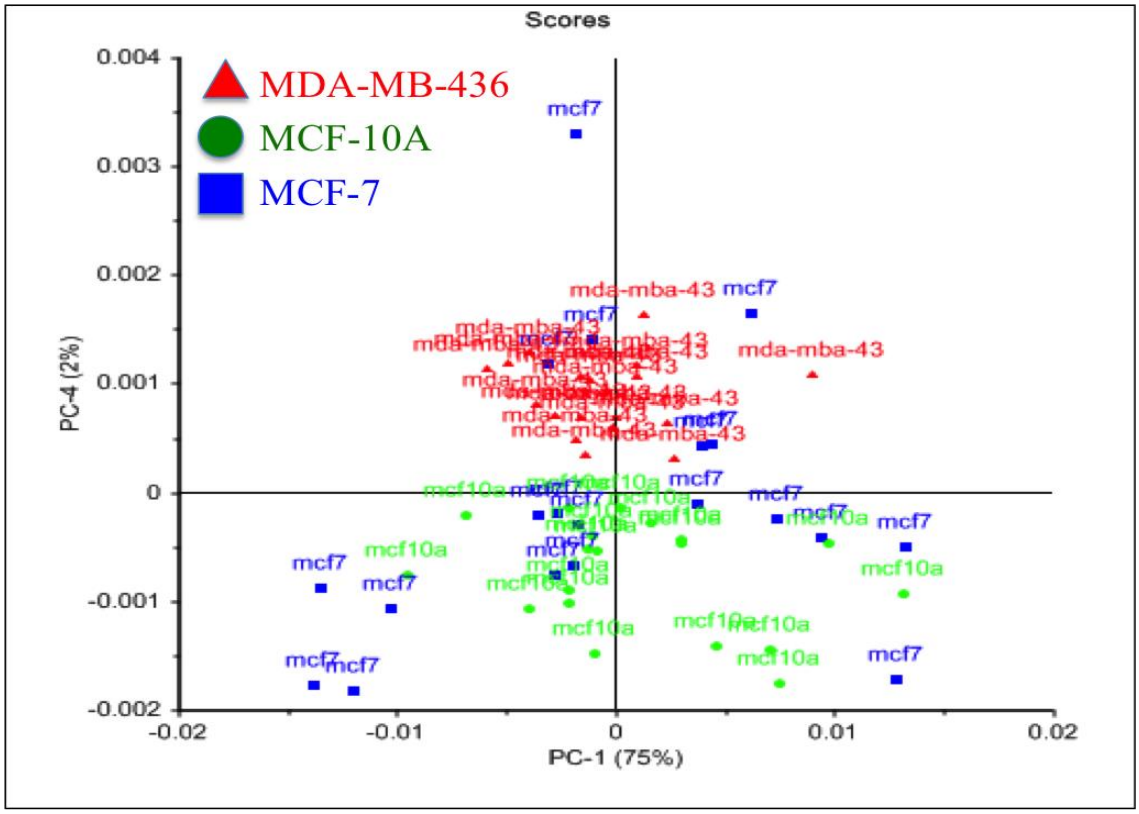

Figure 28: A two-dimensional principal component analysis (PCA) score plot of amide III region for MDA-MB-436 (red), MCF-10A (green) and MCF-7 (blue) cell lines based on first (75%) and fourth (4%) principal components (PC).

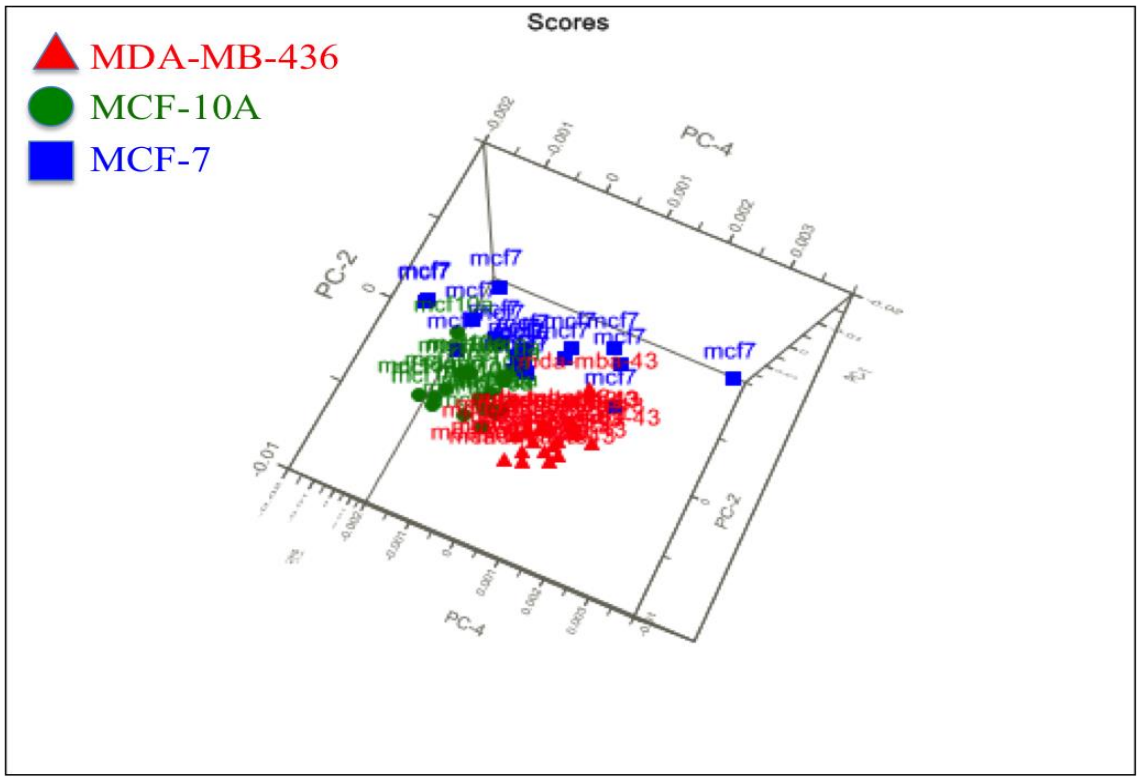

Figure 29: A three-dimensional principal component analysis (PCA) plot of Amide III region for MDA-MB-436 (red), MCF-10A (green) and MCF-7 (blue) cell lines based on first (75%), second(14%) and fourth (2%) principal components (PC).

4.2.2 Linear Discriminant Analysis (LDA)

Linear discriminant Analysis (LDA) models were set up over full spectral range (3300-200 cm^{-1}). Baseline corrections and univector normalization were performed on spectra before setting up LDA models. Five spectra from each group were left out at each for prediction until a total no of twenty spectra from each group were predicted. LDA model of three cell lines was predicted with 100% sensitivity and 91% specificity (Figure 30).

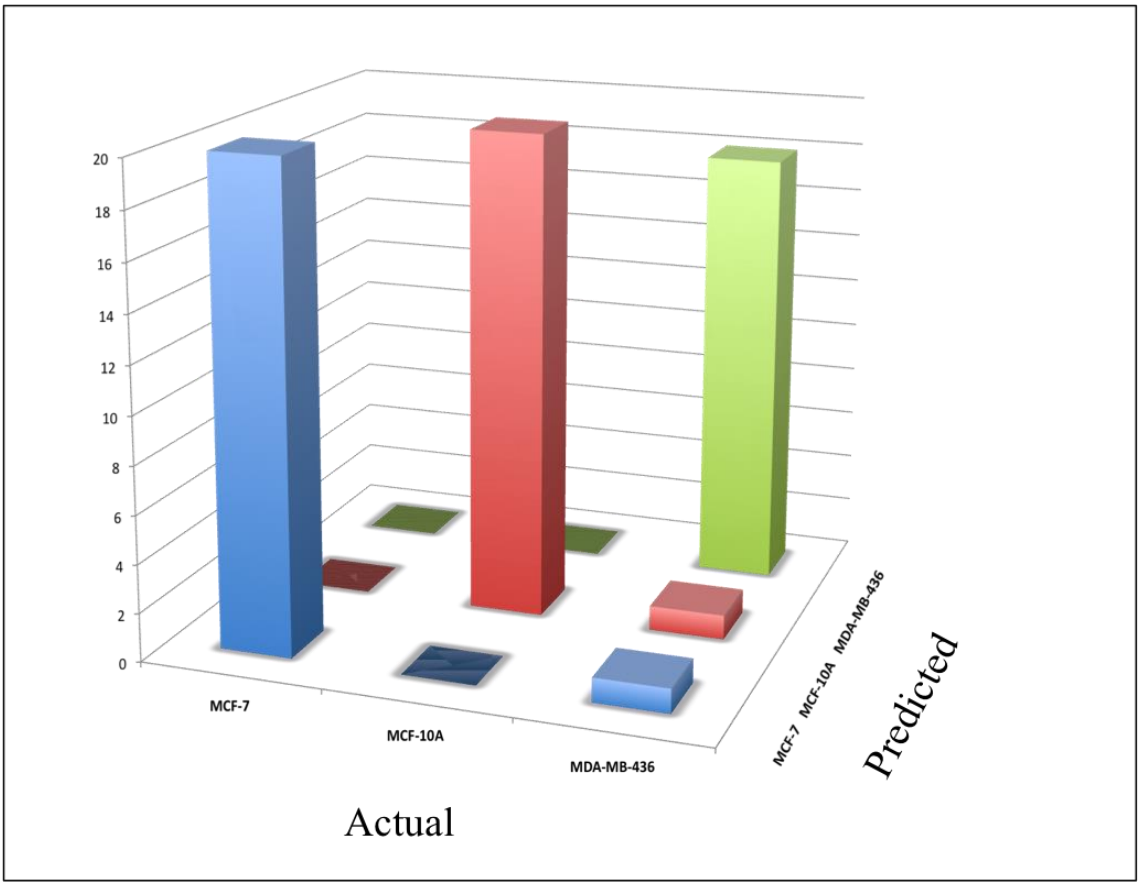

Figure 30: A three-dimensional linear discriminant analysis (LDA) plot for MDA-MB-436 (red), MCF-10A (green) and MCF-7 (blue) cell lines based on predicted sensitivity and specificity.

4.3 Breast cancer cell lines – Discussion

Breast cancer is a most multifarious and heterogeneous disease with different clinical features and different specific treatments. Classical pathology has described structural heterogeneity based on overall morphology and structural organization. But a combination of immunohistochemistry and gene expression profiling studies have understood the disease at molecular level, refining classification based on presence of specific markers such as estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) (Bertos and Park, 2011). This classification method has expressed tumour heterogeneity in to luminal A, luminal B, HER-2 positive and triple negative subgroups. Based on prognostic viewpoint, luminal A and luminal B tumours have given better outcome due to hormone therapy whereas HER-2 positives have developed overall outcome after discovering potential therapeutic candidate such as trastzumab. On the other hand, triple negative tumours have associated with worst prognosis due to lack of any specific marker. Cell lines are extensively used in cancer research as *in vitro* models due to ease to handling, unlimited self-replication and high degree of homogeneity (Holliday and Speirs, 2011). In our study, for the first time, a combination of Raman spectroscopy and multivariate approach was applied to investigate and classify three breast cell lines at cellular level (Talari et al., 2015a).

Spectra obtained for both the normal (MCF-10A) and breast cancer subtypes (MCF-7 and MDA-MB-436) presented major lipid information in the high wavenumber region (3050 - 2800 cm^{-1}) (Figures 9-10) and protein information within the fingerprint region (1800 – 500 cm^{-1}). To demonstrate reproducibility, twenty spectra were collected from each cell line.

The spectra obtained from each subtype showed small visual differences. The concentrations of nucleic acid bases were different in these cell lines and MCF-7 spectra had relative higher amount of DNA bases. Proteins especially in the amide II and III

regions, and lipid ratios were found to be different among these cell lines. The differences in amide peaks might be a result of different amounts of proteins in various structural conformations. These conformations are mainly of the α-helix, unordered and ß sheet structures of the proteins.

Mean Raman spectra of normal (MCF-10A), and two breast cancer cell lines (MDA-MB-436 and MCF-7) showed spectral information about proteins, lipids and nucleic acids. Peaks present at higher wave number regions (2980 to 2800 cm^{-1}) are mainly due to stretching vibrations of lipids and proteins. These peaks are an accumulation of vibrations, such as, $-CH_3$, $-CH_2$ and -CH=CH-. These are considered major spectral markers in the analysis of the disease states. All these cell lines express major lipid peaks at (2940 cm^{-1}, 2921 cm^{-1} and 2948cm^{-1}) with minor changes in intensity. These intensity differences were due to small fluctuations in the concentrations of lipids. Raman spectra of this wavenumber region provide very interesting features regarding the lipid changes in biological systems. These vibrations are extremely useful in assessing lipid peroxidation. At tissue level, it provides crucial information about lipid saturation and may be considered as Raman marker for lipid metabolism. The ratio of different lipid vibrations (-CH=CH- /CH_2) in Raman spectrum will depict disease status of the tissue with respect to lipid metabolism. As the cancer progresses, the intensity will increase due to increased number of unsaturated fatty acids in the cell. This indicates lower degradation of lipids due to decrease in lipid peroxidation. This is the one of the common pointers of rapid proliferating cells. Lipid metabolism in a cell might be affected by the low rate of lipid synthesis, which results in the decrease number of saturated fatty acids. In addition, extreme lipid peroxidation causes higher level of lipid degradation. This is a very good marker for the slow growth behaviour of a cell. Recent investigations have revealed that high amounts of polyunsaturated fatty acids have great influence on transport and signalling pathways, including growth

promoting effects of cell adhesion, cell shape and metastatic behaviour of cancer cells (Ozek et al., 2010c). High wave number vibrations (3900 – 2850cm^{-1}), especially CH_2 stretching vibrations are regarded as best Raman markers to identify the disorder state of lipids in diseased tissue.

Fingerprint region of Raman spectra (1800 to 500 cm^{-1}) highlight different biological components such as proteins, lipids, nucleic acids and fatty acids. The major Raman peaks for Amide I, II and III, C-S stretching, C-H bending, O-P-O vibrations, Phenylalanine, Tyrosine, C-C stretching, CH_2 twisting, CH_2 deformations and C=C stretching of lipids were observed in these cell lines (Rehman, 2012). The other Raman peaks observed in these cell lines were nucleic acid bases such as adenine, thymine, uracil, guanine and cytosine (ring breathing mode), tryptophan (C-C twisting) and phenylalanine (C-C aromatic ring stretching), proline (C-C stretching), glycogen, CH_2, CH_3 deformations of lipids (Talari et al., 2015k). Proteins function as powerful Raman molecular markers in studies of cancer biology. Raman spectra provide biochemical aspects of the protein metabolism such as transcription (RNA synthesis) and translation (protein synthesis) with respect to metastasis. Amide I, II and III bands were observed in all three breast cancer cell lines. These different amide vibrations result from the amide bonds of polypeptide backbones. Amide I band is the result of stretching vibrations of C=O vibrations of the amide group and it represent α-helical structures of proteins. Bending vibrations of N-H of amide group as well as C-N stretching and C-N-H in-plane bending modes of amide groups are revealed in the amide II and III bands respectively. Amide III represents ß sheets and random coils of proteins. Higher levels of RNA and protein content are a fundamental feature of cancer cell.

Peak intensities have shown clear differences among three cell lines. Quantification of peak heights were performed on lipids (2934 cm^{-1}), amide I (1658 cm^{-1}) and amide III (1244 cm^{-1}) regions. Peak heights analysis has shown clear differences between cell

lines. The absolute intensity of each of the region is increased from MCF-10A to MCF-7 to MDA-MB-436. PCA with the whole spectral range and the fingerprint range shows good overall separation between the three cell lines , suggesting that lipids, proteins and nucleic acids might be contributing to the separation. In the PCA of whole spectral region, the first principal component (PC-1) has separated the two cancer cell lines whereas the second principal component (PC-2) has separated MCF-10A (normal) from MCF-7. The PCA of the fingerprint region shows the first PC-1 separates triple negative subtype from the normal cell line and PC-2 separates MCF-7 from MDA-MB-436. A notable feature is that two separate clusters representing a 'normal' class and a 'cancerous/diseased' class do not form. This suggests very large biochemical variation even between the two breast cancer cell lines.

To characterise the biochemical variation in detail, PCA has been further applied to narrower spectral regions such as lipids, amide I and amide III. Some spectral regions are more sensitive to a low number of these components. And as the PCA of whole and fingerprint region have provided sufficient evidence on separation, looking at small spectral regions may provide a better understanding of the chemical differences existing between these breast cell lines. For example, the Amide I band shoulder can give information on the relative concentration of nucleic acids and DNA bases, the high wavenumber region can often be used in conjunction with the 1200-1000- cm^{-1} region to resolve information on the lipid phase and content. This band approach can sometimes provide a clearer picture of the biochemical differences that are of interest to us. Loadings plots for these regions are used in conjunction with the PCA scores to provide information on the spectral features, which are powering the separation between the three cell lines.

4.3.1 Lipids

PCA was performed over the lipid region between 3100 – 2680cm^{-1} (Figures 21-23).

This region often gives information on the relative amount of proteins and lipids. The PC-1 (88%) separates MDA-MB 436 from MCF-7. PC-4 separates breast cancer subtypes from the normal breast cell line. The loadings plots suggest that the v(CH_2) symmetric and asymmetric vibrations of lipids at (2882 cm^{-1}), C-H, CH_2 symmetric vibrations in lipids and proteins (2940 cm^{-1}, 2921 cm^{-1} and 2948cm^{-1}) appear to be powering the separation of the MCF-7 line from the other two lines (Koljenovic et al., 2005a, Sigurdsson et al., 2004, Shetty et al., 2006, Movasaghi et al., 2007d). The cell lines higher in the CH stretching vibrations of proteins (2946cm^{-1}), v_sCH3 of lipids and proteins (2884cm^{-1}) are positive for the separation PC-1 describes and it discriminating MDA-MB-436. Whereas, the CH_3 stretching vibrations (2910cm^{-1}) and CH_2 symmetric stretch of lipids (2886cm^{-1}) are contributing to the separation of the MCF-7 and MCF 10A lines. In each case, the MDA-MB-436 and the MCF 10A cell lines have higher protein contents relative to the MCF-7 line. Whilst both the MDA-MB-436 and MCF 10A are easily resolved from the MCF-7 cluster due to the much lower lipid content, their protein-to-lipid ratio differs significantly. The difference is large enough to further separate MDA-MB-436 from MCF 10A.

4.3.2 Amide I

PCA was used to analyse amide l region between 1800-1530cm^{-1} (Figure 12). The Amide I band gives information on the secondary structure of proteins, nucleic acids/DNA bases and some amino acids. PC-1 has separated the MCF-7 line from rest of the cell lines.

PC-1 loading plot indicates the C=O stretching of amide I of proteins (1687 cm^{-1}), anti-parallel ß sheets of amide I (1670 cm^{-1}), tryptophan or ß sheet of protein (1621 cm^{-1}), C=C of phenylalanine ring vibration, tyrosine (1607 cm^{-1}), tryptophan (1548 cm^{-1}) are higher in MCF-10A and MDA-MB-436 cell lines whereas nucleic acids (1576cm^{-1}) are

higher in MCF-7 cell line. MCF-7 separation from the other cell line is therefore likely based on the higher concentrations of DNA bases and lower protein content seen in this line. The MDA-MB-436 and MCF-10A cell lines contain higher concentrations of protein compared to the MCF-7 line. However, the amount seems to vary between the two. PC-3 and PC-4 loading plots are also separating MCF-7 in a similar manner. Raman peaks of these loading coefficients are 1577cm^{-1} and 1576 cm^{-1} which are assigned to nucleic acids and are higher in the MCF-7 cell line. This variation relative to the DNA base bands is separating MDA-MB-436 from theMCF-10A (Mahadevan-Jansen and Richards-Kortum, 1996, Shetty et al., 2006, Dukor, 2002, Cheng et al., 2005, Lakshmi et al., 2002, Notingher et al., 2004, Stone et al., 2004, Fung et al., 1996).

4.3.3 Amide III

PCA was performed on amide III region (1380-1190 cm^{-1}) to identify chemical differences between the cell lines (Figure 15). The Amide III band gives information on lipids, proteins, nucleic and amino acid contents. PC-2 has separated MCF-7 from rest of the cell lines. MCF-7 separation is largely due to a higher nucleic acid and lipid concentration in this cell line relative to the rest. This is also observed in PC-2, which is describing positive discrimination with the amide III and nucleic acid bases such as adenine and guanine (1337 cm^{-1}), CH_2 deformation of lipids, adenine and cytosine (1258, 1299 and 1304 cm^{-1}), methylene twisting vibrations (1294 cm^{-1}). The difference between the MCF-10A and MDA-MB-436 lines are found in PC-4 where the MDA-MB-436 are clustered away from the normal line due to a higher protein and lipid content (Dukor, 2002, Stone et al., 2004, Chan et al., 2006).

Linear discriminant analysis (LDA) models were setup over full spectral range 3300-200 cm^{-1}. Baseline corrections and Unit Vector Normalisation were performed on spectra before setting up LDA models. Five spectra from each group were left out at

each for prediction until a total number of twenty spectra from each group were predicted. LDA model of three cell lines was predicted with 100% sensitivity and 91% specificity.

Raman spectral data combined with PCAs proves to be an excellent method that allows separation of different types of cell lines based on their lipid, and nucleic acid/DNA, and protein contents, as the MCF-7 cell line appears to be much higher in lipids compared to MDA-MB 436 and MCF 10A and works well to single out this cell line in view of the high wavenumber of amide III bands. It also contains higher nucleic acid/DNA base concentrations relative to the other two cell lines. This might indicate higher metabolic activity or might simply be due to a higher cell density in the data collection volume. Although MCF-7 and MDA-MB-436 are both breast cancer subtypes, the MDA-MB-436 does not appear to contain lipids at a concentration vastly different to those found in the normal MCF-10A cell line. Instead, the difference lies more in the relative protein and amino acid concentrations. Proteins of many different conformations overlap to form the Amide bands. It is possible that the compositions of α-helix, unordered and β-sheet conformations of proteins in the MDA-MB-436 and MCF-10A lines are different. These differences observed in the cell lines may also be useful in identifying chemical changes between the different subtypes of breast cancer although this needs confirmation in a larger panel of cell lines as well as clinical material.

5.1 T47-D Spheroids – Results

T-47 D spheroid cultures were grown few millimetres until 14 days. Initially they are composed of homogeneous population of normal proliferating cells. As the time progressed, the size of spheroid has increased. After 14 days culture, necrotic region was developed in central region of the spheroid and this region was clearly observed in microscope. This region was surrounded by layer of non-proliferating quiescent cells. The external rim of T-47D spheroid remained normal proliferating cells. Based on microscopic observation, three regions were calculated as outer region (normal proliferating region), middle region (hypoxic region) and inner region (necrotic region) and these regions were well established in previous studies. (Sutherland, 1988, Chapman et al., 1981, Carlsson et al., 1979). Raman Spectroscopy has been used as potential tool to evaluate biochemical differences among three regions. T-47D spheroids biochemical snapshots have shown spatial diversity of macromolecules such as proteins, lipids, carbohydrates and nucleic acids. Raman peaks of three regions have differed in peak intensities and peak shifts. Furthermore, three regions of T 47-D spheroid cultures models were classified using multivariate approach such as PCA, cluster analysis and LDA. Chemometric methods have assisted in accurate and reliable classification with improved sensitivity and specificity. Raman spectroscopy has been revealed to be accurate in differentiation of normal and cancerous tissues and different breast cancer cell lines, such as, MCF-7, MCF-10 A and MDA-MB-436 (Talari et al., 2015a). Results presented in this section are to classify three region of T-47D using a combination of Raman spectroscopy and multivariate approach.

A total of 90 spectra were collected from three regions of T-47D spheroids. Mean spectra of each region was extracted and carefully analysed for peak identification, and corresponding chemical interpretation (figure 31). Mean spectra of whole region

(400cm^{-1} - 3600 cm^{-1}) and major peak assignments such as Lipid, amide I, II, III and phenylalanine regions were shown in table (appendix).

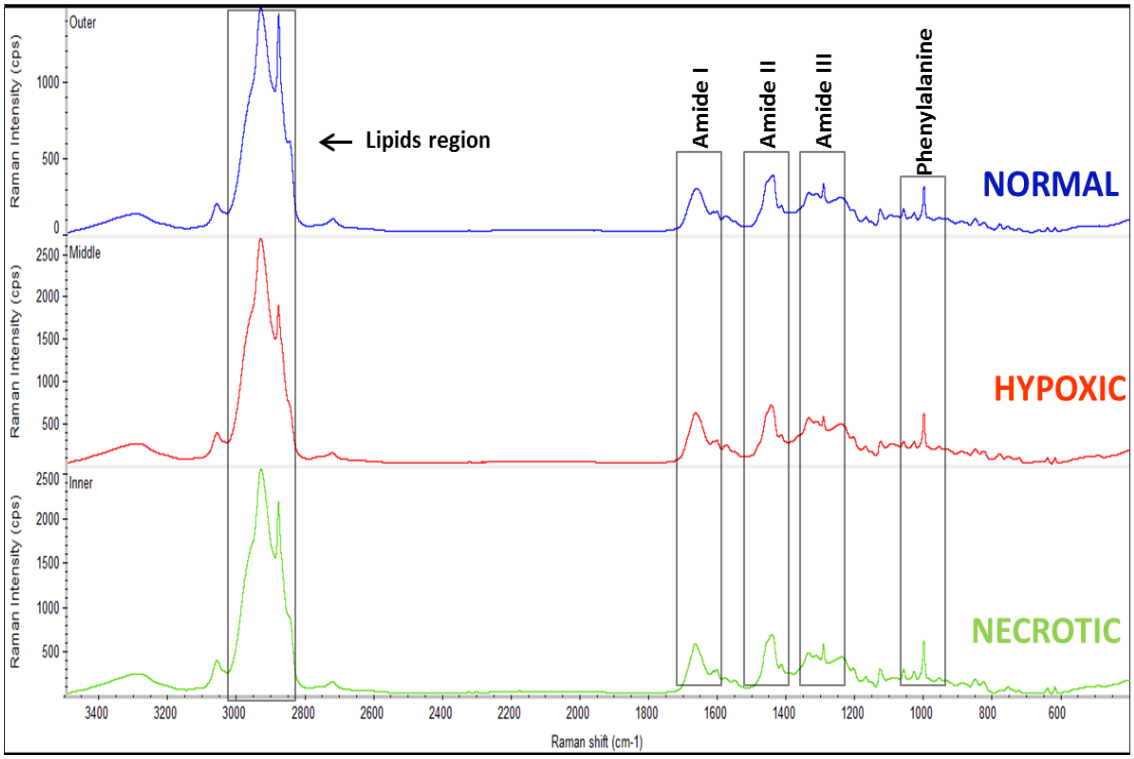

Figure 31: Mean spectra of normal proliferating, hypoxic and necrotic regions of T-47D spheroids

.

5.2 Raman peak analysis

A total of forty spectra were collected from normal proliferating region, hypoxic region and necrotic regions. The mean spectra of these three regions were taken and major biological component regions are mentioned in table. Raman spectra of three regions were analysed qualitatively and quantitatively. The peak assignments for lipid, amide I, II and III region are summarised in table (appendix).

5.2.1 C-H region (3200 - 2700 cm^{-1})

A strong Raman peak arising at 2933 cm^{-1} is characteristic assignment of CH_2 asymmetric stretch of lipids. This band is expressed in normal proliferating region and hypoxic region, whereas, in the necrotic region this peak appears at 2944 cm^{-1}. A weak

Raman band is expressed in all three spheroid regions of spheroids at 3061 cm^{-1}, representing CH stretch of lipids. A strong shoulder peak was observed in all three regions has been attributed to CH_2 symmetric stretch of lipids and CH_2 asymmetric stretch of lipids and proteins. A further small shoulder peak was observed in normal proliferating region at 2849 cm^{-1}, which can be assigned to CH_3 symmetric stretch of lipids (Figure 32). A weak Raman band observed at 2724 cm^{-1} in normal and necrotic regions is assigned to either CH stretching or stretching vibrations of CH, NH and OH groups. This was reported in hypoxic regions at 2725 cm^{-1} (Movasaghi et al., 2007d).

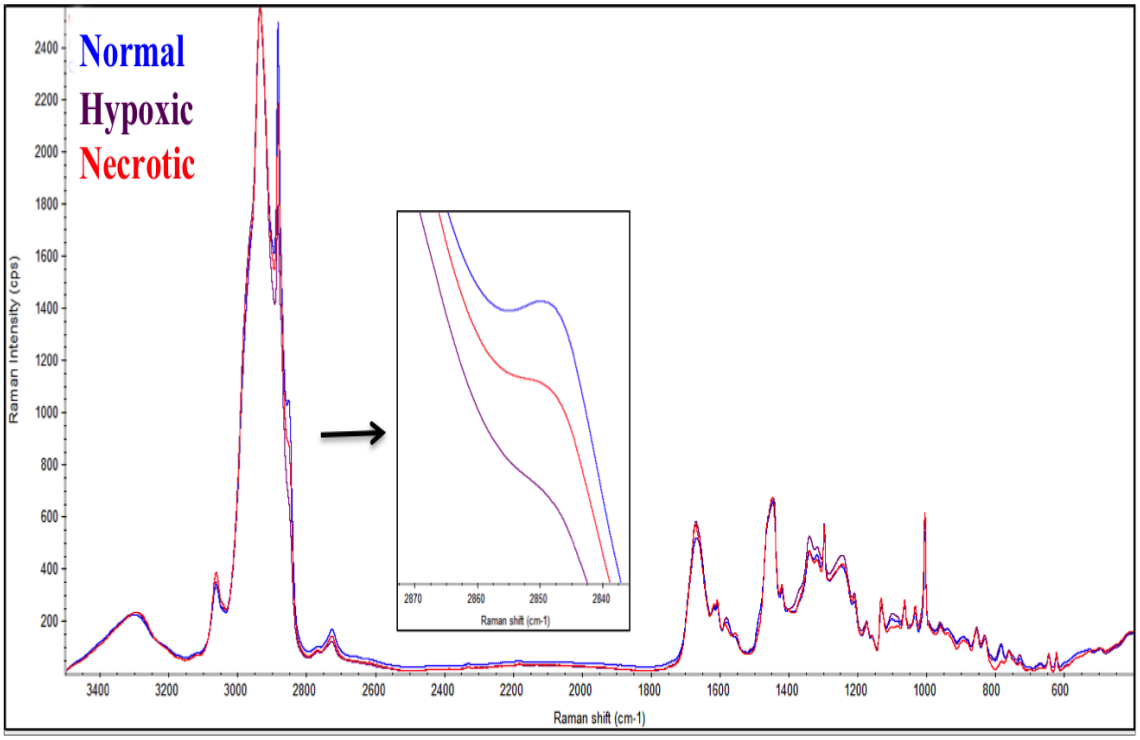

Figure 32: Mean Raman spectra of normal proliferating (blue), hypoxic (dark red) and necrotic (red) regions with major lipid shoulder peak shift.

5.2.2 Amide I region (1672-1350 cm^{-1})

A strong Raman band of amide I was observed at 1666 cm^{-1} (normal), 1667 cm^{-1} (hypoxic) and 1668 cm^{-1} (necrotic) region. This may be due to either; C=C stretching of proteins (β sheets structure of proteins or carbonyl stretches associated with tumours or collagen) (Figure 33). A small Raman peak was observed at 1605 cm^{-1} in hypoxic and necrotic regions, which is a contribution from cytosine (NH_2), ring (C-C) stretch of

phenylalanine, tyrosine, C=C vibrations of proteins, whereas, this band was expressed in normal region at 1604 cm^{-1}. Weak spectral contributions were noticed at 1580 and 1584 cm^{-1}, which are attributed to C-C stretching and C=C of olefinic acids, respectively. Pyrimidine ring of nucleic acid bases was observed at 1579 cm^{-1} in hypoxic region. Weak spectral contribution was observed in necrotic region at 1553 cm^{-1}, which is contribution from CH_2 stretching/CH_3 asymmetric deformation of proteins (Talari et al., 2015k).

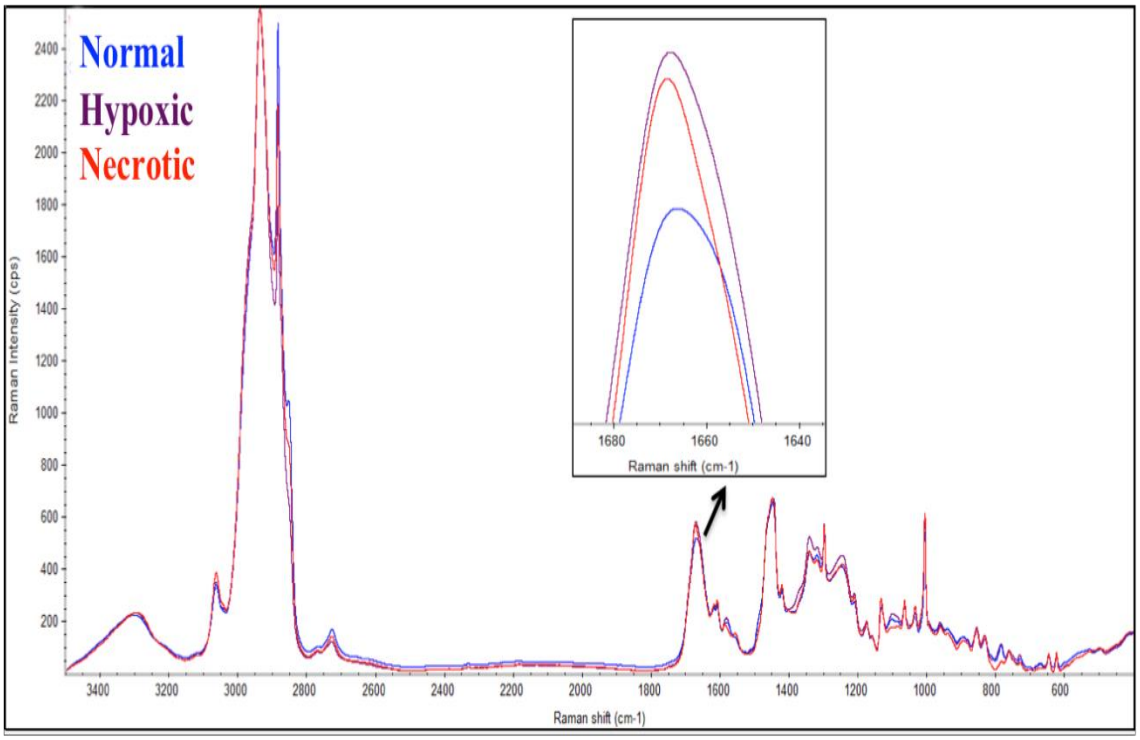

Figure 33: Mean Raman spectra of normal proliferating (blue), hypoxic (dark red) and necrotic (red) regions with major amide I peak shift.

A strong Raman peak was observed at 1441 cm^{-1} in normal region, attributed to CH_2 scissoring and CH_3 bending in lipids, cholesterol and its esters and C-H bending mode of lipids. Hypoxic region expressed a strong Raman peak at 1446 cm^{-1}, characteristic of CH_2 bending mode of proteins and lipids and CH_2 deformation of proteins. A sharp and intense peak observed at 1448 cm^{-1} in hypoxic region due to CH_2CH_3 deformation and CH_2 deformations of collagen. A small shoulder peak was observed in three regions at 1417 cm^{-1} which represents C=C stretching in quinoid (Movasaghi et al., 2007d) .

5.2.3 Amide II region (1240-1350 cm^{-1})

Normal region has shown Raman peaks at 1337 cm^{-1} due to amide III, CH_2 wagging vibrations from glycine back bone and proline side chains, ring breathing mode of adenine and guanine, C-H deformation of proteins and tryptophan. A similar peak was observed at 1339 and 1338 cm^{-1} in hypoxic and necrotic regions respectively. This might be result of C-C stretch of phenyl (1) and C_3-C_3 stretch and C_5-O_5 stretch CHα in-plane bend. The peak arising at 1316 cm^{-1} found in normal and necrotic regions is attributed to ring breathing mode of nucleic acid bases. Bands present at 1294 cm^{-1} (normal and necrotic) and 1295 cm^{-1} (hypoxic) are characteristic of methylene twisting. Asymmetric (PO_2) stretching modes of nucleic acids are found at 1241 cm^{-1} in necrotic region, whereas, this phosphate stretching is found at 1243 cm^{-1} in hypoxic region. A characteristic amide III band was found in normal region at 1245 cm^{-1}. Band present at 1243 cm^{-1} also indicates CH_2 wagging, C-N stretching of collagen, cytosine and thymine (Talari et al., 2015k).

5.2.4 Amide III region (1000-1240 cm^{-1})

A characteristic amide III band was present on all three regions (normal proliferating, hypoxic and necrotic) at 1208 cm^{-1}, which indicates ν(C-C_6H_5), tryptophan, phenylalanine of protein assignment and ring breathing mode of adenine and thymine. A small Raman peak was observed at 1172 cm^{-1}, which represents δ(C-H), tyrosine whereas this peak was observed at 1171 cm^{-1} in both normal and necrotic regions. Peaks present at 1129 and 1130 cm^{-1} are typical assignments of acyl chain vibrations of phospholipids. Molecular vibrations at ν(C-C) skeletal of acyl back bone in lipids, C-C skeletal stretch. Bands present at 1061 and 1062 cm^{-1} are typical assignment of C-C in plane bending and C-C skeletal stretching respectively. Hypoxic region expressed a Raman peak at 1096 cm^{-1}, attributed to Phosphodioxy (PO_2) groups. Normal and necrotic regions expressed C-C vibrational modes of the gauche-bonded chain, amide III and ν(C-C) lipids and fatty acids. All three regions of T-47D spheroids show a

Raman peak at 1031 cm^{-1}, attributed to δ(C-H), p. A sharp and intense peak observed at 1002 cm^{-1} in normal and necrotic regions and 1003 cm^{-1} in hypoxic region is an assignment to ring breathing of aromatic amino acid such as phenylalanine (Figure 34) (Movasaghi et al., 2007d).

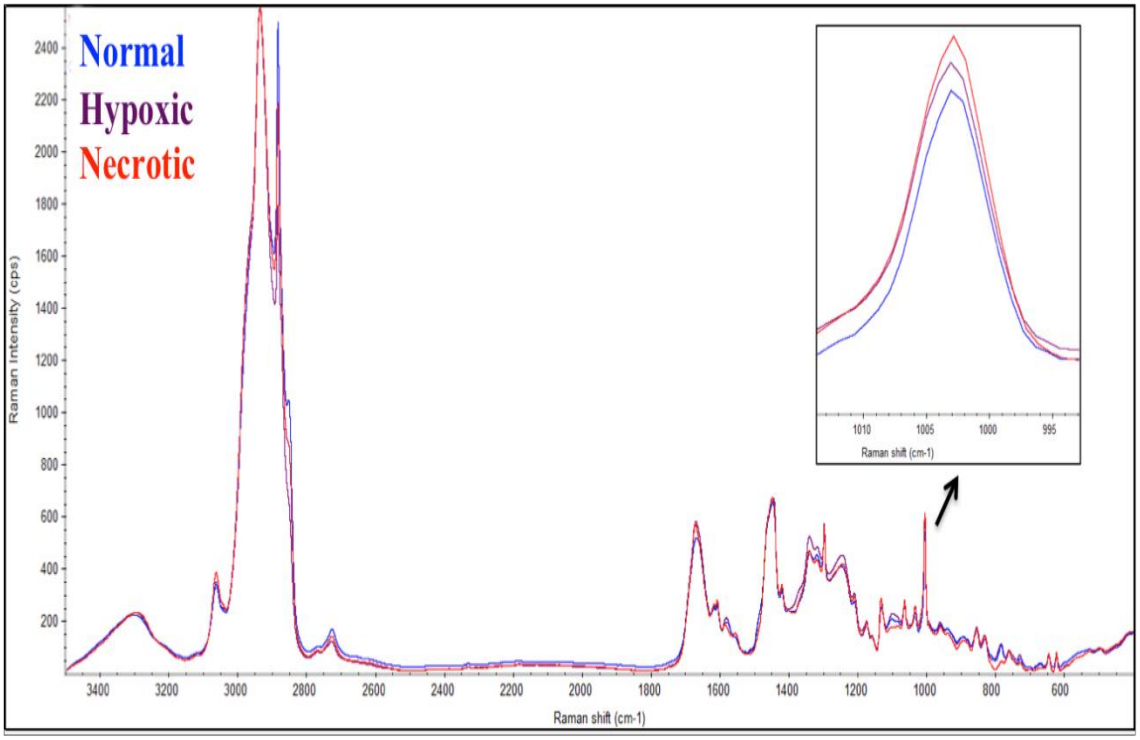

Figure 34: Mean Raman spectra of normal proliferating (blue), hypoxic (dark red) and necrotic (red) regions with phenylalanine peak shift.

5.2.5 Nucleic acid and aminoacid region (600-1000 cm^{-1})

Spectral peaks at 621 and 643 cm^{-1} are typical assignments of C-C twisting mode of phenyl and tyrosine respectively. Hypoxic region shows a Raman band at 727 cm^{-1}, which is associated with C-C stretching and proline (collagen assignment). All three regions of (normal proliferating, hypoxic and necrotic) spheroids expresses indole group containing aminoacid at 758 cm^{-1}. Normal and hypoxic regions express nucleic acid bases, such as, thymine, cytosine and uracil at 782 cm^{-1} whereas necrotic region does not express this peak. A Raman band at 828 cm^{-1} in normal region is typical assignment of out-of-plane ring breathing, tyrosine, O-P-O stretch of phosphodiester. This peak is expressed in hypoxic and necrotic regions at 829 cm^{-1}. All three regions of spheroids

expresses ring-breathing mode of tyrosine, C-C stretch of proline ring and glycogen at 853cm^{-1}. Protein peaks of structural proteins modes were expressed in necrotic regions at 890cm^{-1}, whereas, saccharide spectral peak was reported in normal region at 891 cm^{-1}. Cholesterol peak at 958 cm^{-1} was observed in all three regions of spheroids (figure 35) (Talari et al., 2015k).

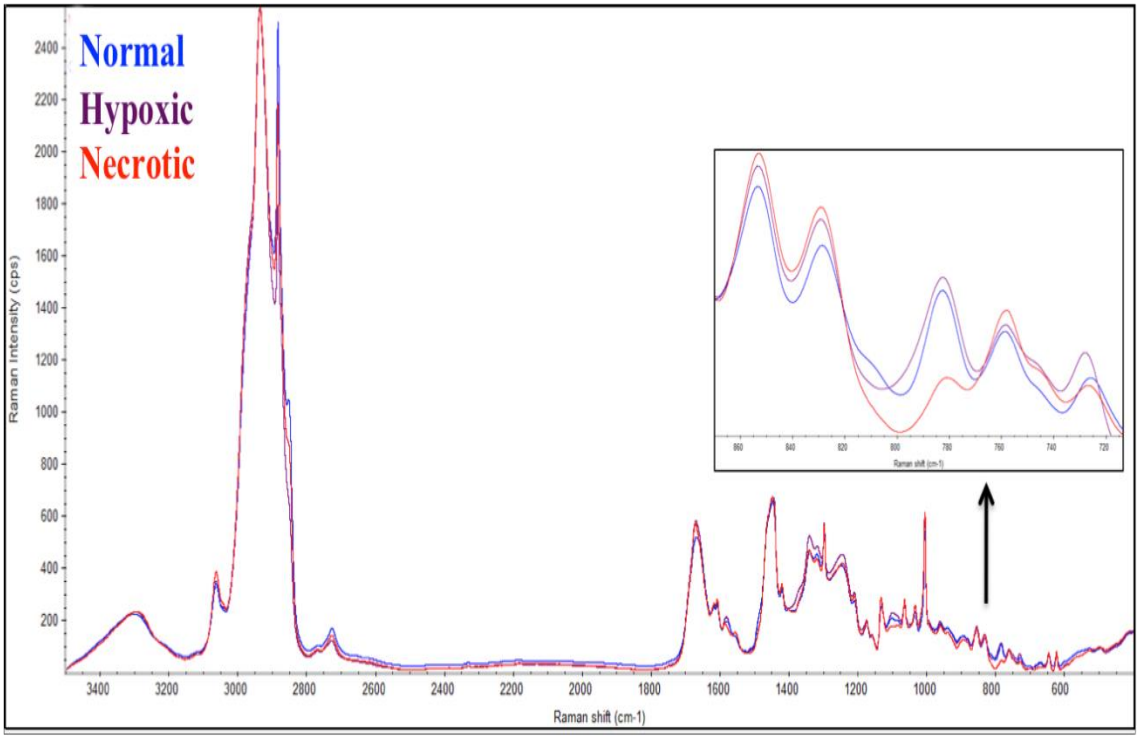

Figure 35: : Mean Raman spectra of normal proliferating (blue), hypoxic (dark red) and necrotic (red) regions with nucleic acid region peak changes.

5.3 Multivariate analysis

The statistical analysis of this section was done in three stages. Firstly, an unsupervised approach, PCA was applied on data matrix of three regions to a lower-dimensional space bridged by the loading vectors. Loading vectors were used to analyze three regions in terms of lipids, proteins and nucleic acids. Secondly, a partial supervised approach, Cluster Analysis was used to study the partition between the three regions in high-wave number region, amide I region and nucleic acid region. Thirdly, a supervised approach, Linear Discriminant Analysis was performed to assess the sensitivity and specificity of spectral features to predict three regions.

PCA was performed on high-wavenumber region (3200-2600 cm^{-1}), fingerprint region (1800-600 cm^{-1}), amide I region (1800-1510 cm^{-1}), amide II region (1510-1390 cm^{-1}), amide III region (1390-1140 cm^{-1}) and nucleic acid region (980-600 cm^{-1}).

5.3.1 High-wavenumber region

PCA was performed on the lipid region between 3200-2600 cm^{-1}. In the first PCA plot of lipid region, PC-2 separated normal proliferating region from necrotic region along positive and negative variables of PC-2 axis, whereas, PC-3 has distributed all three regions along axis line with positive and negative variables without any separation. In the second loading plot, PC-2 has pretty much separated normal proliferating region and necrotic region from hypoxic region, whereas, PC-4 has well separated normal proliferating and necrotic regions from hypoxic region. A three dimensional PCA plot of lipoid region has shown pretty much cluster separation among three regions using PC-2, PC-3 and PC-4 (Figures 36-38).

Loading plot of high-wavenumber region describes the following biochemical snapshot of three regions: Biochemical variation among three regions was increased from PC2 to PC4. Firstly, PC-2 loading plot (blue) has shown majority of Raman spectra of normal proliferating, hypoxic and necrotic regions in high-wavenumber region has shown high lipid contents compared to protein content. PC-4 loading plot (green) has shown that necrotic region and normal proliferating regions have high protein contents and low in lipid contents compared to hypoxic region. PC-3 loading plot (red) hasn't provided significance changes in biochemical differences among three regions. Based on PC-2 and PC-4 loading plot analysis, it is evident that lipid variations between normal and necrotic regions are similar compared to hypoxic region of T-47D spheroids (Figure 39).

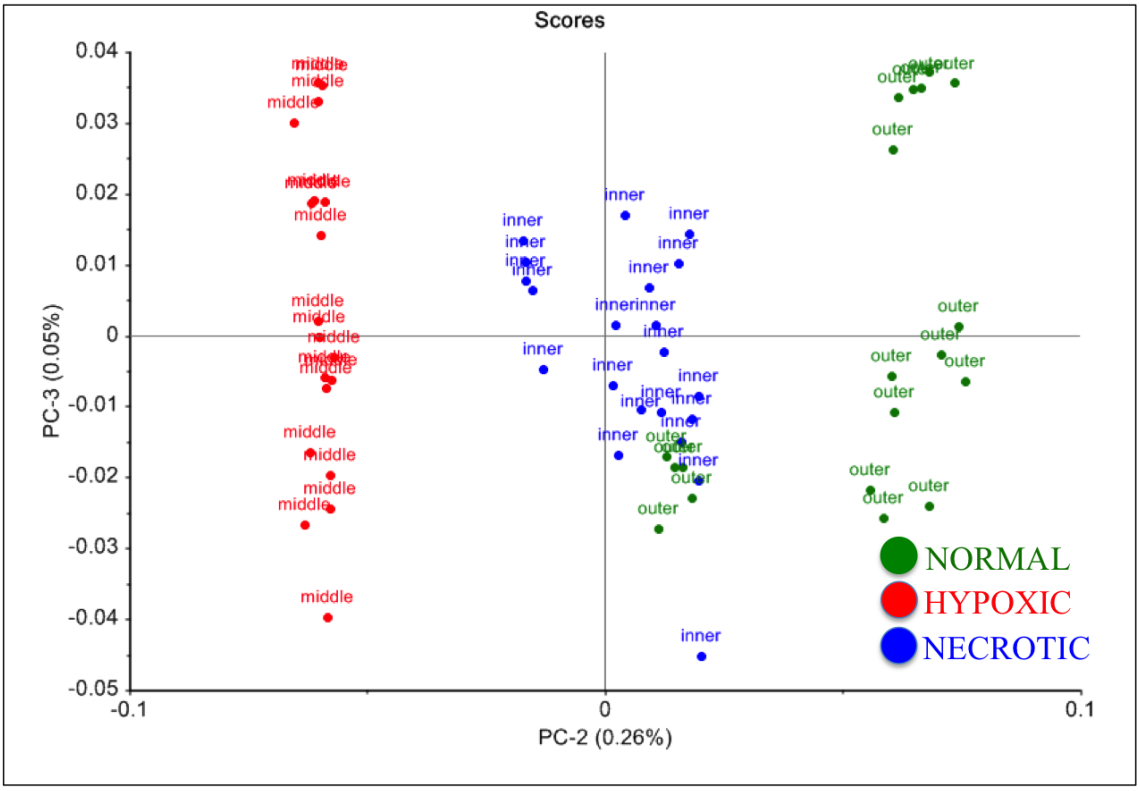

Figure 36: A two-dimensional principal component analysis (PCA) score plot of lipid region for normal proliferating (green), hypoxic (red) and necrotic (blue) regions based on second and third principal components (PC).

