MOLECULAR AND STRUCTURAL MONITORING OF EPITHELIAL MESENCHYMAL IN BREAST TISSUE

SITI NORBAINI BINTI SABTU

FACULTY OF SCIENCE
UNIVERSITI MALAYA
KUALA LUMPUR

2021

University of Malaya

MOLECULAR AND STRUCTURAL MONITORING OF EPITHELIAL MESENCHYMAL IN BREAST TISSUE

SITI NORBAINI BINTI SABTU

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSIC
FACULTY OF SCIENCE
UNIVERSITI MALAYA
KUALA LUMPUR

2021

University of Malaya

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **SITI NORBAINI BINTI SABTU**

Matric No: **17034152/3**

Name of Degree: **DOCTOR OF PHILOSOPHY**

Title of Thesis ("this Work"): **MOLECULAR AND STRUCTURAL MONITORING OF EPITHELIAL MESENCHYMAL IN BREAST TISSUE**

Field of Study: **MEDICAL PHYSICS AND APPLIED RADIATION**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date: 22 March 2021

Subscribed and solemnly declared before,

Witness's Signature Date: 22 March 2021

Name:

Designation:

University of Malaya

MOLECULAR AND STRUCTURAL MONITORING OF EPITHELIAL MESENCHYMAL IN BREAST TISSUE

ABSTRACT

The epithelial mesenchymal transition (EMT) is a crucial process in cancer progression and metastasis. Study of metabolic changes during the EMT process is important in seeking to understand the biochemical changes associated with cancer progression, not least in scoping for therapeutic strategies aimed at targeting EMT. Raman spectroscopy and synchrotron small angle x-ray scattering (SAXS) were used herein to study the metabolic changes associated with EMT in human breast cancer tissue. For Raman analysis, tissue from 23 patients were collected, comprising non-lesional, EMT and non-EMT breast cancer samples. Analysis was made in the fingerprint Raman spectra region (600 -1800 cm^{-1}). Multivariate analysis involving Independent Component Analysis (ICA), Principal Component Analysis (PCA) and Non-Negative Least Square (NNLS) were used to analyse the Raman spectra data. The results show significant differences between EMT and non-EMT cancers in lipid, protein, and nucleic acids. Using SAXS techniques, structural components of the tissues were examined at momentum transfer values between q = 0.2 nm^{-1} and 1.5 nm^{-1}. From the SAXS patterns, axial d-spacing and diffuse scattering intensity were observed to provide the greatest discrimination between the various tissue types. The overall intensity of scattering from cancerous regions is a degree of magnitude greater in cancer-invaded regions.Present results indicate that as aids in tissue diagnosis SAXS are capable of distinguishing areas of invasion by disease as well as delivering further information at the supramolecular level.

Keywords: Epithelial to mesenchymal transition (EMT), Breast Tissue, Diagnostic, Raman spectroscopy, Small Angle X-ray Scattering (SAXS)

University of Malaya

PEMANTAUAN MOLEKUL DAN STRUKTUR EPITELIUM MESENKIM DALAM TISU PAYUDARA

ABSTRAK

Peralihan epitelium ke mesenkim (EMT) adalah proses penting dalam perkembangan kanser dan metastasis. Kajian perubahan metabolik semasa proses EMT adalah penting dalam usaha untuk memahami perubahan biokimia yang berkaitan dengan perkembangan kanser, tidak kurangnya dalam skop untuk strategi terapeutik bertujuan menyasarkan EMT. Oleh kerana potensi untuk kepekaan dan kekhususan yang tinggi, spektroskopi Raman dan sinkrotron kecil sudut x-ray berselerak ('SAXS') telah digunakan di sini untuk mengkaji perubahan metabolik yang berkaitan dengan EMT dalam tisu kanser payudara manusia. Untuk analisis Raman, tisu daripada 23 pesakit telah dikumpulkan, yang terdiri daripada tidak lesi, EMT dan bukan EMT kanser payudara. Analisis telah dibuat di rantau spektrum cap jari Raman (600 -1800 cm^{-1}). Analisis multivariat yang melibatkan Analisis Komponen Bebas ('ICA'), Analisis Komponen Utama ('PCA') dan Persegi Kurang Bukan Negatif ('NNLS') digunakan untuk menganalisis data Raman. Hasilnya menunjukkan perbezaan yang signifikan antara kanser EMT dan bukan EMT pada asid lipid, protein, dan nukleik. Dengan menggunakan teknik SAXS, komponen struktur tisu diperiksa pada nilai pemindahan momentum antara $q = 0.2\ nm^{-1}$ dan $1.5\ nm^{-1}$. Dari corak SAXS, bersama paksi d-jarak dan penyebaran resapan intensiti diperhatikan untuk menyediakan diskriminasi antara pelbagai jenis tisu. Keamatan penyebaran keseluruhan dari kawasan kanser adalah tahap magnitud yang lebih tinggi di kawasan yang diserang kanser. Hasil yang ada menunjukkan bahawa sebagai bantuan dalam diagnosis tisu, SAXS mampu membezakan kawasan pencerobohan oleh penyakit serta menyampaikan maklumat lebih lanjut pada tahap supramolekul.