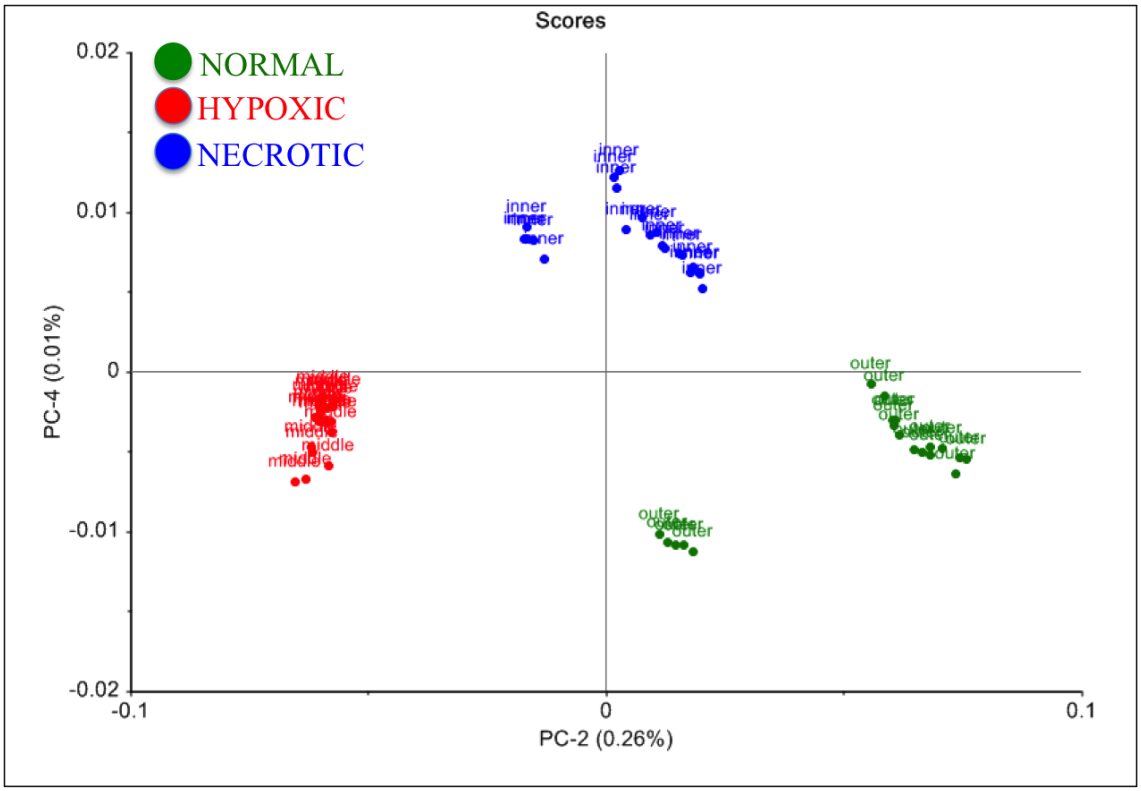

Figure 37: A two-dimensional principal component analysis (PCA) score plot of lipid region for normal proliferating (green), hypoxic (red) and necrotic (blue) regions based on second and fourth principal components (PC).

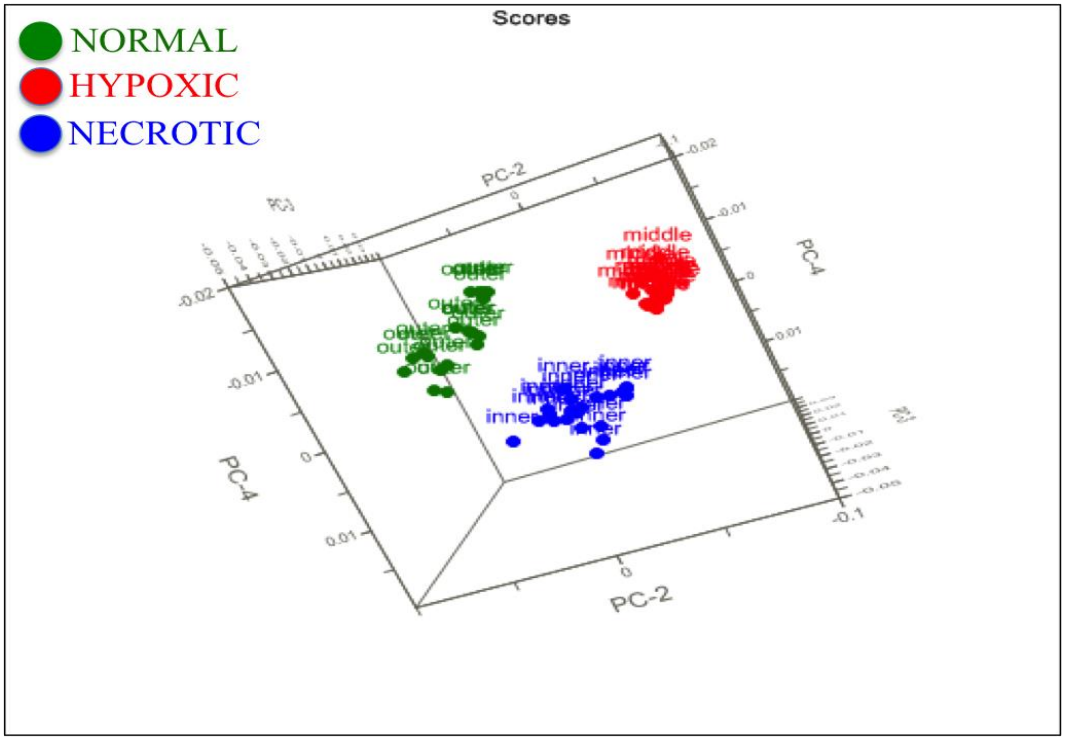

Figure 38: A three-dimensional principal component analysis (PCA) plot of lipid region for normal proliferating (blue), hypoxic (red) and necrotic (green) regions based on second, third and fourth principal components (PC).

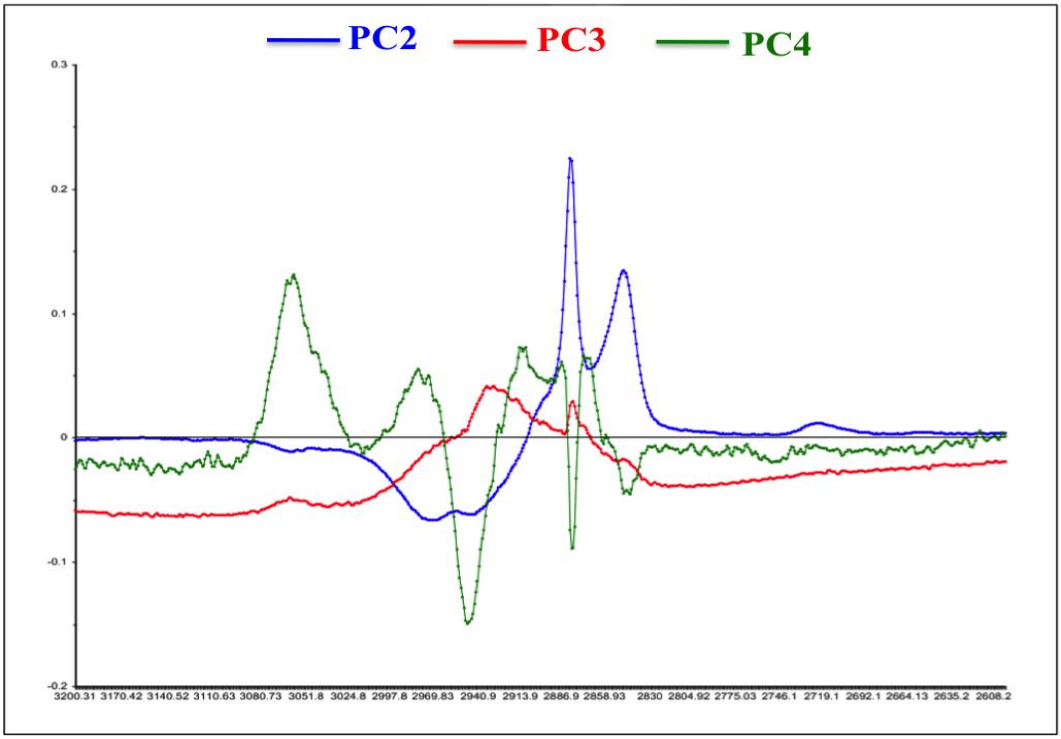

Figure 39: Loading plots of lipid region representing principal components PC-2 (blue), PC-3 (red) and PC-4 (green).

5.3.2 Finger print region

PCA was performed on Fingerprint region between 1800-1600 cm^{-1}. In the first PCA plot of the finger Print region, PC-3 has differentiated normal proliferating region, hypoxic region and necrotic regions partially. PC-5 has pretty much separated normal proliferating region from necrotic region. In the second PCA plot, PC-5 has pretty much separated normal proliferating region from necrotic region, whereas, PC-4 has partially separated normal and necrotic regions from hypoxic region. A three dimensional PCA plot of fingerprint region has shown clear cluster separation among three regions using PC-3, PC-4 and PC-5. PCA loading plot analysis of fingerprint has shown biochemical differences among three regions in terms of lipids, proteins, nucleic acids and carbohydrates. PCA loading plot of fingerprint region has described biochemical snapshot of three regions (Figures 40-42).

Firstly, PC-3 loading plot (blue) has suggested that necrotic region has shown high lipid content (1606 cm^{-1} – C=C bending of lipids, 1438 cm^{-1} – lipids, 1294 cm^{-1} – methylene twisting of lipids) compared to normal proliferating region. It also described that some of hypoxic spectra have shown high lipid content compared to normal proliferation region. PC4 loading plot (red) has shown similar results of PC-3 loading plot in terms of lipid content of necrotic region. PC-3 also indicated that necrotic region has shown less protein content (1668 cm^{-1} – C=O of amide I, 1237 cm^{-1} – amide III and 1004 cm^{-1} - phenylalanine) as well as nucleic acid content (1374 cm^{-1} – adenine, thymine and guanine bases of DNA, 1340 cm^{-1} – nucleic acid modes, 782 cm^{-1} - DNA) compared to normal proliferating region. PC-4 loading plot has shown mixed results in hypoxic region in terms of protein and nucleic acid contents compared to normal and necrotic regions. PC-5 loading plot (green) has shown that normal proliferating region has high in protein content (1669 cm-1 – amide I, 1237 & 1204 cm-1 – amide III, 1131 & 1004 cm-1 – phenylalanine and 756 cm-1 – tryptophan) and nucleic acid content (832

cm-1 - O-P-O stretches of nucleotides) and low in lipid content (Movasaghi et al., 2007a) (Figure 43).

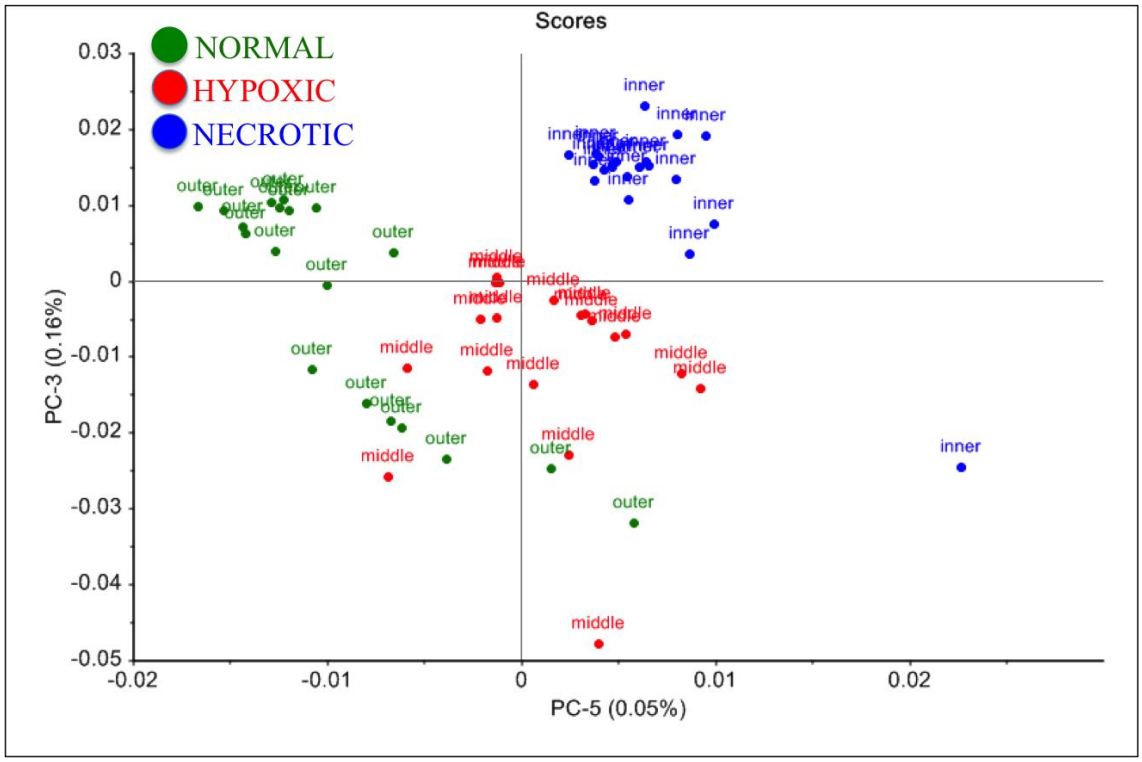

Figure 40: A two-dimensional principal component analysis (PCA) score plot of Fingerprint region for normal proliferating (green), hypoxic (red) and necrotic (blue) regions based on third and fifth principal components (PC).

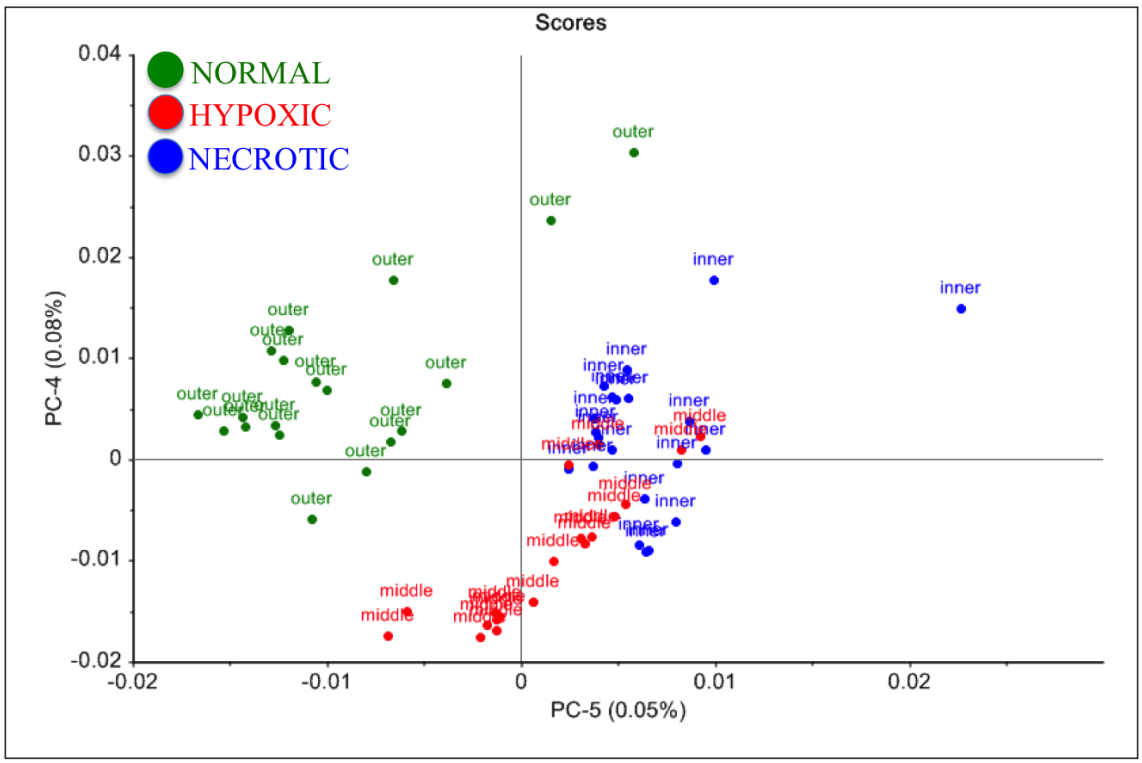

Figure 41: A two-dimensional principal component analysis (PCA) score plot of Fingerprint region for normal proliferating (green), hypoxic (red) and necrotic (blue) regions based on fourth and fifth principal components (PC).

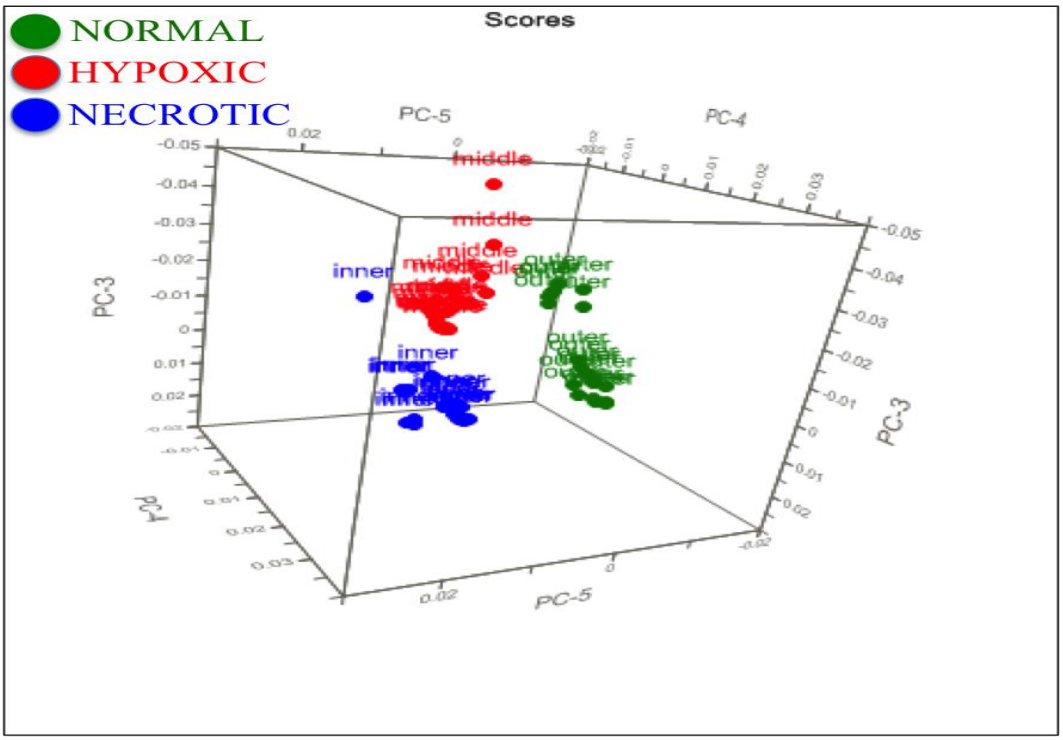

Figure 42: A three-dimensional principal component analysis (PCA) plot of fingerprint region for normal proliferating (normal), hypoxic (red) and necrotic (blue) regions based on third, fourth and fifth principal components (PC).

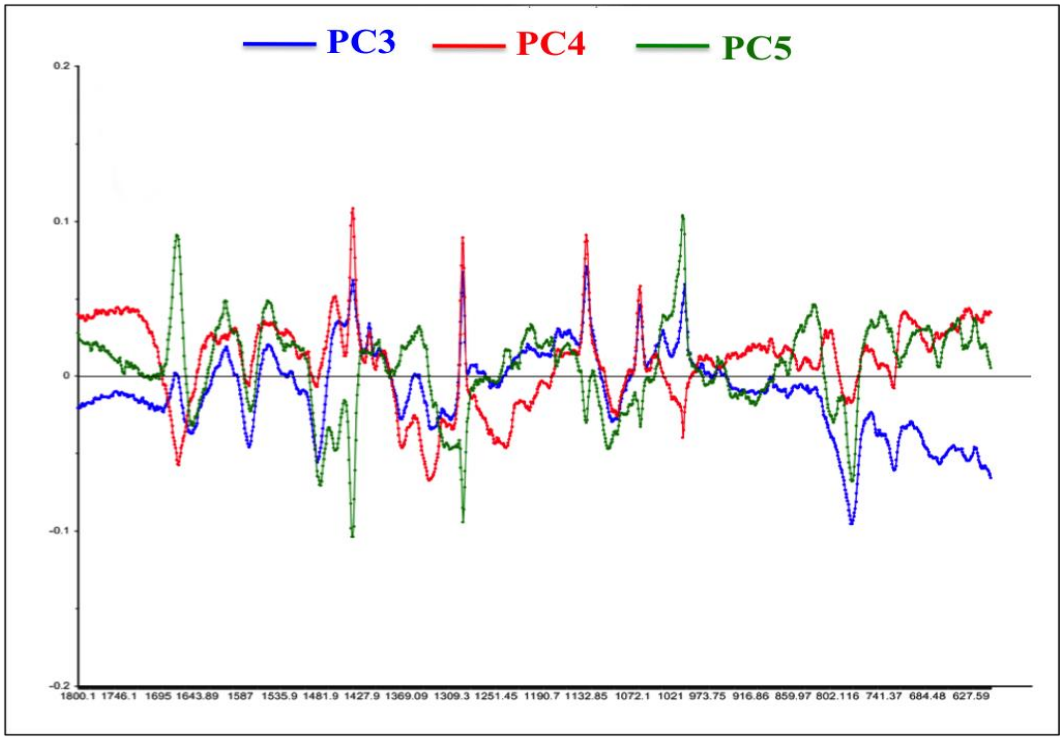

Figure 43: Loading plots of fingerprint region representing principal components PC3 (blue), PC4 (red) and PC5 (green).

5.3.3 Amide I

PCA was performed on Amide region between 1800-1510 cm^{-1}. In the first PCA plot PC-4 has clearly separated normal proliferation region and hypoxic region from necrotic region whereas PC-5 separated hypoxic and necrotic regions from normal proliferating region. In the second PCA plot, PC-5 hasn't contributed much towards separation of normal, hypoxic and necrotic regions, whereas, PC-3 has partially separated normal proliferating region necrotic region. Hypoxic cluster might indicate that amide I region has differed from normal and necrotic region. A three dimensional PCA plot of amide I region has shown cluster separation among three regions using the principal components such as PC-3, PC-4 and PC-5 (Figure 47).

PCA loading plot analysis of amide I region has described following biochemical differences among three regions. Firstly, PC-3 loading plot (blue) has shown low in amide I content (anti-parallel β sheets of amide I at 1670 cm^{-1}) in necrotic and hypoxic

region compared to normal proliferating region. It also indicated high tryptophan content (1560 cm^{-1}) in normal proliferating region. PC-4 loading plot (red) has shown that hypoxic region has high protein contents (α- helix of amide I at 1645 cm^{-1}) and tryptophan contents (1560 cm^{-1}) compared to normal proliferating region. It also described that normal proliferating region has low tryptophan contents (1550 cm^{-1}) compared to hypoxic and necrotic region. PC-5 loading plot (green) has shown that normal proliferating region has high protein contents (amide I at 1670 cm^{-1}) and tryptophan contents (1550 cm^{-1}) compared to hypoxic and necrotic regions (Talari et al., 2015k) (Figure 44-46).

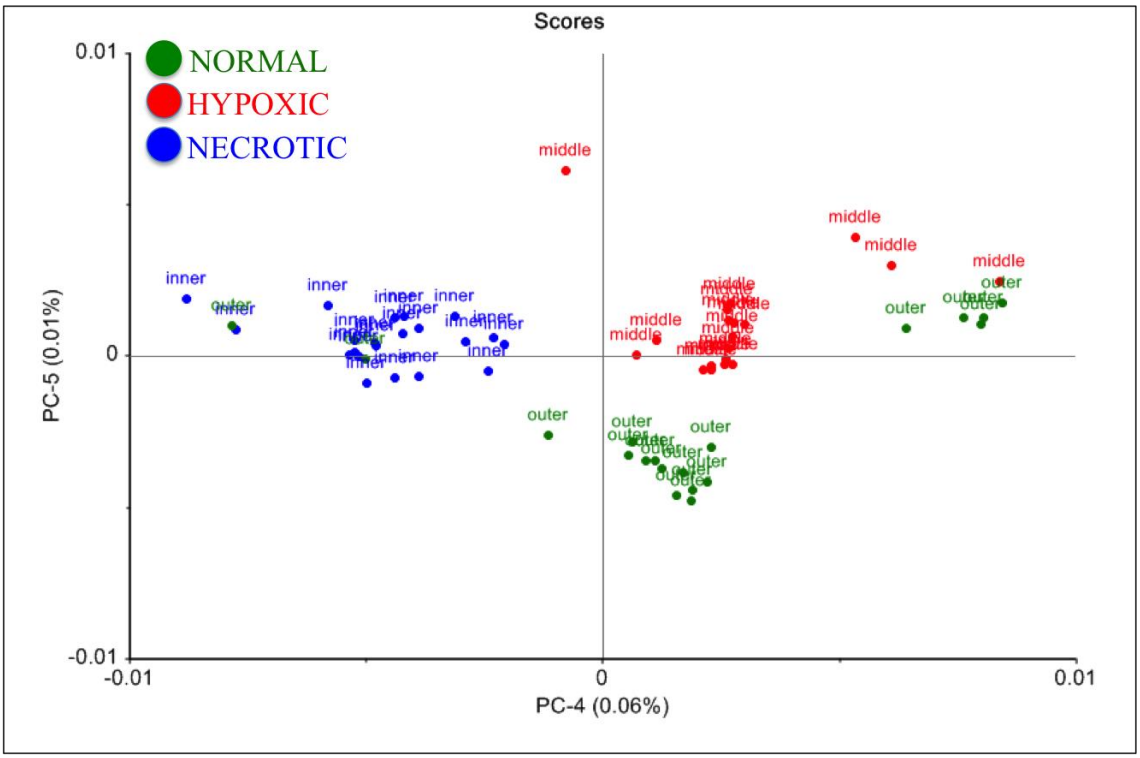

Figure 44: A two-dimensional principal component analysis (PCA) score plot of amide I region for normal proliferating (blue), hypoxic (red) and necrotic (green) regions based on fourth and fifth principal components (PC).

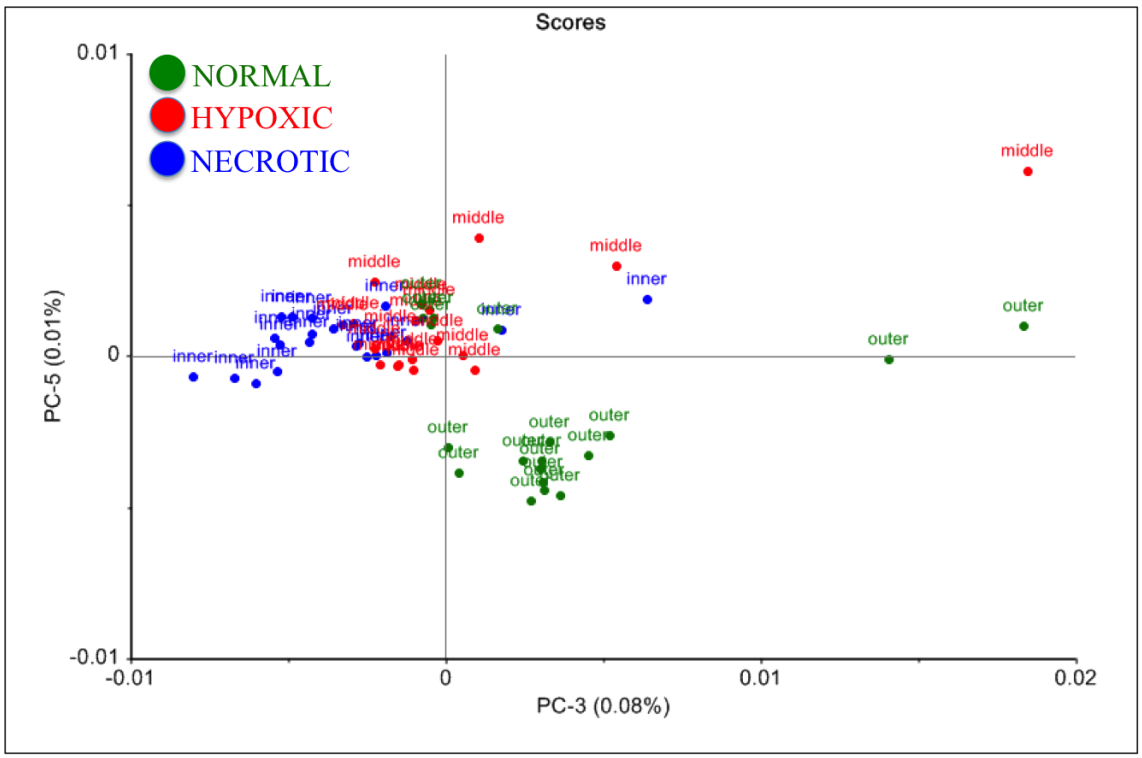

Figure 45: A two-dimensional principal component analysis (PCA) score plot of amide I region for normal proliferating (blue), hypoxic (red) and necrotic (green) regions based on third and fifth principal components (PC).

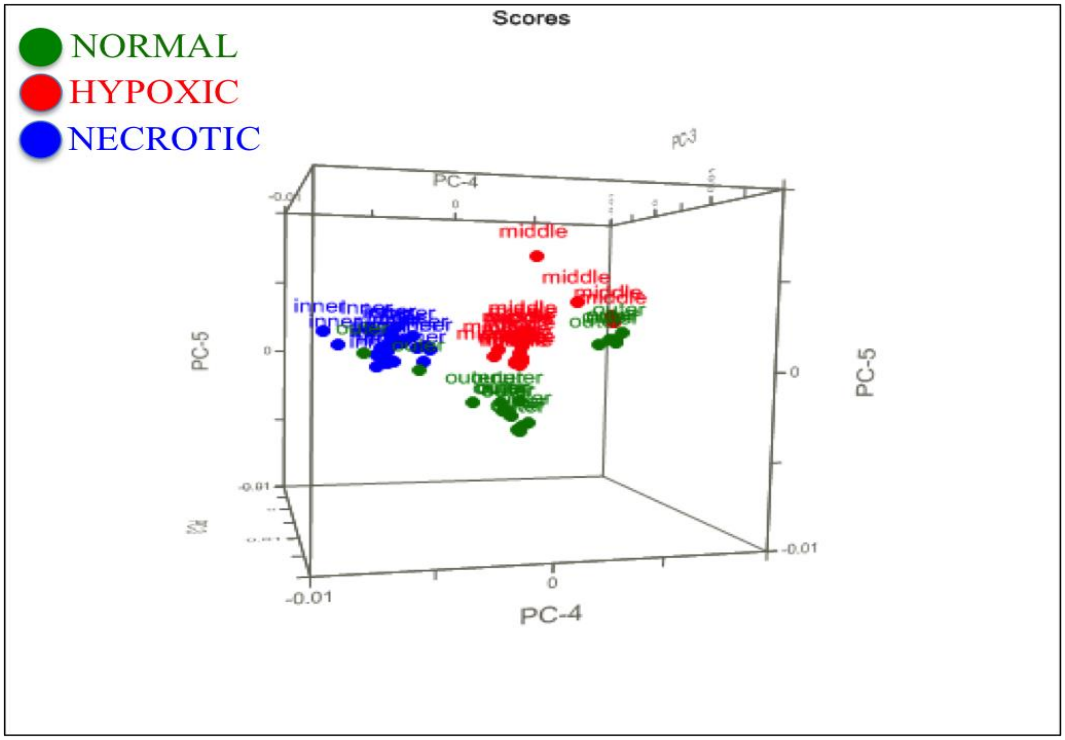

Figure 46: A three-dimensional principal component analysis (PCA) plot of amide I region for normal proliferating (blue), hypoxic (red) and necrotic (green) regions based on third, fourth and fifth principal components (PC).

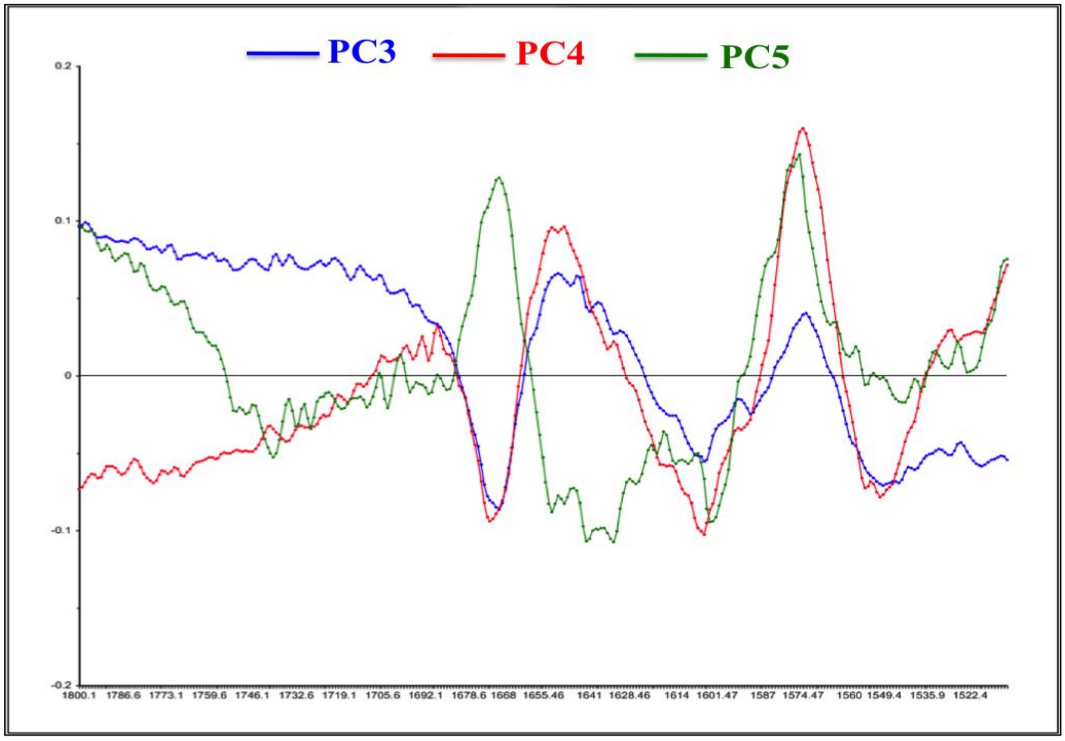

Figure 47: Loading plots of amide I region representing principal components PC3 (blue), PC4 (red) and PC5 (green).

5.3.4 Amide II

PCA was applied on Amide II region between 1510-1390 cm^{-1}. In the first PCA plot, PC-3 has partially separated necrotic region from normal proliferating region and hypoxic region from necrotic region, whereas, PC4 has pretty much separated normal proliferating region from hypoxic region. In the second PCA plot, PC-4 has pretty much separated normal proliferating region from hypoxic region. Necrotic region shares similar amide II content with normal and hypoxic regions. PC-2 hasn't contributed to any separation among three regions. A three dimensional PCA plot of amide II region has shown partial cluster separation among three regions using the principal components PC-2, PC-3 and PC-4 (Figures 48-51).

PCA loading plot analysis of amide II region has described the following biochemical differences among three regions. Firstly, PC-2 loading plot (blue) has shown low in lipid content (lipids at 1438 cm^{-1}) in hypoxic region compared to normal proliferating

region and necrotic regions and does not contribute much towards separation.PC-3 loading plot (red) has shown high in protein content (1484 cm^{-1} – amide II) and deformation of lipids (1440 cm^{-1} – CH, CH_2 and CH_3 deformation of lipids, cholesterol esters and fatty acids) and stretching of lipids (1416 cm^{-1} – C=C stretching of lipids and 1410 cm^{-1} – C-H deformation of lipids) all contribute to separation. PC-4 loading plot (green) has shown that normal proliferating region has high lipid content (lipids at 1438 cm^{-1}) while low in CH vibrations of proteins and lipids (1448 cm^{-1}) N-H plane deformations of proteins (1423 cm^{-1}) compared to hypoxic and necrotic regions (Movasaghi et al., 2007d).

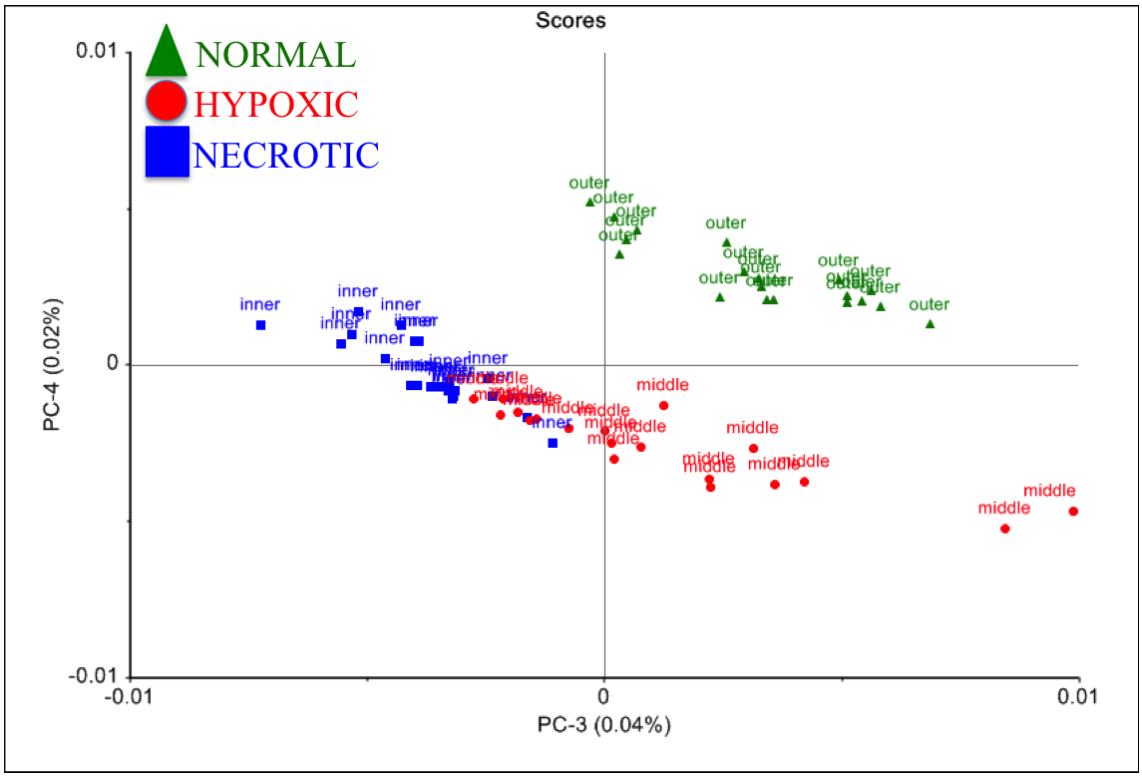

Figure 48: A two-dimensional principal component analysis (PCA) score plot of amide II region for normal proliferating (green), hypoxic (red) and necrotic (blue) regions based on third and fourth principal components (PC).

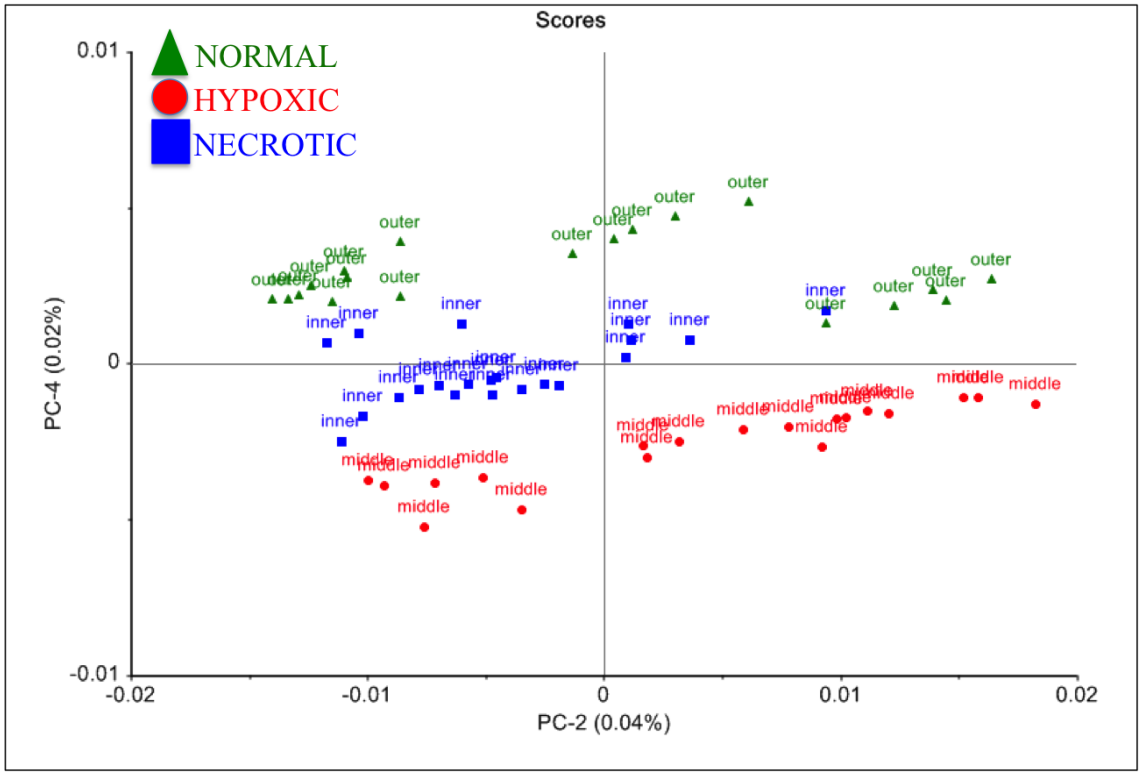

Figure 49: A two-dimensional principal component analysis (PCA) score plot of amide II region for normal proliferating (blue), hypoxic (red) and necrotic (green) regions based on third and fifth principal components (PC).

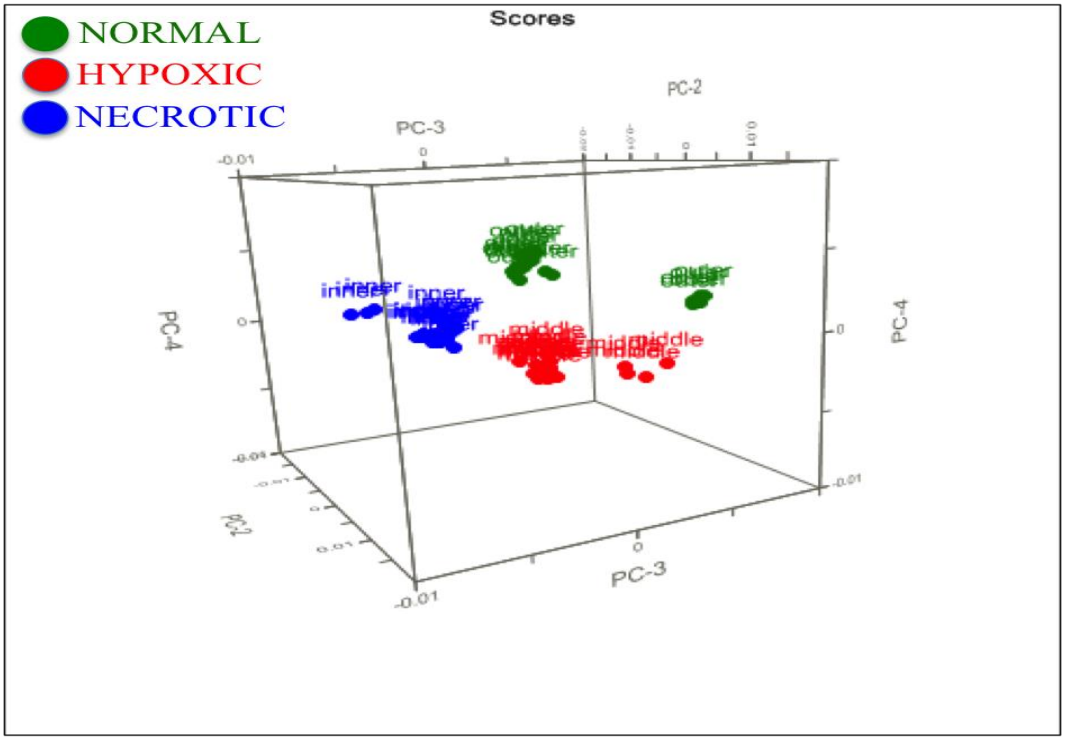

Figure 50: A three-dimensional principal component analysis (PCA) plot of amide II region for normal proliferating (blue), hypoxic (red) and necrotic (green) regions based on third, fourth and fifth principal components (PC).

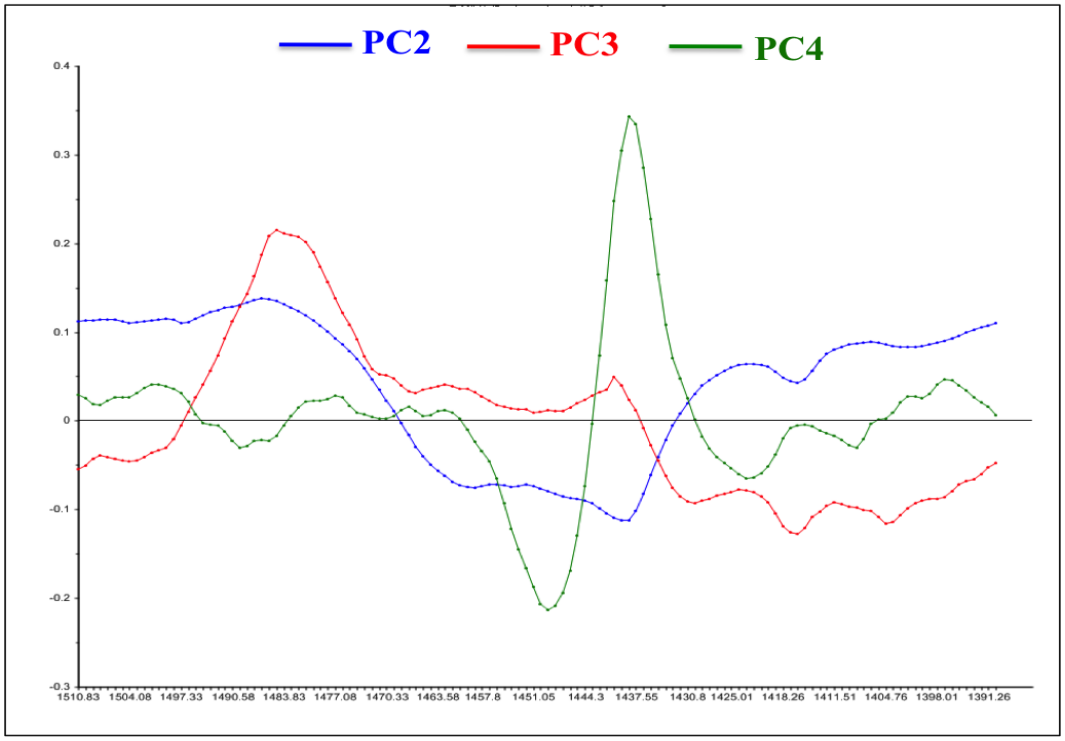

Figure 51: Loading plots of amide II region representing principal components PC2 (blue), PC3 (red) and PC4 (green).

5.3.5 Amide III

PCA was performed on Amide III region between 1390-1140 cm^{-1}. In the first PCA plot, PC-3 positively separated hypoxic region from normal region, whereas, PC-5 has contributed to the separation of normal proliferating region from hypoxic region. In the second PCA plot, PC-4 has separated normal proliferating region from necrotic region. PC-5 has separated hypoxic region positively and normal proliferating region negatively along the axis. A three dimensional PCA plot of amide III region has shown partial cluster separation in normal proliferating region and good separation between necrotic and hypoxic regions using the principal components PC-3, PC-4 and PC-5 (Figures 52-54).

PCA loading plot analysis of amide III region has described following biochemical differences among three regions. Firstly, PC3 loading plot (blue) has shown high methylene twisting in (1294 cm^{-1}) in necrotic region compared to hypoxic region. It also described that PC-3 has shown less nucleic acid content in necrotic region (1374 cm^{-1}) compared to normal proliferating region and necrotic region. PC-4 loading plot (red)

has shown that normal proliferating region has shown lower protein content (1236 and 1204 cm^{-1} – amide III) compared to hypoxic and necrotic regions. PC-5 loading plot (green) has shown that necrotic and hypoxic region have been separated using CH_3 stretching (1371 cm^{-1}), methylene twisting (1294 cm^{-1}), tryptophan & phenylalanine (1209 cm^{-1}) and cytosine and guanine (1177 cm^{-1}) regions compared to normal proliferating regions. PC-5 has also described that normal proliferating region and few spectra of necrotic region have shown less CH3, CH2 twisting and wagging of lipids, collagen content, adenine and thymine of DNA (1257 cm^{-1}) and amide III (1203 cm^{-1}) compared to hypoxic region and remaining spectra of necrotic region(Talari et al., 2015k) (Figure 54).

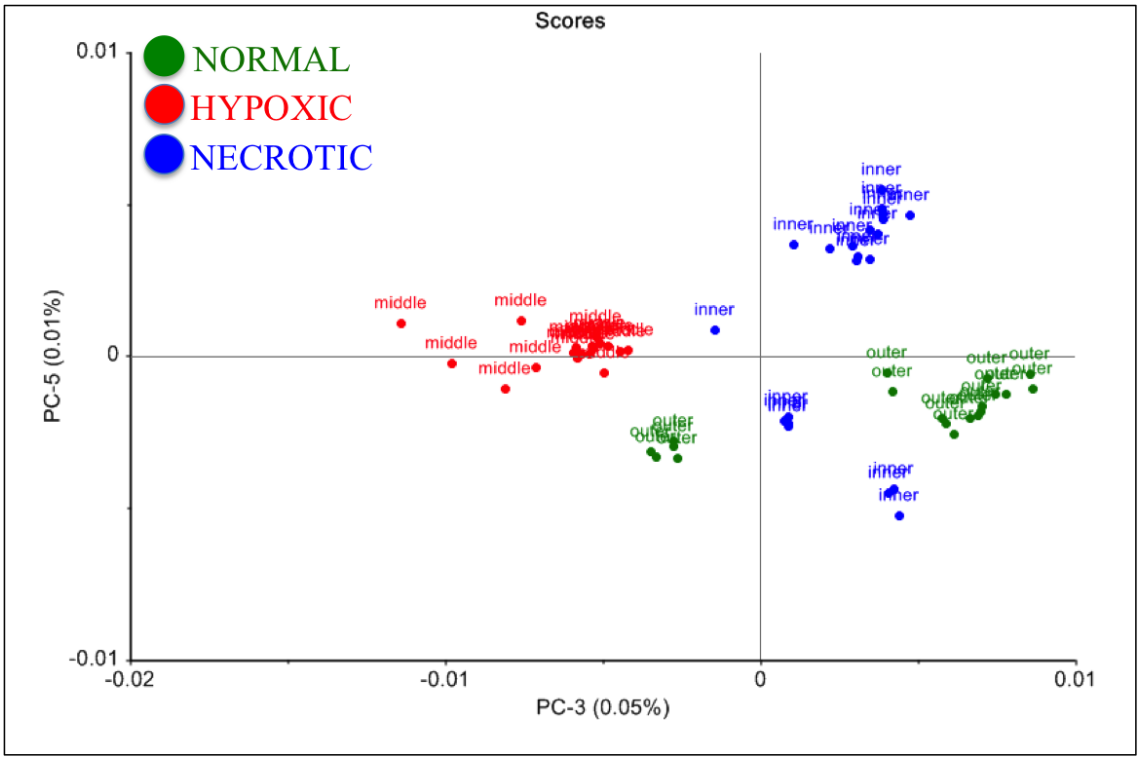

Figure 52: A two-dimensional principal component analysis (PCA) score plot of amide III region for normal proliferating (green), hypoxic (red) and necrotic (blue) regions based on third and fifth principal components (PC).

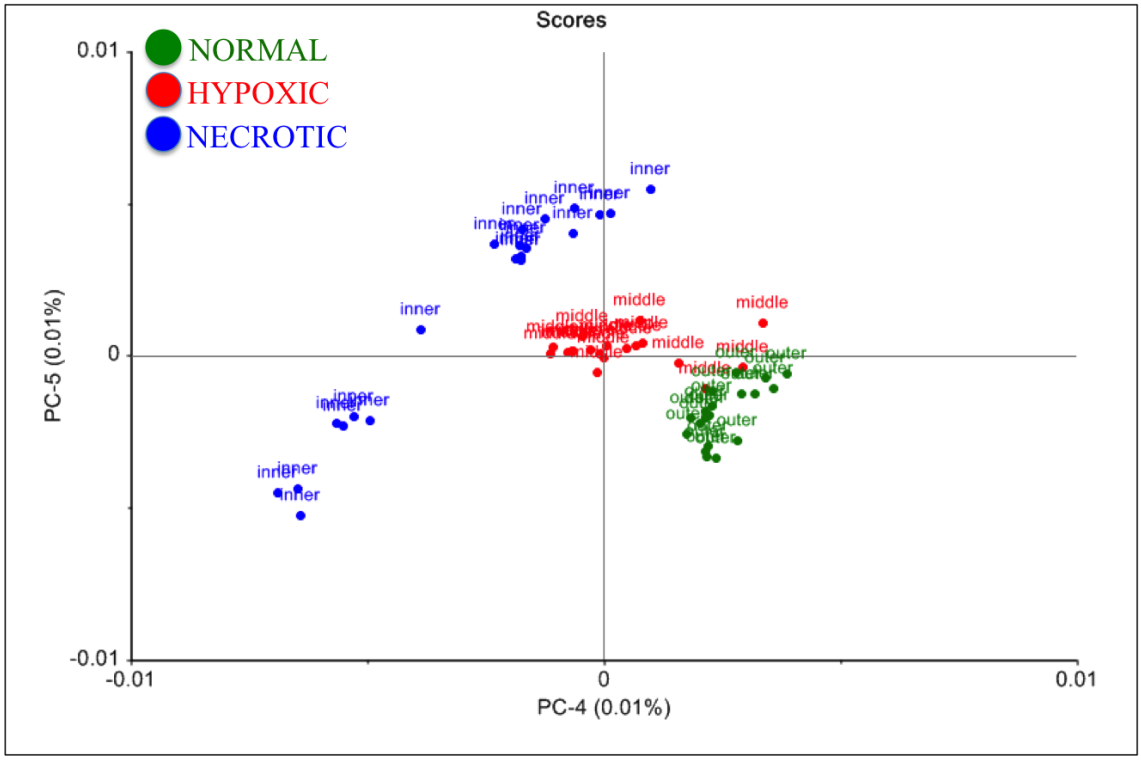

Figure 53: A two-dimensional principal component analysis (PCA) plot of amide III region for normal proliferating (green), hypoxic (red) and necrotic (blue) regions based on third, fourth and fifth principal components (PC).

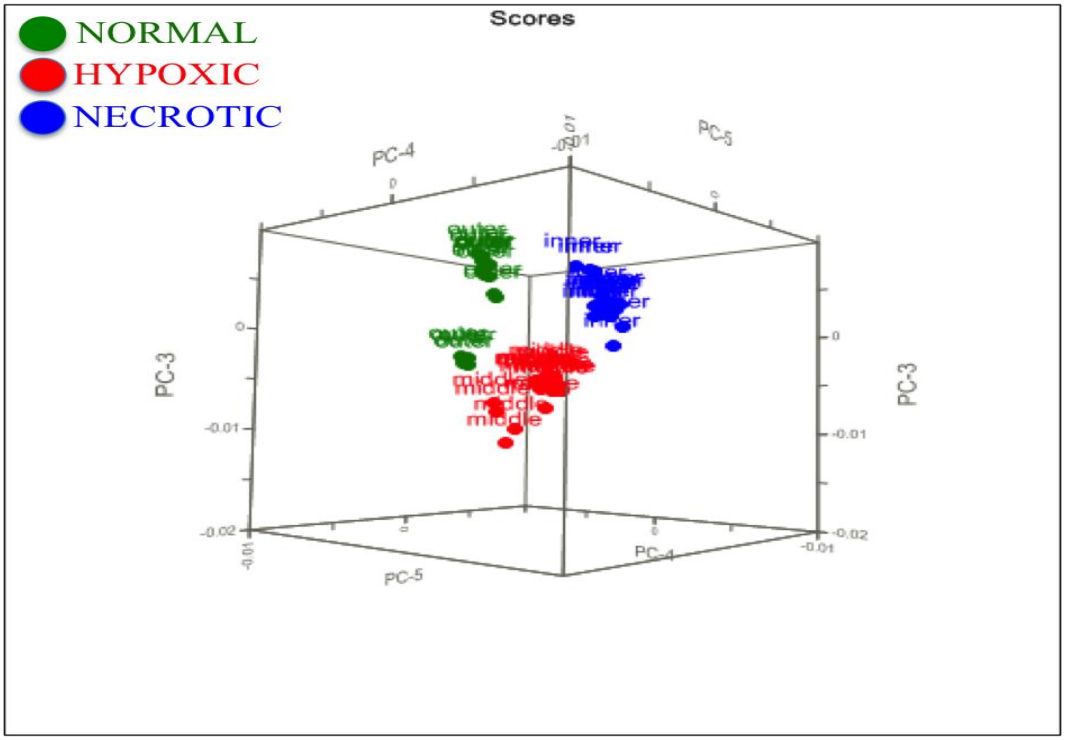

Figure 54: A three-dimensional principal component analysis (PCA) plot of amide III region for normal proliferating (green), hypoxic (red) and necrotic (blue) regions based on third, fourth and fifth principal components (PC).

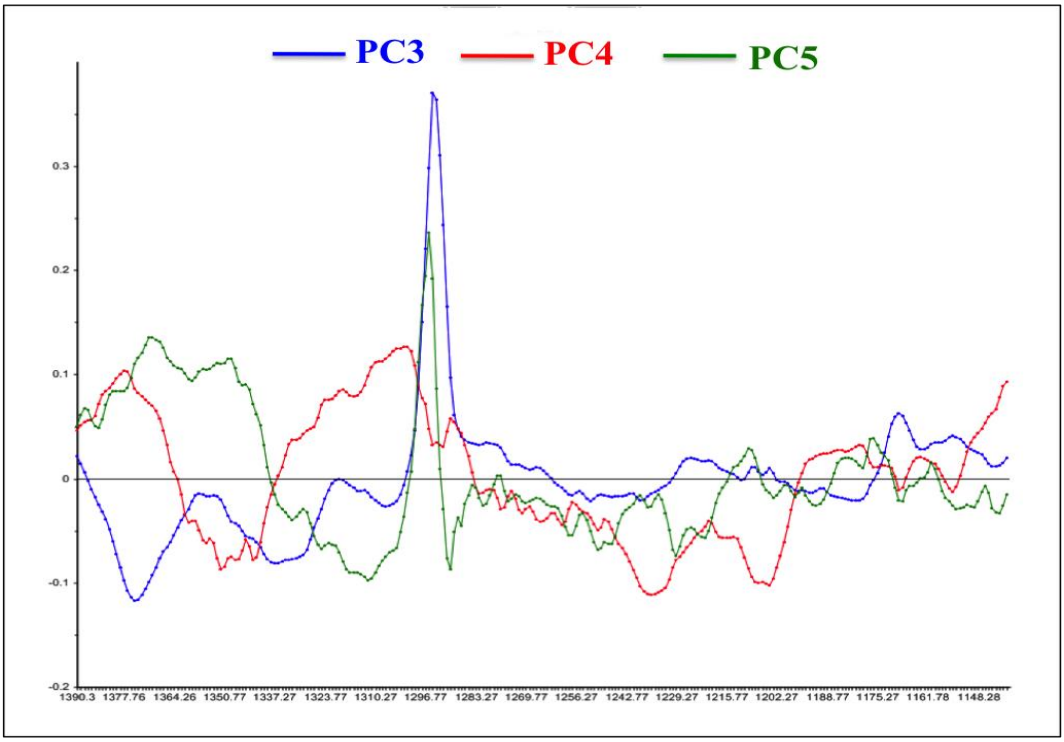

Figure 55: Loading plots of amide III region representing principal components PC3 (blue), PC4 (red) and PC5 (green).

5.3.6 Nucleic Acid Region

PCA was performed on Nucleic Acid region between 980-600 cm^{-1}. In the first PCA plot PC-3 has partially separated necrotic region from hypoxic and normal proliferating regions, whereas, PC-5 has separated normal proliferating region from hypoxic and necrotic regions. In the second PCA plot, PC-5 has partially separated normal proliferating region from hypoxic and necrotic regions, whereas, PC-4 has separated normal proliferating and hypoxic region from necrotic region with lesser extent. A three dimensional PCA plot of nucleic acid region has shown partial cluster separation in normal proliferating region and good separation between necrotic and hypoxic regions using principal components such as PC-3, PC-4 and PC-5 (Figures 56-58).

PCA loading plot analysis of nucleic acid region has described following biochemical differences among three regions of T-47D spheroids. The variation, which is contributed towards separation, is increased from PC-3 to PC-5. Firstly, PC-3 loading plot (blue) has shown that majority of hypoxic region and minority of normal

proliferating region have high purine and pyrimidine bases of DNA and RNA (782 and 728 cm^{-1}), whereas, complete necrotic region and majority of normal proliferating region have shown less lipids (cholesterol at 958 cm^{-1}), proline, hydroxyproline and tyrosine (852 and 830 cm^{-1}), tryptophan (756 cm^{-1}) and tyrosine (643 cm^{-1}). PC4 loading plot (red) has shown that majority of normal proliferating region and hypoxic region have shown high protein content (proline and hydroxyproline – 853, 933 cm^{-1}), tyrosine (643 cm^{-1}) and nucleic acid content (adenine and DNA bases at 782 & 729 cm^{-1}), whereas, necrotic region has shown less methionine aminioacid content (704 and 739 cm^{-1}). PC5 loading plot (green) has shown that necrotic and hypoxic region have shown that high phosphate stitching and cholesterol (959 cm-1), tyrosine and proline (830 cm^{-1}) and adenine (729 cm^{-1}) , whereas, normal proliferating region has shown less glucose (913 cm^{-1}) and phosphodiester groups (810 cm^{-1})(Movasaghi et al., 2007d) (Figure 59).

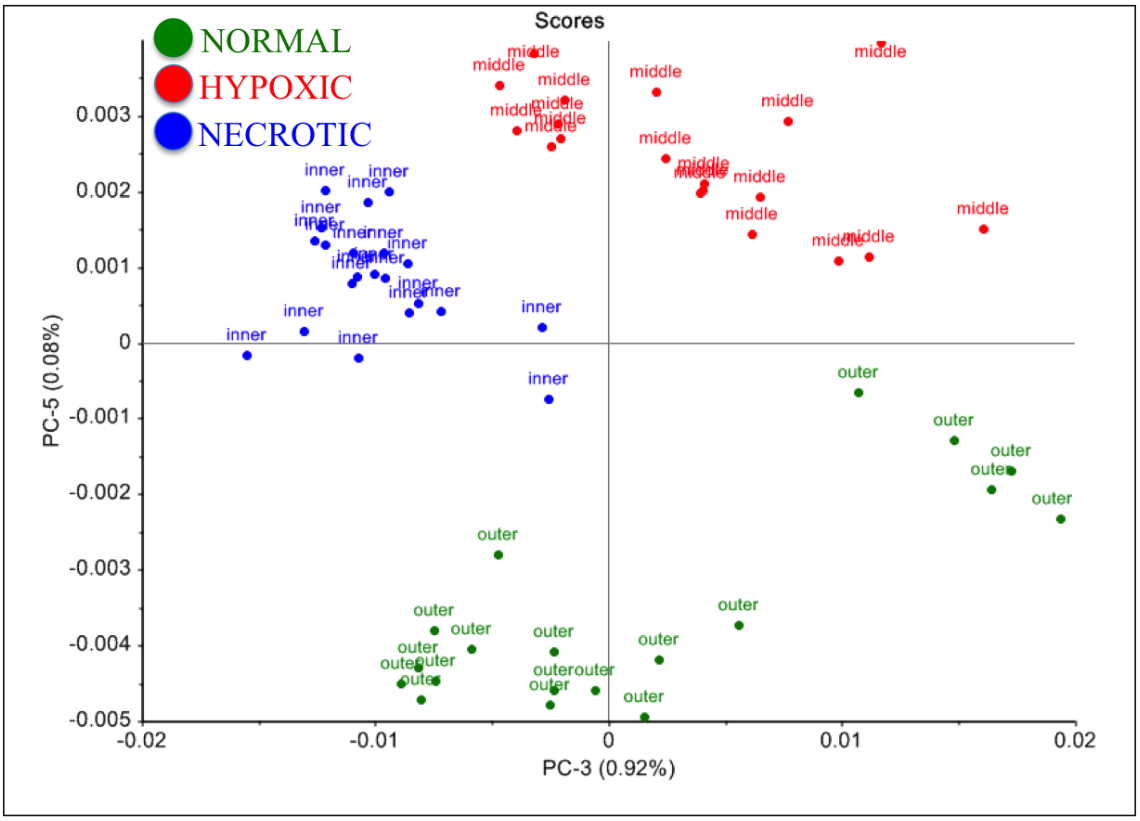

Figure 56: A two-dimensional principal component analysis (PCA) score plot of nucleic acid region for normal proliferating (green), hypoxic (red) and necrotic (blue) regions based on third and fifth principal components (PC).

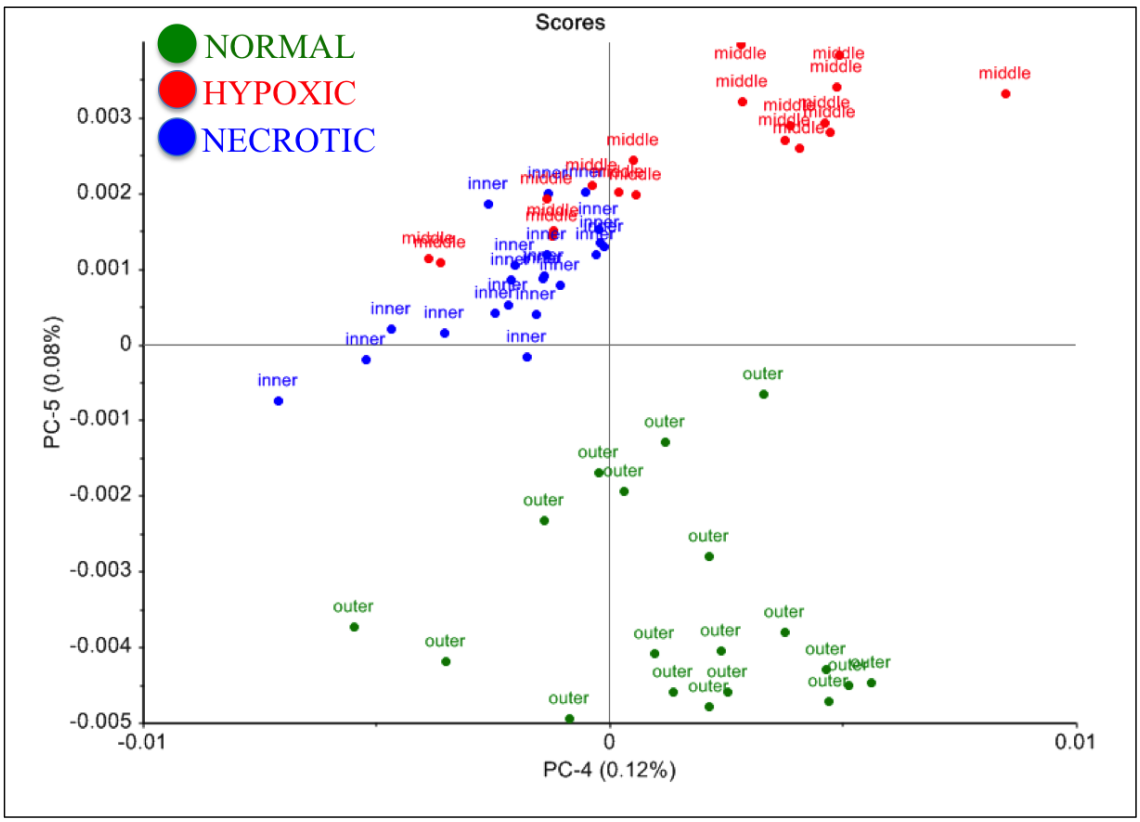

Figure 57: A two-dimensional principal component analysis (PCA) score plot of nucleic acid region for normal proliferating (green), hypoxic (red) and necrotic (blue) regions based on fourth and fifth principal components (PC).

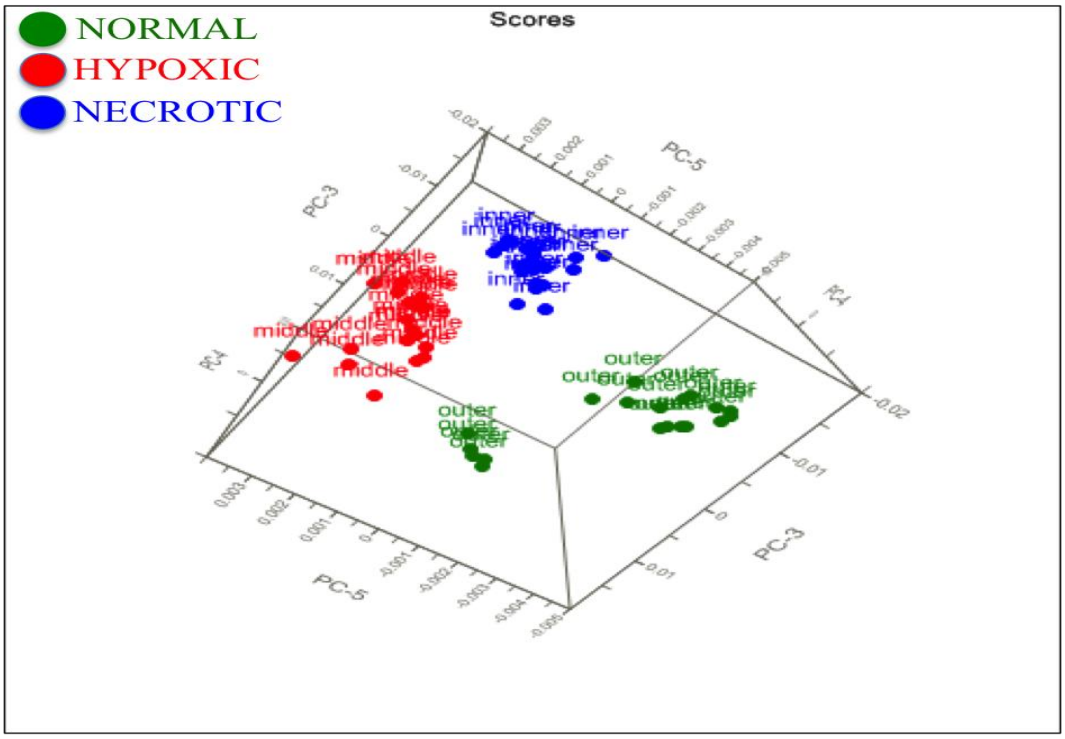

Figure 58: A three-dimensional principal component analysis (PCA) plot of nucleic acid region for normal proliferating (green), hypoxic (red) and necrotic (blue) regions based on third, fourth and fifth principal components (PC).

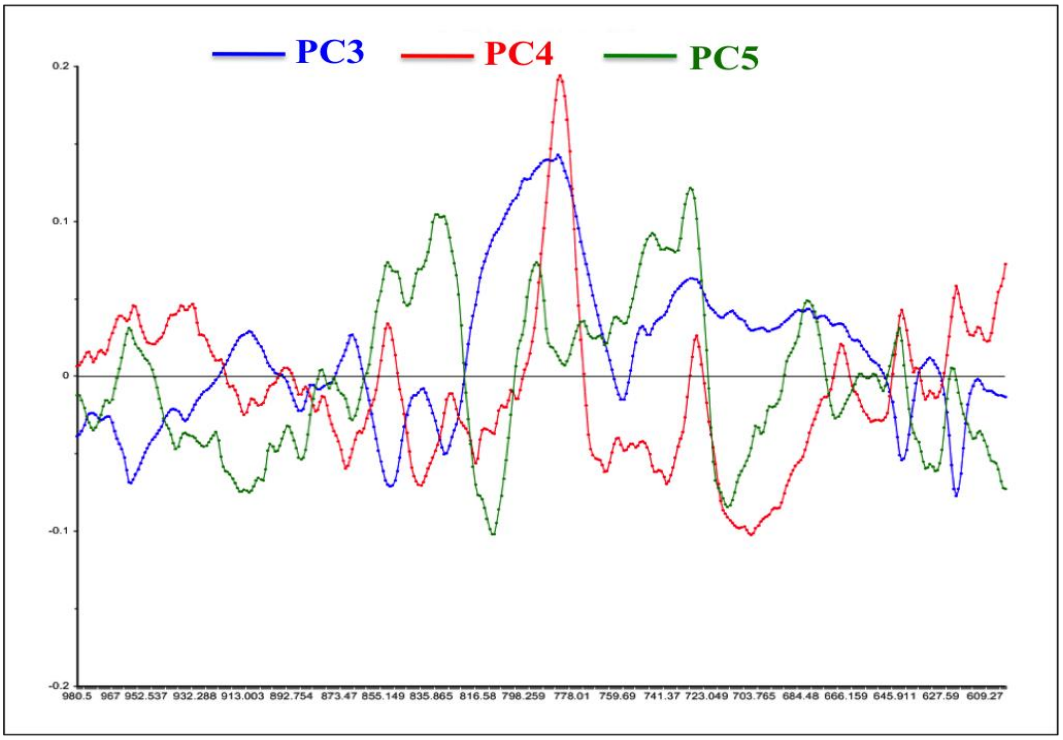

Figure 59: Loading plots of nucleic acid region representing principal components PC3 (blue), PC4 (red) and PC5 (green).

5.3.7 Cluster Analysis

High-wavenumber region (3200-2600 cm^{-1})

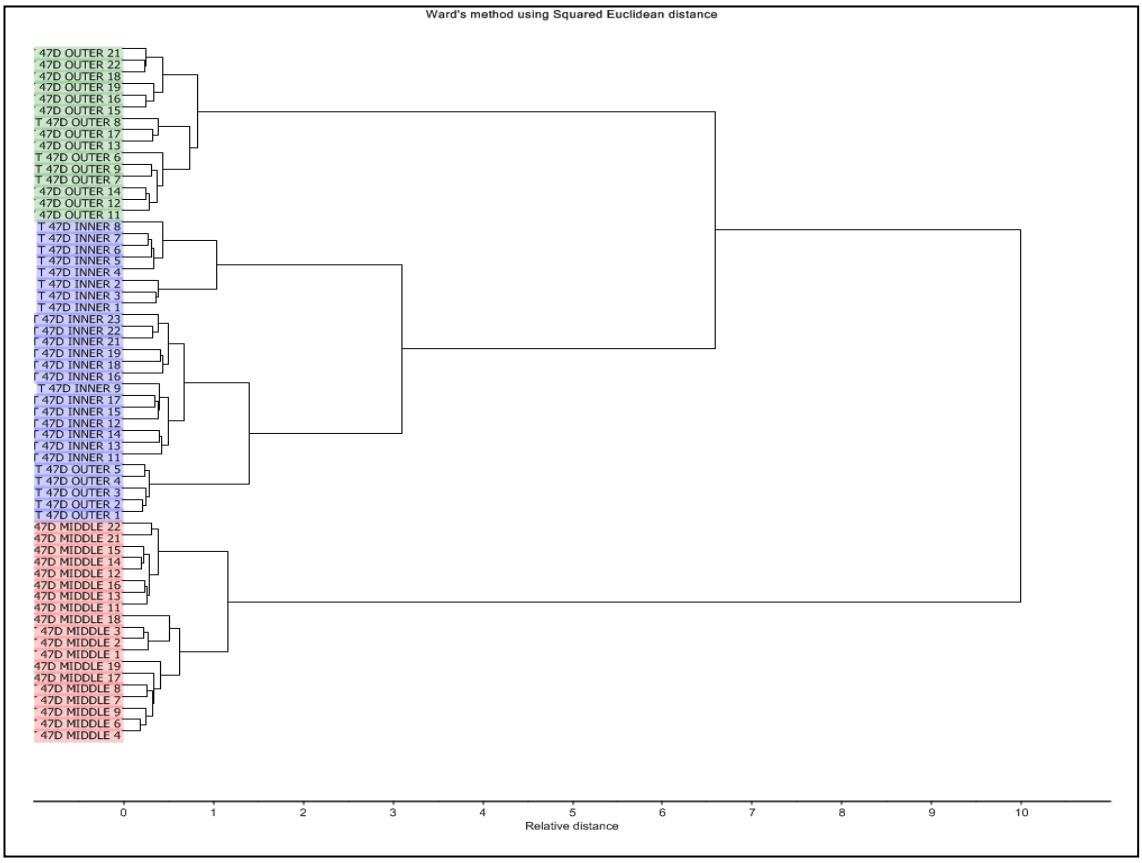

Figure 60: Dendrogram of Cluster Analysis results for normal proliferating region, hypoxic region and necrotic region. CA was performed over the High-wave number region using Wards method's squared Euclidean distance.

Figure 60 has shown the dendrogram of outer (normal proliferating), middle (hypoxic) and inner (necrotic) regions over the high-wavenumber range (3200-2600 cm^{-1}). Three main clusters were formed. Some of the normal proliferating region spectra were mixed with necrotic region. Each region has formed pretty much good cluster and distance measurements has suggested that biochemical snapshot of high-wavenumber region has shown more similarity between normal proliferating and necrotic regions compared to hypoxic region.

Amide I region (1800-1510 cm^{-1})

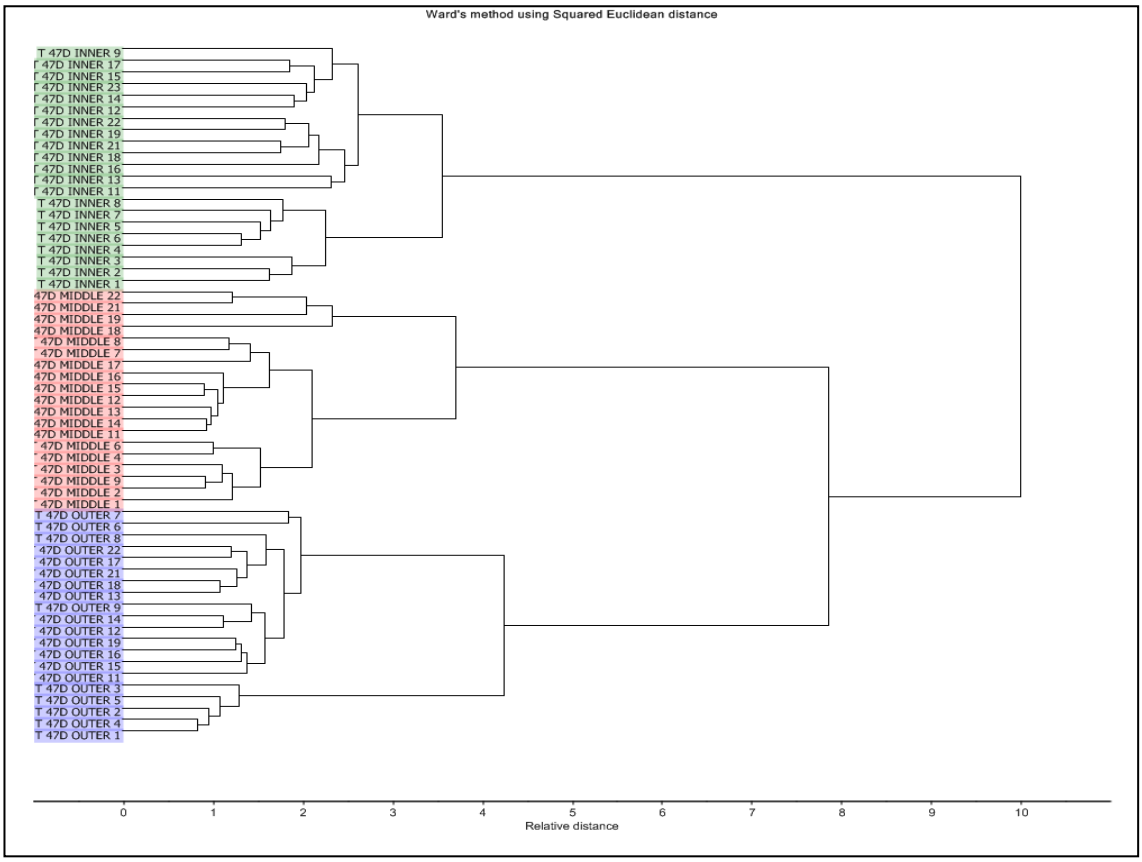

Figure 61: Dendrogram of Cluster Analysis results for normal proliferating region, hypoxic region and necrotic region. CA was performed over the Amide I region using Wards method's squared Euclidean distance.

Figure 61 has shown the dendrogram of outer (normal proliferating), middle (hypoxic) and inner (necrotic) regions over the Amide I range (1800-1510 cm^{-1}). Three main clusters were formed. No spectral mixing was observed within the three regions. Each region formed very individual clusters and distance measurements revealed that biochemical snapshot of Amide I region shows more similarities between normal proliferating and hypoxic regions compared to necrotic region.

Nucleic acid region (980-600 cm^{-1})

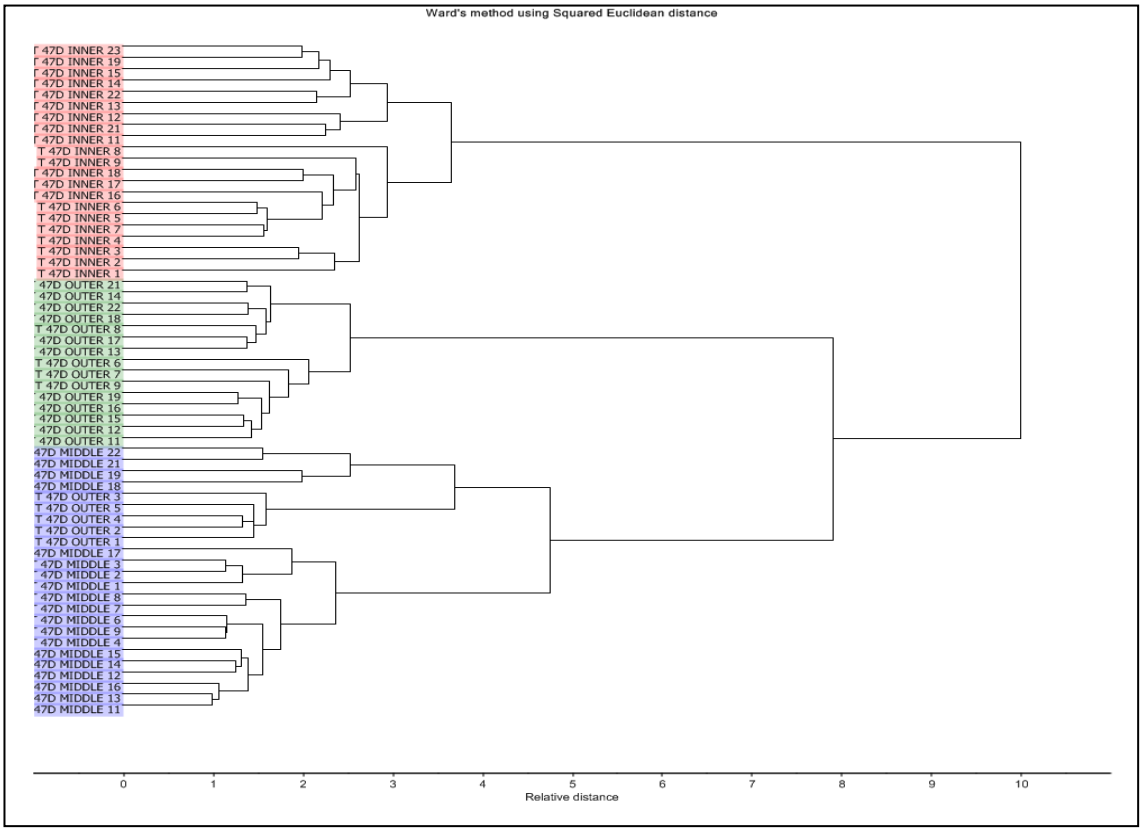

Figure 62: Dendrogram of Cluster Analysis results for normal proliferating region, hypoxic region and necrotic region. CA was performed over the nucleic acid region using Wards method's squared Euclidean distance.

Figure 62 has shown the dendrogram of outer (normal proliferating), middle (hypoxic) and inner (necrotic) regions over the nucleic acid range (980-600 cm^{-1}). Three main clusters were formed. Some of the normal proliferating region spectra were mixed with hypoxic region. Each region formed separate and well-defined clusters. Distance measurements suggest that biochemical snapshot of the nucleic acid region is more similar between normal proliferating and hypoxic regions compared to necrotic region.

5.4 Spheroids – Discussion

This section discusses three aspects of T-47D spheroids. Firstly, a combination of RS and multivariate methods can be used as prospective approach to understand biochemical fingerprint of T-47D spheroids. Secondly, our hypothetical approach ability to differentiate normal proliferating region, hypoxic region and necrotic region with high classification efficacy. Thirdly, discussion of important chemical markers can contribute towards differentiation of three regions of T-47D spheroids. Spectral data analysis is supported by chemometric methods such as PCA, cluster analysis and LDA.

The study of three-dimensional (3D) cell culture models over two-dimensional (2D) monolayers has many advantages. 3D spheroid models mimic tissue specific architecture. It can provide extensive knowledge to understand disease progression and development of new therapeutic drugs (Holliday, 2010). Generally conventional 2D cell culture involves culture of cell lines as monolayers on impermeable plastic surfaces, which often cause loss of phenotypic and functional characteristics. One of the main feature is 'loss of differentiation' and it is mainly happens due to lack of native in vivo 3D confirmation. These monolayers will also preclude cell lines to respond to chemical and molecular gradients in a 3D fashion due to lack of complexity. 3D cell models, particularly spheroid models, have many advantages:

a. Spheroid models will represent 3D architecture of tissues including multicellular arrangement and extracellular matrix deposition that provides spectroscopists extra space to explore chemical pathways,

b. Spheroid models will also provide different cellular microenvironments such as proliferation, quiescent and hypoxic areas allowing one to study chemical differences between these areas,

c. Spheroids provide an opportunity to study cell-cell interactions via tight junctions between cells providing chemical information regarding extracellular matrix (Jaganathan et al., 2014).

A combination of two or more cells in spheroid culture (multicellular spheroids) was originally developed by Sunderland in 1970's. Multicellular spheroids possess well-ordered spherical symmetry and mimic *in vivo* tumour microenvironment. Recent studies in breast cancer have revealed that 3D cell models are ideal source to explore embryo morphogenesis. They have established that 3D breast cancer cell line models allow studying different aspects of cell biology such as proliferation, differentiation, apoptosis and luminal formation (do Amaral et al., 2011, Holliday, 2010).

In our first part of research, a combination of RS and multivariate approach was applied on three different breast cell lines to explore biochemical changes on 2D monolayer cell agar plugs. T-47D models in turn, were prepared to understand biochemical fingerprint of three different regions of spheroids. PCA 2D scores and loading plots were used extensively to understand biochemical nature of spheroids. Although it is impossible to get sensible chemical information from each loading plot but group of loading plots in shorter region is studied in detail to understand biochemical differences.

High-wavenumber region has shown lipid differences in shoulder peak intensities. Hypoxic region has shown lower lipid intensities for CH_3 symmetric stretch of lipids compared to normal and necrotic regions. It is also evident that the nature of lipids present in normal and necrotic region more similar than ones in hypoxic region. Hypoxia, deficient oxygen (O_2) supply, is one of the prominent features of tumour microenvironments. This condition is generated in a cell or tissue due to lack of O_2 required for aerobic respiration. This will lead to development oxidative stress and resulting in production of high lipid peroxidation rate in biological systems (Rauchova

et al., 2012). Mitochondrial metabolism is directly affected during hypoxia because of low amount of O_2 availability at cytochrome oxidases. This condition causes the generation of reactive oxygen species (ROS) in respiratory chain by auto-oxidation. ROS groups are generally known as superoxide (O^{2-}), hydroxyl (HO^-) and peroxyl (RO_2^-) radicals (Naqui et al., 1986). Apart from mitochondrial enzymes, enzymes such as nitric oxide synthase and xanthine oxidase will also enhance production ROS species during hypoxia. ROS play very important role in activation of transcriptional regulators such as hypoxia-inducible factor 1 (HIF-1), nuclear factor kB (NF-kB), activator protein 1 (AP-1) and mitogen-activated protein kinases (MAPK) (Gilkes et al., 2014, Poyton et al., 2009, Semenza, 2000).

Lipid peroxidation in hypoxic region might cause oxidative degradation of lipids. Cell membrane lipids will contribute electrons to ROS groups generate chain reaction (Reed 2011). Chain reaction involves three stages namely generation of fatty acid ROS, reaction of free fatty acid ROS with O_2 to generate peroxyl-fatty acid radical (chain reaction) and radical reaction of two free fatty acid radicals (Esterbauer et al., 1991). Lipid peroxidation directly affects poly-unsaturated fatty acid content of cells because of having high number of reactive hydrogens. As unsaturation increases, lipid content holds high number of multiple double bonds between methylene bridges ($-CH_2-$) and cause cell damages as well (Ronen et al., 1990).

Amide I region has shown clear differences among three different regions of T-47D spheroids. Normal proliferating region has shown low intensity compared to hypoxic and necrotic regions. Figure 33 has shown that hypoxic region has different collagen content compared to normal and necrotic regions. Amide I peak position of normal region also suggests more α-helix form proteins compared to other regions too. Loading

plot results also suggest that normal proliferating region is separated from hypoxic and necrotic regions due to low amide I and high tryptophan contents.

Amide I represent majority of collagen content in breast tissues. Collagen offers structural integrity and strength to human body. The major protein content of Extracellular matrix (ECM) is fibrous proteins which include collagens, elastins, laminins and fibronectins. ECM is made up of 90% of collagen proteins and protein contents of human body collagen. ECM plays very important role in tumour progression and metastasis(Conklin and Keely, 2012). It is made up of different proteins that can control tissue development, homeostasis and disease progression. Breast cancer studies have revealed that protein metabolism related to ECM remodelling is enhanced during disease. Histological studies have identified excessive ECM accumulations in solid tumours. The prominent feature of ECM deposition in tumour tissue is accumulation of collagen(Conklin and Keely, 2012).

Central dogma of protein biology includes three major processes, such as, replication (Synthesis of DNA from DNA), transcription (synthesis of RNA from DNA) and translation (synthesis of protein from mRNA). Hypoxia plays an important role in transcriptional status of collagen gene expression. Research in collagen biosynthesis has shown interesting facts about hypoxia influence. For example, under hypoxic conditions, increased expressions of type I procollagen genes are reported in fibroblasts cultures. Moreover, hypoxic conditions have also enhanced type I, II and IV procollagen levels in lung and pulmonary cells (Gilkes et al., 2014) (Berg et al., 1998, Falanga et al., 1993).

During hypoxia, cancer cell secrete a growth factor known as hypoxia-inducible factor (HIF-1). HIF-1 plays a vital role in production of growth factors from tumour cells. Collagen deposition in hypoxic environment is regulated by three kinds of cells namely

fibroblasts, macrophages and mesenchymal cells. HIF-1 supports the recruitment of macrophages and fibroblasts at hypoxic regions. Macrophages generate growth factors that help fibroblasts to motivate collagen deposition (Wynn and Barron, 2010). Under hypoxic conditions, mesenchymal stem cells will also contribute towards collagen deposition. HIFs control the maturation of collagen fibril by modifying enzymes such as prolyl hydroxylase, lysyl oxidase, lox like protein 2 and 4. A combination of these pathways enhances collagen maturation and production that may help tumour cells to further metastatic dissemination. Spectral differences in peak shifts and peak intensities among three regions have suggested that Amide I position and tryptophan content is sensitive to separate all three regions of T-47D spheroids (Gordon and Hahn, 2010, Gilkes et al., 2014).

Collagen constitutes the majority of pathological breast tissue proteins. The major amino acids of collagen are proline, valine, glycine and phenylalanine. Collagen synthesis involves multiple steps, firstly endoplasmic reticulum host post-translational modifications, such as, hydroxylations and glycosylations. Amino acids, such as, proline, hydroxyproline, lysine and hydroxylysine play important role in thermal stability of the collagen structures ((Myllyharju and Kivirikko, 2004). In Figure 35, hypoxic and necrotic regions have shown high concentration of amino acid and nucleic acid contents compared to normal region. Loading plots of PCA over this region also suggested that these regions were sensitive particularly to necrotic versus hypoxic areas. The high concentration of nucleic acid and amino acid content might suggest that protein and genetic metabolic pathways would assist hypoxic cells to face necrosis due to stress. Along with proline and hydroxyproline intensities, phenylalanine has also shown relative higher intensities in hypoxic and necrotic regions compared to normal region. Hypoxic and necrotic regions have shown higher amino acid and nucleic acid contents and this might suggest that more nucleic acid and amino acids were

synthesized. This may be due to fact that some genetic pathways are changing in the cancer cells in hypoxic and necrotic regions that may arrive from hypoxic stress in T-47D spheroids.

Dendrogram of high-wavenumber region has also confirmed that lipids in normal and necrotic regions have shown more similarities compared to hypoxia region. Amide I region is unique in each region and it may suggest that conformational changes are distinctive to each region. Similarity between normal and hypoxic region may suggest that amide I region is undergoing changes during hypoxic condition and it might be completely different from necrotic region. Nucleic acid region has shown more similarities between normal proliferating region and hypoxic regions compared to necrotic regions whereas loading plots suggest that hypoxic and necrotic regions are sensitive in separation based on nucleic acid content. This may suggest that nucleic acid content is unique in both regions and hypoxic region is related to survival region whereas necrotic region adopts apoptosis due to stress.

In conclusion, Raman spectra combined to multivariate approach proves to be an exceptional approach that allows separation of normal proliferating, hypoxic and necrotic regions based on lipids, amide I, III and nucleic acid content. These differences observed in three different regions might be useful in identification of chemical changes associated stress or strain faced by each region progressing towards necrosis. Loading plots suggested that normal proliferating region is separated with low amide I content and high tryptophan content compared to hypoxic and necrotic regions. Peak intensity and peak shifts have suggested that amide I content is unique in each region with regards to conformation and quantity. Amide III region, especially nucleic acid and amino acid content, is particularly sensitive to necrotic and hypoxic regions. This might be due to stress or strain associated with hypoxic and necrotic regions.

Chapter 6

6.1 Tissue Micro Array (TMA) breast biopsies – Results

In this section, full spectral analysis of luminal a, luminal b, HER2 positive and triple negative breast cancer subtypes were conducted. Raman Spectroscopy has been used as prospective approach to evaluate biochemical differences among four subtypes of tissue microarray samples. Intra biopsy and inter biopsy chemical changes were reported using combination of chemometric methods. Each biopsy has produced unique biochemical snapshots regarding spatial diversity of macromolecules such as proteins, lipids, carbohydrates and nucleic acids. Raman spectra of each subtype have differed in peak intensities and peak shifts. Furthermore, all four subtypes were classified using multivariate approach such as PCA, cluster analysis and LDA. Chemometric methods have assisted in accurate and reliable classification with improved sensitivity and specificity.