Kata kunci: Peralihan epitel ke mesenkim (EMT), Tisu Payudara, Diagnostik, spektroskopi Raman, Penyebaran Sinar X Sudut Kecil (SAXS)

University of Malaya

ACKNOWLEDGEMENTS

This thesis represents the hard work not only from me but from the whole research team. It is a pleasure to thank those who made this project possible. Primarily, I am heartily thankful to my first supervisor, Dr Siti Fairus Abdul Sani, whose encouragement, guidance and support from the initial until now, enabled me to develop an understanding of the subject and providing me with such a great opportunity to complete this research work. I have been extremely lucky to have a supervisor who cared so much about my work, and who responded to my questions and queries so promptly. Also, I would like to extend my gratitude to my second supervisor Prof. Dr Zurina Osman for her continued support and encouragement. This work would not have been possible without the guidance and professional advices and discussion from all research collaborators, in particular Professor Datuk Dr. Looi Lai Meng from the Department of Pathology, University of Malaya and Professor David Bradley from the Sunway University, Malaysia. I am also indebted to the scientific and technical staff of the Synchrotron Light Research Institute, Nakhon Ratchasima, Thailand for their valuable help in all respects during the completion of the project. I offer my regards and blessings to all of those who supported me especially my family (my mother, sisters, brother, nephews and niece) and all my friends for their continuous understanding who become a constant source of moral and emotional support. Thank you very much! Lastly, I take this opportunity to express my deep gratitude to the University of Malaya for a research grant of University of Malaya - Bantuan Kecil Penyelidikan (BK093-2016), and the Ministry of Higher Education, Malaysia for support under the Fundamental Research Grant Scheme (FRGS) FP032-2017A.

University of Malaya

TABLE OF CONTENTS

ABSTRACT iii

ABSTRAK iv

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vi

LIST OF FIGURES x

LIST OF TABLES xvi

LIST OF SYMBOLS AND ABBREVIATIONS xviii

LIST OF APPENDICES xxi

CHAPTER 1: INTRODUCTION **1**

1.1 Research Background 1

1.2 Research Objectives 5

1.3 Thesis Structure 6

CHAPTER 2: LITERATURE REVIEW **9**

2.1 Breast Cancer 9

2.1.1 Epithelial-to-Mesenchymal Transition (EMT) 11

2.1.2 EMT in Breast Cancer 15

2.1.3 Collagen 18

2.2 Raman Spectroscopy 19

2.2.1 The Applications of Raman Spectroscopy in Cancer Detection 22

2.2.2 The Application on Raman Spectroscopy on Breast Cancer 23

2.2.3 The Application on Raman Spectroscopy on Detection of EMT in Breast Cancer 31

2.3 Small Angle Scattering (SAS) 34

University of Malaya

2.3.1 Small Angle X-ray Scattering (SAXS) 34

2.3.2 Bragg's Law 37

2.3.3 SAXS in Breast Tissue 39

2.3.4 Axial (Meridional Peak Features) 41

CHAPTER 3: METHODOLOGY 45

3.1 Breast tissue samples 45

3.1.1 Raman Spectroscopy Measurement 45

3.1.2 Small Angle X-ray Scattering (SAXS) Measurement 47

3.2 Breast tissue samples preparation 49

3.2.1 Preparation of breast cancer tissue into FFPE blocks 49

3.2.2 Dewaxing of the FFPE Blocks 50

3.2.3 Determination of EMT status by immunohistochemistry (IHC) and scoring 51

3.2.4 Determination of tumour size, histological grade, estrogen receptor (ER), progesterone receptor (PR) and HER2 status 56