TMA Biopsies slide de-wax and H&E staining:

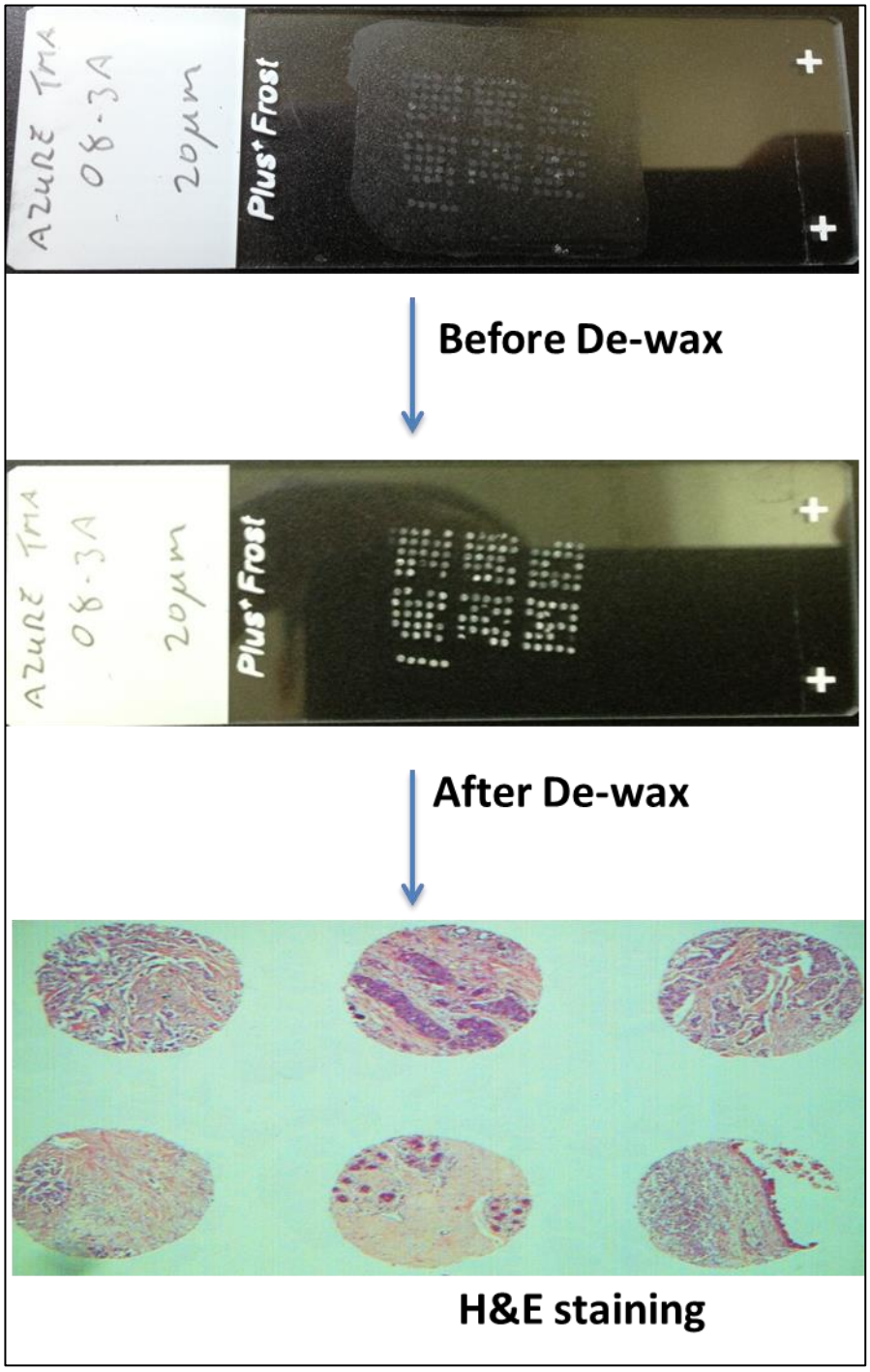

Figure 63: TMA Biopsies slide de-waxing and staining

Few H&E stained images of TMA biopsies (Scale bar indicates 100 µm):

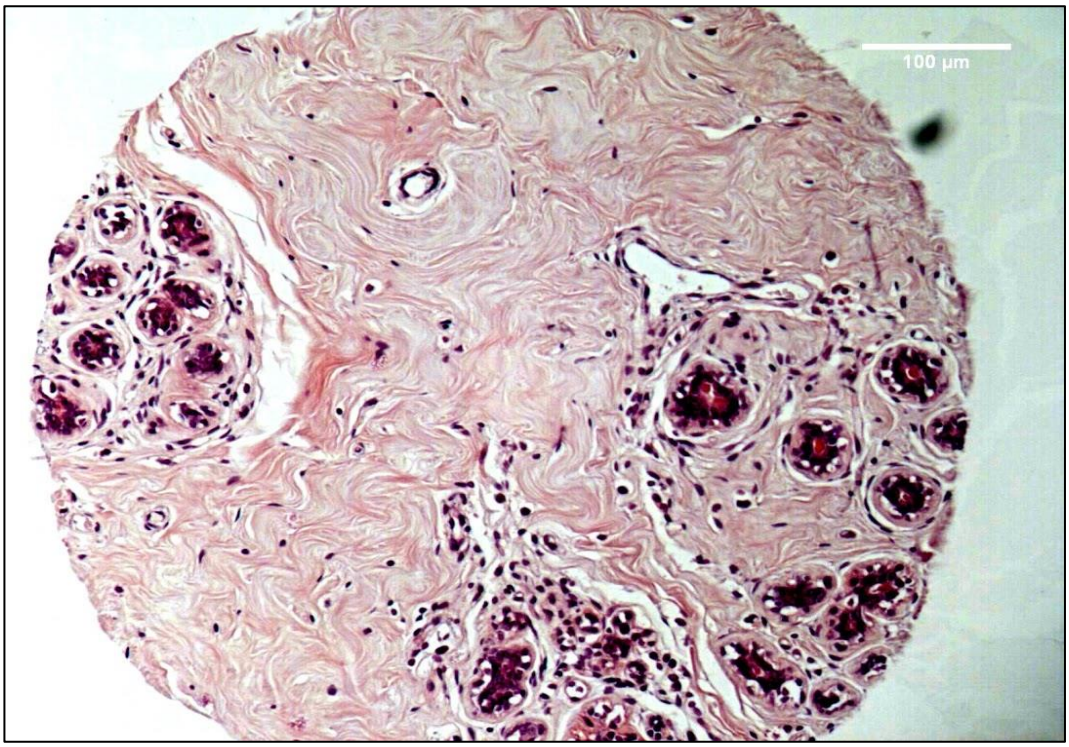

Figure 64: Ductal carcinoma NST of TMA biopsy

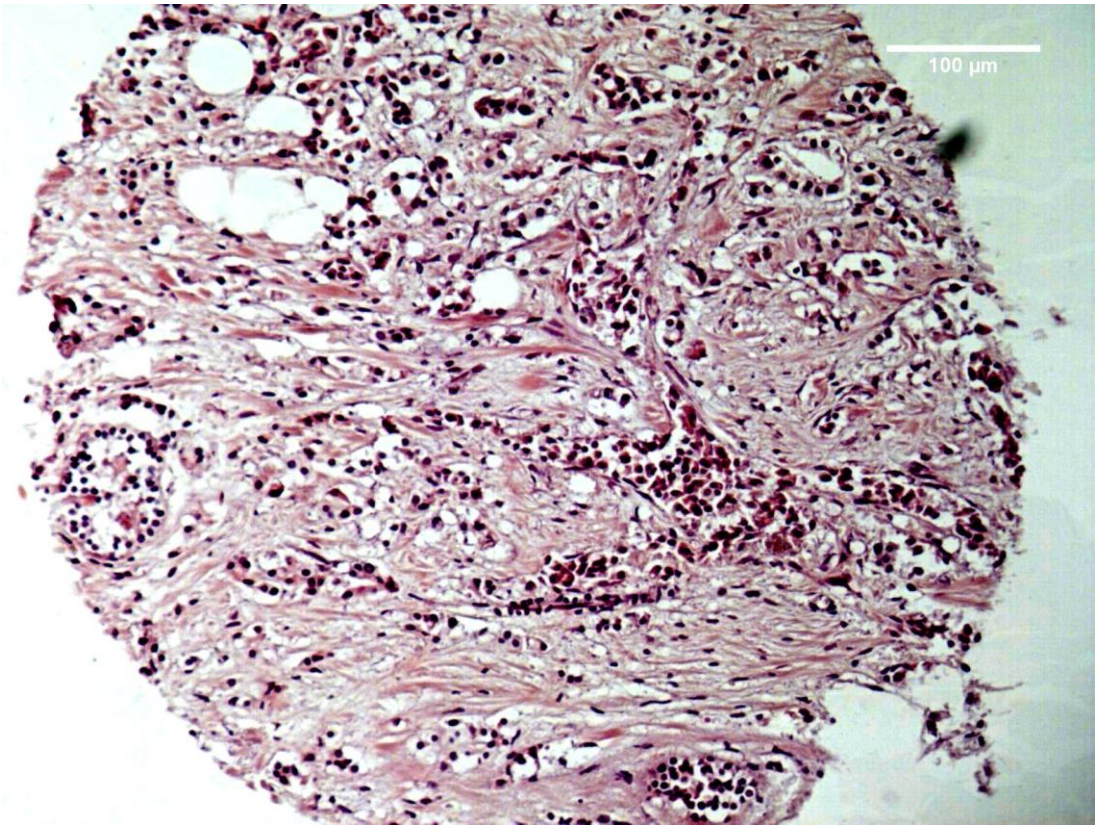

Figure 65: Ductal carcinoma NST of TMA biopsy

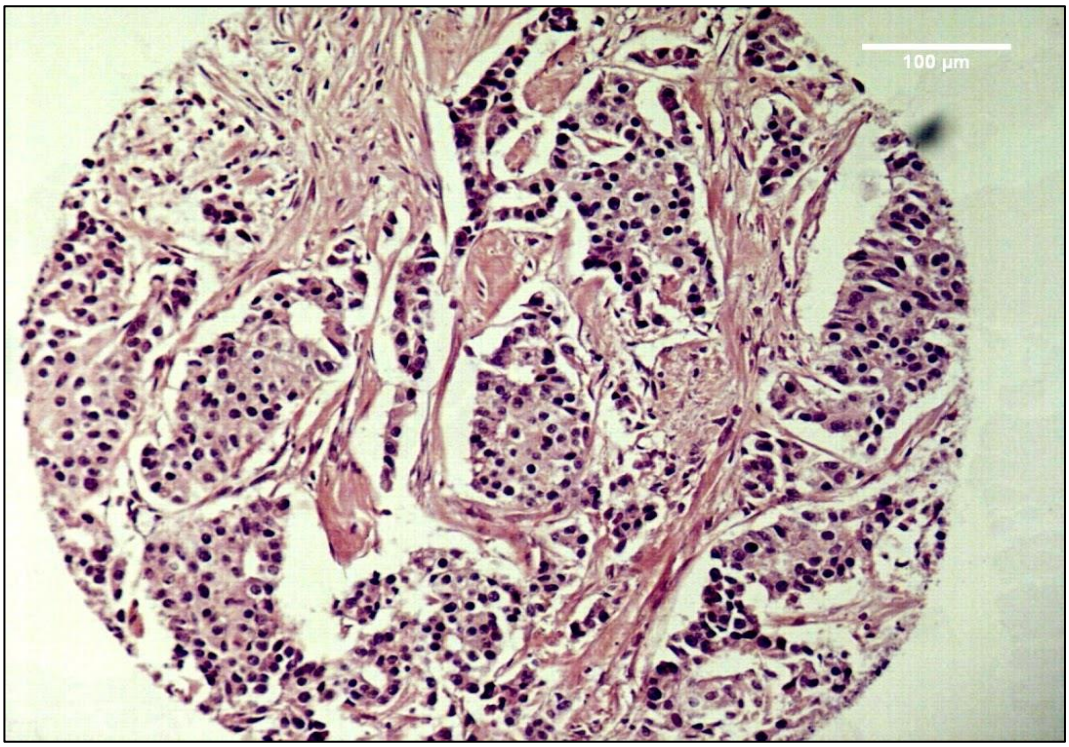

Figure 66: Ductal carcinoma of TMA biopsy

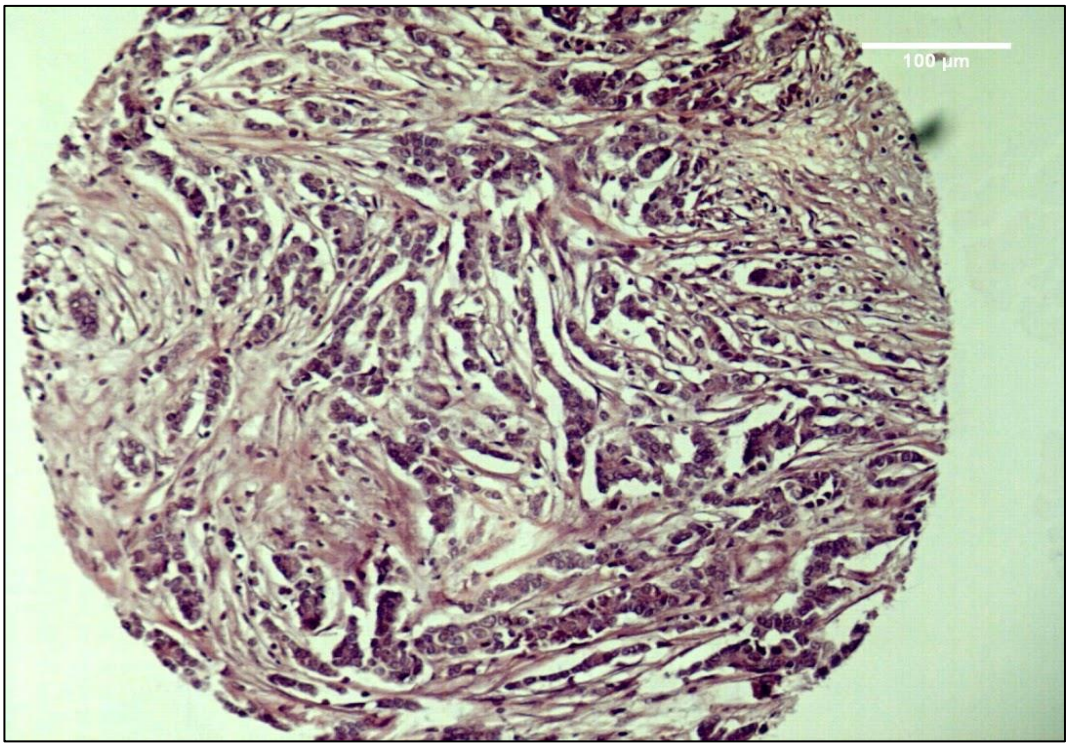

Figure 67: Ductal carcinoma NST of TMA biopsy

6.2 Raman peak analysis

Four hundred spectra were collected from luminal A, luminal B, HER2 and triple negative subtypes and mean spectra of each subtype was extracted (Figure 69). The band numbers across the whole spectral range of each subtype and their corresponding biological component assignments are summarised in table (appendix). Peak assignments for the lipid region, amide I, amide II and amide III regions are discussed in the next section.

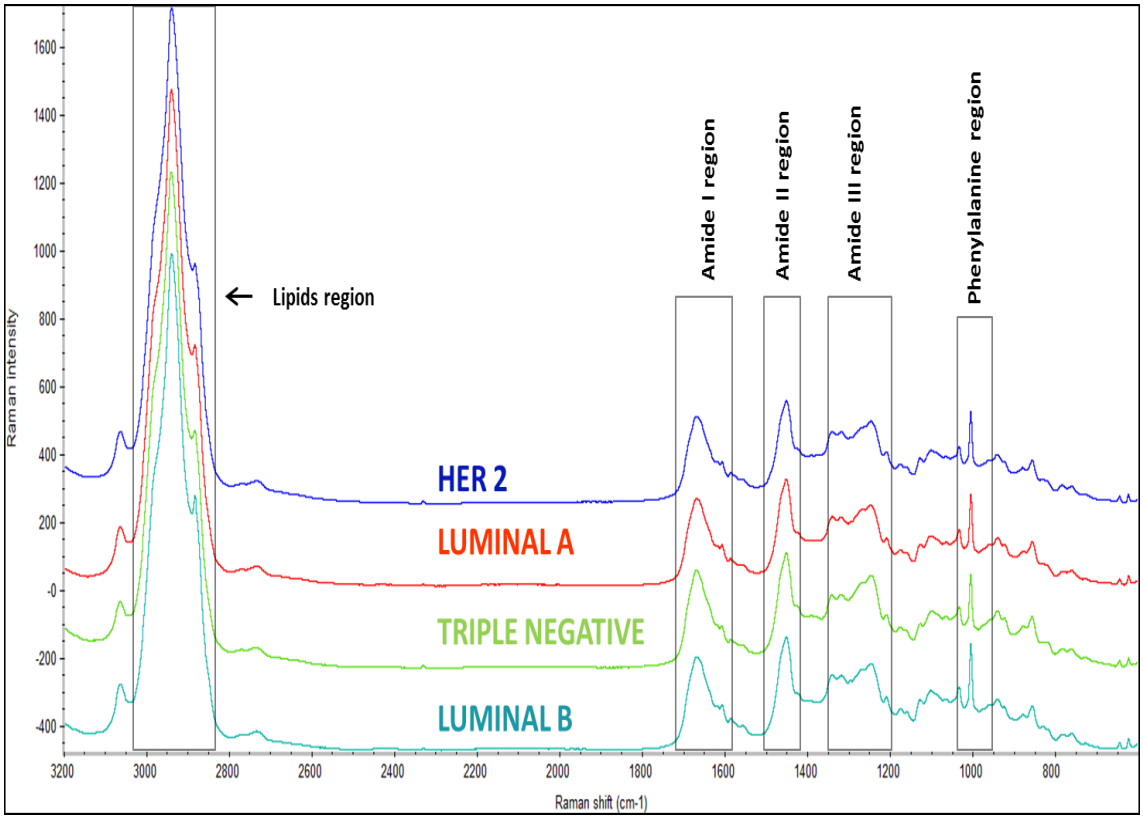

Figure 68: Mean spectra of luminal a, luminal b, HER2 positive and triple negative subtypes of tissue microarray breast cancer biopsies.

6.2.1 C-H region (3200 - 2700 cm^{-1})

A small band at 3061 cm^{-1} was observed in all four subtypes representing C-H stretch of lipids. A strong Raman peak was observed at 2936 cm^{-1} in luminal B and HER2 positive, attributed to $\nu(CH_3)$ of lipids whereas similar peak was observed in luminal A at 2937 cm^{-1} and in triple negative at 2938 cm^{-1} corresponding to $\nu(CH_3)$ of proteins and CH stretching of lipids and proteins. A strong shoulder peak was observed in three

subtypes at (luminal A, HER2 positive and triple negative) at 2881cm^{-1} attributed to CH_2 symmetric stretch of lipids and CH_2 asymmetric stretch of lipids and proteins. A similar shoulder peak was observed in luminal B at 2880 cm^{-1} corresponding to CH_2 stretching of lipids (Figures 70-72).

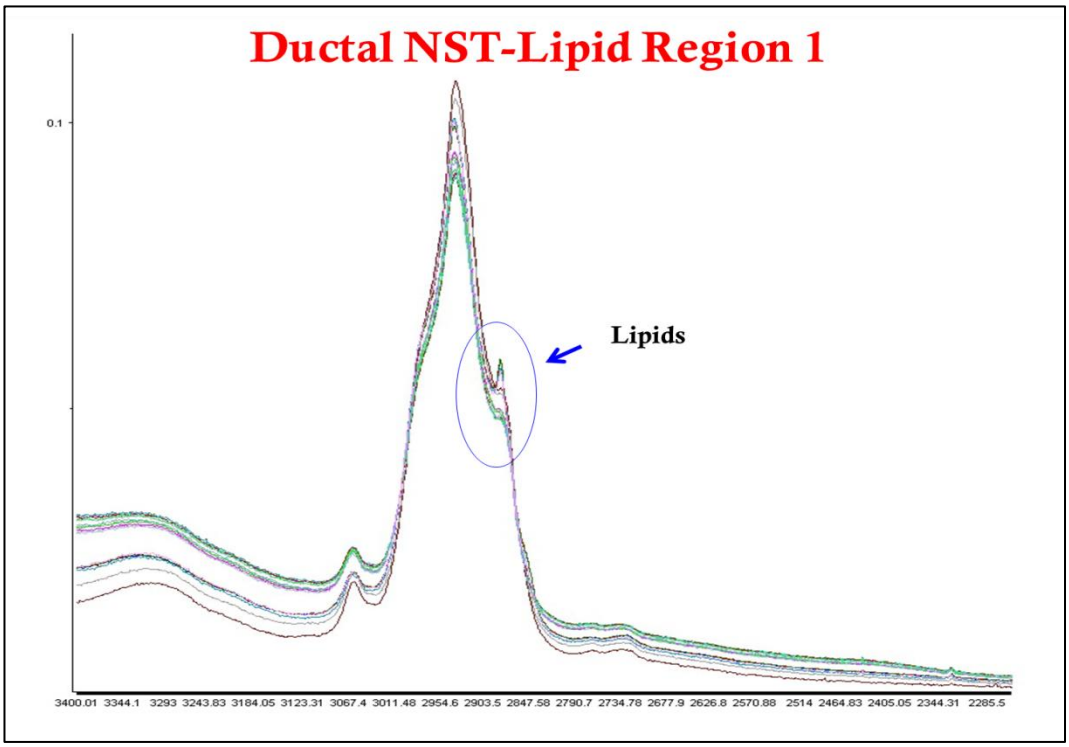

Figure 69: Lipid intensity differences at high-wavenumber region in individual ductal NST biopsy.

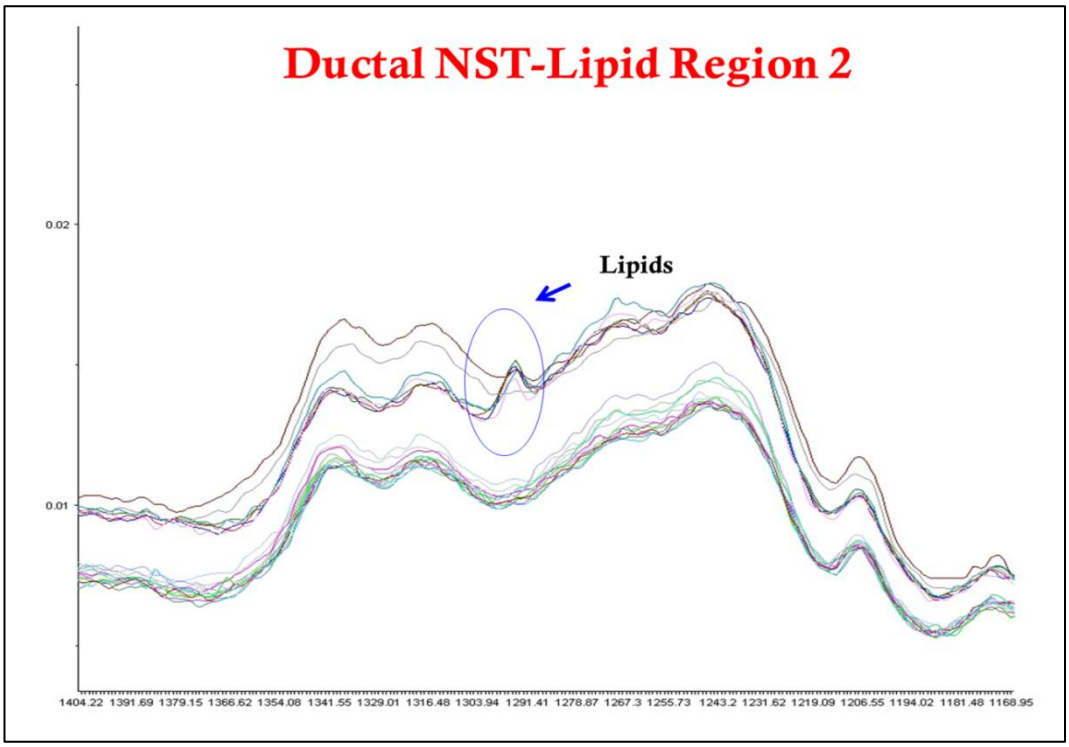

Figure 70: Lipid peak intensity differences at amide II region in individual ductal NST biopsy.

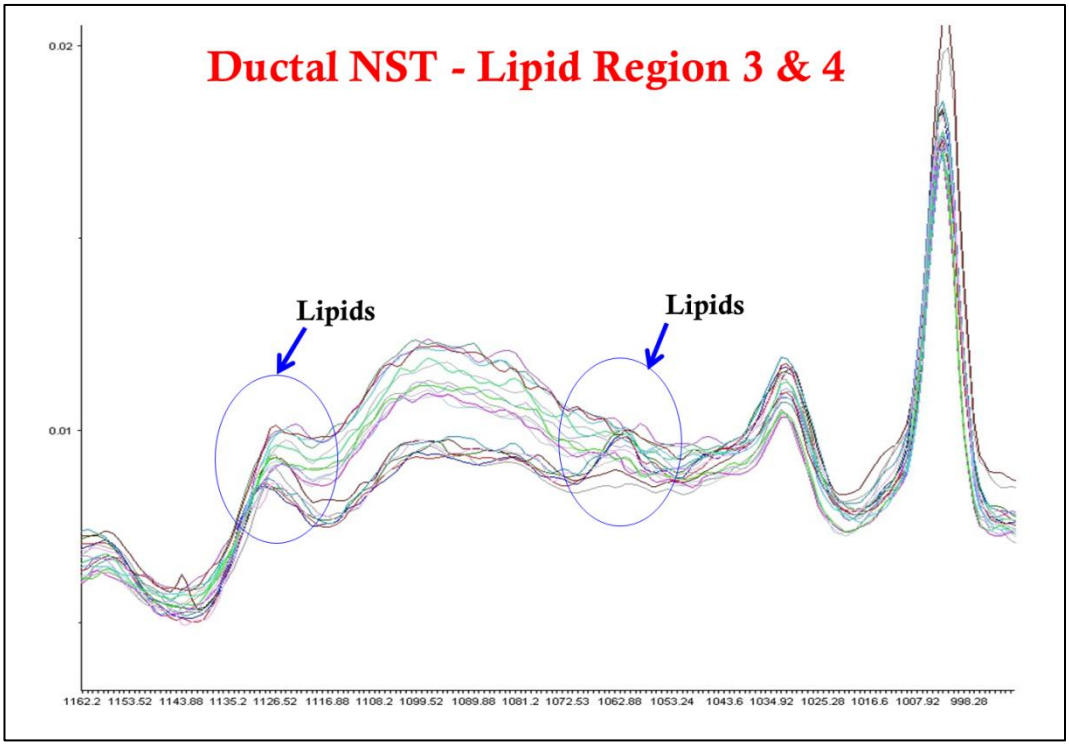

Figure 71: Lipid peak intensity differences at amide III region in individual ductal NST biopsy.

6.2.2 Amide I & II region (1670-1350 cm^{-1})

A strong Raman peak of amide I was observed at 1665 cm^{-1} in luminal A and B, whereas this peak is shifted to 1666 cm^{-1} in HER2 and 1667 in triple negative subtypes. Collagen is the prominent contributor of this peak in breast cancer tissues. This is due to either C=C stretching of proteins, α-helical structure of proteins or carbonyl stretches of tumours or collagen. A small Raman peak was observed at 1605 cm^{-1} in triple negative, whereas this peak is shifted in luminal A, B and HER2 subtypes. This peak is attributed to cytosine (NH_2), ring C-C stretch of phenylalanine, tyrosine, C=C vibrations of proteins. Weak spectral contributions were noticed at 1584 and 1583 cm^{-1}, which are attributed to C-C stretching and C=C bending mode of phenylalanine of proteins. C-H vibration of proteins and lipids were expressed in all four subtypes at 1449 cm^{-1}. Triple negative subtype have shown Raman peak at 1339 cm^{-1}, attributed to C-C stretch of phenyl (I) and C_3-C_3 stretch and C_5-O_5 stretch CHα in-plane bend. The other three subtypes have expressed this peak at 1338 cm^{-1}. Luminal A and Triple negative have shown Raman peak at 1317 cm^{-1}, whereas luminal B and HER2 have shown Raman band at 1316 cm^{-1}. The assignment of this band is purine base such as guanine (Figures 73-74).

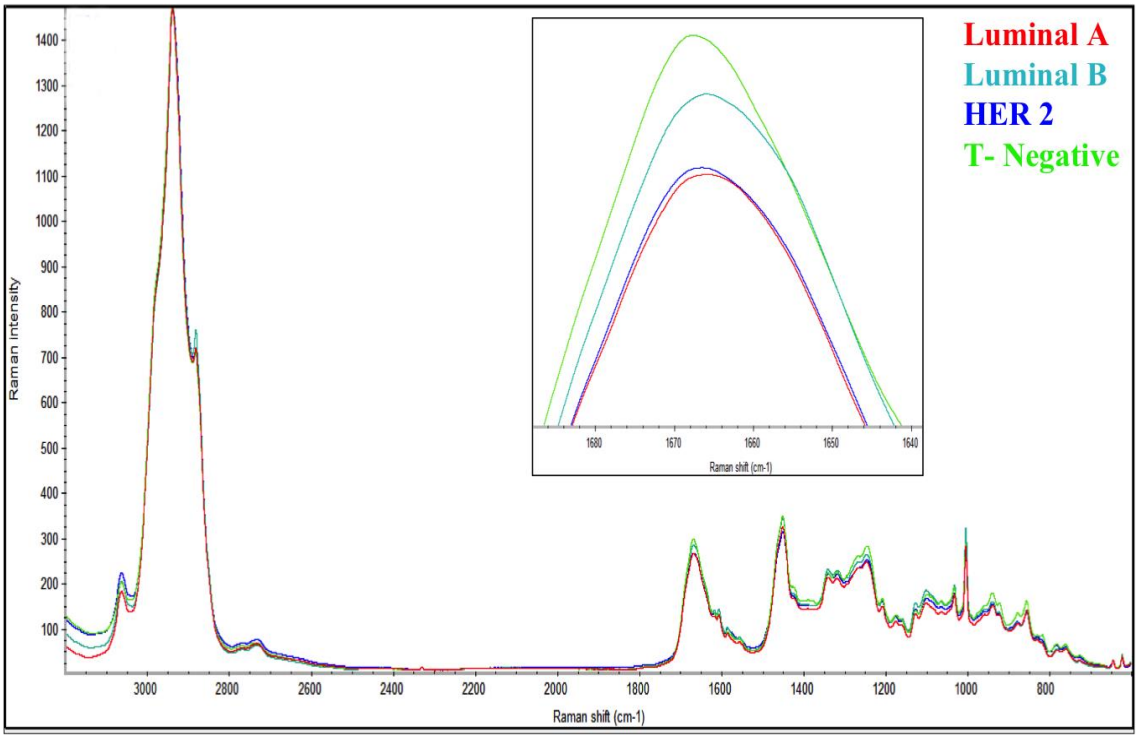

Figure 72: Amide I peak intensities and shifts of different subtypes of breast cancer.

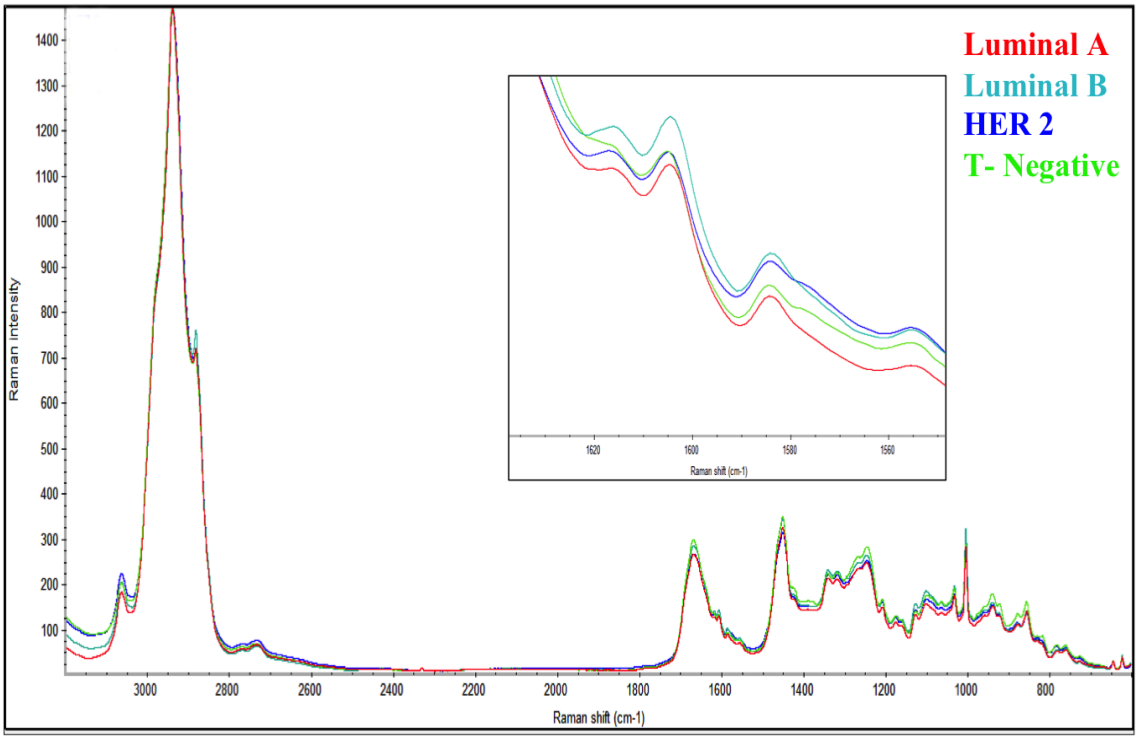

Figure 73: Amide I shoulder peak intensities and shifts of different subtypes of breast cancer.

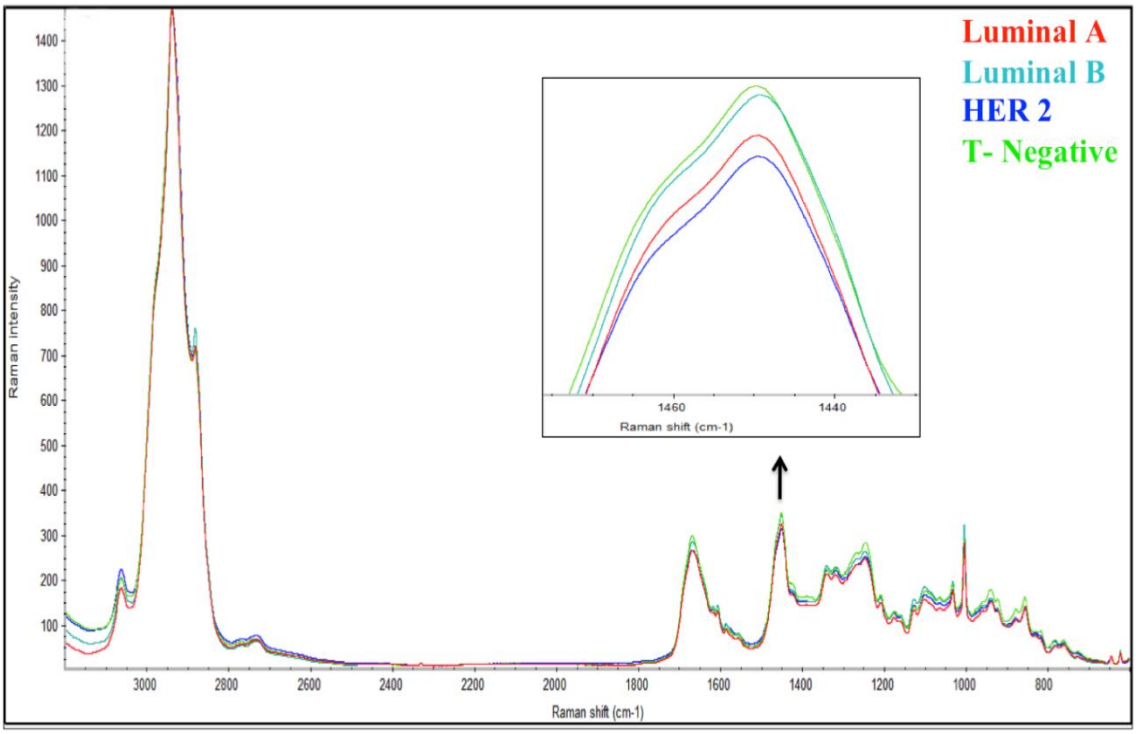

Figure 74: Amide II peak intensities and shifts of different subtypes of breast cancer.

6.2.3 Amide III region (1250-1170 cm^{-1})

Amide III peak has been present in all four types at 1245 cm^{-1}. Amide III represents stretching vibrations of C-N and bending vibrations of N-H. It denotes disordered protein confirmation. All four subtypes have shown Raman peak at 1207 cm^{-1} attributed to hydroxyproline and tyrosine. Triple negative subtype has shown very weak Raman peak at 1174 cm^{-1} representing δ(C-H) and tyrosine. This peak is shifted in luminal A and HER2 subtype at 1173 cm^{-1} representing tyrosine of collagen type I, cytosine and guanine bases. The same peak is shifted in luminal B at 1172 cm^{-1} attributed to δ(C-H) and tyrosine. All four subtypes have reported shoulder peak at 1125 cm^{-1} attributed to C-C of nucleic acid peaks (Figure 76-77).

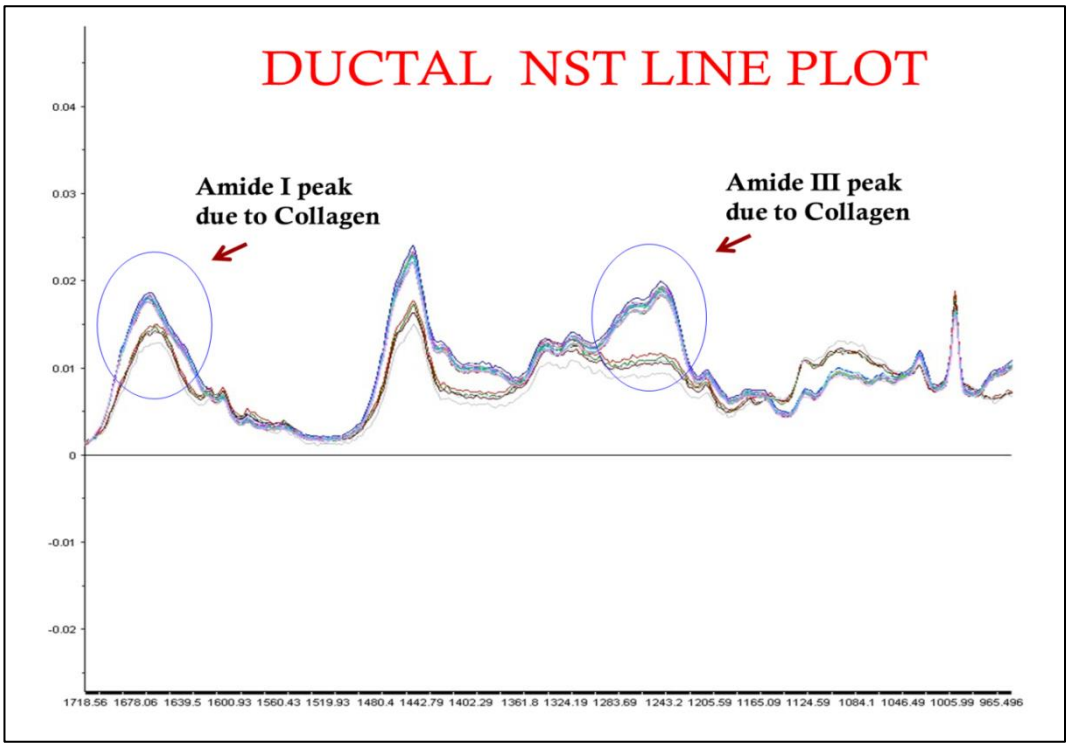

Figure 75: Collagen differences at amide I and III regions in individual ductal NST biopsy.

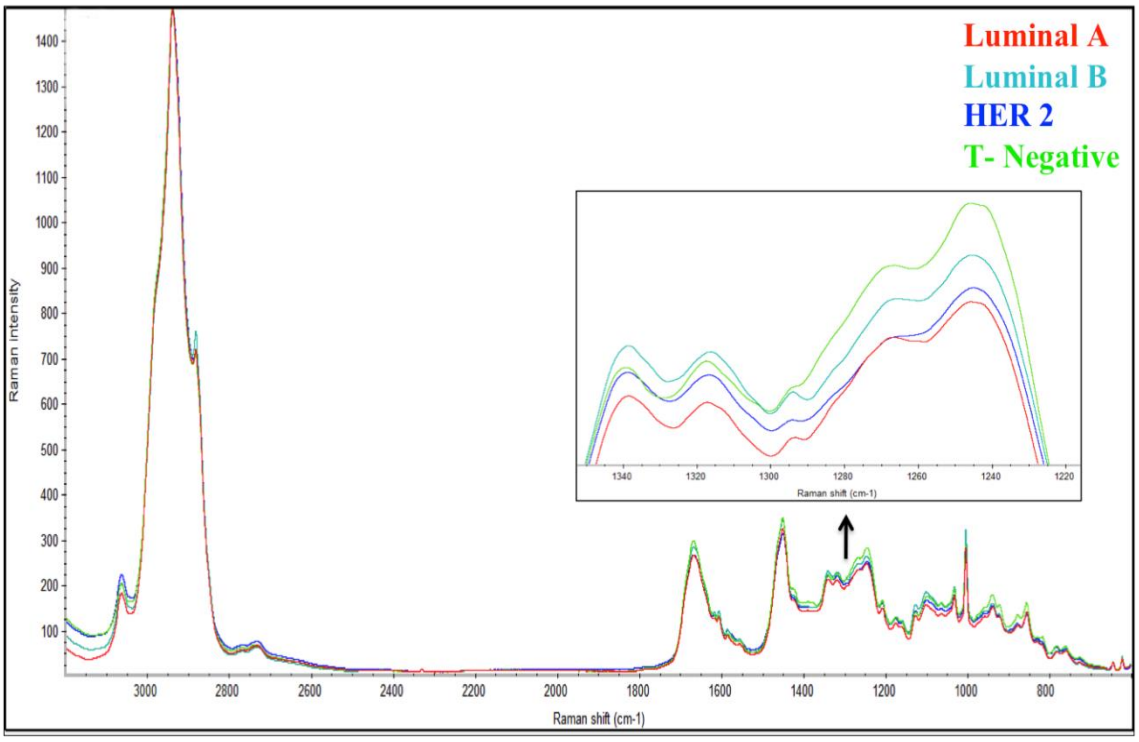

Figure 76: Amide III peak intensities and shifts of different subtypes of breast cancer.

6.2.4 Phenylalanine (1170-1000 cm^{-1})

A characteristic amide III band was present in luminal A subtype at 1100 cm^{-1}, which indicates (C-C) vibration modes of gauche-bonded chain and *ν*(C-C) of lipids and fatty

acids. A medium Raman peak was observed in luminal B and HER2 subtypes at 1099 cm^{-1} attributed (ν(C-N)) and this peak shifted to 1097 cm^{-1} in triple negative subtypes. A medium Raman peak is observed in triple negative subtype at 1063 cm^{-1} and this peak is shifted in luminal A at 1062 cm^{-1}. HER2 and triple negative subtypes have shown medium Raman peak at 1032 cm^{-1} attributed to CH_2 and CH_3 bending modes of collagen and phospholipids. This peak is shifted in luminal A and B at 1031 cm^{-1}. A sharp, intense and typical peak has been observed at 1002 cm^{-1} (luminal and HER2 subtypes) 1003 cm^{-1} (triple negative) is a characteristic assignment of ring breathing of aromatic amino acid such as phenylalanine (Figure 78).

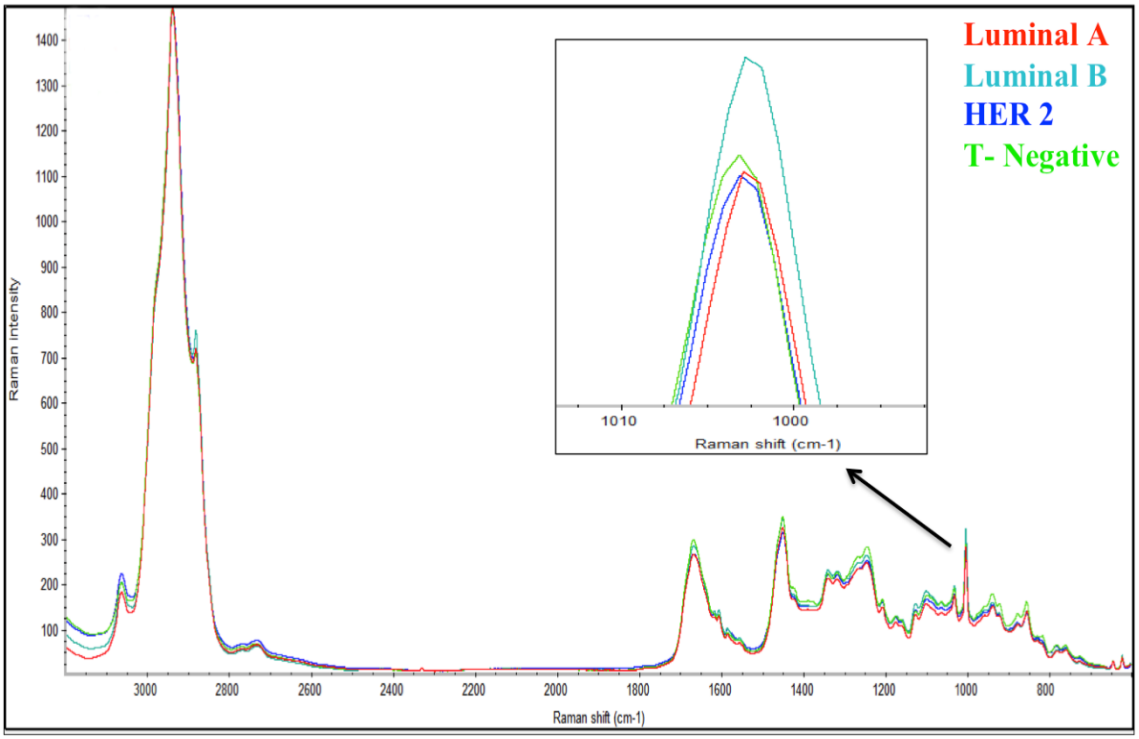

Figure 77: Phenylalanine peak intensities and shifts of different subtypes of breast cancer.

6.2.5 Nucleic acid and amino acid region (1000 - 600 cm^{-1})

Luminal and triple negative subtypes have expressed a medium Raman peak at 938 cm^{-1} attributed to proline, hydroxyproline, ν(C-C) skeletal collagen backbone and C-C stretch of polypeptide backbones whereas luminal A subtype expressed at 937 cm^{-1}. A medium Raman peak observed in triple negative subtypes at 921 cm^{-1} attributed to

proline ring, glucose and lactic acid molecules. A medium Raman peak observed in luminal B, HER2 and triple negative subtypes at 877 cm^{-1} is attributed to C-C-N^{+} symmetric stretch of lipids, C-O-C ring of carbohydrates. All four subtypes of breast cancer expressed Raman peak at 854 cm^{-1} is attributed to C-O-C skeletal mode of α-anomers (polysaccharides, pectin), ring breathing mode of tyrosine. HER2 subtype has expressed Raman peak at 782 cm^{-1} attributed to DNA, Thymine, Cytosine, Uracil, and RNA. Triple negative subtype have expressed at 781 cm^{-1} attributed to ring breathing mode of cytosine and uracil, whereas luminal B expressed 780 cm^{-1}. A very weak Raman peak observed in HER2 and triple negative subtypes at 759 cm^{-1} attributed to Tryptophan and ethanolamine group of phosphatidylethanolamine. Luminal subtypes have expressed Raman peak at 758 cm^{-1} attributed to tryptophan. A weak Raman peak at 643 cm^{-1} in luminal B, HER2 and triple negative subtype attributed to C-C twisting mode of tyrosine and it is shifted to 642 cm^{-1} in luminal A subtypes. HER2 and triple negative subtypes have expressed C-C twisting mode of phenylalanine and disulphide bridges of cysteine amino acid content. This peak is shifted to 620 cm^{-1} in luminal subtypes.