3.3 Experimental Method: Raman Spectroscopy 60

3.3.1 Instrumentation of Raman Spectroscopy 60

3.3.2 Raman Spectroscopy Measurement 63

3.3.3 Analysis of Raman Spectra 64

3.3.3.1 Pre-processing data 64

3.3.3.2 Multivariate analysis 65

3.4 Experimental Method: Small Angle X-ray Scattering (SAXS) 68

3.4.1 Instrumentation of Synchrotron radiation 68

3.4.2 Experimental Setup: Synchrotron source 69

3.4.2.1 Photon source and beamline optics 71

University of Malaya

3.4.3 Analysis of SAXS Spectra 73

3.4.3.1 Image Acquisition and Pre-processing 78

3.4.4 Statistical analysis 79

CHAPTER 4: RESULTS AND DISCUSSIONS 80

4.1 Raman Spectroscopy Data Analysis 80

4.1.1 Pre-processing data 80

4.1.2 Multivariate analysis: Paraffin removal process 84

4.1.2.1 FFPE dewaxed samples 85

4.1.2.2 FFPE waxed samples 92

4.1.3 Metabolite Characterisation of non-lesional, EMT and non-EMT tissues 99

4.1.3.1 Lipid 99

4.1.3.2 Protein 108

4.1.3.3 Nucleic Acids 115

4.1.4 Statistical Analysis 118

4.1.4.1 Principal Component Analysis (PCA) 118

4.2 Small Angle X-ray Scattering (SAXS) Data Analysis 122

4.2.1 SAXS analysis of 16 classified breast tissues 122

4.2.2 SAXS analysis of 51 blinded breast tissues 127

4.2.3 Statistical Analysis 137

CHAPTER 5: CONCLUSION 141

5.1 Raman Spectroscopy 141

5.2 SAXS 144

5.3 Future works 145

University of Malaya

REFERENCES.....148

LIST OF PUBLICATIONS AND PAPERS PRESENTED 164

APPENDIX 168

University of Malaya

LIST OF FIGURES

Figure 2.1 : Ductal and lobular cancer. The circles' area are showed the invasive carcinoma cells ("breast cancer: idc (invasive ductal carcinoma) – causes, symptoms and cure," 2017)…… 9

Figure 2.2 : The EMT process ("Epithelial – mesenchymal transition," 2017)……..... 12

Figure 2.3 : The EMT and MET in tumour invasion.……... 15

Figure 2.4 : The order of arrangement of fibrillar collagen in breast tissue. Image extracted from Sabeena Sidhu, 2009……..... 19

Figure 2.5 : The Rayleigh and Raman scattering processes…….....… 21

Figure 2.6 : Small angle x-ray scattering geometry ……..... 36

Figure 2.7 : Construction of the scattering vector, q with incoming photon, ki and the scattered photon, kf……..... 37

Figure 2.8 : Diagram shows the phase shift which causes constructive (left) and destructive (right) interferences. Image extracted from https:// commons. wikimedia. org/wiki/File :Loi _de _ bragg.png…… 38

Figure 2.9 : Diagram shows the x-ray diffraction. Image extracted from https://en.wikipedia.org/wiki/Xray_crystallography#/media /File:Bragg_diffraction_2.svg……. 39

Figure 2.10 : Typical one-dimensional profile extracted from a SAXS image of healthy tissue of present work. A fat peak is found in range 3 from 1.0 - 2.0 nm-1. The 3rd and 5th order peaks can be found in range 1 from 0.2 - 0.6 nm-1…….. 42

Figure 3.1 : Photomicrographs of five breast cancer samples illustrating their E-cadherin and Vimentin scores, captured at comparable magnifications. Positive IHC expression is depicted by brown colour staining of the cells. Images of samples EMT-11, EMT-20 and EMT-23 illustrate the strong Vimentin staining of stromal cells which are not considered for scoring of cancer cells. EMT-20 illustrates the difference in morphology of vimentin-positive tumour cells at the right lower triangle of the picture (which show

University of Malaya

the cancer characteristics of cellular pleomorphism and enlarged nuclei) from vimentin-positive stromal (benign) cells (which are elongated and have small bland nuclei). The column at the right indicates the designated EMT status of the illustrated samples….. 53

Figure 3.2 : Schematic diagram of Raman spectrometer and a resulting Raman spectrum…….. 61

Figure 3.3 : The components of Renishaw inVia Raman spectroscopy. Image extracted from https://www.ccmr.cornell.edu/wpcontent/uploads/sites/2/2016/08/Wire4_Training_Modules_Compilation-Part-1.pdf……. 62

Figure 3.4 : The procedure of pre-processing of Raman spectra…… 65

Figure 3.5 : The (a) raw and (b) pre-processed data of Raman spectra……..... 65

Figure 3.6 : The multivariate analysis for both groups of Raman spectra data... 66

Figure 3.7 : Set-up of the Beamline 1.3W: SAXS at Synchrotron Light Research Institute (SLRI) in Thailand..... 67

Figure 3.8 : Experimental setup of the SAXS experiment. Image extracted from Sabeena Sidhu, 2009..... 71

Figure 3.9 : Schematic diagram of the SAXS beamline. BM: bending magnet, WB slit: white beam slit, DMM: double multilayer monochromator, FM: focusing mirror, S1- S3: slits, IC: Ionization chamber, S: sample, BS: beam stop, PD: photodiode, CCD: charge-coupled device detector. Image extracted from Siriwart Soontaranon, 2011. 71

Figure 3.10 : Example of a SAXS pattern obtained from the 2-Dimensional detector 76

Figure 4.1 : The pre-processed Raman spectra from all the samples obtained from the average of all the samples. Figure label number 1 is for FFPE dewaxed samples and figure label number 2 is for FFPE without dewaxed samples. (Black line: Non-lesional; Red line: EMT; Green line: Non-EMT) . Note: the x-axis represents the Raman shift and y-axis is Raman intensity. The lines are offset for clarity…… 81

Figure 4.2 : The spectra of pure paraffin……. 82

Figure 4.3 : The Raman spectra of all types of samples following extraction of the paraffin wax signal. Figure label number 1 is for FFPE dewaxed samples and figure label number 2 is for FFPE without dewaxed samples. (Black line: Non-lesional; Red line: EMT; Green line: Non-EMT). Note: the x-axis represents the Raman shift and y-axis is Raman intensity. The lines are offset for clarity. 83

University of Malaya

Figure 4.4 : The comparison between pure paraffin, FFPE waxed and FFPE dewaxed samples. The lines are offset for clarity. 84

Figure 4.5 : Principal Component (PC) loadings for (a) non-lesional, (b) EMT and (c) non-EMT of FFPE dewaxed samples. The lines are offset for clarity. 87

Figure 4.6 : The mean of Raman spectra of (a) non-lesional, (b) EMT and (c) non-EMT breast tissue after paraffin removal process. 88

Figure 4.7 : PC loadings for (a) EMT and (b) non-EMT tissues of FFPE waxed samples. The lines are offset for clarity. 94

Figure 4.8 : The mean of Raman spectra of FFPE waxed of (a) EMT and (b) non-EMT breast tissue after paraffin removal process. 95

Figure 4.9 : Summary of CH group of lipid vibrational bands observed for dewaxed samples (Group 1) in non-lesional, EMT and non-EMT breast cancer tissue. The black colour trace represents the non-lesional tissues, red colour represents EMT tissues and green colour represents non-EMT tissues. Comparison of the relative intensity of CH group lipid Raman bands in dewaxed samples between non-lesional, EMT and non-EMT breast tissue samples are depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. (*) denotes a statistically significant difference between the groups reported by the ANOVA test. 102

Figure 4.10 : Summary of phospholipid vibrational bands observed for dewaxed samples (Group 1) in non-lesional, EMT and non-EMT breast cancer tissue. The black colour trace represents the non-lesional tissues, red colour represents EMT tissues and green colour represents non-EMT tissues. Comparison of the relative intensity of phospholipid Raman bands in dewaxed samples between non-lesional, EMT and non-EMT breast tissue samples are depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. (*) denotes a statistically significant difference between the groups reported by the ANOVA test. 103

Figure 4.11 : Summary of C-C stretching of lipid vibrational bands observed for dewaxed samples (Group 1) in non-lesional, EMT and non-EMT breast cancer tissue. The black colour trace represents the non-lesional tissues, red colour represents EMT tissues and green colour represents non-EMT tissues. Comparison of the relative intensity of C-C stretching lipid Raman bands in dewaxed samples between

University of Malaya

non-lesional, EMT and non-EMT breast tissue samples are depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. No statistically significant differences between the groups are reported by the ANOVA test……..... 104