6.3 Multivariate analysis

PCA was performed on luminal A, B, HER2 positive and triple negative subtype spectra in order to focus on variability existing among subtypes. In this process, we extracted loading plots from each score plot. These describe the amount of variation for each variable for a given principal component. Loading plots not only provide variations of data sets but also provide correlation of loading point and position of the spectra. Examination of loading plots will offer information regarding origin of variability with in the dataset and based on this information we can explore biochemical components and its contribution towards variations among subtypes.

6.3.1 Full spectral range

PCA was performed on full spectral range (3200-600 cm^{-1}) of four subtypes. Four subtypes formed really dense clusters and are not clearly separated. In the first score plot, PC-2 has differentiated luminal B from rest of the subtypes but separation of other subtypes were not much clear. In the second score plot every subtype was grouped at both positive and negative sides of PC-3 axis indicating large inner group variation. PC-2 and PC-3 have described little or no discrimination among these subtypes. In both PCA score plots, majority spectra of luminal A and luminal B have shown positive and negative extremes, respectively. A three dimensional PCA plot of full spectral range has shown good cluster formation in luminal B, HER2 positive and triple negative subtypes and weak cluster formation in luminal A subtype. A 3D PCA plot has shown good separation in luminal A and luminal B subtypes, whereas, HER2 positive and triple negative subtypes are overlapped using the principal components PC-1, PC-2 and PC-3 (Figures 79-81).

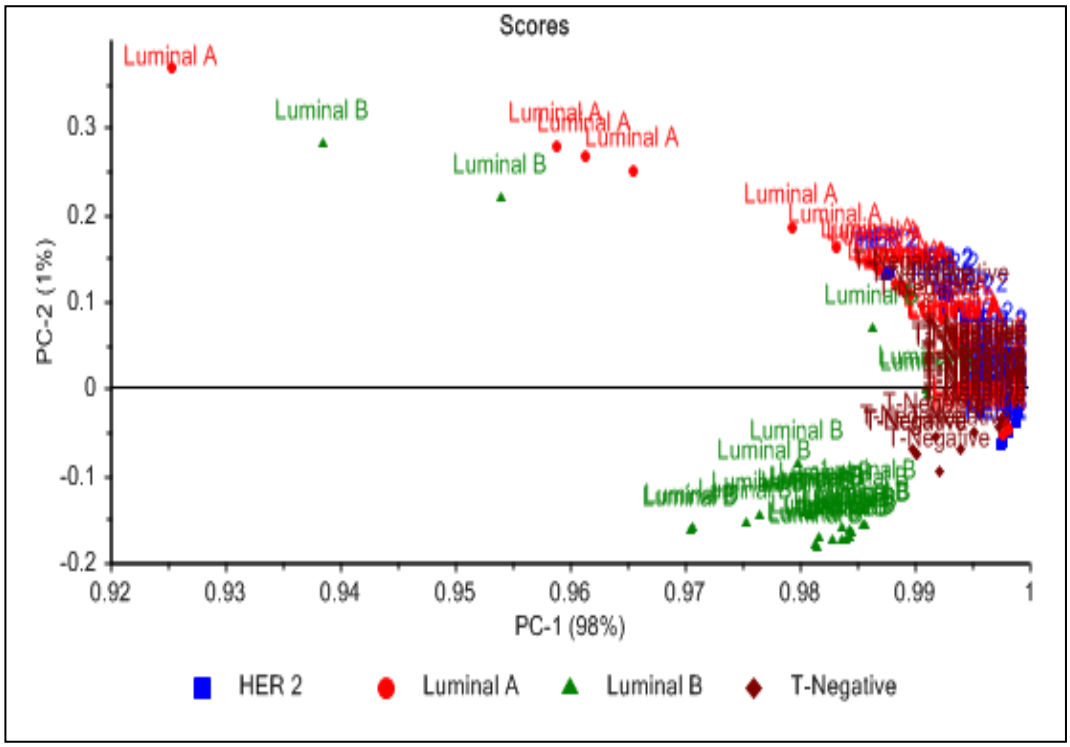

Figure 78 A two-dimensional principal component analysis (PCA) score plot of full spectral range for luminal A (red), luminal B (green), HER2 (blue) and T-Negative (dark red) based on first and second principal components (PC).

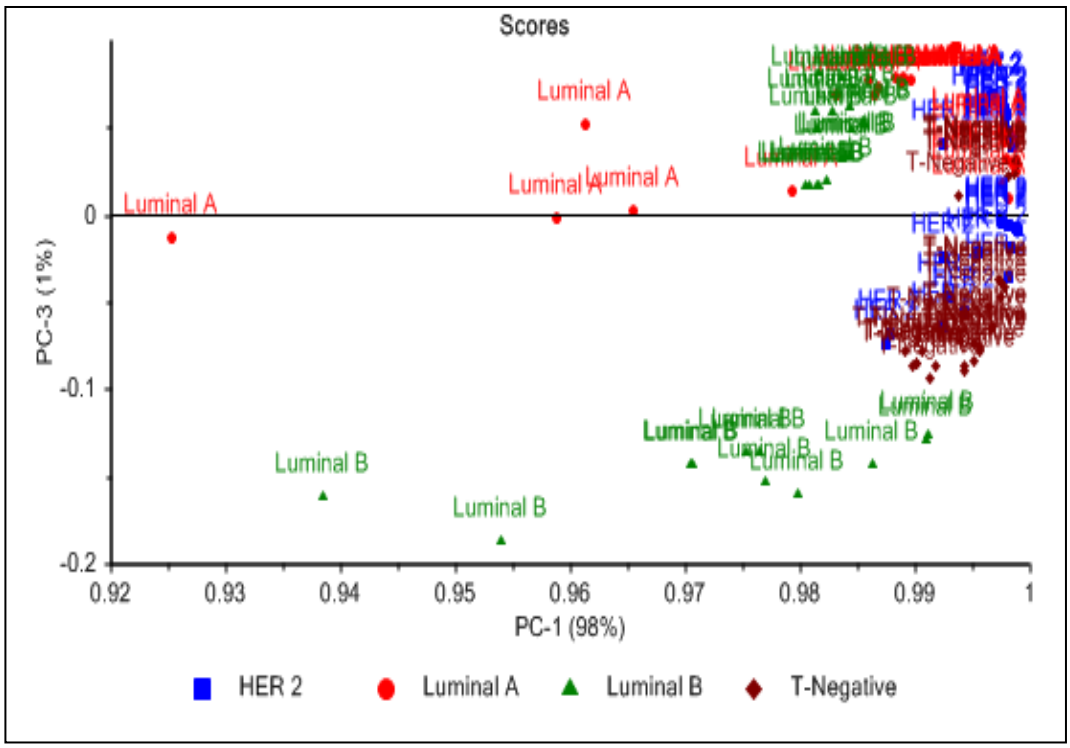

Figure 79: A two-dimensional principal component analysis (PCA) score plot of full spectral range for luminal A (red), luminal B (green), HER2 (blue) and T-Negative (dark red) based on first and third principal components (PC).

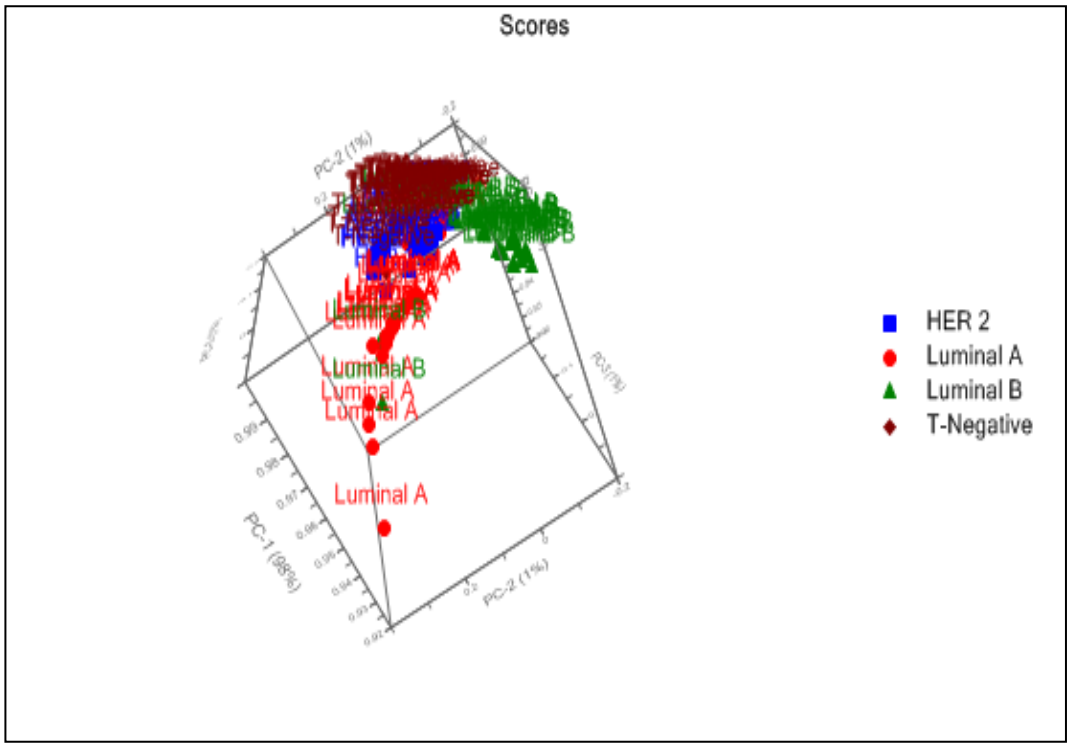

Figure 80: A three-dimensional principal component analysis (PCA) plot of full spectral range for luminal A (red), luminal B (green), HER2 (blue) and T-Negative (dark red) based on first, second and third principal components (PC).

PCA loading plot analysis of full spectral range has described that PC-2 loading plot (red) has differentiated majority of luminal B subtypes from other subtypes. It is difficult to interpret loadings information in full spectral range. Because variation associated with each PC is dependent on so many peaks being in unison with regards to a vast number of different features, such as, peak position, shift, width and height. Based on this, we extended PCA analysis to narrower regions, such as, lipids, amide I, III and nucleic acid regions to identify biochemical variations within subtypes (Figure 82).

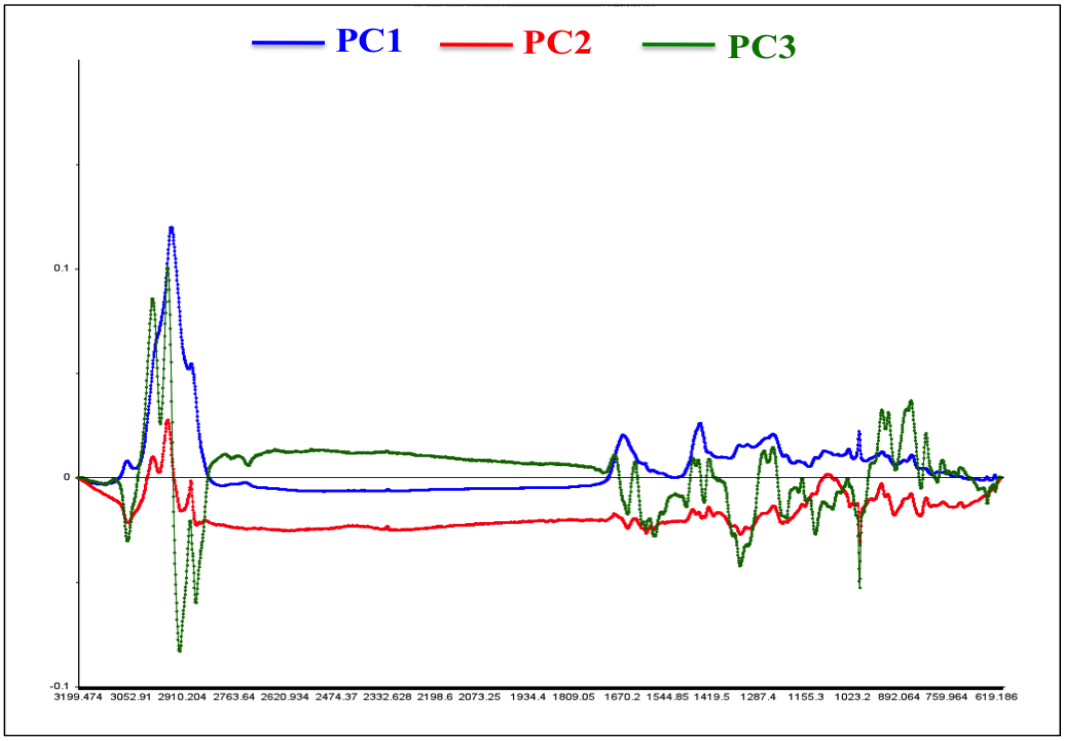

Figure 81: A loading plot of principal component analysis (PCA) for nucleic acid region representing first (blue), second (red) and third (green) principal components (PC).

6.3.2 Lipids

PCA analysis was performed on lipid region (3200 -2600 cm^{-1}). In the first score plot, PC2 has pretty much separated luminal A from triple negatives, whereas, HER2 positive and luminal B was dispersed. Notably, luminal A were grouped at positive sides on the PC-2 axis, whereas, triple negatives were on both axes. PC-3 has shown whole distribution subtype clusters on positive side of axis whereas a subset of luminal B has distributed on negative side of the axis. The second score plot has shown huge mixture of subtype spectra along PC-3 axis rather than clusters. A three dimensional PCA plot of high-wavenumber range has shown good cluster formation in HER2 and triple negative subtypes and weak cluster formation in luminal A and B subtypes. A 3D PCA plot has shown good separation in luminal B subtype whereas luminal A, HER2 and triple negative subtypes are overlapped using the principal components PC-1, PC-2 and PC-3 (Figures 83-85).

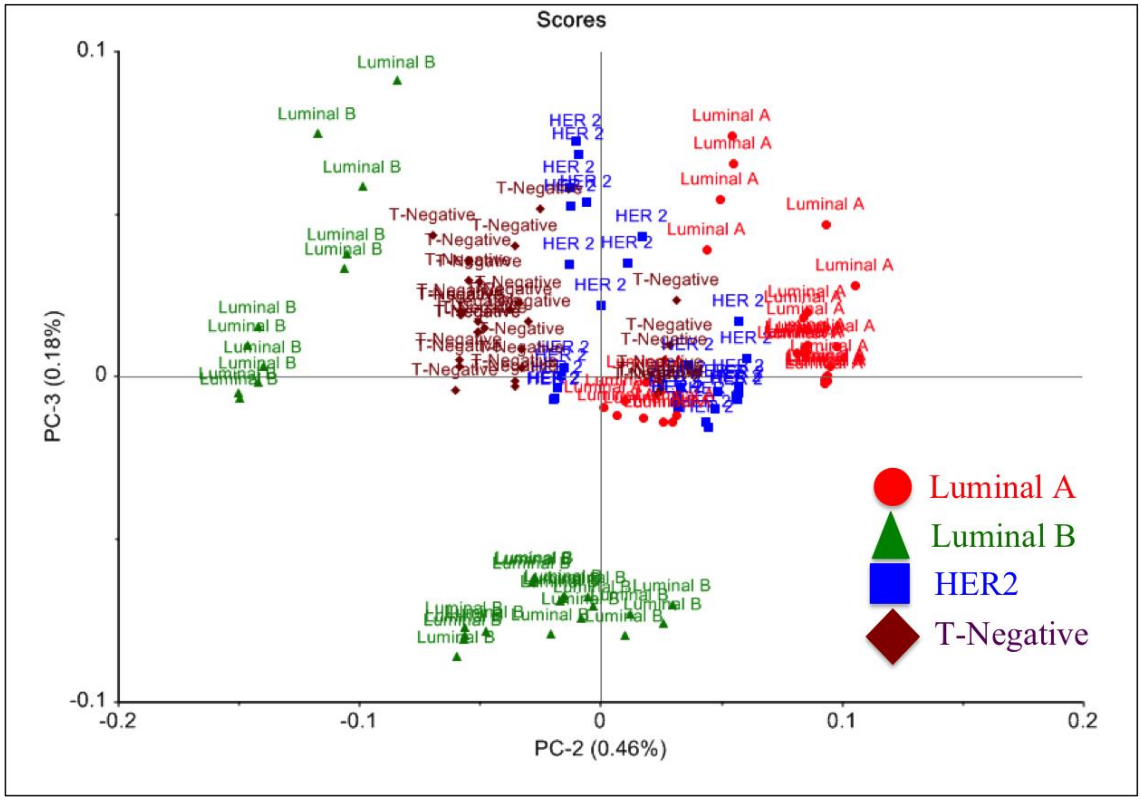

Figure 82: A two-dimensional principal component analysis (PCA) score plot of lipid region for luminal A (red), luminal B (green), HER2 (blue) and T-Negative (dark red) based on second and third principal components (PC).

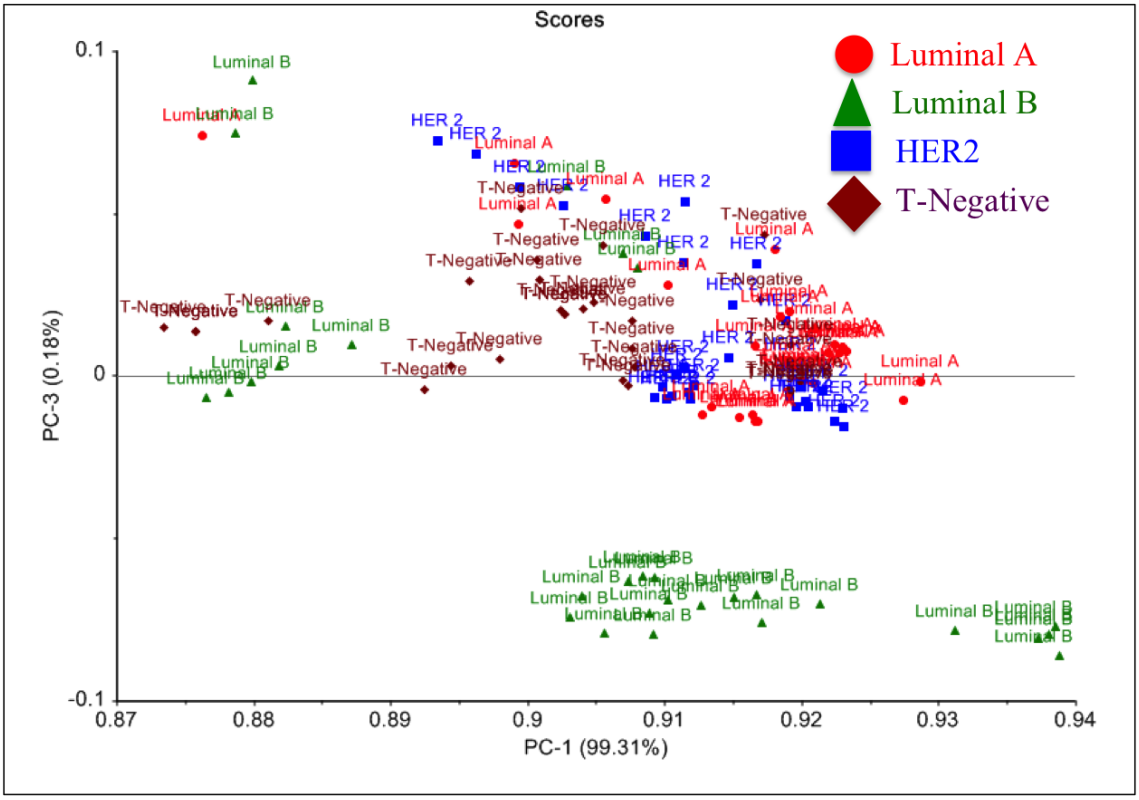

Figure 83: A two-dimensional principal component analysis (PCA) score plot of lipid region for luminal A (red), luminal B (green), HER2 (blue) and T-Negative (dark red) based on first and third principal components (PC).

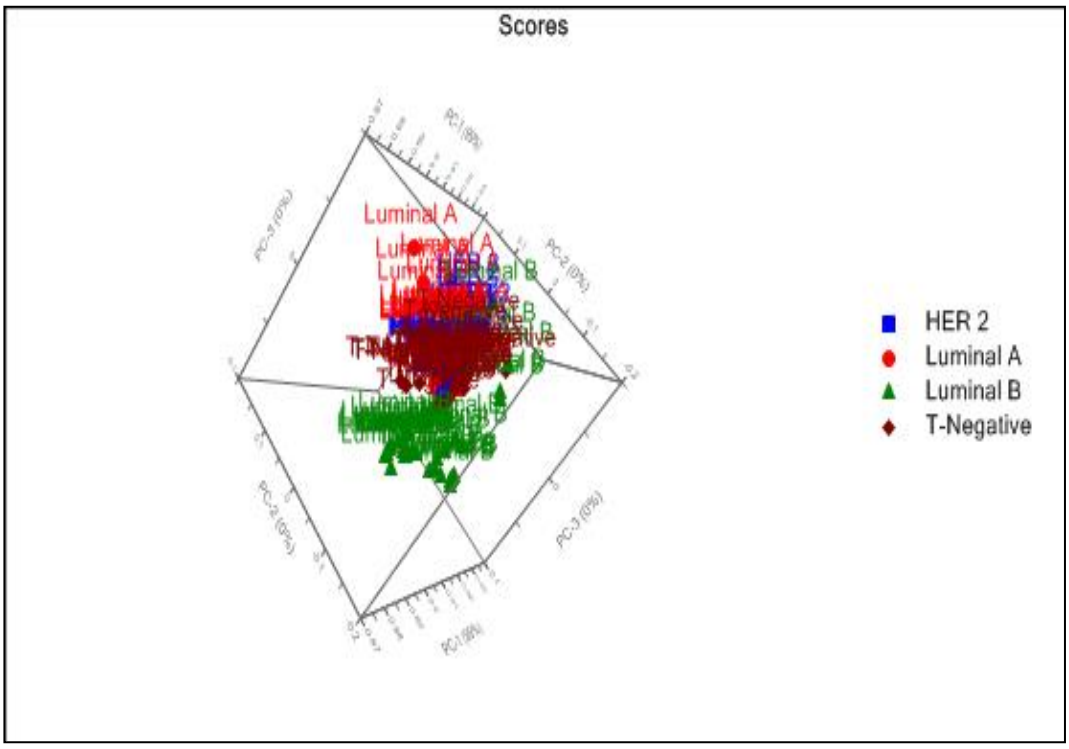

Figure 84: A three-dimensional principal component analysis (PCA) plot of lipid region for luminal A (red), luminal B (green), HER2 (blue) and T-Negative (dark red) based on first, second and third principal components (PC).

PCA loading plot analysis of high-wavenumber region has described biochemical differences among four subtypes. Firstly, PC-1 loading plot (blue) describes variations in protein specific symmetric CH_3 vibrations of all subtypes. PC-2 loading plot (red) has pretty much positively separated luminal A and majority of triple negatives subtype on 2991 and 2848 cm^{-1} (CH_3 symmetric vibration of lipids and proteins) compared to remaining subtypes. PC-2 has also negatively separated triple negative, luminal B subtypes and minority of HER2 subtype with 3060 cm^{-1}, 2914 cm^{-1} (CH stretching of proteins) and 2869 cm^{-1} (CH_2 symmetric and asymmetric stretch of lipids, and CH_2 asymmetric stretch of proteins). PC-3 loading plot (green) has shown that majority of all subtypes and except minority of luminal B has positively separated with 2918 cm^{-1} (CH stretching of proteins) and 2881 cm^{-1} (CH_2 symmetric and asymmetric stretching vibrations) (Figure 86).

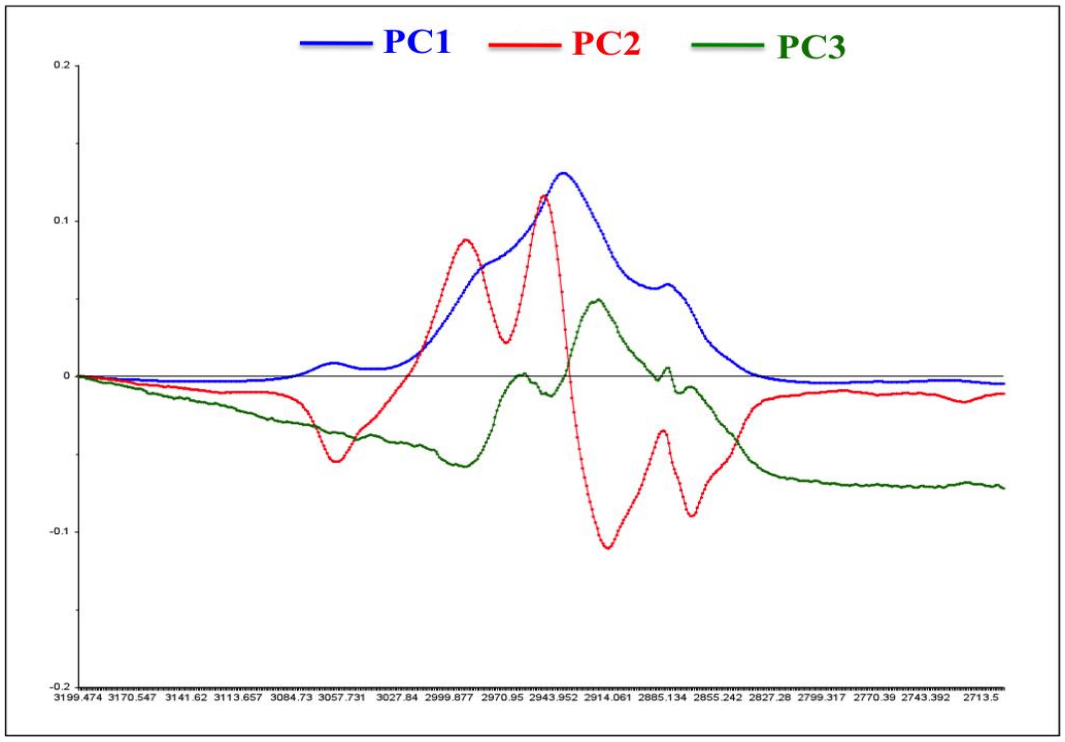

Figure 85: A loading plot of principal component analysis (PCA) for lipid region representing first (blue), second (red) and third (green) principal components (PC).

6.3.3 Amide I

PCA was performed over amide I region at 1750 to 1500 cm^{-1}. The first score plot has shown overall distribution of four subtypes clusters. PC-2 describes a variation that is considerably different between luminal A and luminal B. PC-3 has seen almost four subtype cluster distribution both negative and positive side of the axis. The second score plot has also shown pretty different to first plot. PC-1 and PC-2 that are good for separating luminal A from luminal B, PC-3 doesn't really separate any subtype except may be triple negative from the rest and PC-4 certainly separates triple negative from the rest. Interestingly, PC-2 has pretty much separated luminal B from luminal A from luminal B. PC-3 doesn't really separate any subtype except may be triple negative from rest of the subtypes. PC-4 certainly separates triple negative from the rest of subtypes. A three dimensional PCA plot of amide I range has shown good cluster formation for luminal A, HER2 positive and triple negative subtypes and weak cluster formation in luminal B subtype. A 3D PCA plot has shown good separation in luminal A, HER2 positive and triple negative subtypes whereas luminal B is overlapped with luminal A using the principal components PC-1, PC-3 and PC-4 (Figures 87-90).

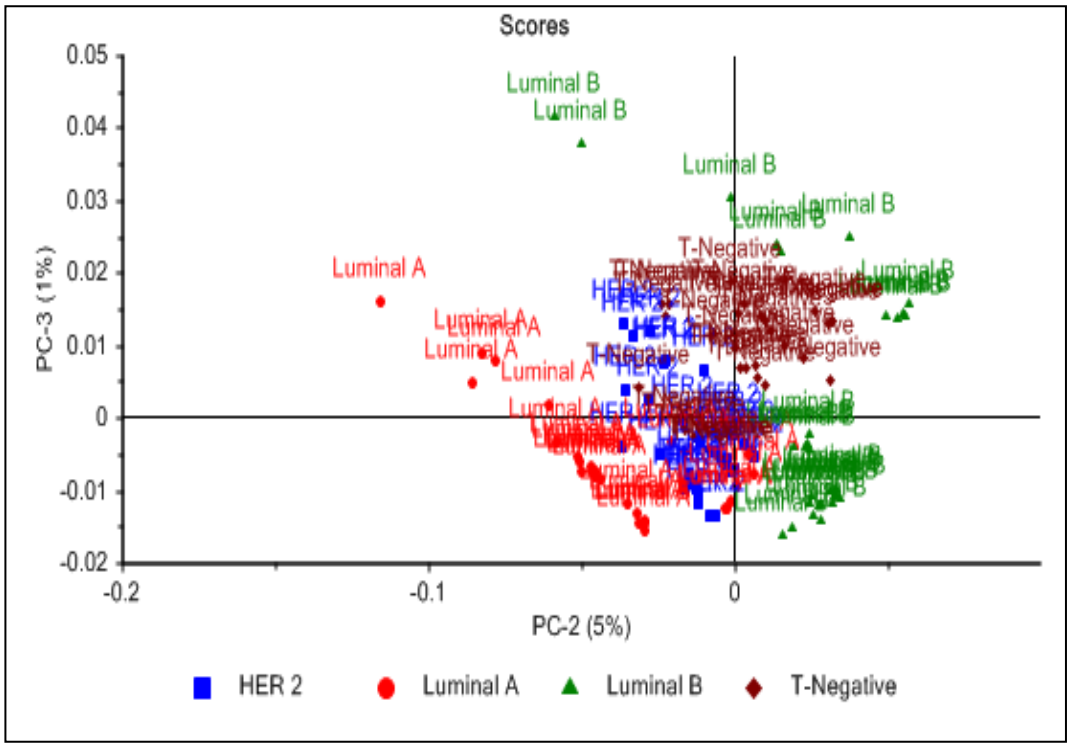

Figure 86: A two-dimensional principal component analysis (PCA) score plot of amide I region for luminal A (red), luminal B (green), HER2 (blue) and T-Negative (dark red) based on second and third principal components (PC).

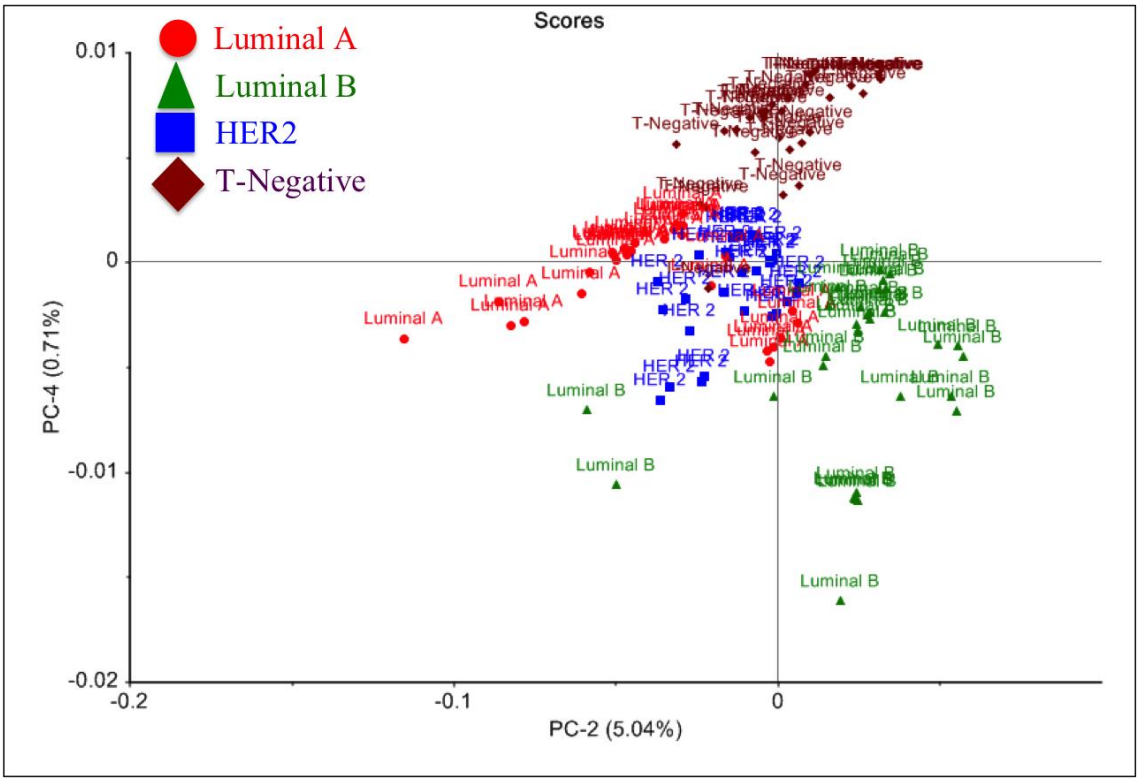

Figure 87: A two-dimensional principal component analysis (PCA) score plot of lipid region for luminal A (red), luminal B (green), HER2 (blue) and T-Negative (dark red) based on second and third principal components (PC).

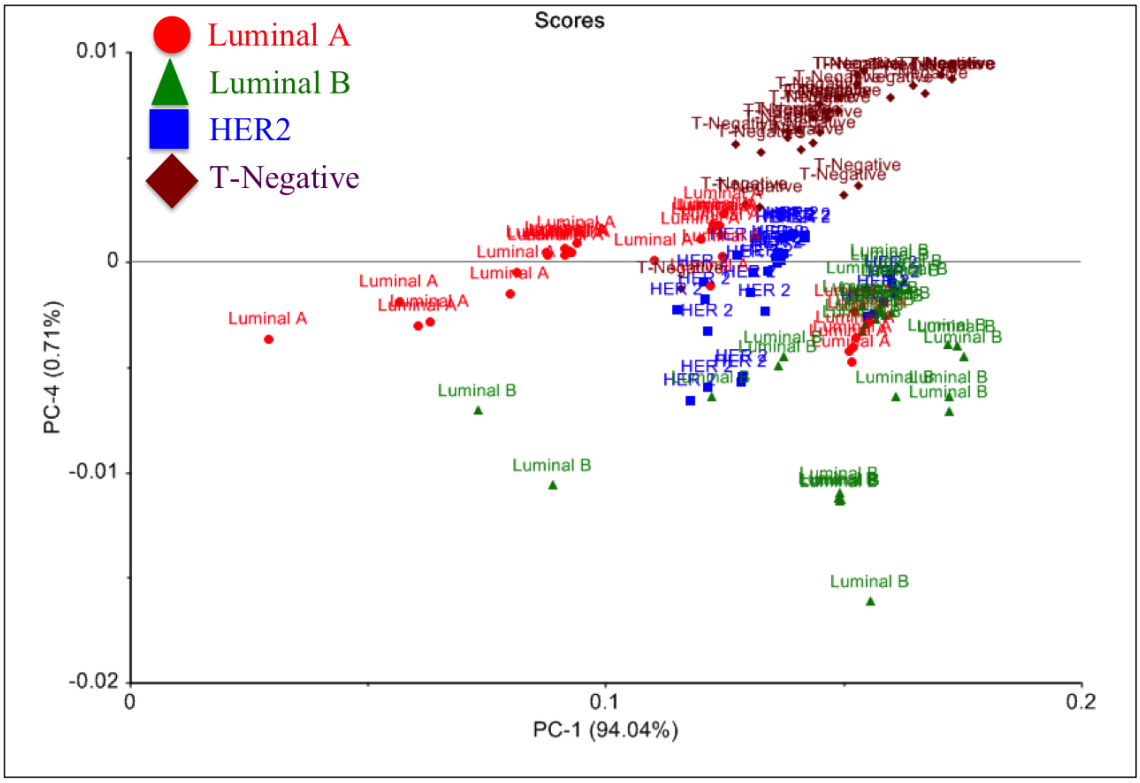

Figure 88: A two-dimensional principal component analysis (PCA) score plot of amide I region for luminal A (red), luminal B (green), HER2 (blue) and T-Negative (dark red) based on first and fourth principal components (PC).

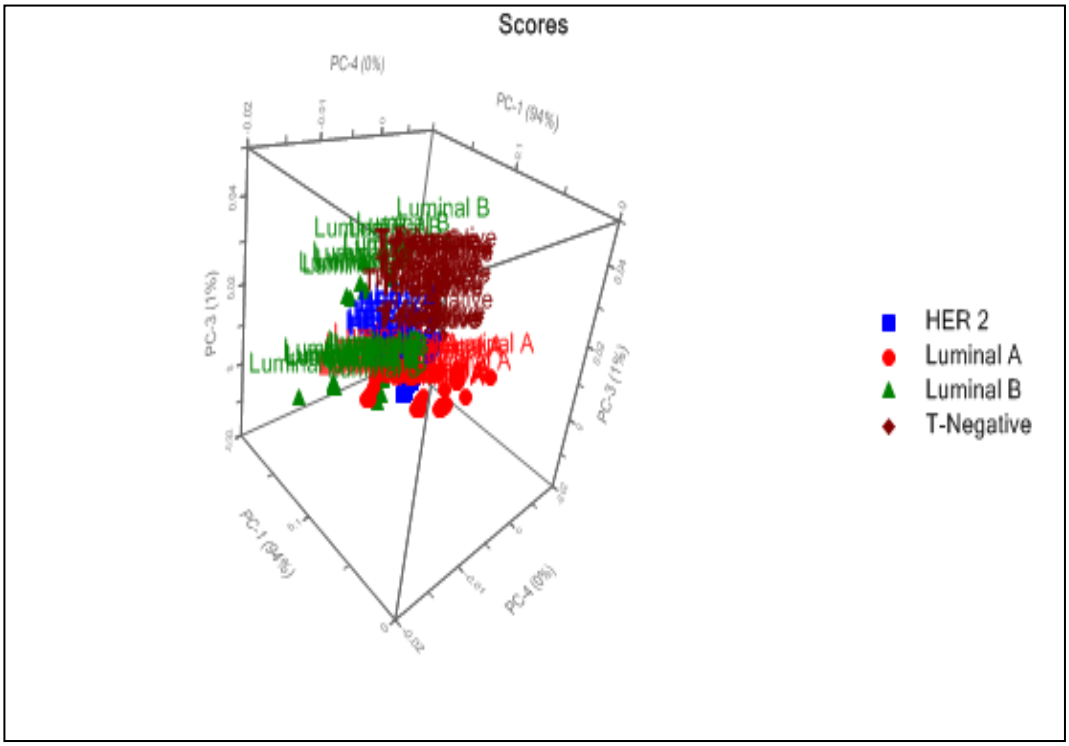

Figure 89: A three-dimensional principal component analysis (PCA) plot of amide I region for luminal A (red), luminal B (green), HER2 (blue) and T-Negative (dark red) based on first, third and fourth principal components (PC).

PCA loading plot analysis of amide I region has described biochemical differences among four subtypes. Firstly, PC-1 loading plot (blue) has sensitive in protein conformation of different subtypes. Every PC of this region is sensitive to protein conformation. Collagen at 1667 cm^{-1} (either unordered or β sheets) is positively separated triple negative and majority of HER2 subtypes. PC-3 loading plot (red) has pretty much sensitive in 1654 cm^{-1} (α-helical confirmation of collagen), 1614 cm^{-1} (tyrosine), 1604 cm^{-1} (phenylalanine and tyrosine) and 1576 cm^{-1} (guanine) and 1634 cm^{-1} (amide I). PC-4 loading plot (green) has shown that majority of triple negative, luminal A and HER2 subtypes are positively separated at 1671 (β sheet structures of protein conformation) and 1575 (ring breathing mode of DNA bases), and some of the luminal Band HER2 and luminal A are negatively separated at 1601 (phenylalanine) and 1547 (proline). Collagen was important for separating luminal A and luminal B from rest of the subtypes but nucleic acids and amino acids were more important in separating triple negatives from rest of the subtypes. In a nut shell, PCA of amide I region is sensitive to protein conformation of different subtypes (Figure 91).

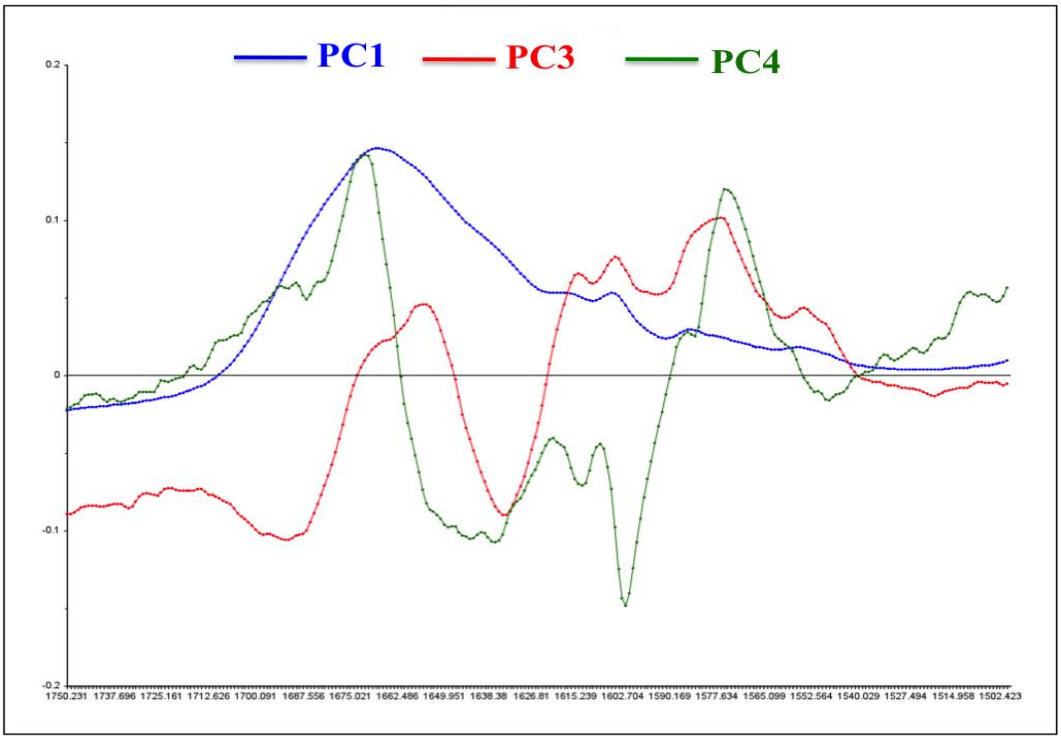

Figure 90: A loading plot of principal component analysis (PCA) for amide I region representing first (blue), third (red) and fourth (green) principal components (PC).

6.3.4 Amide III

PCA was employed on amide III region at 1400- 1200 cm^{-1}. In the first score plot, PC-2 has pretty much separated luminal A from triple negative subtype, whereas, HER2 positive and luminal B are widely distributed. All four subtypes were randomly distributed along the PC-3 axis both positive and negative side. A three dimensional PCA plot of Amide III range has shown better cluster formation in triple negative subtypes and weak cluster formation in luminal A, B and HER2 positive subtypes. A 3D PCA plot has shown a weak separation in luminal A whereas luminal B, HER2 positive and triple negative subtypes are overlapped using the principal components PC-1, PC-2 and PC-3 (Figures 92-95).

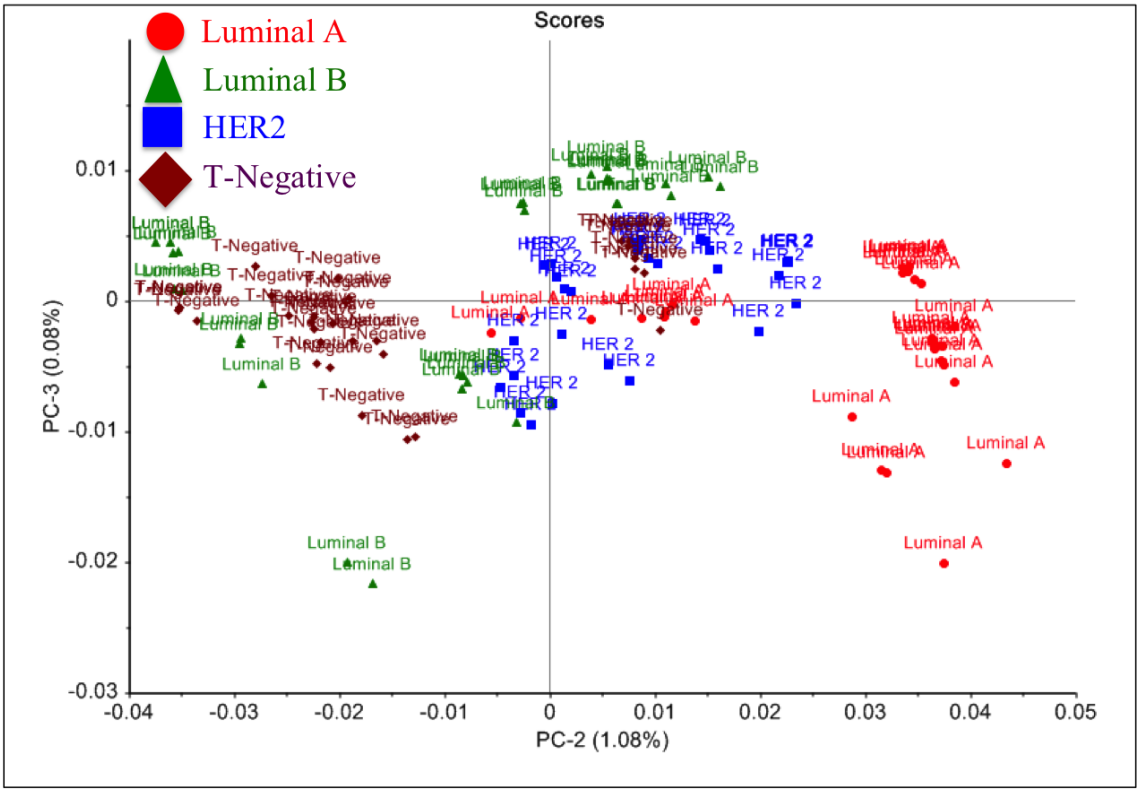

Figure 91: A two-dimensional principal component analysis (PCA) score plot of amide III region for luminal A (red), luminal B (green), HER2 positive (blue) and T-Negative (dark red) based on first and fourth principal components (PC).

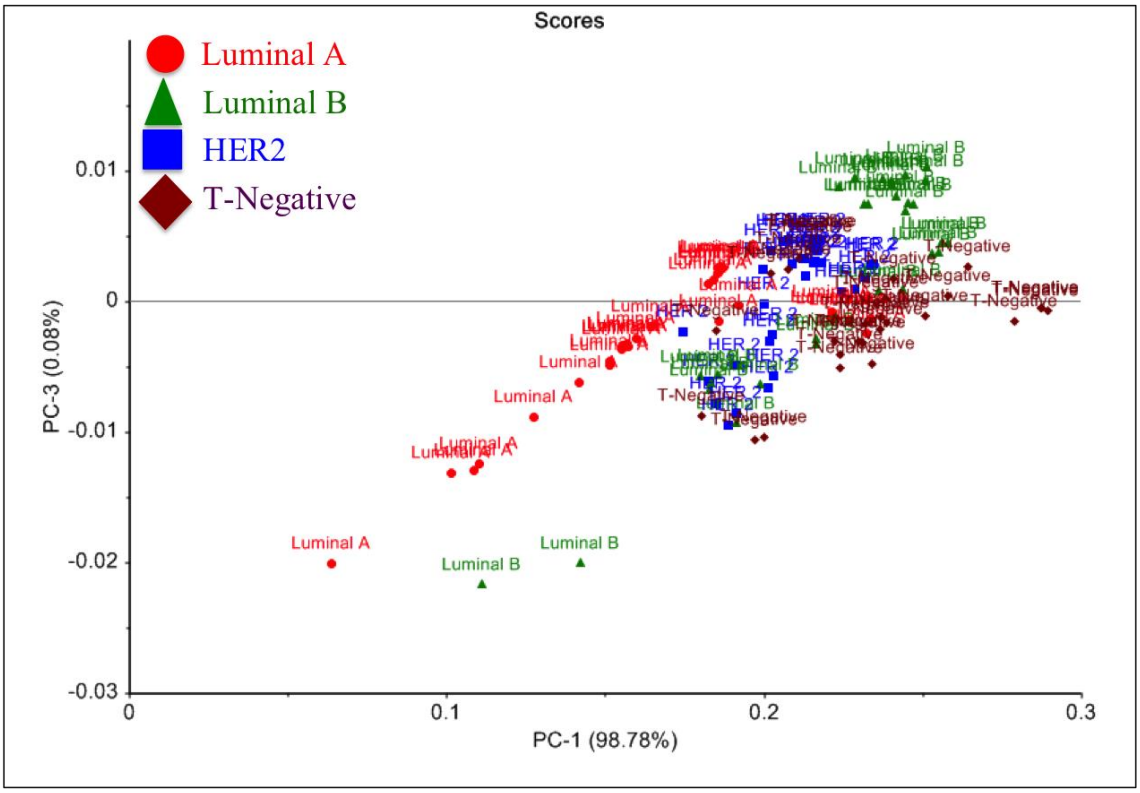

Figure 92: A two-dimensional principal component analysis (PCA) score plot of amide III region for luminal A (red), luminal B (green), HER2 positive (blue) and T-Negative (dark red) based on first and third principal components (PC).

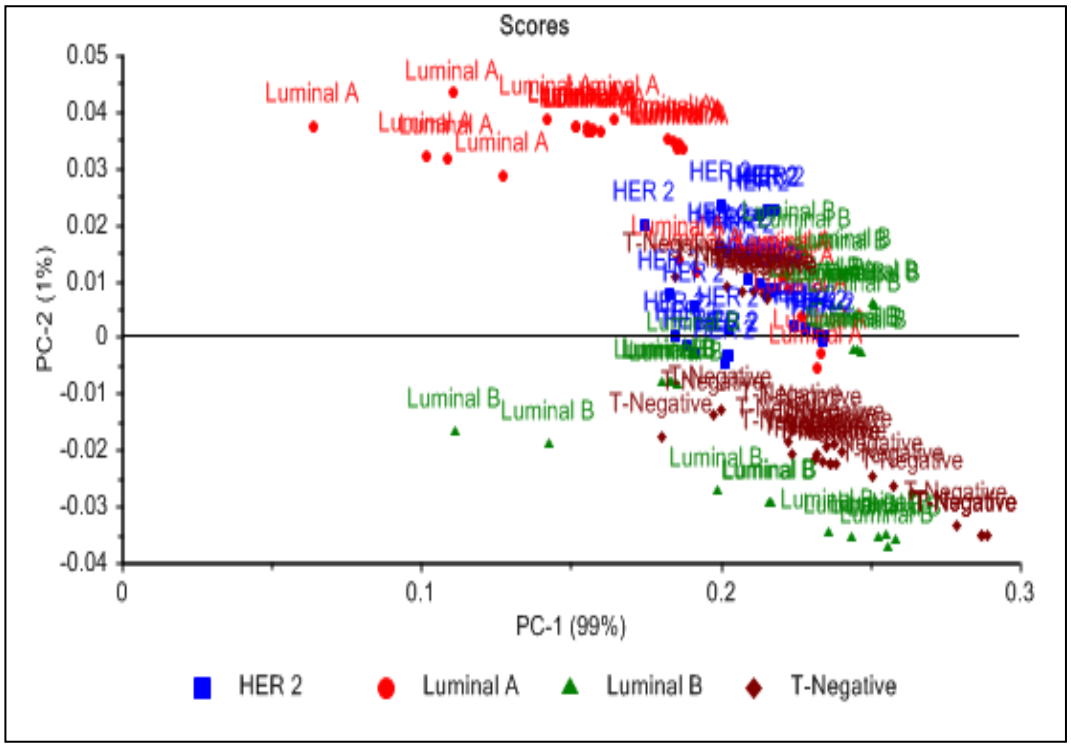

Figure 93: A two-dimensional principal component analysis (PCA) score plot of amide III region for luminal A (red), luminal B (green), HER2 positive (blue) and T-Negative (dark red) based on first and second principal components (PC).

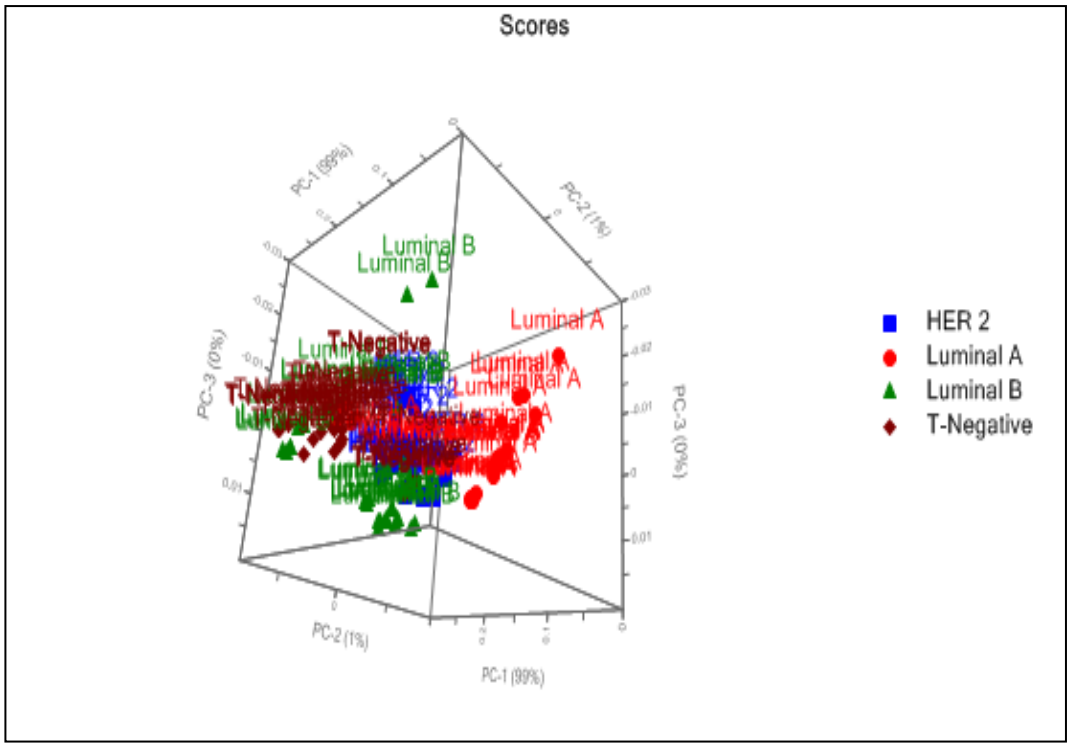

Figure 94: A three-dimensional principal component analysis (PCA) plot of amide III region for luminal A (red), luminal B (green), HER2 positive (blue) and T-Negative (dark red) based on first, second and third principal components (PC).

PCA loading plot analysis of Amide III region has described biochemical differences among four subtypes. Firstly, PC-2 loading plot (red) describes a variation that is sensitive to protein conformation of different subtypes. It has positively separated majority of luminal B, HER2 and triple negative subtypes at 1269 and 1244 and negatively at 1336 cm^{-1} (purine bases and collagen assignment). PC-3 loading plot (green) has pretty much sensitive in 1377 cm^{-1} (lipid assignments), 1339 cm^{-1} (C-C stretch of phenylalanine), 1279 cm^{-1} (amide III of α helix) and 1277 cm^{-1} (amide III of α helix) (Figure 96).

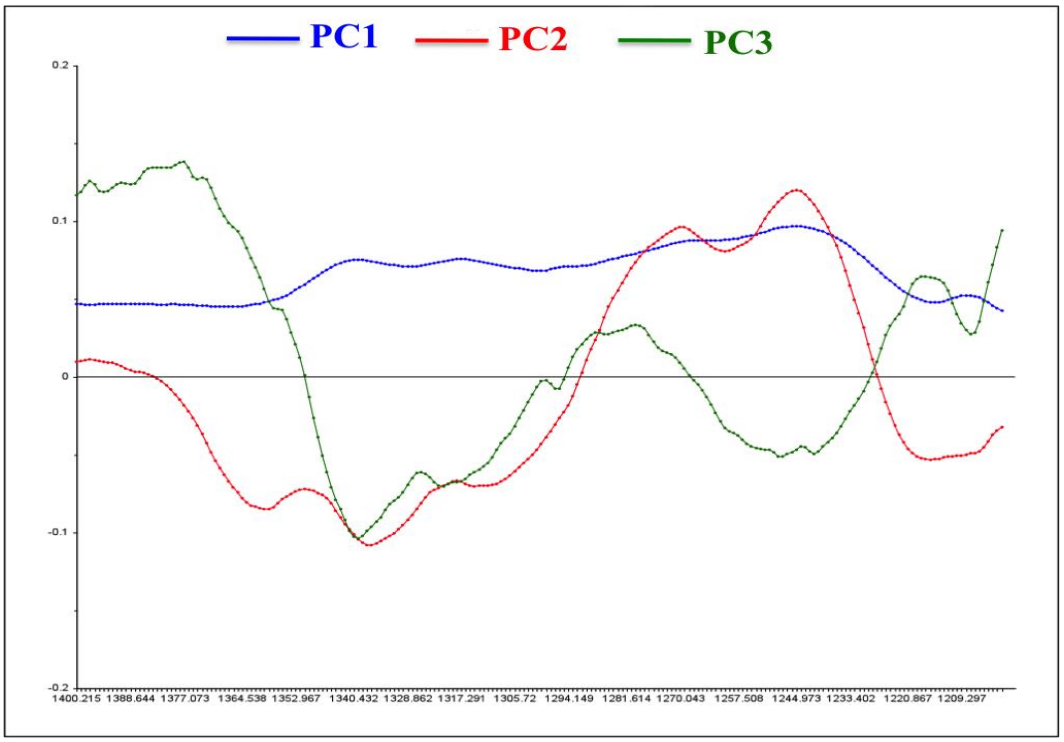

Figure 95: A loading plot of principal component analysis (PCA) for amide III region representing first (blue), second (red) and third (green) principal components (PC).

6.3.5 Nucleic Acid region

PCA was performed on nucleic acid region at 980 – 610 cm^{-1}. The first and third score plot has shown clear separation between triple negative and luminal A subtypes, whereas, HER 2 and luminal B were distributed on both sides of PC-2 axis. Interestingly, each subtype can be observed as almost subsets of clusters except few

spectra of luminal A and luminal B. The second score plot has shown indiscriminate separation of four subtypes across the PC-3 axis. A three dimensional PCA plot of nucleic acid range has shown weak two-subset cluster formation in HER2, three-subset cluster formation in triple negative subtype and weak cluster formation in luminal A and B subtype. A 3D PCA plot has shown good separation in luminal A subtype, whereas, luminal B, HER2 and triple negative subtypes are overlapped using the principal components PC-1, PC-2 and PC-3 (Figures 97-100).

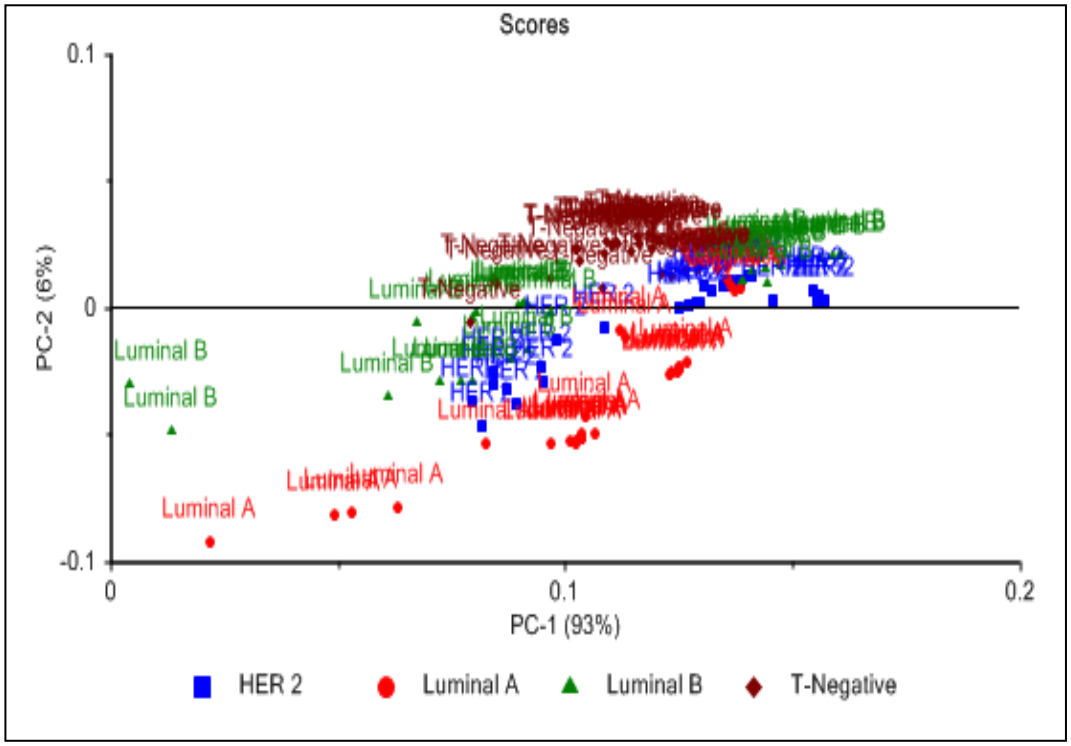

Figure 96: A two-dimensional principal component analysis (PCA) score plot of nucleic acid region for luminal A (red), luminal B (green), HER2 (blue) and T-Negative (dark red) based on first and second principal components (PC).

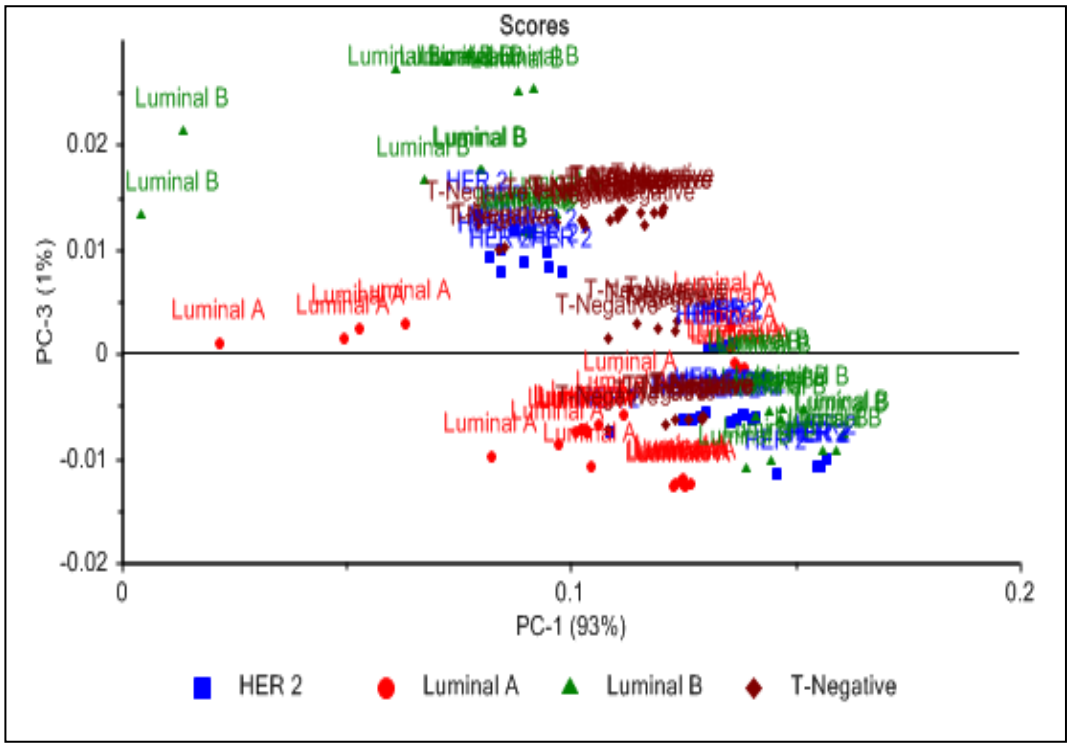

Figure 97: A two-dimensional principal component analysis (PCA) score plot of nucleic acid region for luminal A (red), luminal B (green), HER2 (blue) and T-Negative (dark red) based on first and third principal components (PC).

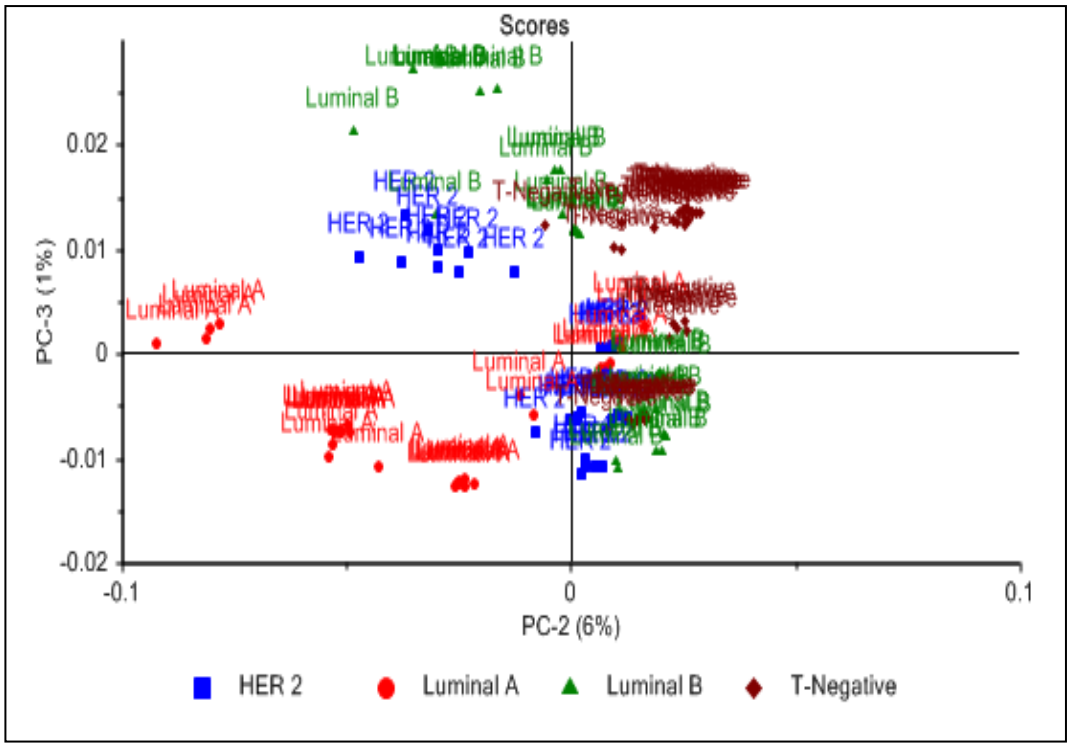

Figure 98: A two-dimensional principal component analysis (PCA) score plot of nucleic acid region for luminal A (red), luminal B (green), HER2 (blue) and T-Negative (dark red) based on second and third principal components (PC).

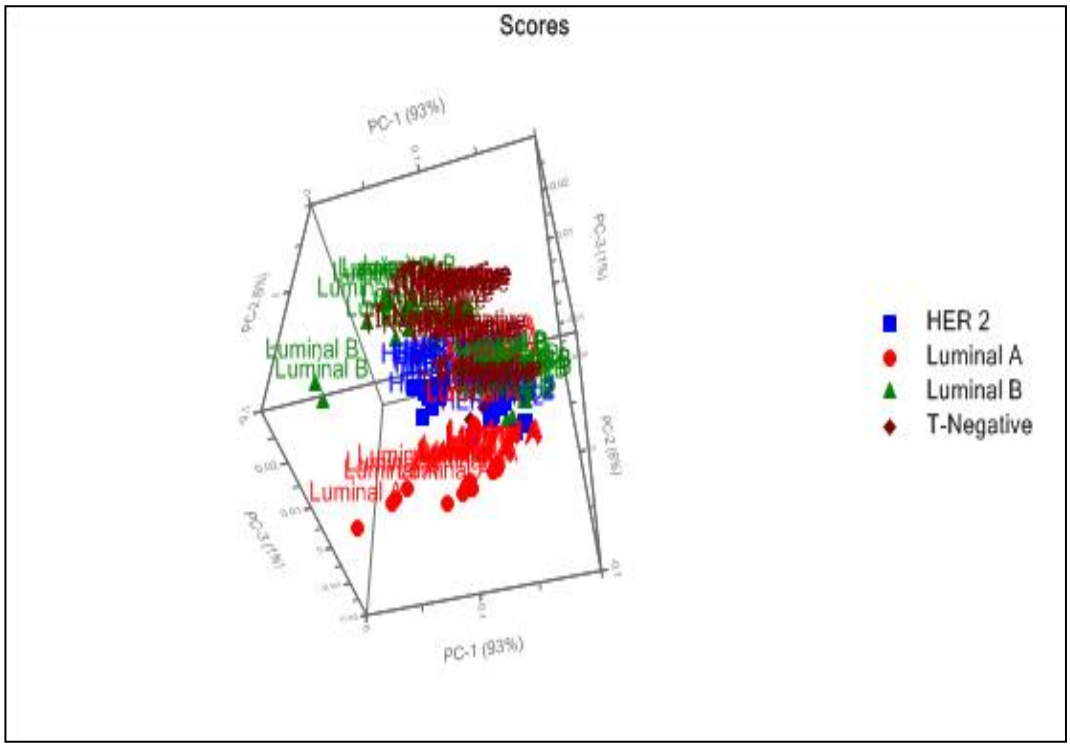

Figure 99: A three-dimensional principal component analysis (PCA) plot of nucleic acid region for luminal A (red), luminal B (green), HER2 (blue) and T-Negative (dark red) based on first, second and third principal components (PC).

PCA loading plot analysis of nucleic acid region has described biochemical differences among four subtypes. Firstly, PC-1 loading plot (blue) is sensitive in 938, 921, 877 and 855 cm^{-1}regions. PC2 loading plot (red) is pretty much sensitive in 939 and 827 cm^{-1}. PC-3 loading plot (green) has shown positive separation at 960 cm^{-1} (phosphate vibrations), 903 cm^{-1} (C-C skeletal stretchings), 829 cm^{-1} (O-P-O stretching of DNA and RNA), 783 cm^{-1} (thymine, uracil and cytosine), 758 cm^{-1} (tryptophan and ethanolamine of phosphatidylethanolamine), 643 cm^{-1} (C-C twisting mode of tyrosine), and 621 cm^{-1} (disulphide bridges of cysteine content). It has negatively separated for 920 cm^{-1} (proline/glucose/lactic acid), 870 cm^{-1} (proline/hydroxyproline/valine and polysaccharides), 814 cm^{-1} (proline/hydroxyproline/tyrosine/ phosphodiester of RNA), and 770 cm^{-1} (sugar backbones or cyclic ring of nucleic acid bases) (Figure 101).

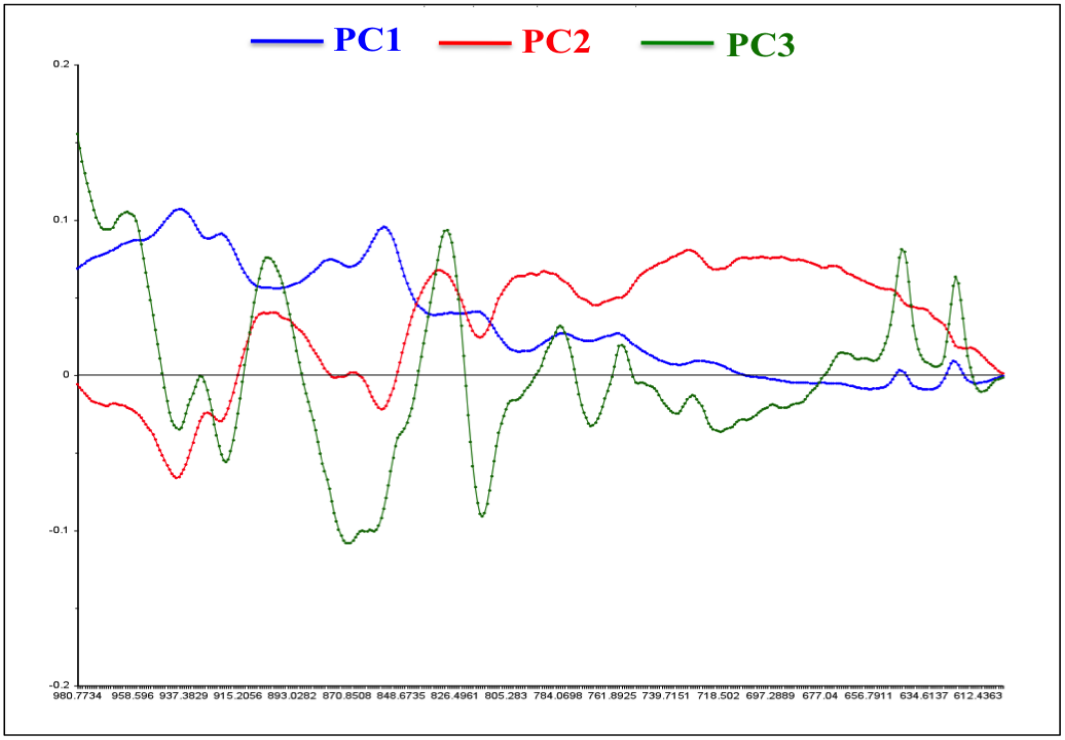

Figure 100: A loading plot of principal component analysis (PCA) for nucleic acid region representing first (blue), second (red) and third (green) principal components (PC).

6.3.6 Cluster Analysis

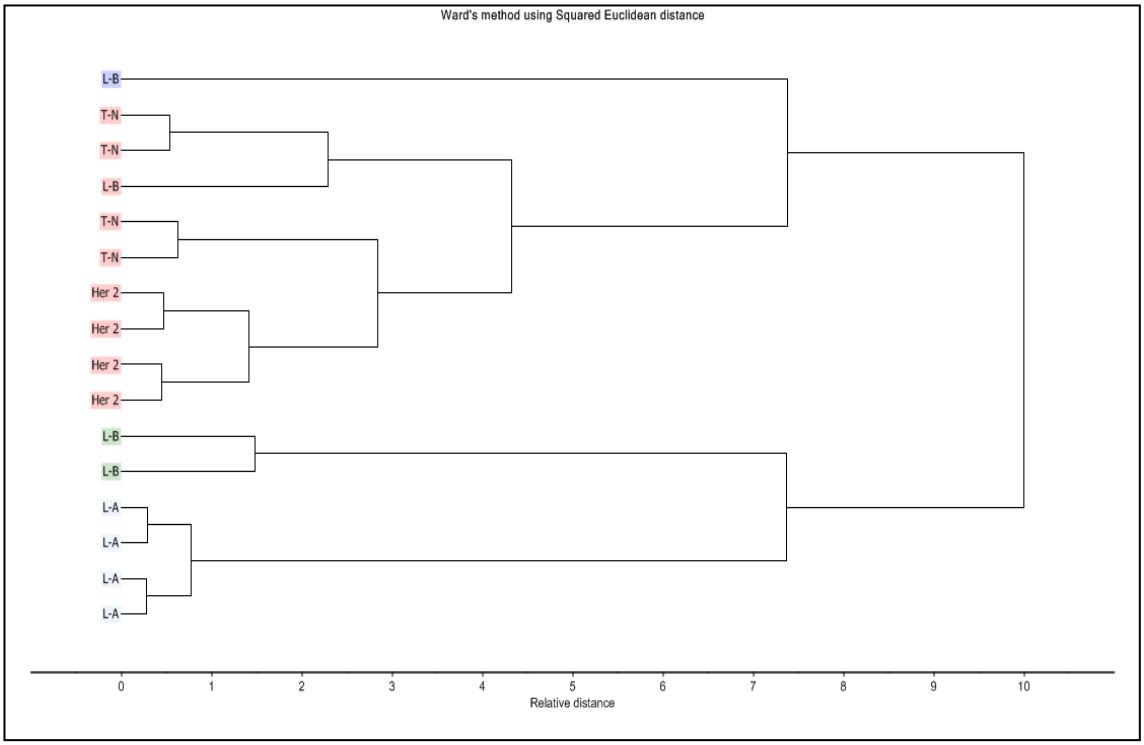

Figure 101: Dendrogram of Cluster Analysis results for luminal A (L-A), luminal B (L-B), HER2 (Her 2) and T-Negative (T-N) subtype (four sets of each subtype) samples. CA was performed over the full spectral range using Wards method's Squared Euclidean distance

Figure 102 has shown the dendrogram of luminal A (L-A), luminal B (L-B), HER2 (Her 2) and T-Negative (T-N) subtype (four sets of each subtype) samples over full spectral range (3200-400 cm^{-1}). Four main clusters were formed. Some of the luminal B spectra were mixed with triple negative and HER2 subtypes. Each region has formed pretty much good cluster and distance measurements has suggested that biochemical snapshot of full spectral region has shown more similarity between luminal A and subset of luminal B, and HER2 and triple negative subtype. One subset of luminal B has shown similarity with triple negative subtype. Distance measurements have shown luminal A and B chemically more distant from HER2 and triple negative subtypes.

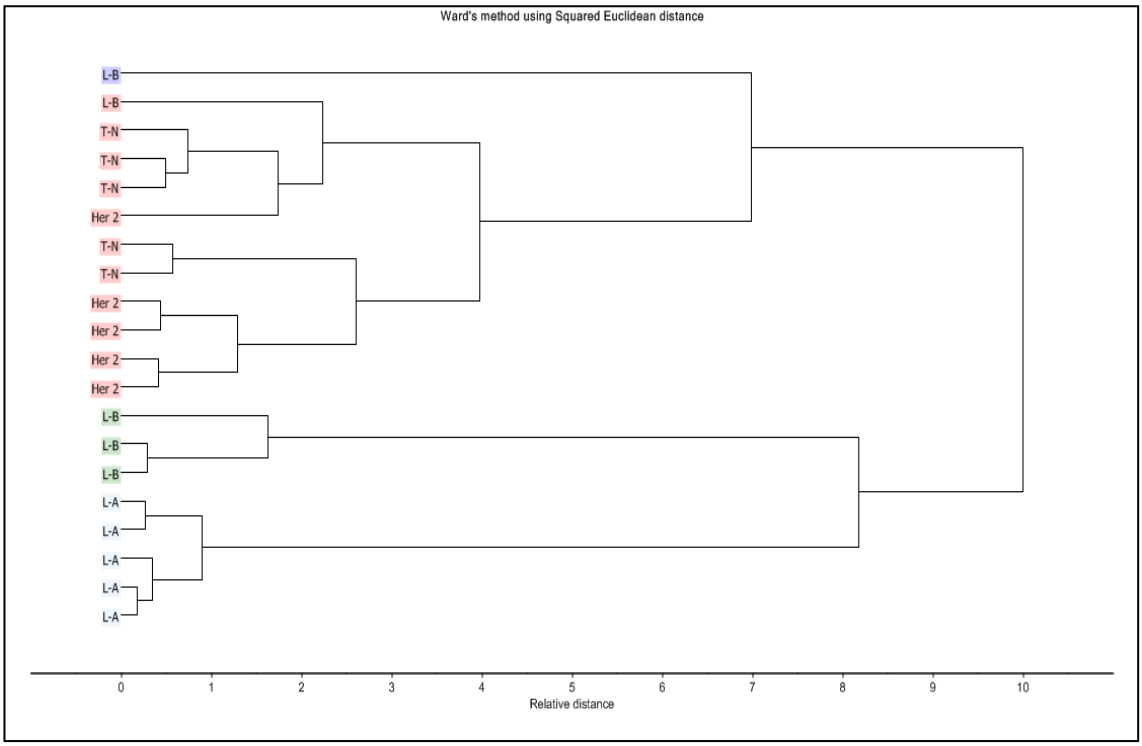

Figure 102: Dendrogram of Cluster Analysis results for luminal A (L-A), luminal B (L-B), HER2 (Her 2) and T-Negative (T-N) subtype (four sets of each subtype) samples. CA was performed over the full spectral range using Wards method's using Squared Euclidean distance.

Figure 103 has shown the dendrogram of luminal A (L-A), luminal B (L-B), HER2 (Her 2) and T-Negative (T-N) subtype (five sets of each subtype) samples over full spectral range (3200-400 cm^{-1}). Four main clusters were formed. This was done to check luminal B chemical nature. Each region has formed pretty much good cluster and distance measurements has suggested that biochemical snapshot of full spectral region has shown more similarity between luminal A and subset of luminal B, and HER2 and triple negative subtype. One subset of luminal B has shown similarity with triple negative subtype. Distance measurements have shown luminal A and B chemically more distant from HER2 and triple negative subtypes.

6.3.7 Linear Discrimination Analysis (LDA)

A total 120 spectra (30 spectra each subtype) was selected and processed using baseline correction and normalization for predictive classification with LDA. LDA models were

setup over the full spectral range 3200-600 cm^{-1}. Five samples from each group were left out at each pass until a total number of 30 spectra from each group were tested against the training group. LDA fusion matrix of four subtypes of TMA breast biopsies were mentioned in Table 17.

Luminal A subtype was predicted with a sensitivity of 70% where 8 spectra were predicted as Luminal B and 1 spectra predicted as HER2 positive subtype. None of the Luminal A subtype spectra were misclassified in the triple negative subtype. Luminal B subtype was predicted with sensitivity of 100% where none of the luminal B spectra were predicted wrongly. HER2 positive subtype was predicted with sensitivity of 90% with 3 spectra were misclassified as luminal B. Triple negative subtype was predicted with sensitivity of 96.67% where only one spectrum was predicted as luminal B.

Table 17: LDA fusion matrix of Luminal A, B, HER2 positive and triple negative subtypes

	Actual	HER2	Luminal A	Luminal B	Triple Negative
Predicted					
HER2		27	1	0	0
Luminal A		0	21	0	0
Luminal B		3	8	30	1
Triple Negative		0	0	0	29
Total		30	30	30	30

6.4 TMA Biopsies - Discussion

Breast Cancer diagnosis takes two steps, namely screening approaches followed by biopsy procedures. Screening methods such as mammography, ultrasound and magnetic resonance imaging, and biopsy procedures such as fine needle, core needle and surgical biopsies are in lead. A combination of histopathology, immunohistochemical methods and gene profile approaches are considered as gold standard methods in breast cancer subtype identification. Diagnosis approaches have limitations such as surgical inventions; false positives, time delays, pain and trauma of patient and risk of infections have allowed researchers to explore new non-invasive, reagent-free and less painful approaches (Haka et al., 2005). Raman spectroscopy, a vibrational spectroscopic approach, not only provides real time biochemical fingerprint of tissues but also understanding the disease as it progresses. Raman spectra can acquire in less time and spectroscopic approach can also eliminates intraobserver and inerobserver variability that commonly related with histopathological diagnosis (Swain and Stevens, 2007).

In our study, for the first time, a combination of Raman and multivariate approach have been applied on different breast cancer biopsies to identify chemical changes associated with different subtypes of breast cancer. Characteristic Raman markers were used to differentiate luminal A, B, HER2 positive and triple negative subtypes of breast cancer and in addition intra and inter biochemical changes of biopsies using different spectral features, was also identified.

Breast Cancer biopsies which we were analysed in our study were mainly ductal carcinoma No Special Type (NST). Multiple genetic mutations and protein dysfunction are the main cause of breast cancer (Rakha et al., 2008). It is a well-known fact that cancer tissues have higher cell proliferation and metabolic activity that result in changes of concentration and oxidation states of different chemical species. The major biological activities observed in cancer cells are loss of differentiation, nuclear enlargement

especially increasing in genetic material (Kang, 2002). Spectral changes were identified especially at lipids, amide I and III regions among luminal A, B, HER2 and triple negatives subtypes. Furthermore, subtypes were also classified using supervised and unsupervised algorithms. Significant spectral variations in terms of peak shifts, shapes and intensities in four subtypes of breast biopsies were observed. Raman studies have also identified chemical differences in individual biopsies as well.

6.4.1 High-wavenumber region – Lipids prospective view

High-wavenumber region of Raman spectra represents stretching vibrations of lipids and proteins. Biochemical fingerprint of this region deals with metabolic status of fatty acids and hydrophobic interactions. The major molecular vibrations are symmetric and asymmetric stretches of CH_3, CH_2 and -CH=CH-. These vibrations represent lipid metabolism and are helpful in assessing lipid peroxidation in biological systems (Talari et al., 2015a). Deregulated lipid biosynthesis is a characteristic feature of cancerous tissues. Lipid and protein peaks are expressed in triple negative and luminal A at 2938 and 2937 cm^{-1}, respectively, whereas, in HER2 positive and luminal B is expressed at 2936 cm^{-1}. Changes in peak intensities and shifts are a result of differences in structural conformations and concentration of lipids in different subtypes of cancerous biopsies. Lipid synthesis plays a vital role in cancer development and progression (Hirsch et al., 2010). Lipid requirements of mammalian tissues are fulfilled by uptake of fatty acids and lipoproteins from the blood stream. Fatty acids and cholesterol production is restricted to a subset of tissues such as lactating breast, liver and adipose tissue. Nevertheless, reactivation of lipid synthesis has been reported in tumorous tissues (Ozek et al., 2010a).

Initial studies demonstrated that cancerous tissues use *de novo* lipogenesis to produce lipids, including fatty acids and phospholipids (Swinnen et al., 2002). Type of lipid content in breast tumour tissue is related to liver tissue in terms of fatty acid levels.

Although tumour tissue is capable of uptaking lipids from the tissue environment, initial studies have concluded that *de novo* lipogenesis plays vital role in rapid proliferation of cancer cells (Milgraum et al., 1997). In recent years, combination of gene expression and genome scale metabolic studies of breast cancer has revealed two important concepts of lipid synthesis in cancerous tissues (Hilvo et al., 2011). Firstly, biosynthesis of fatty acids is a characteristic feature of early stages of tumour formation because these play important role in rapid proliferation of cancer cell. Secondly, advanced stages of cancer have shown antioxidant nature for detoxification of reactive oxygen species (ROS)(Germain et al., 1998).

Raman peak at 2938 cm^{-1} in triple negative subtype is located at 2937 and 2936 cm^{-1} in luminal A and luminal B & HER2 subtypes respectively. Peak position differences in luminal A and triple negative have shown variations in stretching vibrations of lipids in CH and CH_3. Peak shifts have differentiated luminal A and triple negative from HER2 and luminal B. Raman band at 2881 cm^{-1}, which are assignments of CH_2 symmetric and asymmetric stretch of lipids are found in luminal A, HER2 and triple negative subtypes. CH_2 stretching lipids of high-wave number of luminal B was reported at 2880 cm^{-1}. Lipid biosynthesis is correlated with saturated lipid content in cells. Lipid peroxidation is directly proportional to lipid degradation and synthesis of lipids is decreased in low number saturated fatty acids (Ozek et al., 2010a). Low amount of saturated lipids also represents slow growth behaviour of cell. Unsaturated fatty acid content is directly proportional to the cancer progress. High-unsaturated fatty acid content is present in metastatic stages (Bartsch et al., 1999). This Raman approach has even identified stromal lipid peak intensity differences in individual TMA biopsies (Figures 70-72). These lipid intensity differences and structural information were helpful in understanding these four subtypes of breast cancer as well classification.

6.4.2 High-wavenumber region – Proteins prospective view

High-wavenumber region (2940 – 2930 cm^{-1}) plays an important role in understanding cancer at biochemical molecular level. Biochemical fingerprint of this region deals with denaturation and conformational changes of proteins. Normal breast tissue is dominated by lipid predominant peaks because adipose tissue is distributed around lobules and ducts. Cell and nuclear membranes also contribute to lipid peaks due to their triglyceride composition (Shafer-Peltier et al., 2002). The cancerous breast tissue is subjugated by protein content and high-wavenumber is useful to study protein vibrations in terms of amino acid nature. Increase in protein content and its contribution towards carcinogenesis is evident in breast cancer and nearly 30% cases are having amplified gene as protein product in HER2 positive subtypes (Olayioye, 2001). The main difficulty in this region is the interpretation of Raman bands. The major problem is that most of the lipid and fatty acid bands overlap with protein bands. Overlap is generally between CH of lipids with CH_3, CH_2 and CH of amino acid side chains. Previous Raman spectroscopic studies on based breast cancer studies have reported that this region includes not only aliphatic and aromatic amino acids but also other amino acids such as histidine, threonine and proline (Surmacki et al., 2013).

Fingerprint region of all subtypes spectra represents a combination of protein, lipid and nucleic acid information. Majority of information comes from proteins distributed in three regions, namely amides I, II and III. Protein peak assignments for C=O stretchings, C-H bendings, CH_2 deformations and CH_2 twisting provide secondary structural information. Peak assignments for ring breathing modes of aromatic amino acids such as phenylalanine, tyrosine and tryptophan provide information on side chain vibrations of proteins. Nucleic acids represent molecular vibrations of DNA and RNA bases, and phosphate backbones. Purine bases such as adenine and guanine, pyrimidines bases such as thymine, cytosine and uracil, and phosphate PO_3^{2-} represent

breathing modes and asymmetric and symmetric phosphate stretchings. Lipids are also present in fingerprint region as CH_2 deformations and C=C stretching of lipids. In a nutshell, fingerprint region plays vital role in understanding chemical changes associated with different subtypes using building blocks of biological molecules of tissues (Talari et al., 2015a, Talari et al., 2015k).

Protein metabolism plays a key role in cancer biology in terms of cell differentiation and cell proliferation, hence, considered as potential Raman markers in disease identification. The second and third steps of central dogma, such as, transcription (DNA to mRNA synthesis) and translation (mRNA to protein synthesis) signatures were captured in fingerprint region through molecular vibrations (Thomas Jr, 1999). Aromatic amino acids such as tryptophan and tyrosine plays crucial role in various metabolic processes and are required for rapidly proliferating cancer cells. Previous studies have reported that high amount of tryptophan in cancerous tissues is also evident and tryptophan peak can be specifically observed in Raman spectrum at 1583 cm^{-1} (Takeuchi, 2003). Normal breast tissue has shown characteristic tryptophan between amide I and amide II and peak intensity is increased as cancer progresses in tissues. Peak intensity of tryptophan is increased in cancerous tissues. Tryptophan peak intensity differences and peak shifts were observed in four subtypes. Luminal B have shown tryptophan peak at 1583 cm^{-1}, whereas, luminal A, HER2 positive and triple negative subtypes have observed in peak shift. Peak intensity is similar between HER2 and luminal B subtypes whereas luminal A has shown low intensity and triple negative is further lowest. Significant tryptophan changes were observed and luminal B and HER2 have 10 fold increase compared to triple negative and 5 fold increase compared to luminal A. Protein content changes will be helpful in different subtypes of breast tissues. Normal breast tissue has shown less tryptophan content compared to cancerous tissues. Tryptophan Peak shift was observed in luminal B subtype and intensity

difference identified in luminal A and triple negative subtypes because they are lower in intensity. Recent studies have confirmed that tryptophan metabolic pathway has significant role in cancer progression. Indoleamine 2,3-dioxygenase (IDO), rate-limiting enzyme of tryptophan metabolic pathway, expression profile is associated with advancement of breast cancer. Raman can be helpful in identification of tryptophan signatures in different subtypes of breast cancer.

Proteins peaks, majority of them amide peaks, result from carbonyl stretchings from polypeptide backbones. Amide I represents protein secondary structural information and the mainly contributor to 80% C=O stretching vibrations. The information provided by Amide I peak is mainly dependent on their position and shape. Triple negative subtype has shown amide I peak with high intensity next to Luminal B subtype whereas Luminal A and HER2 subtypes have shown less amide I peak intensities (Figure73). The higher intensity of amide I represent higher amount of collagen. Ductal carcinoma in situ and fibroadenoma samples usually shows large number of cells compared to other lesions of the breast. Increasing cell nucleus size is one of the important factors in cancer. Pathologists consider higher nuclear to cytoplasm ratio is the best way to diagnose the disease. Amide I peak appeared to 1667 cm^{-1} in triple negative subtype and it represents β-sheets of collagen protein. The structural modes of malignant breast tissue proteins have shown this peak in previous studies. (Hanlon et al., 2000) Amide I vibrations of tumour proteins were shifted to 1666 cm-1 in HER2 subtype and to 1665 cm^{-1} at luminal A and B. SERS studies of saliva have also described this peak as C=O stretching vibrations, whereas, spectra of brain tumours have attributed this peak to C=C stretching vibrations (Farquharson et al., 2005). Lung cancer studies have shown that Raman peak at 1666 cm^{-1} was only present in cancerous tissue and it is assigned to collagen. In this study it was suggested that collagen content plays important role not only in breast cancer but also other cancers as well. Collagen content can be used as

quantitative Raman markers in disease identification especially in subtype classification (Kaminaka et al., 2001) and our studies also confirmed the importance of collagen contents both in identification of cancer and classification of subtypes.

6.4.3 Collagen

Collagen is a chief component of connective tissue made up of tropocollagen subunits. Each subunit is made up of three polypeptide chains of about 1000 amino acids and whole collagen is made up of at least 28 genetically distinct species. Each polypeptide chain represents 'Glycine – X – Y' repeat where X and Y represents proline and 4-hydroxyproline respectively. N-H group of Glycine amino acid forms hydrogen bond with C=O group of proline. In addition, water and hydroxyproline also form hydrogen bonds to stabilize the collagen network. These bonds in collagen help to maintain a triple helical conformation. Based on supra molecular structures, collagen is classified into different subgroups and are summarized in Table 18.

Table 18: Different types of collagen

S.No	Name of collagen	Type of collagen
1	Fibril-forming collagens	I, II, III, V, XI, XXIV and XXVII
2	Fibril-associated collagens with interrupted triple helices (FACITs)	IX, XII, XIV, XVI, XIX, XX, XXI, XXII and XXVI
3	Collagen with transmembrane domain	XIII, XVII, XXIII and XXV
4	Hexagonal network forming collagens	X
5	Basement membrane collagen	IV
6	Multiplexins	XV and XVIII
7	Beaded filaments-forming collagens	VI
8	Anchoring fibrils-forming collagens	VII

Due to desmoplastic reaction, stroma of neoplastic breast tissue starts to accumulate dense fibrous tissue possessing newly formed ECM components, and predominantly, collagen. Previous studies has established that invasion sites of ductal infiltrating carcinoma have shown accumulation of type I and III collagens whereas protein extraction studies have revealed that type V collagen increases by nearly 10 % in DIC compared to normal breast tissue. In recent years, gene expression studies have explored tumour associated collagen structures (TACS) related to breast cancer progression(Provenzano et al., 2008, Campagnola et al., 2001). Optical approaches such as Multiphoton and second harmonic generation microscopy show that TACS is important in cancer progression using mouse tumour explants and microarrays. TACS 1 is a limited density collagen and usually appears at small tumour foci. TACS 2 is tangentially oriented collagen towards the smooth boundary of tumour and TACS 3 is perpendicularly oriented towards irregular invasive tumour boundary. TACS 3 is described as a consistent and powerful marker in disease identification especially in triple negatives (Provenzano et al., 2006). This marker is going to be considered as an independent prognosis key irrespective of tumour grade, size and receptor status. In recent years, second harmonic generation (SHG) microscopy is used to quantify collagen amount and evaluated abnormal collagen fibrils to study malignancy in different stages of breast cancer. Apart from SHG microscopy, X-ray scattering approaches have also proved that collagen content and its structural associated changes differ in different stages of breast cancer and also predicted invasion directions in terms of cancer spreading (Conklin and Keely, 2012, Campagnola et al., 1999).

Multivariate approach (PCA) was applied on whole spectral range and much overlap was observed in luminal A, HER2 and triple negative subtypes. Luminal B subtype has separated very well using PC2 and it formed very good cluster, and this might be due to less intervariation in biopsies. Overlapping of three subtypes might indicate that

biopsies among these groups possess many similarities in whole spectral range and this lead us to go for a closer look to explore small spectral regions. Lipids, amide I, amide III and nucleic acid regions were explored using PCA. This is the first time, breast cancer biopsy data was explored based on subtype classification. Previous studies have identified malignant tissues from normal and benign in various cancers such as lung cancer (Huang et al., 2003, Kaminaka et al., 2001), skin cancer (Cheng et al., 2005, de Oliveira et al., 2010), brain cancer (Kirsch et al., 2010), gastro intestinal cancer (Tan et al., 2003, Shetty et al., 2006) and oral cancer (Malini et al., 2006b). Recent studies involving combination of Raman and multivariate approaches have classified normal from malignant tissue in lung cancer (Li et al., 2012)

Nuclear magnetic resonance studies have revealed that phospholipids and mitochondrial metabolism were affected in breast cancer cells and phosphocholine content was increased nearly 20 fold in primary breast cancer cell lines and nearly 30 fold in metastatic breast cell lines. In addition to that, genetic alterations increase choline transport and enhance the synthesis of phosphocholine and betaine and decline the choline derived ether lipids in breast cancer cells. Decreased levels of ATP, phosphocreatine and influx of pyruvate have clearly shown that mitochondrial metabolism was damaged in breast cancer cells. Genetic alterations increased choline transport and enhanced the synthesis of phosphocholine and betaine and decline the choline derived ether lipids in breast cancer cells.

Initially, Raman studies identified changes in fatty acid content and β carotene between normal cancerous biopsies. Later studies proved that protein changes in normal and ductal carcinoma were observed as peak shifts. Normal breast tissue has shown C-H protein vibration at 1439 cm^{-1}, whereas, ductal carcinoma NST samples have shown this peak at 1449 cm^{-1}. Here, all four subtypes have expressed this peak at 1449 cm^{-1} and it

clearly shows that all of biopsies were in malignant state. This wavenumber shift is probably indicative of large lipid accumulation in the malignant state compared to normal breast.

The spectral region between 1400-800 cm^{-1} has represented six different amino acids such as glycine, alanine, proline, tyrosine, valine and phenylalanine. These amino acids are basic backbones of primary structure of proteins. The principal component of breast tissue is collagen and it is mainly composed of glycine, valine, proline and phenylalanine. It is clear that the intensity of these amino acids is increased in Triple negative subtype compared to the other subtypes. Alanine, glycine, proline and tyrosine peak intensities were observed high in triple negative subtypes and low in Luminal B subtypes. The order in which these amino acid peak intensities decrease is triple negative to luminal A to HER2 positive to luminal B. Collagen peak intensity variation is observed in amide I and amide III regions within individual biopsies (Figure 76).

The intensity of collagen among different breast cancer subtype was much varied; this difference was due to various concentration of collagen among the investigated subtypes. Our Raman spectral data has shown that invasive triple negative cancer has the highest intensity of collagen among all other subtypes. Raman spectra also reported high collagen associated amino acids in fibroadenoma and invasive lobular carcinomas (Haka et al., 2005). Based on this relative higher amount of collagen content in triple negative subtype compared to other subtypes have indicated that this is one of the most pathogenic subtype, Raman could be in useful in identification of new chemical pathogenic marker in terms of collagen content.

Fingerprint range of Raman spectra not only provides major protein information at amide I, II and III region but also provides biochemical information regarding amino acids sidechains, functional groups of aromatic amino acids, such as, phenylalanine,

tyrosine and tryptophan, functional groups of carotenoids, acidic residues and sulphur containing residues from cysteine and methionine. Amide I region of triple negative subtype has shown β- sheets modes and these might be coming from structural modes of tumour. A majority of carbonyl stretches and a minority of C=C stretchings vibrations might be a major source for this peak.

Amide I peak is shifted gradually towards lower wavenumber in HER2 and luminal subtypes. Maiti et al reported that amide I peak at 1667 cm^{-1} has contributed from β sheet structures. A sharp peak also represents increased hydrogen bonding in β sheets of proteins (Figure 97-99). Three shoulder peaks were observed in amide I region namely 1621, 1605 and 1584 cm^{-1}. The first two shoulder peaks attributed from amino acids and, later from aminoacids and nucleic acids. Raman peak at 1605 cm^{-1} in triple negatives, 1604 cm^{-1} at HER2 and luminal subtypes have attributed to aromatic aminoacids such as phenylalanine and tyrosine. CARS based colorectal tissue studies have suggested that Raman peak at 1584 cm^{-1} is attributed to nucleic acids, whereas, RS based nasopharyngeal tissue studies attributed to C=C bending mode of phenylalanine (Krafft et al., 2009a, Lau et al., 2003a). Previous studies have reported that these peak intensities are decreased in cancerous tissue compared to normal. HER2 and luminal B subtypes have shown increased intensities compared to luminal A and triple negative subtypes.

Earlier breast cancer studies have reported significant and reproducible peak shifts in amide II region. Raman Spectra of normal breast tissue have expressed amide II peak at 1439 cm^{-1}, whereas, invasive ductal carcinoma biopsies at 1449 cm^{-1}. All four subtypes have expressed this peak at 1449 cm^{-1} with peak intensity differences. This is due to CH_2 deformations of fatty acid and proteins (Figure 75).

Fatty acid composition varies in breast cancerous tissues. Triglyceride content is reduced nearly more than half in cancerous tissues compared to normal, whereas, phospholipid content is increased four folds in cancerous biopsies. Overall reduced lipid content in cancerous biopsies was reported. Triple negative and luminal B subtypes have shown increased peak intensities compared to luminal A and HER2 subtypes. Phospholipid metabolic studies have revealed that membrane lipids, such as, phosphatidylethanolamine and phosphatidylcholine plays vital role in breast cancers. Membrane lipid content might assist breast tumour invasion as well. Low levels phosphatidylethanolamine act as marker in early prediction of visceral metastasis. A later phase of metastasis was predominated by low levels of phosphatidylcholine (Frank et al., 1995a).

Amide III region has shown three major peaks at 1339, 1317, 1245 and 1207 cm^{-1}. Triple negative subtype has shown amide III peak at 1339 cm^{-1}, whereas, remaining subtypes have expressed at 1338 cm^{-1}. Normal mice mammary gland haven't expressed Raman peak at 1339 cm^{-1} but mammary tumours have expressed and it is attributed to hydrated α-helix δ (N-H) and *ν*(C-N) of elastin proteins.

It might also be attributed to purine bases (adenine and guanine) of nucleic acid content (Kast et al 2007). Raman peak at 1317 cm^{-1} represents $\nu(C_{\alpha}\text{-}H)$, CH_2 of aliphatic amino acids. All four subtypes have shown Amide III peak (Figure 77) at 1245 cm^{-1} and shoulder peak at 1207 cm^{-1} attributed to hydroxyproline and tyrosine. Amide III region represents major vibrations such as N-H bending and C-N stretching and minor vibrations of C=O in-plane bending and C=C stretching vibrations (Socrates, 2004). Protein conformation in amide III region (Figure 100) is more complex because it relies on side chain of amino acids, where N-H bending is responsible for many modes of vibrations in this region. Polypeptide backbone and amino acid side-chain vibrations

significantly vary in this region making it difficult to interpret secondary structure of proteins (Mikhonin et al., 2004). Overall, higher intensity of amide III is observed in triple negative subtype and higher protein content with tyrosine and hydroxyproline was observed in HER2 subtype. Amide III region has shown triple negative and luminal B has higher protein content than HER2 and luminal A subtype.

A strong phenylalanine peak was observed at 1003 cm^{-1} for triple negative subtype and at 1002 cm^{-1} in luminal and HER2 subtypes. Luminal B has showed the highest intensity of phenylalanine compared to other subtypes (Figure 102). The region between 1000-600 cm^{-1} has shown significant collagen associated amino acid content (Frank et al., 1995a). Proline and tyrosine are observed in higher quantities in triple negative subtypes compared to luminal and HER2 subtypes. It is well-known fact that an increase in collagen content is a key marker in carcinogenesis (Haka et al., 2005). In the case of breast cancer, desmoplastic reaction leads to deposition of collagen known as reactive fibrosis. Collagen deposition is a stromal indicator to invasive carcinoma. Two prominent peaks observed around 642 and 620 cm^{-1} representing C-C twisting of tyrosine and C-S stretching vibrations. Peak intensities for these two peaks differed in all four subtypes of breast cancer.

Chapter 7

7.1 Conclusions

A combination of Raman spectroscopy and chemometrics methods was applied for the first time in three levels of breast cancer studies. In the first phase, normal and breast cancer cell lines were studied in 2D level to explore potentiality of Raman spectroscopy in chemical fingerprint and classification. In the second level, the hierarchy has increased and RS was applied on 3D spheroid models to investigate chemical changes in normal proliferating region, hypoxic region and necrotic regions of T-47D spheroids. In the third level, RS was further employed on tissue microarray (TMA) breast biopsies to identify chemical changes associated with different subtypes of breast cancer. The conclusions drawn from each study is described as follows.

7.1.1 Breast cancer cell lines (2D approach):

Cell lines are used as *in vitro* models in breast cancer research. Cell lines have shown various advantages including unlimited self-replication potential, high degree of homogeneity and ease of handling. One normal and two breast cancer subtypes were used in this study with the aim to identify chemical differences among breast cancer cell lines using dispersive Raman spectroscopy combined with chemometrics.

RS was used to characterise and differentiate two breast cancer and one normal breast cell lines (MDA-MB-436, MCF-7 and MCF-10A) and spectra of the cell lines revealed basic differences in the concentration of biochemical compounds, such as, lipids, nucleic acids and proteins.

Raman peaks were found to differ in intensity and principal component analysis (PCA) was able to identify variations that lead to accurate and reliable separation of the three cell lines. Linear discriminant analysis (LDA) model of three cell lines was predicted with 100% sensitivity and 91% specificity.

Raman spectral data combined with PCAs proves to be an excellent method that allows separation of different types of cell lines based on their lipid, and nucleic acid/DNA, and protein contents, as the MCF-7 cell line appears to be much higher in lipids compared to MDA-MB 436 and MCF 10A and works well to single out this cell line in view of the high wavenumber of amide III bands. It also contains higher nucleic acid/DNA base concentrations relative to the other two cell lines. This might indicate higher metabolic activity or might simply be due to a higher cell density in the data collection volume. Although MCF-7 and MDA-MB-436 are both breast cancer subtypes, the MDA-MB-436 does not appear to contain lipids at a concentration vastly different to those found in the normal MCF-10A cell line. Instead, the difference lies more in the relative protein and amino acid concentrations. Proteins of many different conformations overlap to form the Amide bands. It is possible that the compositions of α-helix, unordered and β-sheet conformations of proteins in the MDA-MB-436 and MCF-10A lines are different. These differences observed in the cell lines may also be useful in identifying chemical changes between the different subtypes of breast cancer although this needs confirmation in a larger panel of cell lines as well as clinical material (Talari et al 2015).

7.1.2 T-47D spheroids (3D approach)

The study of three-dimensional (3D) cell culture models over two-dimensional (2D) monolayers has several advantages. 3D spheroid models mimic tissue specific architecture. It can provide extensive knowledge to understand disease progression and development of new therapeutic drugs. Generally conventional 2D cell culture involves culture of cell lines as monolayers on impermeable plastic surfaces, which often cause loss of phenotypic and functional characteristics. One of the main features is ‘loss of differentiation’ and it mainly happens due to lack of native in vivo 3D confirmation. These monolayers will also preclude cell lines to respond to chemical and molecular

gradients in a 3D fashion due to lack of complexity. 3D cell models, particularly spheroid models, have many advantages:

a. Spheroid models will represent 3D architecture of tissues including multicellular arrangement and extracellular matrix deposition that provides spectroscopists extra space to explore chemical pathways,

b. Spheroid models will also provide different cellular microenvironments such as proliferation, quiescent and hypoxic areas allowing one to study chemical differences between these areas,

c. Spheroids provide an opportunity to study cell-cell interactions via tight junctions between cells providing chemical information regarding extracellular matrix (Jaganathan et al., 2014).

RS studies were extended to single cells to multiple cell spheroids. Human breast cancer cell lines (T-47D) were grown as spheroids and a combination of RS and Cluster analysis were employed to understand biochemical fingerprint and differentiation of normal proliferating, hypoxic and necrotic regions of spheroids. These variations may be useful in identifying new spectral markers and further understanding of cancer metabolism.

Raman spectra combined to multivariate approach prove to be an exceptional approach that allows separation of normal proliferating, hypoxic and necrotic regions based on lipids, Amide I, Amide III and nucleic acid content. These differences observed in three different regions might be useful in identification of chemical changes associated stress or strain faced by each region progressing towards necrosis. Loading plots suggested that normal proliferating region is separated with low amide I content and high tryptophan content compared to hypoxic and necrotic regions. Peak intensity and peak shifts have suggested that Amide I content is unique in each region with regards

conformation and quantity. Amide III region, especially nucleic acid and amino acid content, is particularly sensitive to necrotic and hypoxic regions. This might be due to stress or strain associated with hypoxic and necrotic regions.

7.1.3 TMA breast cancer biopsies

Human breast biopsies on TMA slide were analysed using RS and Chemometrics approaches. Biopsies were classified as luminal A, luminal B, HER2 positive and triple negative subtypes to understand chemical changes associated with breast cancer subtypes. Supervised and unsupervised algorithms were applied on biopsy data to explore intra and inter dataset biochemical changes associated with lipids, collagen and nucleic acid content.

TMA breast subtypes have shown huge chemical heterogeneity not only in different subtypes but also with in each individual subtype. Fatty acids, amide I and III content have contributed many variations among four subtypes of breast biopsies. Lipid region PCA plots have shown good separation for luminal B subtype from remaining subtypes, whereas amide I plots have shown considerable differences between luminal A and B. Amide III plots have considerably differentiated luminal A from triple negative subtype, whereas nucleic acid content plots have shown clear separation between triple negative and luminal A subtypes. Cluster analysis on full spectral range have revealed that each subtype has formed good cluster and shown more similarity between luminal A and subset of luminal B, and HER2 and triple negative subtype. One subset of luminal B has shown similarity with triple negative subtype. Distance measurements have shown luminal A and B chemically more distant from HER2 and triple negative subtypes. LDA predicted specificity accuracy of luminal A, B HER2 and triple negative subtypes were 70%, 100%, 90% and 96.7%, respectively.

7.2 Future Work

Breast Cancer diagnosis takes two steps, namely, screening approaches followed by biopsy procedures. Screening methods such as mammography, ultrasound and magnetic resonance imaging, and biopsy procedures such as fine needle, core needle and surgical biopsies are in the lead. A combination of histopathology, immunohistochemical methods and gene profile approaches are considered as gold standard methods in breast cancer subtype identification. Limitations associated with diagnostic approaches such as surgical intervention, time delays, pain and trauma to the patient and risk of infections have lead researchers to explore new non-invasive, reagent-free and less painful approaches. Raman spectroscopy, a vibrational spectroscopic approach, not only provides real time biochemical fingerprint of tissues but can also provide an understanding of the aggressiveness of the disease. Raman spectra can be acquired in less time and can also eliminate intraobserver and interobserver variability that is commonly related with histopathological diagnosis. This PhD work has explained the potentiality of a Raman approach in identifying chemical changes at the cellular level, spheroid level and biopsy level.

- Future studies involve increasing the number of TMA biopsies and employing Raman spectroscopic studies on different subtypes, which could help in understanding clinical and histologic growth patterns of subtypes.
- Raman studies in this work were helpful in differentiating breast cancer cell lines, different areas of breast spheroids and in the identification of chemical differences between breast subtypes. Therefore it would be valuable to employ to study monitoring of the disease, which would certainly helpful in understanding cancers chemical pathways.

- Area mapping of a larger number of biopsies of different subtypes will be helpful to track changes in collagen and lipid profiles in each subtypes and to establish a larger database based on these profiles.
- This work has identified chemical differences starting from the cellular level to tissue level. It would be useful to identify changes that arise in the chemical fingerprints following drug treatment. This would further aid understanding of the chemical nature of the interaction of drugs on cancer pathways.
- Raman has great potential in margin assessment of cancer tissues. It can further help clinicians to establish borders during breast cancer surgery.

This PhD work has evaluated the Raman approach in identification of chemical fingerprint from breast cancer cell lines level to microarray biopsy level. This optical approach helps to understand structural and chemical changes of biological macromolecules in disease and it could be helpful in not only in cancer diagnosis but also to understand the chemical nature of metastatic stages as well. Development of Real-time Raman systems with hand held Raman probes in breast cancer studies has the potential to collect diagnostic information in terms of molecular markers such as proteins, lipids and nucleic acids. These systems may provide immediate feedback to clinicians during biopsy procedures and significantly improve patient care. Moreover, Raman approach has potential in improving diagnostic accuracies and early disease diagnosis as well as long term monitoring during therapy.

References

ABRAMCZYK, H., BROZEK-PLUSKA, B., SURMACKI, J., JABLONSKA-GAJEWICZ, J. & KORDEK, R. (2012) Raman 'optical biopsy' of human breast cancer. *Progress in Biophysics & Molecular Biology,* 108**,** 74-81.

ABRAMCZYK, H., BROZEK-PLUSKA, B., SURMACKI, J., JABLONSKA, J. & KORDEK, R. (2011) The label-free Raman imaging of human breast cancer. *Journal of Molecular Liquids,* 164**,** 123-131.

ALBERTS, B. J., A. LEWIS, J. MORGAN, D. RAFF, M. ROBERTS, K. WALTER, P. (2015) *Molecular Biology of the Cell, Sixth Edition*.

ALEXANDER, D. D., MORIMOTO, L. M., MINK, P. J. & CUSHING, C. A. (2010) A review and meta-analysis of red and processed meat consumption and breast cancer. *Nutrition Research Reviews,* 23**,** 349-365.

ALI, S. & COOMBES, R. C. (2002) Endocrine-responsive breast cancer and strategies for combating resistance. *Nature Reviews Cancer,* 2.

ALLAIN, L. R. & VO-DINH, T. (2002) Surface-enhanced Raman scattering detection of the breast cancer susceptibility gene BRCA1 using a silver-coated microarray platform. *Analytica Chimica Acta,* 469**,** 149-154.

ANDRUS, P. G. L. & STRICKLAND, R. D. (1998) Cancer grading by Fourier transform infrared spectroscopy. *Biospectroscopy,* 4.

BANKAPUR, A., KRISHNAMURTHY, R. S., ZACHARIAH, E., SANTHOSH, C., CHOUGULE, B., PRAVEEN, B., VALIATHAN, M. & MATHUR, D. (2012) Micro-Raman Spectroscopy of Silver Nanoparticle Induced Stress on Optically-Trapped Stem Cells. *Plos One,* 7.

BANWELL, C. N. & MCCASH, E. M. (1983) *Fundamentals of molecular spectroscopy*, McGraw-Hill London.

BARR, H., DIX, T. & STONE, N. (1998) Optical spectroscopy for the early diagnosis of gastrointestinal malignancy. *Lasers in Medical Science,* 13.

BARTSCH, H., NAIR, J. & OWEN, R. W. (1999) Dietary polyunsaturated fatty acids and cancers of the breast and colorectum: emerging evidence for their role as risk modifiers. *Carcinogenesis,* 20**,** 2209-2218.

BASSETT, L., LIU, T., GIULIANO, A. & GOLD, R. (1991) The prevalence of carcinoma in palpable vs impalpable, mammographically detected lesions. *AJR. American journal of roentgenology,* 157**,** 21-24.

BEEBE, K. R. & KOWALSKI, B. R. (1987) An introduction to multivariate calibration and analysis. *Analytical Chemistry,* 59**,** 1007A-1017A.

BERG, J. T., BREEN, E. C., FU, Z. X., MATHIEU-COSTELLO, O. & WEST, J. B. (1998) Alveolar hypoxia increases gene expression of extracellular matrix proteins and platelet-derived growth factor-B in lung parenchyma. *American Journal of Respiratory and Critical Care Medicine,* 158**,** 1920-1928.

BERGHOLT, M. S., ZHENG, W., LIN, K., HO, K. Y., TEH, M., YEOH, K. G., SO, J. B. Y. & HUANG, Z. (2011a) Characterizing variability in in vivo Raman spectra of different anatomical locations in the upper gastrointestinal tract toward cancer detection. *Journal of biomedical optics,* 16**,** 037003.

BERGHOLT, M. S., ZHENG, W., LIN, K., HO, K. Y., TEH, M., YEOH, K. G., SO, J. B. Y. & HUANG, Z. (2011b) Combining near-infrared-excited autofluorescence and Raman spectroscopy improves in vivo diagnosis of gastric cancer. *Biosensors and Bioelectronics,* 26**,** 4104-4110.

BERTOS, N. R. & PARK, M. (2011) Breast cancer - one term, many entities? *Journal of Clinical Investigation,* 121**,** 3789-3796.

BINOY, J., ABRAHAM, J. P., JOE, I. H., JAYAKUMAR, V. S., PETTIT, G. R. & NIELSEN, O. F. (2004) NIR-FT Raman and FT-IR spectral studies and ab initio

calculations of the anti-cancer drug combretastatin-A4. *Journal of Raman Spectroscopy,* 35.

BRANCALEON, L., DURKIN, A. J., TU, J. H., MENAKER, G., FALLON, J. D. & KOLLIAS, N. (2001) In vivo Fluorescence Spectroscopy of Nonmelanoma Skin Cancer¶. *Photochemistry and Photobiology,* 73**,** 178-183.

CAMERLINGO, C., DELFINO, I., PERNA, G., CAPOZZI, V. & LEPORE, M. (2011) Micro-Raman Spectroscopy and Univariate Analysis for Monitoring Disease Follow-Up. *Sensors,* 11**,** 8309-8322.

CAMPAGNOLA, P. J., CLARK, H. A., MOHLER, W. A., LEWIS, A. & LOEW, L. M. (2001) Second-harmonic imaging microscopy of living cells. *Journal of biomedical optics,* 6**,** 277-286.

CAMPAGNOLA, P. J., LEWIS, A. & LOEW, L. M. (1999) High-resolution nonlinear optical imaging of live cells by second harmonic generation. *Biophysical journal,* 77**,** 3341-3349.

Cancer Research UK. (2012) Breast cancer statistics. [Online] Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer. [Accessed 11th October 2015].

Cancer Research UK. (2014) TNM breast cancer staging. [Online] Available from: http://www.cancerresearchuk.org/about-cancer/type/breast cancer/treatment/tnm-breast-cancer-staging. [Accessed 11th October 2015].

CARLSSON, J., STÅLNACKE, C.-G., ACKER, H., HAJI-KARIM, M., NILSSON, S. & LARSSON, B. (1979) The influence of oxygen on viability and proliferation in cellular spheroids. *International Journal of Radiation Oncology* Biology* Physics,* 5**,** 2011-2020.

CARTAXO, S. B., DE ABRANCHES OLIVEIRA SANTOS, I. D., BITAR, R., OLIVEIRA, A. F., FERREIRA, L. M., MARTINHO, H. S. & MARTIN, A. A.

(2010) FT-Raman spectroscopy for the differentiation between cutaneous melanoma and pigmented nevus. *Acta Cirurgica Brasileira,* 25**,** 351-356.

CARTER, R. A. B., MARTIN, A. A., NETTO, M. M. & SOARES, F. A. (2004) FT-Raman spectroscopy study of human breast tissue. *Conference on Biomedical Vibrational Spectroscopy and Biohazard Detection Technologies.* San Jose, CA.

CAZAP, E., BUZAID, A. C. & GARBINO, C. (2008) Breast Cancer in latin america. *Cancer Supplement,* 113**,** 2359-2365.

CHAN, J. W., TAYLOR, D. S., ZWERDLING, T., LANE, S. M., IHARA, K. & HUSER, T. (2006) Micro-Raman spectroscopy detects individual neoplastic and normal hematopoietic cells. *Biophysical Journal,* 90.

CHAPMAN, J., FRANKO, A. & SHARPLIN, J. (1981) A marker for hypoxic cells in tumours with potential clinical applicability. *British Journal of Cancer,* 43**,** 546.

CHEN, S.-J., ZHANG, Y., YE, X.-P., HU, K., ZHU, M.-F., HUANG, Y.-Y., ZHONG, M. & ZHUANG, Z.-F. (2014) Study of the molecular variation in pre-eclampsia placenta based on micro-Raman spectroscopy. *Archives of Gynecology and Obstetrics*, 290**,** 943-946.

CHENG, J.-X. & XIE, X. S. (2004) Coherent anti-Stokes Raman scattering microscopy: instrumentation, theory, and applications. *The Journal of Physical Chemistry B,* 108**,** 827-840.

CHENG, W. T., LIU, M. T., LIU, H. N. & LIN, S. Y. (2005) Micro-Raman spectroscopy used to identify and grade human skin pilomatrixoma. *Microscopy Research and Technique,* 68.

CHOWDARY, M., KUMAR, K. K., KURIEN, J., MATHEW, S. & KRISHNA, C. M. (2006) Discrimination of normal, benign, and malignant breast tissues by Raman spectroscopy. *Biopolymers,* 83**,** 556-569.

CONKLIN, M. W. & KEELY, P. J. (2012) Why the stroma matters in breast cancer: insights into breast cancer patient outcomes through the examination of stromal biomarkers. *Cell adhesion & migration,* 6**,** 249-260.

CONROY, J., RYDER, A. G., LEGER, M. N., HENNESSEY, K. & MADDEN, M. G. (2005) Qualitative and quantitative analysis of chlorinated solvents using Raman spectroscopy and machine learning. *Opto-Ireland 2005 Conference.* Dublin, IRELAND.

CRANE, N. J., POLFER, E., ELSTER, E. A., POTTER, B. K. & FORSBERG, J. A. (2013) Raman spectroscopic analysis of combat-related heterotopic ossification development. *Bone,* 57**,** 335-342.

Dawson, S. J., Rueda, O. M., Aparicio, S., & Caldas, C. (2013). A new genome-driven integrated classification of breast cancer and its implications. *The EMBO journal*, *32*(5), 617-628

DAY, N. (1991) Screening for breast cancer. *British medical bulletin,* 47**,** 400-415.

DE OLIVEIRA, A. F., SANTOS, I., CARTAXO, S. B., BITAR, R. A., ENOKIHARA, M., MARTINHO, H. D., MARTIN, A. A. & FERREIRA, L. M. (2010) Differential diagnosis in primary and metastatic cutaneous melanoma by FT-Raman spectroscopy. *Acta Cirurgica Brasileira,* 25**,** 434-439.

DEBRUIN, L. S. & JOSEPHY, P. D. (2002) Prospectives on the chemical etiology of breast cancer. *Environmental Health Perspectives,* 110**,** 119-128.

DHAR, A., JOHNSON, K. S., NOVELLI, M. R., BOWN, S. G., BIGIO, I. J., LOVAT, L. B. & BLOOM, S. L. (2006) Elastic scattering spectroscopy for the diagnosis of colonic lesions: initial results of a novel optical biopsy technique. *Gastrointestinal endoscopy,* 63**,** 257-261.

DO AMARAL, J. B., REZENDE-TEIXEIRA, P., FREITAS, V. M. & MACHADO-SANTELLI, G. M. (2011) MCF-7 Cells as a Three-Dimensional Model for the

Study of Human Breast Cancer. *Tissue Engineering Part C-Methods,* 17**,** 1097-1107.

DRAUX, F., JEANNESSON, P., BELJEBBAR, A., TFAYLI, A., FOURRE, N., MANFAIT, M., SULÉ-SUSO, J. & SOCKALINGUM, G. D. (2009) Raman spectral imaging of single living cancer cells: a preliminary study. *Analyst,* 134**,** 542-548.

DUKOR, R. K. (2002) *Vibrational Spectroscopy in the Detection of Cancer*.

ESTERBAUER, H., SCHAUR, R. J. & ZOLLNER, H. (1991) CHEMISTRY AND BIOCHEMISTRY OF 4-HYDROXYNONENAL, MALONALDEHYDE AND RELATED ALDEHYDES. *Free Radical Biology and Medicine,* 11**,** 81-128.

EVANS, C. L., XU, X., KESARI, S., XIE, X. S., WONG, S. T. & YOUNG, G. S. (2007) Chemically-selective imaging of brain structures with CARS microscopy. *Optics Express,* 15**,** 12076-12087.

FALANGA, V., MARTIN, T. A., TAKAGI, H., KIRSNER, R. S., HELFMAN, T., PARDES, J. & OCHOA, M. S. (1993) LOW-OXYGEN TENSION INCREASES MESSENGER-RNA LEVELS OF ALPHA-1 (I) PROCOLLAGEN IN HUMAN DERMAL FIBROBLASTS. *Journal of Cellular Physiology,* 157**,** 408-412.

FARLEY, J., SHIN, H., BRAY, F., FORMAN, D., MATHERS, C. & PARKIN, D. (2010) GLOBOCAN 2008: cancer incidence and mortality worldwide. *Lyon France: international Agency for Research on Cancer*.

FARQUHARSON, S., SHENDE, C., INSCORE, F. E., MAKSYMIUK, P. & GIFT, A. (2005) Analysis of 5-fluorouracil in saliva using surface-enhanced Raman spectroscopy. *Journal of Raman Spectroscopy,* 36.

FENG, S., CHEN, R., LIN, J., PAN, J., WU, Y., LI, Y., CHEN, J. & ZENG, H. (2011a) Gastric cancer detection based on blood plasma surface-enhanced Raman

spectroscopy excited by polarized laser light. *Biosensors and Bioelectronics,* 26**,** 3167-3174.

FENG, S., PAN, J., WU, Y., LIN, D., CHEN, Y., XI, G., LIN, J. & CHEN, R. (2011b) Study on gastric cancer blood plasma based on surface-enhanced Raman spectroscopy combined with multivariate analysis. *Science China-Life Sciences,* 54**,** 828-834.

FERLAY, J., SHIN, H. R., BRAY, F., FORMAN, D., MATHERS, C. & PARKIN, D. M. (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer,* 127**,** 2893-2917.

FERRARO, J. R., NAKAMOTO, K. & BROWN, C. W. (2003) Chapter 6 - Biochemical and Medical Applications. IN BROWN, J. R. F. N. W. (Ed.) *Introductory Raman Spectroscopy (Second Edition).* San Diego, Academic Press.

FLEISCHMANN, M., HENDRA, P. J. & MCQUILLAN, A. (1974) Raman spectra of pyridine adsorbed at a silver electrode. *Chemical Physics Letters,* 26**,** 163-166.

FRANK, C. J., MCCREERY, R. L. & REDD, D. C. (1995a) Raman spectroscopy of normal and diseased human breast tissues. *Analytical chemistry,* 67**,** 777-783.

FRANK, C. J., MCCREERY, R. L. & REDD, D. C. B. (1995c) RAMAN-SPECTROSCOPY OF NORMAL AND DISEASED HUMAN BREAST TISSUES. *Analytical Chemistry,* 67.

FRANK, C. J., MCCREERY, R. L. & REDD, D. C. B. (1995e) RAMAN-SPECTROSCOPY OF NORMAL AND DISEASED HUMAN BREAST TISSUES. *Analytical Chemistry,* 67**,** 777-783.

FRANK, C. J., REDD, D. C. B., GANSLER, T. S. & MCCREERY, R. L. (1994) CHARACTERIZATION OF HUMAN BREAST BIOPSY SPECIMENS WITH NEAR-IR RAMAN-SPECTROSCOPY. *Analytical Chemistry,* 66**,** 319-326.

FUNG, M. F. K., SENTERMAN, M. K., MIKHAEL, N. Z., LACELLE, S. & WONG, P. T. T. (1996) Pressure-tuning Fourier transform infrared spectroscopic study of carcinogenesis in human endometrium. *Biospectroscopy,* 2.

GALLIER, S., GORDON, K. C., JIMENEZ-FLORES, R. & EVERETT, D. W. (2011) Composition of bovine milk fat globules by confocal Raman microscopy. *International Dairy Journal,* 21**,** 402-412.

GEDDES, D. T. (2007) Inside the lactating breast: The latest anatomy research. *Journal of Midwifery & Womens Health,* 52.

GERMAIN, E., CHAJÈS, V., COGNAULT, S., LHUILLERY, C. & BOUGNOUX, P. (1998) Enhancement of doxorubicin cytotoxicity by polyunsaturated fatty acids in the human breast tumor cell line MBA-MB-231: Relationship to lipid peroxidation. *International journal of cancer,* 75**,** 578-583.

GILKES, D. M., SEMENZA, G. L. & WIRTZ, D. (2014) Hypoxia and the extracellular matrix: drivers of tumour metastasis. *Nature Reviews Cancer,* 14**,** 430-439.

GONZALEZ-SOLIS, J. L., RODRIGUEZ-LOPEZ, J., MARTINEZ-ESPINOSA, J. C., FRAUSTO-REYES, C., JAYE-SUAREZ, L. F., AGUILAR-LEMARROY, A. C., VARGAS-RODRIGUEZ, H. & MARTINEZ-CANO, E. (2010) Detection of Cervical Cancer Analyzing Blood Samples with Raman Spectroscopy and Multivariate Analysis. *23rd International Congress on Laser Medicine/IALMS Courses/3rd Biannual Congress of the International-Photo-Therapy-Association.* Florence, ITALY.

GORDON, M. K. & HAHN, R. A. (2010) Collagens. *Cell and Tissue Research,* 339**,** 247-257.

HAKA, A. S., SHAFER-PELTIER, K. E., FITZMAURICE, M., CROWE, J., DASARI, R. R. & FELD, M. S. (2005) Diagnosing breast cancer by using Raman spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 12371-12376.

HAKA, A. S., VOLYNSKAYA, Z., GARDECKI, J. A., NAZEMI, J., SHENK, R., WANG, N., DASARI, R. R., FITZMAURICE, M. & FELD, M. S. (2009) Diagnosing breast cancer using Raman spectroscopy: prospective analysis. *Journal of Biomedical Optics,* 14.

HANLON, E. B., MANOHARAN, R., KOO, T. W., SHAFER, K. E., MOTZ, J. T., FITZMAURICE, M., KRAMER, J. R., ITZKAN, I., DASARI, R. R. & FELD, M. S. (2000) Prospects for in vivo Raman spectroscopy. *Physics in Medicine and Biology,* 45**,** R1-R59.

HARPER, A. P., KELLY-FRY, E. & NOE, J. S. (1981) Ultrasound breast imaging—the method of choice for examining the young patient. *Ultrasound in medicine & biology,* 7**,** 231-237.

HAYNES, C. L., MCFARLAND, A. D. & DUYNE, R. P. V. (2005) Surface-enhanced Raman spectroscopy. *Analytical Chemistry,* 77**,** 338 A-346 A

.

HILVO, M., DENKERT, C., LEHTINEN, L., MÜLLER, B., BROCKMÖLLER, S., SEPPÄNEN-LAAKSO, T., BUDCZIES, J., BUCHER, E., YETUKURI, L. & CASTILLO, S. (2011) Novel theranostic opportunities offered by characterization of altered membrane lipid metabolism in breast cancer progression. *Cancer research,* 71**,** 3236-3245.

HIRSCH, H. A., ILIOPOULOS, D., JOSHI, A., ZHANG, Y., JAEGER, S. A., BULYK, M., TSICHLIS, P. N., LIU, X. S. & STRUHL, K. (2010) A transcriptional signature and common gene networks link cancer with lipid metabolism and diverse human diseases. *Cancer cell,* 17**,** 348-361.

HOLLIDAY, D. L. (2010) Three-dimensional In Vitro Model of Breast Cancer: Toward Replacing the Need for Animal Experiments. *Atla-Alternatives to Laboratory Animals,* 38**,** 41-44.

HOLLIDAY, D. L. & SPEIRS, V. (2011) Choosing the right cell line for breast cancer research. *Breast Cancer Research,* 13.

HU, Z. Y., FAN, C., OH, D. S., MARRON, J. S., HE, X. P., QAQISH, B. F., LIVASY, C., CAREY, L. A., REYNOLDS, E., DRESSLER, L., NOBEL, A., PARKER, J., EWEND, M. G., SAWYER, L. R., WU, J. Y., LIU, Y. D., NANDA, R., TRETIAKOVA, M., ORRICO, A. R., DREHER, D., PALAZZO, J. P., PERREARD, L., NELSON, E., MONE, M., HANSEN, H., MULLINS, M., QUACKENBUSH, J. F., ELLIS, M. J., OLOPADE, O. I., BERNARD, P. S. & PEROU, C. M. (2006) The molecular portraits of breast tumors are conserved across microarray platforms. *Bmc Genomics,* 7.

HUANG, N. Y., SHORT, M., ZHAO, J. H., WANG, H. Q., LUI, H., KORBELIK, M. & ZENG, H. S. (2011) Full range characterization of the Raman spectra of organs in a murine model. *Optics Express,* 19**,** 22892-22909.

HUANG, Z. W., LUI, H., MCLEAN, D. I., KORBELIK, M. & ZENG, H. S. (2005) Raman spectroscopy in combination with background near-infrared autofluorescence enhances the in vivo assessment of malignant tissues. *Photochemistry and Photobiology,* 81.

HUANG, Z. W., MCWILLIAMS, A., LUI, H., MCLEAN, D. I., LAM, S. & ZENG, H. S. (2003) Near-infrared Raman spectroscopy for optical diagnosis of lung cancer. *International Journal of Cancer,* 107.

ISHIBE, N., HANKINSON, S. E., COLDITZ, G. A., SPIEGELMAN, D., WILLETT, W. C., SPEIZER, F. E., KELSEY, K. T. & HUNTER, D. J. (1998) Cigarette smoking, cytochrome P450 1A1 polymorphisms, and breast cancer risk in the nurses' health study. *Cancer Research,* 58**,** 667-671.

JAGANATHAN, H., GAGE, J., LEONARD, F., SRINIVASAN, S., SOUZA, G. R., DAVE, B. & GODIN, B. (2014) Three-Dimensional In Vitro Co-Culture Model of Breast Tumor using Magnetic Levitation. *Scientific Reports,* 4.

JONSSON, G., STAAF, J., VALLON-CHRISTERSSON, J., RINGNER, M., HOLM, K., HEGARDT, C., GUNNARSSON, H., FAGERHOLM, R., STRAND, C., AGNARSSON, B. A., KILPIVAARA, O., LUTS, L., HEIKKILA, P., AITTOMAKI, K., BLOMQVIST, C., LOMAN, N., MALMSTROM, P.,

OLSSON, H., JOHANNSSON, O. T., ARASON, A., NEVANLINNA, H., BARKARDOTTIR, R. B. & BORG, A. (2010) Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics. *Breast Cancer Research,* 12.

KAH, J. C. Y., KHO, K. W., LEE, C. G. L. & RICHARD, C. J. (2007) Early diagnosis of oral cancer based on the surface plasmon resonance of gold nanoparticles. *International journal of nanomedicine,* 2**,** 785.

KAMEMOTO, L. E., MISRA, A. K., SHARMA, S. K., GOODMAN, M. T., LUK, H., DYKES, A. C. & ACOSTA, T. (2010) Near-Infrared Micro-Raman Spectroscopy for in Vitro Detection of Cervical Cancer. *Applied Spectroscopy,* 64**,** 255-261.

KAMINAKA, S., ITO, T., YAMAZAKI, H., KOHDA, E. & HAMAGUCHI, H. (2002) Near-infrared multichannel Raman spectroscopy toward real-time in vivo cancer diagnosis. *Journal of Raman Spectroscopy,* 33.

KAMINAKA, S., YAMAZAKI, H., ITO, T., KOHDA, E. & HAMAGUCHI, H. O. (2001) Near-infrared Raman spectroscopy of human lung tissues: possibility of molecular-level cancer diagnosis. *Journal of Raman Spectroscopy,* 32.

KANG, D.-H. (2002) Oxidative stress, DNA damage, and breast cancer. *AACN Advanced Critical Care,* 13**,** 540-549.

KATSAMA, A., SOURVINOS, G., ZACHOS, G. & SPANDIDOS, D. A. (2000) Allelic loss at the BRCA1, BRCA2 and TP53 loci in human sporadic breast carcinoma. *Cancer Letters,* 150**,** 165-170.

KAWABATA, T., KIKUCHI, H., OKAZAKI, S., YAMAMOTO, M., HIRAMATSU, Y., YANG, J., BABA, M., OHTA, M., KAMIYA, K., TANAKA, T. & KONNO, H. (2011) Near-Infrared Multichannel Raman Spectroscopy with a 1064 nm Excitation Wavelength for Ex Vivo Diagnosis of Gastric Cancer. *Journal of Surgical Research,* 169**,** E137-E143.

KELLER, M. D., KANTER, E. M., LIEBER, C. A., MAJUMDER, S. K., HUTCHINGS, J., ELLIS, D. L., BEAVEN, R. B., STONE, N. & MAHADEVAN-JANSEN, A. (2008) Detecting temporal and spatial effects of epithelial cancers with Raman spectroscopy. *Disease Markers,* 25**,** 323-337.

KIM, K. W., KIM, S. K., KIM, M. S. & SUH, S. W. (1986) A RAMAN-SPECTROSCOPIC STUDY ON 5'-RGMP GEL AND SELF-AGGREGATE. *Biopolymers,* 25**,** 753-762.

KIRSCH, M., SCHACKERT, G., SALZER, R. & KRAFFT, C. (2010) Raman spectroscopic imaging for in vivo detection of cerebral brain metastases. *Analytical and Bioanalytical Chemistry,* 398**,** 1707-1713.

KNEIPP, J., SCHUT, T. B., KLIFFEN, M., MENKE-PLUIJMERS, M. & PUPPELS, G. (2003) Characterization of breast duct epithelia: a Raman spectroscopic study. *Vibrational spectroscopy,* 32**,** 67-74.

KOEHLER, M., MACHILL, S., SALZER, R. & KRAFFT, C. (2009) Characterization of lipid extracts from brain tissue and tumors using Raman spectroscopy and mass spectrometry. *Analytical and Bioanalytical Chemistry,* 393**,** 1513-1520.

KOLJENOVIC, S., SCHUT, T. B., VINCENT, A., KROS, J. M. & PUPPELS, G. J. (2005a) Detection of meningioma in dura mater by Raman spectroscopy. *Analytical Chemistry,* 77.

KOLJENOVIĆ, S., SCHUT, T. C. B., KROS, J. M., VAN DEN BERGE, H. J. & PUPPELS, G. J. (2002) Discriminating vital tumor from necrotic tissue in human glioblastoma tissue samples by Raman spectroscopy. *Laboratory Investigation,* 82**,** 1265-1277.

KOLJENOVIC, S., SCHUT, T. C. B., WOLTHUIS, R., DE JONG, B., SANTOS, L., CASPERS, P. J., KROS, J. M. & PUPPELS, G. J. (2005f) Tissue characterization using high wave number Raman spectroscopy. *Journal of Biomedical Optics,* 10.

KONIG, K., BREUNIG, H. G., BUCKLE, R., KELLNER-HOFER, M., WEINIGEL, M., BUTTNER, E., STERRY, W. & LADEMANN, J. (2011) Optical skin biopsies by clinical CARS and multiphoton fluorescence/SHG tomography. *Laser Physics Letters,* 8**,** 465-468.

KRAFFT, C., DIETZEK, B. & POPP, J. (2009a) Raman and CARS microspectroscopy of cells and tissues. *Analyst,* 134**,** 1046-1057.

KRAFFT, C., NEUDERT, L., SIMAT, T. & SALZER, R. (2005) Near infrared Raman spectra of human brain lipids. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy,* 61.

KRAFFT, C., RAMOJI, A. A., BIELECKI, C., VOGLER, N., MEYER, T., AKIMOV, D., RÖSCH, P., SCHMITT, M., DIETZEK, B. & PETERSEN, I. (2009b) A comparative Raman and CARS imaging study of colon tissue. *Journal of biophotonics,* 2**,** 303-312.

KRAFFT, C., STEINER, G., BELEITES, C. & SALZER, R. (2009c) Disease recognition by infrared and Raman spectroscopy. *Journal of Biophotonics,* 2**,** 13-28.

KRAFFT, C., STEINER, G., BELEITES, C. & SALZER, R. (2009f) Disease recognition by infrared and Raman spectroscopy. *Journal of biophotonics,* 2**,** 13-28.

LAKOWICZ, J. R. (2013) *Principles of fluorescence spectroscopy*, Springer Science & Business Media.

LAKSHMI, R. J., KARTHA, V. B., KRISHNA, C. M., SOLOMON, J. G. R., ULLAS, G. & DEVI, P. U. (2002) Tissue Raman spectroscopy for the study of radiation damage: Brain irradiation of mice. *Radiation Research,* 157.

LARRAONA-PUY, M., GHITA, A., ZOLADEK, A., PERKINS, W., VARMA, S., LEACH, I. H., KOLOYDENKO, A. A., WILLIAMS, H. & NOTINGHER, I.

(2009) Development of Raman microspectroscopy for automated detection and imaging of basal cell carcinoma. *Journal of Biomedical Optics,* 14.

LARRAONA-PUY, M., GHITA, A., ZOLADEK, A., PERKINS, W., VARMA, S., LEACH, I. H., KOLOYDENKO, A. A., WILLIAMS, H. & NOTINGHER, I. (2011) Discrimination between basal cell carcinoma and hair follicles in skin tissue sections by Raman micro-spectroscopy. *Journal of Molecular Structure,* 993**,** 57-61.

LASKA, J. & WIDLARZ, J. (2005) Spectroscopic and structural characterization of low molecular weight fractions of polyaniline. *Polymer,* 46.

LAU, D. P., HUANG, Z., LUI, H., MAN, C. S., BEREAN, K., MORRISON, M. D. & ZENG, H. (2003a) Raman spectroscopy for optical diagnosis in normal and cancerous tissue of the nasopharynx—preliminary findings. *Lasers in surgery and medicine,* 32**,** 210-214.

LAU, D. P., HUANG, Z. W., LUI, H., ANDERSON, D. W., BEREAN, K., MORRISON, M. D., SHEN, L. & ZENG, H. S. (2005) Raman spectroscopy for optical diagnosis in the larynx: Preliminary findings. *Lasers in Surgery and Medicine,* 37.

LAU, D. P., HUANG, Z. W., LUI, H., MAN, C. S., BEREAN, K., MORRISON, M. D. & ZENG, H. S. (2003b) Raman spectroscopy for optical diagnosis in normal and cancerous tissue of the nasopharynx - Preliminary findings. *Lasers in Surgery and Medicine,* 32.

LEE, P. & MEISEL, D. (1982) Adsorption and surface-enhanced Raman of dyes on silver and gold sols. *The Journal of Physical Chemistry,* 86**,** 3391-3395.

LI, C. I., URIBE, D. J. & DALING, J. R. (2005) Clinical characteristics of different histologic types of breast cancer. *British Journal of Cancer,* 93**,** 1046-1052.

LI, X., YANG, T., LI, S. & YU, T. (2011) Surface-enhanced Raman spectroscopy differences of saliva between lung cancer patients and normal people.

Conference on Clinical and Biomedical Spectroscopy and Imaging II. Munich, GERMANY.

LI, X. Z., YANG, T. Y. & DING, J. H. (2012) Surface Enhanced Raman Spectroscopy (SERS) of Saliva for the Diagnosis of Lung Cancer. *Spectroscopy and Spectral Analysis,* 32**,** 391-393.

LI, Y., WEN, Z.-N., LI, L.-J., LI, M.-L., GAO, N. & GUO, Y.-Z. (2010) Research on the Raman spectral character and diagnostic value of squamous cell carcinoma of oral mucosa. *Journal of Raman Spectroscopy,* 41**,** 142-147.

LIN-VIEN, D., COLTHUP, N. B., FATELEY, W. G. & GRASSELLI, J. G. (1991) *The handbook of infrared and Raman characteristic frequencies of organic molecules,* Elsevier.

LIN, D., FENG, S., PAN, J., CHEN, Y., LIN, J., CHEN, G., XIE, S., ZENG, H. & CHEN, R. (2011a) Colorectal cancer detection by gold nanoparticle based surface-enhanced Raman spectroscopy of blood serum and statistical analysis. *Optics express,* 19**,** 13565-13577.

LIN, D., LIN, J., WU, Y., FENG, S., LI, Y., YU, Y., XI, G., ZENG, H. & CHEN, R. (2011b) Investigation on the interactions of lymphoma cells with paclitaxel by Raman spectroscopy. *Spectroscopy-an International Journal,* 25**,** 23-32.

LINDEMAN, G. J. & VISVADER, J. E. (2011) Hereditary Breast Cancer Genetics-From Clinical Curiosities to Mainstream Paradigms PREFACE. *Journal of Mammary Gland Biology and Neoplasia,* 16**,** 1-2.

LYNG, F. M., FAOLAIN, E. O., CONROY, J., MEADE, A. D., KNIEF, P., DUFFY, B., HUNTER, M. B., BYRNE, J. M., KELEHAN, P. & BYRNE, H. J. (2007) Vibrational spectroscopy for cervical cancer pathology, from biochemical analysis to diagnostic tool. *Experimental and Molecular Pathology,* 82.

MADHAVI, M., MADHAVI, K. & JITHAN, A. V. (2012) Preparation and in vitro/in vivo characterization of curcumin microspheres intended to treat colon cancer. *Journal of pharmacy & bioallied sciences*, 4.

MAGEE, N. D., VILLAUMIE, J. S., MARPLE, E. T., ENNIS, M., ELBORN, J. S. & MCGARVEY, J. J. (2009) Ex vivo diagnosis of lung cancer using a Raman miniprobe. *The Journal of Physical Chemistry B,* 113**,** 8137-8141.

MAHADEVAN-JANSEN, A., RICHARDS-KORTUM, R. & IEEE (1997) Raman spectroscopy for cancer detection: A review. *International Conference of the IEEE Engineering-in-Medicine-and-Biology-Society.* Chicago, Il.

MAHADEVAN-JANSEN, A. & RICHARDS-KORTUM, R. R. (1996) Raman spectroscopy for the detection of cancers and precancers. *Journal of Biomedical Optics,* 1**,** 31-70.

MALINI, R., VENKATAKRISHNA, K., KURIEN, J., M PAI, K., RAO, L., KARTHA, V. & KRISHNA, C. M. (2006a) Discrimination of normal, inflammatory, premalignant, and malignant oral tissue: a Raman spectroscopy study. *Biopolymers,* 81**,** 179-193.

MALINI, R., VENKATAKRISHNA, K., KURIEN, J., PAI, K. M., RAO, L., KARTHA, V. B. & KRISHNA, C. M. (2006b) Discrimination of normal, inflammatory, premalignant, and malignant oral tissue: A Raman spectroscopy study. *Biopolymers,* 81.

MANOHARAN, R., SHAFER, K., PERELMAN, L., WU, J., CHEN, K., DEINUM, G., FITZMAURICE, M., MYLES, J., CROWE, J., DASARI, R. R. & FELD, M. S. (1998) Raman spectroscopy and fluorescence photon migration for breast cancer diagnosis and imaging. *Photochemistry and Photobiology,* 67**,** 15-22.

MAZUREK, S. & SZOSTAK, R. (2006) Quantitative determination of captopril and prednisolone in tablets by FT-Raman spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis,* 40.

MIKHONIN, A. V., AHMED, Z., IANOUL, A. & ASHER, S. A. (2004) Assignments and conformational dependencies of the amide III peptide backbone UV resonance Raman bands. *The Journal of Physical Chemistry B,* 108**,** 19020-19028.

MILGRAUM, L. Z., WITTERS, L. A., PASTERNACK, G. A. & KUHAJDA, F. P. (1997) Enzymes of the fatty acid synthesis pathway are highly expressed in in situ breast carcinoma. *Clinical Cancer Research,* 3**,** 2115-2120.

MIN, Y. K., YAMAMOTO, T., KOHDA, E., ITO, T. & HAMAGUCHI, H. (2005) 1064 nm near-infrared multichannel Raman spectroscopy of fresh human lung tissues. *Journal of Raman Spectroscopy,* 36.

MOLCKOVSKY, A., SONG, L.-M. W. K., SHIM, M. G., MARCON, N. E. & WILSON, B. C. (2003) Diagnostic potential of near-infrared Raman spectroscopy in the colon: differentiating adenomatous from hyperplastic polyps. *Gastrointestinal endoscopy,* 57**,** 396-402.

MOVASAGHI, Z., REHMAN, S. & REHMAN, I. U. (2007a) Raman spectroscopy of biological tissues. *Applied Spectroscopy Reviews,* 42.

MOVASAGHI, Z., REHMAN, S. & REHMAN, I. U. (2007d) Raman spectroscopy of biological tissues. *Applied Spectroscopy Reviews,* 42**,** 493-541.

MOVASAGHI, Z., REHMAN, S. & REHMAN, I. U. (2008) Fourier transform infrared (FTIR) spectroscopy of biological tissues. *Applied Spectroscopy Reviews,* 43**,** 134-179.

MYLLYHARJU, J. & KIVIRIKKO, K. I. (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends in Genetics,* 20**,** 33-43.

NAQUI, A., CHANCE, B. & CADENAS, E. (1986) REACTIVE OXYGEN INTERMEDIATES IN BIOCHEMISTRY. *Annual Review of Biochemistry,* 55**,** 137-166.

NATRAJAN, R., WEIGELT, B., MACKAY, A., GEYER, F. C., GRIGORIADIS, A., TAN, D. S. P., JONES, C., LORD, C. J., VATCHEVA, R., RODRIGUEZ-PINILLA, S. M., PALACIOS, J., ASHWORTH, A. & REIS, J. S. (2010) An integrative genomic and transcriptomic analysis reveals molecular pathways and networks regulated by copy number aberrations in basal-like, HER2 and luminal cancers. *Breast Cancer Research and Treatment,* 121**,** 575-589.

NAUMANN, D. (1998) Infrared and NIR raman spectroscopy in medical microbiology. *Conference on Infrared Spectroscopy - New Tool in Medicine.* San Jose, Ca.

NAWAZ, H., BONNIER, F., MEADE, A. D., LYNG, F. M. & BYRNE, H. J. (2011) Comparison of subcellular responses for the evaluation and prediction of the chemotherapeutic response to cisplatin in lung adenocarcinoma using Raman spectroscopy. *Analyst,* 136**,** 2450-2463.

NEUGEBAUER, U., CLEMENT, J. H., BOCKLITZ, T., KRAFFT, C. & POPP, J. (2010) Identification and differentiation of single cells from peripheral blood by Raman spectroscopic imaging. *Journal of Biophotonics,* 3**,** 579-587.

NOTINGHER, I., GREEN, C., DYER, C., PERKINS, E., HOPKINS, N., LINDSAY, C. & HENCH, L. L. (2004) Discrimination between ricin and sulphur mustard toxicity in vitro using Raman spectroscopy. *Journal of the Royal Society Interface,* 1.

NYSTRÖM, L., WALL, S., RUTQVIST, L., LINDGREN, A., LINDQVIST, M., RYDEN, S., ANDERSSON, J., BJURSTAM, N., FAGERBERG, G. & FRISELL, J. (1993) Breast cancer screening with mammography: overview of Swedish randomised trials. *The Lancet,* 341**,** 973-978.

O FAOLAIN, E., HUNTER, M. B., BYRNE, J. M., KELEHAN, P., MCNAMARA, M., BYRNE, H. J. & LYNG, F. M. (2005) A study examining the effects of tissue processing on human tissue sections using vibrational spectroscopy. *Vibrational Spectroscopy,* 38.

OLAYIOYE, M. A. (2001) Update on HER-2 as a target for cancer therapy: intracellular signaling pathways of ErbB2/HER-2 and family members. *Breast Cancer Res,* 3**,** 385-389.

OREL, S. (2001) MR imaging of the breast. *Magnetic resonance imaging clinics of North America,* 9**,** 273-88, v.

OSAWA, M., CHALMERS, J. & GRIFFITHS, P. (2002) Handbook of Vibrational Spectroscopy. *Wiley, Chichester,* 1**,** 785.

OSHIMA, Y., SHINZAWA, H., TAKENAKA, T., FURIHATA, C. & SATO, H. (2010) Discrimination analysis of human lung cancer cells associated with histological type and malignancy using Raman spectroscopy. *Journal of Biomedical Optics,* 15.

OYAMA, T., KOIBUCHI, Y. & MCKEE, G. (2004) Core needle biopsy (CNB) as a diagnostic method for breast lesions: comparison with fine needle aspiration cytology (FNA). *Breast Cancer,* 11**,** 339-342.

OZEK, N. S., TUNA, S., ERSON-BENSAN, A. E. & SEVERCAN, F. (2010a) Characterization of microRNA-125b expression in MCF7 breast cancer cells by ATR-FTIR spectroscopy. *Analyst,* 135**,** 3094-3102.

OZEK, N. S., TUNA, S., ERSON-BENSAN, A. E. & SEVERCAN, F. (2010c) Characterization of microRNA-125b expression in MCF7 breast cancer cells by ATR-FTIR spectroscopy. *Analyst,* 135.

PARISE, C. A., BAUER, K. R., BROWN, M. M. & CAGGIANO, V. (2009) Breast Cancer Subtypes as Defined by the Estrogen Receptor (ER), Progesterone Receptor (PR), and the Human Epidermal Growth Factor Receptor 2 (HER2) among Women with Invasive Breast Cancer in California, 1999-2004. *Breast Journal,* 15**,** 593-602.

POLOMSKA, M., KUBISZ, L., KALAWSKI, R., OSZKINIS, G., FILIPIAK, R. & MAZUREK, A. (2010) Fourier Transform Near Infrared Raman Spectroscopy in Studies on Connective Tissue. *Acta Physica Polonica A,* 118**,** 136-140.

POYTON, R. O., CASTELLO, P. R., BALL, K. A., WOO, D. K. & PAN, N. (2009) Mitochondria and Hypoxic Signaling A New View. *Hypoxia and Consequences from Molecule to Malady,* 1177**,** 48-56.

PROVENZANO, P. P., ELICEIRI, K. W., CAMPBELL, J. M., INMAN, D. R., WHITE, J. G. & KEELY, P. J. (2006) Collagen reorganization at the tumor-stromal interface facilitates local invasion. *BMC medicine,* 4**,** 38.

PROVENZANO, P. P., INMAN, D. R., ELICEIRI, K. W., KNITTEL, J. G., YAN, L., RUEDEN, C. T., WHITE, J. G. & KEELY, P. J. (2008) Collagen density promotes mammary tumor initiation and progression. *BMC medicine,* 6**,** 11.

RAKHA, E. A., EL-SHEIKH, S. E., KANDIL, M. A., EL-SAYED, M. E., GREEN, A. R. & ELLIS, I. O. (2008) Expression of BRCA1 protein in breast cancer and its prognostic significance. *Human pathology,* 39**,** 857-865.

RANIERO, L., CANEVARI, R. A., RAMALHO, L. N. Z., RAMALHO, F. S., DOS SANTOS, E. A. P., BITAR, R. A., JALKANEN, K. J., MARTINHO, H. S. & MARTIN, A. A. (2011) In and ex vivo breast disease study by Raman spectroscopy. *Theoretical Chemistry Accounts,* 130**,** 1239-1247.

RAUCHOVA, H., VOKURKOVA, M. & KOUDELOVA, J. (2012) Hypoxia-Induced Lipid Peroxidation in the Brain During Postnatal Ontogenesis. *Physiological Research,* 61**,** S89-S101.

REEVES, G. K., PIRIE, K., BERAL, V., GREEN, J., SPENCER, E., BULL, D. & MILLION WOMEN STUDY, C. (2007) Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. *British Medical Journal,* 335.

REHMAN, I. U., MOVASAGHI, Z. AND REHMAN, S. (2012) *Vibrational Spectroscopy for Tissue Analysis*, CRC Press, Taylor & Francis Group.

REHMAN, S., MOVASAGHI, Z., TUCKER, A. T., JOEL, S. P., DARR, J. A., RUBAN, A. V. & REHMAN, I. U. (2007) Raman spectroscopic analysis of breast cancer tissues: identifying differences between normal, invasive ductal carcinoma and ductal carcinoma in situ of the breast tissue. *Journal of Raman Spectroscopy,* 38**,** 1345-1351.

REYNOLDS, P. (2013) Smoking and Breast Cancer. *Journal of Mammary Gland Biology and Neoplasia,* 18**,** 15-23.

ROBERT, B. (2009) Resonance Raman spectroscopy. *Photosynthesis research,* 101**,** 147-155.

RODRIGUES, M., WEERSINK, R. A. & WHELAN, W. M. (2010) Assessment of thermal coagulation in ex-vivo tissues using Raman spectroscopy. *Journal of Biomedical Optics,* 15**,** 068001-068001-6.

RONEN, S. M., STIER, A. & DEGANI, H. (1990) NMR-STUDIES OF THE LIPID-METABOLISM OF T47D HUMAN BREAST-CANCER SPHEROIDS. *Febs Letters*, 266.

RUIZ-CHICA, A. J., MEDINA, M. A., SANCHEZ-JIMENEZ, F. & RAMIREZ, F. J. (2004) Characterization by Raman spectroscopy of conformational changes on guanine-cytosine and adenine-thymine oligonucleotides induced by aminooxy analogues of spermidine. *Journal of Raman Spectroscopy,* 35.

SAHA, A., BARMAN, I., DINGARI, N. C., MCGEE, S., VOLYNSKAYA, Z., GALINDO, L. H., LIU, W., PLECHA, D., KLEIN, N., DASARI, R. R. & FITZMAURICE, M. (2011) Raman spectroscopy: a real-time tool for identifying microcalcifications during stereotactic breast core needle biopsies. *Biomedical Optics Express,* 2**,** 2792-2803.

SATHULURI, R. R., YOSHIKAWA, H., SHIMIZU, E., SAITO, M. & TAMIYA, E. (2011) Gold Nanoparticle-Based Surface-Enhanced Raman Scattering for Noninvasive Molecular Probing of Embryonic Stem Cell Differentiation. *Plos One,* 6.

SATO, H., YAMAMOTO, Y. S., MARUYAMA, A., KATAGIRI, T., MATSUURA, Y. & OZAKI, Y. (2009) Raman study of brain functions in live mice and rats: A pilot study. *Vibrational Spectroscopy,* 50**,** 125-130.

SAUER, G. R., ZUNIC, W. B., DURIG, J. R. & WUTHIER, R. E. (1994) FOURIER-TRANSFORM RAMAN-SPECTROSCOPY OF SYNTHETIC AND BIOLOGICAL CALCIUM PHOSPHATES. *Calcified Tissue International,* 54**,** 414-420.

SCHULZ, H. & BARANSKA, M. (2007) Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vibrational Spectroscopy,* 43.

SCOCCIANTI, C., LAUBY-SECRETAN, B., BELLO, P. Y., CHAJES, V. & ROMIEU, I. (2014) Female Breast Cancer and Alcohol Consumption A Review of the Literature. *American Journal of Preventive Medicine,* 46**,** S16-S25.

SEBAG, J., NIE, S., REISER, K. & YU, N. T. (1993) RAMAN-SPECTROSCOPY CHARACTERIZATION OF DIABETES EFFECTS ON HUMAN VITREOUS IN DIABETIC-RETINOPATHY. *Conf on Ophthalmic Technologies 3.* Los Angeles, Ca.

SEBALLOS, L., ZHANG, J. Z. & SUTPHEN, R. (2005) Surface-enhanced Raman scattering detection of lysophosphatidic acid. *Analytical and Bioanalytical Chemistry,* 383.

SEMENZA, G. L. (2000) HIF-1: mediator of physiological and pathophysiological responses to hypoxia. *Journal of Applied Physiology,* 88**,** 1474-1480.

SHA, M. Y., XU, H., NATAN, M. J. & CROMER, R. (2008) Surface-enhanced Raman scattering tags for rapid and homogeneous detection of circulating tumor cells in

the presence of human whole blood. *Journal of the American Chemical Society,* 130**,** 17214-17215.

SHAFER-PELTIER, K. E., HAKA, A. S., FITZMAURICE, M., CROWE, J., MYLES, J., DASARI, R. R. & FELD, M. S. (2002) Raman microspectroscopic model of human breast tissue: implications for breast cancer diagnosis in vivo. *Journal of Raman Spectroscopy,* 33**,** 552-563.

SHAO, J., LIN, M., LI, Y., LI, X., LIU, J., LIANG, J. & YAO, H. (2012) In Vivo Blood Glucose Quantification Using Raman Spectroscopy. *Plos One,* 7.

SHAPIRO, A., GOFRIT, O. N., PIZOV, G., COHEN, J. K. & MAIER, J. (2011) Raman Molecular Imaging: A Novel Spectroscopic Technique for Diagnosis of Bladder Cancer in Urine Specimens. *European Urology,* 59**,** 106-112.

SHAW, R. A. & MANTSCH, H. H. (1999) Vibrational biospectroscopy: from plants to animals to humans. A historical perspective. *Journal of Molecular Structure,* 481.

SHETTY, G., KENDALL, C., SHEPHERD, N., STONE, N. & BARR, H. (2006) Raman spectroscopy: elucidation of biochemical changes in carcinogenesis of oesophagus. *British Journal of Cancer,* 94.

SIGURDSSON, S., PHILIPSEN, P. A., HANSEN, L. K., LARSEN, J., GNIADECKA, M. & WULF, H. C. (2004) Detection of skin cancer by classification of Raman spectra. *Ieee Transactions on Biomedical Engineering,* 51.

SILVEIRA, L., SATHAIAH, S., ZANGARO, R. A., PACHECO, M. T. T., CHAVANTES, M. C. & PASQUALUCCI, C. A. G. (2002) Correlation between near-infrared Raman spectroscopy and the histopathological analysis of atherosclerosis in human coronary arteries. *Lasers in Surgery and Medicine,* 30.

SOCRATES, G. (2004) *Infrared and Raman characteristic group frequencies: tables and charts,* John Wiley & Sons.

SORLIE, T., PEROU, C. M., TIBSHIRANI, R., AAS, T., GEISLER, S., JOHNSEN, H., HASTIE, T., EISEN, M. B., VAN DE RIJN, M., JEFFREY, S. S., THORSEN, T., QUIST, H., MATESE, J. C., BROWN, P. O., BOTSTEIN, D., LONNING, P. E. & BORRESEN-DALE, A. L. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 10869-10874.

STONE, N., KENDALL, C., SHEPHERD, N., CROW, P. & BARR, H. (2002) Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers. *Journal of Raman Spectroscopy,* 33.

STONE, N., KENDALL, C., SMITH, J., CROW, P. & BARR, H. (2004) Raman spectroscopy for identification of epithelial cancers. *Faraday Discussions,* 126.

SU, L., SUN, Y. F., CHEN, Y., CHEN, P., SHEN, A. G., WANG, X. H., JIA, J., ZHAO, Y. F., ZHOU, X. D. & HU, J. M. (2012) Raman spectral properties of squamous cell carcinoma of oral tissues and cells. *Laser Physics,* 22**,** 311-316.

SURMACKI, J., MUSIAL, J., KORDEK, R. & ABRAMCZYK, H. (2013) Raman imaging at biological interfaces: applications in breast cancer diagnosis. *Molecular cancer,* 12**,** 48.

SUTHERLAND, R. M. (1988) Cell and environment interactions in tumor microregions: the multicell spheroid model. *Science,* 240**,** 177-184.

SWAIN, R. & STEVENS, M. (2007) Raman microspectroscopy for non-invasive biochemical analysis of single cells. *Biochemical Society Transactions,* 35**,** 544-550.

SWINNEN, J. V., ROSKAMS, T., JONIAU, S., VAN POPPEL, H., OYEN, R., BAERT, L., HEYNS, W. & VERHOEVEN, G. (2002) Overexpression of fatty acid synthase is an early and common event in the development of prostate cancer. *International journal of cancer,* 98**,** 19-22.

TAKEUCHI, H. (2003) Raman structural markers of tryptophan and histidine side chains in proteins. *Biopolymers,* 72**,** 305-317.

TALARI, A., EVANS, C., HOLEN, I., COLEMAN, R. & REHMAN, I. U. (2015a) Raman spectroscopic analysis differentiates between breast cancer cell lines. *Journal of Raman Spectroscopy,* 46**,** 421-427.

TALARI, A. C. S., MOVASAGHI, Z., REHMAN, S. & REHMAN, I. U. (2015k) Raman Spectroscopy of Biological Tissues. *Applied Spectroscopy Reviews,* 50**,** 46-111.

TAN, Y. Y., SHEN, A. G., ZHANG, J. W., WU, N., FENG, L., WU, Q. F., YE, Y., HU, J. M. & IEEE (2003) Design of auto-classifying system and its application in roman spectroscopy diagnosis of gastric carcinoma. *2003 International Conference on Machine Learning and Cybernetics, Vols 1-5, Proceedings*.

The Royal College of Pathologists. (2015) Pathology: Datasets and Tissue Pathways. [Online] Available from: http://www.rcpath.org/publications-media/publications/datasets/datasets-TP.htm. [Accessed 11th October 2015]

THISTLETHWAITE, J. & STEWART, R. A. (2007) Clinical breast examination for asymptomatic women: Exploring the evidence. *Australian family physician,* 36**,** 145-149.

THOMAS JR, G. J. (1999) Raman spectroscopy of protein and nucleic acid assemblies. *Annual review of biophysics and biomolecular structure,* 28**,** 1-27.

TOLLES, W. M., NIBLER, J., MCDONALD, J. & HARVEY, A. (1977) A review of the theory and application of coherent anti-Stokes Raman spectroscopy (CARS). *Applied Spectroscopy,* 31**,** 253-271.

TORRE, L. A., BRAY, F., SIEGEL, R. L., FERLAY, J., LORTET-TIEULENT, J. & JEMAL, A. (2015) Global Cancer Statistics, 2012. *Ca-a Cancer Journal for Clinicians,* 65**,** 87-108.

UTZINGER, U., HEINTZELMAN, D. L., MAHADEVAN-JANSEN, A., MALPICA, A., FOLLEN, M. & RICHARDS-KORTUM, R. (2001) Near-infrared Raman spectroscopy for in vivo detection of cervical precancers. *Applied Spectroscopy,* 55.

VAN DE VEN, M., VAN LANGEN, H., VERWER, W. & LEVINE, Y. K. (1984) RAMAN SPECTRA AND CONFORMATIONS OF N HEXADECYNOIC FATTY-ACIDS AND THEIR POTASSIUM SALTS. *Chemistry and Physics of Lipids,* 34**,** 185-200.

VOGLER, N., MEYER, T., AKIMOV, D., LATKA, I., KRAFFT, C., BENDSOE, N., SVANBERG, K., DIETZEK, B. & POPP, J. (2010) Multimodal imaging to study the morphochemistry of basal cell carcinoma. *Journal of biophotonics,* 3**,** 728-736.

VOLYNSKAYA, Z., HAKA, A. S., BECHTEL, K. L., FITZMAURICE, M., SHENK, R., WANG, N., NAZEMI, J., DASARI, R. R. & FELD, M. S. (2008) Diagnosing breast cancer using diffuse reflectance spectroscopy and intrinsic fluorescence spectroscopy. *Journal of biomedical optics,* 13**,** 024012-024012-9.

WANG, H., FU, Y., ZICKMUND, P., SHI, R. & CHENG, J.-X. (2005) Coherent anti-stokes Raman scattering imaging of axonal myelin in live spinal tissues. *Biophysical Journal,* 89**,** 581-591.

WEIGELT, B., GEYER, F. C. & REIS, J. S. (2010) Histological types of breast cancer: How special are they? *Molecular Oncology,* 4**,** 192-208.

WEINFURTNER, R. J., PATEL, B., LARONGA, C., LEE, M. C., FALCON, S. L., MOONEY, B. P., YUE, B. & DRUKTEINIS, J. S. (2014) Magnetic Resonance Imaging-Guided Core Needle Breast Biopsies Resulting in High-Risk Histopathologic Findings: Upstage Frequency and Lesion Characteristics. *Clinical breast cancer*.

WOLFF, A. C., HAMMOND, M. E. H., SCHWARTZ, J. N., HAGERTY, K. L., ALLRED, D. C., COTE, R. J., DOWSETT, M., FITZGIBBONS, P. L., HANNA, W. M., LANGER, A., MCSHANE, L. M., PAIK, S., PEGRAM, M.

D., PEREZ, E. A., PRESS, M. F., RHODES, A., STURGEON, C., TAUBE, S. E., TUBBS, R., VANCE, G. H., DE VIJVER, M. V., WHEELER, T. M. & HAYES, D. F. (2007) American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Archives of Pathology & Laboratory Medicine,* 131**,** 18-43.

WYNN, T. A. & BARRON, L. (2010) Macrophages: Master Regulators of Inflammation and Fibrosis. *Seminars in Liver Disease,* 30**,** 245-257.

YAZDI, Y., RAMANUJAM, N., LOTAN, R., MITCHELL, M. F., HITTELMAN, W. & RICHARDS-KORTUM, R. (1999) Resonance Raman spectroscopy at 257 nm excitation of normal and malignant cultured breast and cervical cells. *Applied spectroscopy,* 53**,** 82-85.

YU, C., GESTL, E., ECKERT, K., ALLARA, D. & IRUDAYARAJ, J. (2006) Characterization of human breast epithelial cells by confocal Raman microspectroscopy. *Cancer detection and prevention,* 30**,** 515-522.

ZENONE, F., LEPORE, M., PERNA, G., CARMONE, P., DELFINO, I., GAETA, G. M. & CAPOZZI, V. (2007) Oral pathology diagnosis by means of micro-Raman spectroscopy on biopsies and blood serum - art. no. 64250V. *Conference on Lasers in Dentistry XIII.* San Jose, CA.

ZHOU, Y., LIU, C.-H., SUN, Y., PU, Y., BOYDSTON-WHITE, S., LIU, Y. & ALFANO, R. R. (2012) Human brain cancer studied by resonance Raman spectroscopy. *Journal of biomedical optics,* 17**,** 116021-116021.

ZHU, C., PALMER, G. M., BRESLIN, T. M., HARTER, J. & RAMANUJAM, N. (2008) Diagnosis of breast cancer using fluorescence and diffuse reflectance spectroscopy: a Monte-Carlo-model-based approach. *Journal of biomedical optics,* 13**,** 034015-034015-15.

Appendix

Publications:

- Raman spectroscopic analysis differentiates between breast cancer cell lines **A.C.S. Talari**, C.A. Evans, I Holen, R.E. Coleman and I. U. Rehman. *Journal of Raman Spectroscopy (JRS),* 46, 421-427, 2015.
- Raman Spectroscopy of Biological Tissues. **A.C.S. Talari**, Z. Movasaghi, S. Rehman and I. U. Rehman. *Applied Spectroscopy Reviews* Volume 50, Issue 1, 46-111, 2015.

Manuscripts in progress

- Raman spectroscopic analysis differentiates normal proliferating, hypoxic and necrotic regions of T-47D spheroids. **A.C.S. Talari**, C.A. Evans, I Holen, R.E. Coleman and I. U. Rehman.
- Raman spectroscopic evaluation of different subtypes of tissue micro array breast tissue samples **A.C.S. Talari**, C.A. Evans, I Holen, R.E. Coleman and I. U. Rehman.
- Raman mapping of 2D and 3D breast models **A.C.S. Talari**, C.A. Evans, I Holen, R.E. Coleman and I. U. Rehman.

Poster Presentations

- 4-7 June 2012: 8th NCRI conference, Liverpool, UK.
- 27 April 2013: 5NEONC Breast Cancer Symposium, Staff House, University of Hull, UK.
- 12-14 May 2013: 5th WMRIF (World Materials Research Institutes Forum) Symposium, EMPA AKADEMIE.

- 22 March 2013: Inaugural cancer research forum (ICRF), The Ridge Ranmoor Student Village Sheffield, UK
- 20 May 2014: 2nd year poster presentation, Department of material sciences and engineering, University of Sheffield, UK

Conferences and Seminars

- Research in Progress (RIP) seminars at Kroto Research Institute (KRI), Sheffield and Weekly Seminars of Research skills and communications
- **Raman Analysis of Different subtypes of Breast Cancer cell lines**

  Talari A. C. S., Coleman. R.E. and Rehman I. U. *8th NCRI Cancer Conference BT Convention Centre, Liverpool L3 4FP UK 4-7 November 2012.*
- **Raman Analysis of Different subtypes of Breast Cancer cell lines** Rehman S., **Talari A. C. S.** and Rehman I.U. *5NEONC Breast Cancer Symposium, Staff House, University of Hull HU6 7RX UK 27 April 2013*
- **Spectroscopy of Cancer: Raman Analysis of Different Subtypes of Breast Cancer Tissues and Cell Lines Talari A. C. S**., Evans, C.A., Holen, I., Coleman. R.E. and Rehman I. U. *5th WMRIF (World Materials Research Institutes Forum) Symposium, EMPA AKADEMIE, 12-14 May 201*.
- 207th IRDG meeting (16 Oct 2014) **on From Cancer diagnostics to drug release: vibrational spectroscopy in biomedical sciences**. AstraZeneca, Macclesfield.
- Guest speaker in seminar on **Spectroscopy in Biomaterials and Tissue Engineering** at Kroto Research Institute, University of Sheffield, conducted by Thermo Fisher Scientific (8th July 2015).

EPSRC Studentship

- I have received fully funded EPSRC studentship during my PhD programme.

Raman peak and its assignments

Peak No	Assignment	References
618	C-C twisting of proteins	(Chan et al., 2006)
620	C-C twist aromatic ring	(O Faolain et al., 2005)
621	C-C twisting mode of Phenylalanine	(Stone et al., 2004, Stone et al., 2002, Notingher et al., 2004)
637	*v*(C-S) gauche of Methionine	(Shetty et al., 2006)
642	C-C twist mode of Tyrosine	(Sathuluri et al., 2011)
643	C-C twisting mode of Tyrosine	(Stone et al., 2004, Cheng et al., 2005, Binoy et al., 2004)
671	Ring breathing mode of Guanine	(Kim et al., 1986)
722	DNA	(Binoy et al., 2004)
725	Ring breathing mode of Adenine	(Chan et al., 2006)
727	C-C stretching, Proline (collagen assignment)	(Frank et al., 1995e)
726	C-S (protein), CH_2 rocking, Adenine	(Stone et al., 2004)
746	Ring breathing mode of Thymine	(Chan et al., 2006)
758	*v*s indole ring breathing of Tryptophan	(Chen et al., 2014)
759	Tryptophan, Ethanolamine group, Phosphatidylethanolamine	(Stone et al., 2004, Krafft et al., 2005)
779	Ring vibration of nucleic acids	(Lyng et al., 2007)
780	Ring breathing mode of Uracil	(Farquharson et al., 2005)
782	DNA, Thymine, Cytosine, Uracil, RNA	(Binoy et al., 2004, Notingher et al.,

		2004)
821	*ν*(CC) backbone of collagen	(Crane et al., 2013)
828	Ring breathing mode of Tyrosine, Phosphodiester, Tyrosine / O-P-O stretch DNA	(Stone et al., 2002, Stone et al., 2004, Ruiz-Chica et al., 2004)
829	Aliphatic and aromatic deformation of CCH	(Larraona-Puy et al., 2011)
847	Monosaccharides (α-glucose), (C-O-C) skeletal mode, disaccharides (maltose),	(Shetty et al., 2006)
852	Proline, Hydroxyproline, Tyrosine, Tyrosine ring breathing, Glycogen	(Cheng et al., 2005, Binoy et al., 2004, Chan et al., 2006)
853	C-C stretch of Proline ring, Tyrosine ring breathing mode, Glycogen	(Binoy et al., 2004, Stone et al., 2002, Stone et al., 2004)
854	(C-O-C) skeletal mode of α-anomers (polysaccharides, pectin), ring breathing mode of tyrosine	(Shetty et al., 2006, Notingher et al., 2004)
875	Antisymmetric stretch vibrations of Choline group, Phospholipids, Phosphatidylcholine, Sphingomyelin	(Krafft et al., 2005)
877	C-C-N+ symmetric stretch of lipids, C-O-C ring of carbohydrates	(Notingher et al., 2004)
895	Phosphodiester, deoxyribose	(Ruiz-Chica et al., 2004)
890	Protein peaks, structural protein modes of tumours, β-anomers	(Hanlon et al., 2000, Dukor, 2002, Utzinger et al., 2001, Krafft et al., 2005)
891	Saccharide band (overlaps with acyl band)	(Krafft et al., 2005)
892	Backbone, C-C skeletal	(Chan et al., 2006)

921	Proline ring, glucose, lactic acid	(Kaminaka et al., 2002)
932	Skeletal, C-C, α-heix	(Stone et al., 2004)
936	*ν*(C-C) of α-helix conformation for proteins	(Bergholt et al., 2011a)
937	Proline (Collagen Type I), amino acid side chain vibrations of Proline and Hydroxyproline, C-C vibration of collagen backbone, glycogen, *ν*(C-C) residues of α helix of proteins, C-C stretching of proteins, C-O-C glycosides of carbohydrates	(Cheng et al., 2005, Frank et al., 1995e, Binoy et al., 2004, Lakshmi et al., 2002, Notingher et al., 2004)
938	Proline, Hydroxyproline, *ν*(C-C) skeletal of collagen backbone, C-C stretch backbone	(Cheng et al., 2005, Stone et al., 2004)
958	Carotenoids, cholesterol	(Stone et al., 2004)
1000	Phenylalanine, bound and free NADH	(Malini et al., 2006b)
1002	C-C aromatic ring stretching, Phenylalanine, Phenylalanine	(O Faolain et al., 2005, Malini et al., 2006b)
1003	Symmetric ring breathing mode of phenylalanine, C-C skeletal stretch	(Frank et al., 1995e) (Chan et al., 2006)
1031	CH_3 CH_2 bending modes of collagen and phospholipids, phenylalanine of collagen, proline, C-H in-plane bending mode of phenylalanine, C-H stretching of proteins	(Huang et al., 2003, Stone et al., 2002, Dukor, 2002, Chan et al., 2006, Stone et al., 2004)
1032	CH_3 CH_2 bending modes of collagen and phospholipids, Phenylalanine of collagen, Proline,	(Huang et al., 2003, O Faolain et al., 2005, Cheng et al., 2005, Frank et al., 1995e)
1039	Phenylalanine	(Ferraro et al., 2003)

1058	C-N stretching of amide III	(Bankapur et al., 2012)
1061	C-C in-plane bending	(O Faolain et al., 2005)
1062	C-C skeletal stretching	(Seballos et al., 2005)
1063	C-C skeletal stretch	(Koljenovic et al., 2005a)
1075	Symmetric stretching vibrations of phosphate group	(Sauer et al., 1994)
1096	Phosphodioxy (PO_2^-) groups	(Krafft et al., 2005)
1097	*v*(C-C), phospholipids	(Polomska et al., 2010)
1099	*v*(C-N)	(Cheng et al., 2005)
1100	C-C vibration mode of the gauche-bonded chain, amide III, *v*(C-C)- Lipids, fatty acids	(Seballos et al., 2005, Lakshmi et al., 2002, Shetty et al., 2006)
1124	v(C-C) skeletal of acyl backbone in lipids	(Cheng et al., 2005)
1125	Glucose, *v*(C-O) +*v*(C-C), disaccharides sucrose	(Shao et al., 2012)
1126	Paraffin, *v*(C-C) skeletal of acyl backbone of lipids, C-N stretching of proteins, *v*(C-O) *v*(C-C), disaccharides, sucrose	(Dukor, 2002, Cheng et al., 2005, Chan et al., 2006, Shetty et al., 2006)
1129	*v*(C-C) skeletal of acyl backbone in lipid lipids	(Cheng et al., 2005)
1130	C-C skeletal stretch transformation, Phospholipid structural changes (trans versus gauche isomerism), Acyl chains	(O Faolain et al., 2005, Andrus and Strickland, 1998, Krafft et al., 2005)
1156	C-C, C-N stretching of proteins	(Chan et al., 2006)
1153	Carbohydrates	(Dukor, 2002)

1170	C-H in-plane bending mode of tyrosine	(Stone et al., 2002, Stone et al., 2004)
1171	Tyrosine (collagen type I), (CH) Phenylalanine	(Cheng et al., 2005, Lakshmi et al., 2002)
1172	δ(C-H), Tyrosine (protein assignment)	(Huang et al., 2003)
1173	Cytosine, Guanine, Tyrosine (collagen type I)	(Ruiz-Chica et al., 2004, Cheng et al., 2005)
1174	Tyrosine, phenylalanie, C-H bendings protein	(Chan et al., 2006)
1207	Hydroxyproline, Tyrosine	(Raniero et al., 2011)
1208	ν(C-C_6H_5), tryptophan, phenylalanine, Adenine, Thymine, Amide III	(Huang et al., 2003, Huang et al., 2005, Chan et al., 2006)
1241	Asymmetric (PO_2^-) stretching modes of nucleic acids of malignant tissues	(Cheng et al., 2005)
1243	Amide III, asymmetric PO_2^- stretching modes of Phosphodiester, CH_2 wagging and C-N stretching of collagen, Pyrimidine bases such as thymine and cytosine	(Dukor, 2002, Cheng et al., 2005, Schulz and Baranska, 2007, Stone et al., 2004)
1244	δ(NH), *ν*(C-N) and *ν*(C-C) of amide III proteins	(Movasaghi et al., 2007d)
1245	Amide III	(Shetty et al., 2006)
1295	Amide III	(Carter et al., 2004)
1310	Lipid specific peaks	(Sigurdsson et al., 2004)
1316	Guanine	(Ruiz-Chica et al., 2004)
1317	Guanine (B,Z-maker)	(Ruiz-Chica et al., 2004)
1318	Guanine (ring breathing modes of the	(Chan et al., 2006, Lakshmi et al.,

	DNA/RNA bases. C-H deformation of proteins, Amide III of α-helix	2002)
1334	Tryptophan	(Rodrigues et al., 2010)
1337	Amide III, and CH_2 wagging vibrations from glycine backbone and proline side chains, Ring breathing mode of Adenine and Guanine bases, C-H deformation of proteins	(Dukor, 2002)
1338	Adenine	(Lin et al., 2011b)
1339	C-C stretch of Phenyl (I) and C_3-C_3 stretch and C_5-O_5 stretch CHα in-plane bend	(Fung et al., 1996)
1441	CH_2 scissoring & CH_3 bending in lipids, cholesterol and its esters, C-H bending mode of lipids	(Lau et al., 2005, Silveira et al., 2002)
1442	Fatty acids, CH_2 bending mode, CH_3, CH_2 deformation of collagen assignment, Triglycerides	(Hanlon et al., 2000), (Mahadevan-Jansen et al., 1997), (Frank et al., 1995e), (Silveira et al., 2002)
1446	CH_2 bending mode of proteins & lipids, CH_2 deformation	(Stone et al., 2004, Stone et al., 2002, Frank et al., 1995e)
1448	CH_2 CH_3 deformation, CH_2 deformation, CH_2, collagen	(Cheng et al., 2005, Frank et al., 1995c, Malini et al., 2006b, Kaminaka et al., 2001)
1449	C-H vibration of proteins and lipids	(Notingher et al., 2004)
1517	B-carotene accumulation (C-C stretching mode)	(Silveira et al., 2002)
1553	Tryptophan, ν(C=C), tryptophan (protein assignment), ν(C=C), porphyrin	(Huang et al., 2003)

1576	Nucleic acid mode	(Mahadevan-Jansen et al., 1997, Barr et al., 1998)
1577	Bound and free NADH, IgG	(Malini et al., 2006b, Lakshmi et al., 2002)
1579	Pyrimidine ring of nucleic acid bases	(Stone et al., 2004)
1580	C-C stretching	(Laska and Widlarz, 2005)
1582	δ(C=C), Phenylalanine,	(Huang et al., 2003)
1583	C=C bending mode of phenylalanine	(Lau et al., 2003b)
1584	C=C bending mode of phenylalanine, C=C olefinic stretch	(Lau et al., 2003b, O Faolain et al., 2005)
1604	Cytosine (NH2), Phenylalanine, tyrosine	(Naumann, 1998, Chan et al., 2006)
1605	Cytosine (NH_2), ring C-C stretch of Phenyl, phenylalanine, tyrosine, C=C (protein)	(Ruiz-Chica et al., 2004, Schulz and Baranska, 2007, Chan et al., 2006, Lakshmi et al., 2002)
1616	C=C stretching mode of tyrosine and tryptophan	(Stone et al., 2004, Stone et al., 2002)
1655	Amide I of collagen, C=C of lipids in normal tissues, ν(C=O) amide I, α-helix, collagen	(Cheng et al., 2005, Malini et al., 2006b, Huang et al., 2003, Tan et al., 2003)
1658	Amide I (α-helix)	(Lakshmi et al., 2002)
1659	Amide I vibration of collagen like proteins, C=O stretching vibration of α-polypeptides, cholesterol peak of atherosclerotic spectrum	(Min et al., 2005, Shaw and Mantsch, 1999, Silveira et al., 2002)
1665	Amide I of collagen	(Cheng et al., 2005, Frank et al., 1995c)

1666	Collagen	(Kaminaka et al., 2001)
1667	Protein band, C=C stretching vibrations, α-helical structure of Amide I, structural protein modes of tumours, carbonyl stretch C=O vibrations	(Hanlon et al., 2000, Dukor, 2002, Mahadevan-Jansen et al., 1997, Huang et al., 2003, Utzinger et al., 2001, Farquharson et al., 2005)
1668	Amide I	(Huang et al., 2003)
2722	Combination of s(CH_2)+ t(CH_2) of fatty acids	(Van De Ven et al., 1984)
2724	C-H stretches, stretching vibrations of CH, NH and OH groups	(Conroy et al., 2005, Krafft et al., 2005)
2725	C-H stretches, stretching vibrations of CH, NH and OH groups	(Krafft et al., 2005, Conroy et al., 2005)
2726	v(C-H) stretches, Stretching vibrations of CH, NH and OH groups	(Gallier et al., 2011) (Krafft et al., 2005)
2731	C-H stretches, stretching vibrations of CH, NH and OH groups	(Conroy et al., 2005, Krafft et al., 2005)
2732	C-H stretches, stretching vibrations of CH, NH and OH groups	(Conroy et al., 2005, Krafft et al., 2005)
2734	C-H stretches, stretching vibrations of CH, NH and OH groups	(Conroy et al., 2005, Krafft et al., 2005)
2849	CH_3 symmetric stretch of lipids	(Koljenovic et al., 2005a)
2880	CH_2 stretching of lipids	(Mazurek and Szostak, 2006)
2881	CH_2 asymmetric stretch of lipids and proteins	(Koljenovic et al., 2005f)
2932	CH_2 asymmetric stretch of lipids and proteins	(Koljenovic et al., 2005a)

2933	CH_2 symmetric stretch	(Koljenovic et al., 2005a)
2936	*v*(CH_3) of lipids	(Sigurdsson et al., 2004, Shetty et al., 2006)
2937	CH_3 stretching of proteins, CH stretching of proteins and lipids	(Sigurdsson et al., 2004, Shetty et al., 2006)
2934	CH_2 asymmetric stretch	(Koljenovic et al., 2005f)
2938	CH stretch of lipids and proteins	(Sigurdsson et al., 2004, Shetty et al., 2006)
3060	CH olefinic stretch	(Camerlingo et al., 2011)
3061	CH olefinic stretch (lipids)	(Camerlingo et al., 2011), (Zenone et al., 2007)
3062	Aromatic C-H stretching	(Sebag et al., 1993)

TMA Biopsy data

Patients trial number	TMA location	ER status	PR status	HER2 status	Histological grade	Tumour pathology
59	G02	1	3	4	2	Lobular
92	H07	1	1	4	1	Ductal NST
104	H08	1	1	4	3	Ductal NST
119	A01	1	3	4	2	Ductal NST
235	H09	1	1	4	2	Mixed, tubular/lobular
237	H10	1	1	4	1	Ductal NST
250	A03	1	3	4	2	Ductal NST
251	H11	1	1	1	3	Ductal NST
255	G03	1	3	4	3	Ductal NST
282	G04	1	1	4	2	Lobular
290	A04	1	3	4	1	Ductal NST
303	H13	2	2	4	2	Ductal NST
306	A05	1	3	4	3	Ductal NST
324	H14	1	1	2	2	Ductal NST
334	H15	1	1	2	2	Ductal NST
351	G05	1	3	4	3	Ductal NST
365	H16	1	1	3	1	Ductal NST
366	H17	1	1	2	1	Ductal NST
433	I01	1	1	2	1	Ductal NST
435	I03	1	2	2	3	Ductal NST
455	E04	1	2	1	2	Ductal NST
481	G07	1	3	4	2	Ductal NST
502	I02	1	1	4	3	Ductal NST
504	E13	1	1	4	2	Ductal NST
510	I04	1	1	4	3	Ductal NST
565	I05	1	3	2	2	Ductal NST
607	I07	1	3	9876	3	Ductal NST
643	I08	1	3	3	3	Ductal NST

653	G08	2	3	2	3	Ductal NST
669	A07	1	3	4	2	Ductal NST
768	I09	1	1	4	2	Ductal NST
769	I10	2	2	4	3	Ductal NST
784	E05	1	3	4	2	Mucinous
790	E14	1	1	4	2	Ductal NST
797	B07	2	3	4	3	Ductal NST
815	I11	1	1	4	1	Ductal NST
816	C14	1	1	3	2	Ductal NST
822	A08	1	3	4	1	Ductal NST
868	C15	1	1	4	2	Ductal NST
871	A09	1	3	4	1	Lobular
872	C16	2	2	4	2	Ductal NST
877	C17	1	1	4	2	Ductal NST
894	I13	1	3	2	3	Lobular
902	G09	1	3	4	2	Ductal NST
904	B08	1	3	4	3	Ductal NST
910	I14	1	1	1	3	Ductal NST
916	I15	1	2	2	2	Ductal NST
919	I16	1	1	2	2	Ductal NST
942	D01	1	1	4	2	Mixed, ductal/lobular
974	D02	1	1	4	3	Mixed, ductal/lobular
979	D03	1	1	2	3	Lobular
995	B09	1	3	4	3	Ductal NST
1046	I17	1	3	2	3	Ductal NST
1070	J01	2	1	1	3	Ductal NST
1072	E15	1	1	4	3	Ductal NST
1080	E07	1	3	4	2	Ductal NST
1084	B10	2	3	4	3	Ductal NST
1096	J02	1	1	3	2	Ductal NST
1120	J03	1	2	1	2	Ductal NST
1123	J04	1	1	4	2	Ductal NST
1144	J05	2	2	1	2	Ductal NST
1158	G10	1	3	4	3	Ductal NST

1165	J07	2	2	2	3	Ductal NST
1170	J08	1	1	2	1	Ductal NST
1183	J09	2	2	1	3	Ductal NST
1212	J10	1	1	2	1	Ductal NST
1229	G11	1	3	4	2	Ductal NST
1240	J11	2	2	4	2	Ductal NST
1254	J13	1	1	4	2	Lobular
1256	J14	1	1	2	2	Ductal NST
1274	D04	2	2	1	3	Ductal NST
1294	G13	1	3	4	3	Ductal NST
1335	B11	2	3	4	3	Ductal NST
1379	G14	1	2	4	3	Ductal NST
1412	J15	1	1	4	2	Ductal NST
1414	J16	1	1	2	3	Ductal NST
1415	J17	1	1	4	3	Ductal NST
1426	G15	1	3	4	3	Ductal NST
1429	G16	1	3	4	2	Ductal NST
1441	E08	1	3	4	3	Ductal NST
1474	B13	1	3	4	2	Ductal NST
1482	E03	1	3	4	9876	Missing
1494	E09	1	3	1	3	Ductal NST
1508	K01	1	1	4	2	Ductal NST
1520	B14	2	3	2	3	Ductal NST
1547	E10	2	2	2	3	Ductal NST
1569	D05	2	2	2	3	Ductal NST
1607	K02	1	3	3	3	Ductal NST
1616	E16	2	2	1	3	Ductal NST
1622	E17 and G01	1	1	4	2	Other
1635	K03	1	1	4	3	Ductal NST
1648	D07	2	2	1	3	Ductal NST
1650	B15	1	3	4	3	Ductal NST
1692	B16	1	3	4	2	Ductal NST
1693	B17	1	3	4	3	Ductal NST

1709	D08	1	1	1	3	Ductal NST
1747	K04	1	1	4	2	Ductal NST
1751	D09	1	2	2	3	Ductal NST
1764	A02	1	3	4	2	Ductal NST
1795	C01	1	3	4	3	Ductal NST
1866	K05	1	1	3	2	Ductal NST
1939	E11	1	3	3	2	Mixed, ductal/lobular
1948	D10	1	1	2	3	Ductal NST
1973	G17	1	3	4	2	Lobular
1981	K07	1	1	3	2	Lobular
1987	K08	1	1	2	2	Ductal NST
2000	K09	1	1	4	2	Ductal NST
2022	D11	2	2	1	2	Ductal NST
2028	K10	1	1	2	3	Ductal NST
2082	D13	1	1	4	2	Ductal NST
2139	K11	1	2	3	2	Ductal NST
2148	C02	1	3	4	3	Ductal NST
2162	K13	1	1	2	3	Ductal NST
2235	D14	1	1	1	2	Ductal NST
2244	K14	1	1	2	3	Ductal NST
2246	D15	1	1	2	3	Ductal NST
2251	D16	1	1	1	3	Ductal NST
2289	C03	2	3	4	3	Ductal NST
2303	H01	1	3	4	2	Ductal NST
2390	K15	1	1	3	3	Ductal NST
2401	K16	1	1	2	2	Ductal NST
2419	A10	1	1	2	4	Ductal NST
2427	D17	2	2	1	2	Ductal NST
2482	C04	1	3	2	3	Ductal NST
2496	K17	1	2	4	3	Ductal NST
2715	C05	1	3	2	3	Ductal NST
2716	H02	2	1	4	3	Ductal NST
2750	A11	1	2	2	4	Ductal NST
2767	C07	1	3	2	3	Ductal NST

2773	A13	1	1	4	1	Tubular
2792	C08	1	3	2	3	Ductal NST
2811	A14	2	2	2	2	Ductal NST
2844	E01	1	1	2	3	Ductal NST
2857	C09	2	3	2	3	Ductal NST
2858	B05	1	1	1	1	Ductal NST
2894	H03	1	3	4	2	Mixed, ductal/lobular
2933	A15	2	2	2	3	Ductal NST
2940	H04	1	3	4	2	Mixed, ductal/lobular
2974	C10	1	3	2	3	Ductal NST
3007	C11	1	3	1	3	Ductal NST
3020	A16	1	1	4	2	Lobular
3051	C13	1	3	1	3	Ductal NST
3098	H05	1	3	4	2	Ductal NST
3180	A17	1	1	4	3	Ductal NST
3204	E02	2	2	1	3	Ductal NST
3209	B01	1	1	4	2	Ductal NST
3234	B02	1	1	4	2	Ductal NST
3239	B03	1	1	4	2	Ductal NST
3289	B04	1	1	2	2	Ductal NST