Figure 4.12 : Summary of CH group lipid vibrational bands observed for waxed samples (Group 2) in EMT and non-EMT breast cancer tissue. The red colour represents EMT tissues and green represents non-EMT tissues. Comparison of the relative intensity of CH group lipid Raman bands in waxed samples between EMT and non-EMT breast tissue samples is depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. (*) denotes a statistically significant difference between the groups reported by the t-test. 105

Figure 4.13 : Summary of phospholipid vibrational bands observed for waxed samples (Group 2) in EMT and non-EMT breast cancer tissue. The red colour represents EMT tissues and green represents non-EMT tissues. Comparison of the relative intensity of phospholipid Raman bands in waxed samples between EMT and non-EMT breast tissue samples is depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. (*) denotes a statistically significant difference between the groups reported by the t-test. 106

Figure 4.14 : Summary of C-C stretching vibrational bands observed for waxed samples (Group 2) in EMT and non-EMT breast cancer tissue. The red colour represents EMT tissues and green represents non-EMT tissues. Comparison of the relative intensity of C-C stretching lipid Raman bands in waxed samples between EMT and non-EMT breast tissue samples is depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. No statistically significant difference between the groups reported by the t-test. 107

Figure 4.15: : Summary of Amide III vibrational bands observed for dewaxed samples (Group 1) of non-lesional, EMT and non-EMT breast cancer tissue. The black colour represents the non-lesional tissues, while red represents EMT tissues and green represents non-EMT tissues. Comparison of relative intensity of Amide III Raman bands

University of Malaya

for dewaxed samples of non-lesional, EMT and non-EMT breast tissue is depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. Note: the x-axis represents the tissue type and y-axis is Raman intensity. No statistically significant difference between the groups reported by the ANOVA test. 109

Figure 4.16 : Summary of CH group protein vibrational bands observed for dewaxed samples (Group 1) of non-lesional, EMT and non-EMT breast cancer tissue. The black colour represents the non-lesional tissues, while red represents EMT tissues and green represents non-EMT tissues. Comparison of relative intensity of protein Raman bands for dewaxed samples of non-lesional, EMT and non-EMT breast tissue is depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. Note: the x-axis represents the tissue type and y-axis is Raman intensity. (*) denotes a statistically significant difference between the groups reported by the ANOVA test. 110

Figure 4.17 : Amide I of non-lesional tissue…… 111

Figure 4.18 : Summary of Amide III vibrational bands observed for waxed samples (Group 2) of EMT and non-EMT breast cancer tissue. The colour red represents EMT tissues and green represents non-EMT tissues. Comparison of the relative intensity of protein Raman bands for waxed samples of EMT and non-EMT breast tissue us depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. Note: the x-axis represents the tissue type and y-axis is Raman intensity. No statistically significant differences between the groups are reported by the t-test. 112

Figure 4.19 : Summary of CH group protein vibrational bands observed for waxed samples (Group 2) of EMT and non-EMT breast cancer tissue. The colour red represents EMT tissues and green represents non-EMT tissues. Comparison of the relative intensity of protein Raman bands for waxed samples of EMT and non-EMT breast tissue us depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. Note: the x-axis represents the tissue type and y-axis is Raman

University of Malaya

intensity. No statistically significant differences between the groups are reported by the t-test. 113

Figure 4.20 : Summary of nucleic acid vibrational bands observed for dewaxed samples (Group 1) in non-lesional, EMT and non-EMT breast cancer tissue. The colour black represents the non-lesional tissues, while red represents EMT tissues and green represents non-EMT tissues. Comparison of the relative intensity of nucleic acid Raman bands for dewaxed samples in non-lesional, EMT and non-EMT breast cancer tissue is depicted in the bar chart. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. Note: the x-axis represents the tissue type and y-axis is Raman intensity. No statistically significant differences between the groups are reported by the ANOVA test. 116

Figure 4.21 : The nucleic acid vibrational bands of O-P-O stretch observed for waxed samples (Group 2) in non-EMT breast cancer tissue. Each column represents average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. The x-axis represents the wavenumber while the y-axis represents the intensity. No statistically significant differences between the groups are reported by the t-test..... 117

Figure 4.22 : PCA plots of all studied samples; non-lesional, EMT and non-EMT breast tissues of both dewaxed (A and B) and waxed (C) samples. 120

Figure 4.23 : Typical scattering profiles of four groups of breast tissue analyzed in present work…… 123

Figure 4.24 : Individual axial d-spacing (nm) points for the 16 groups of classified tissues, including non-lesional, benign, non-EMT (labelled as EMT negative) and EMT (labelled as EMT positive). Note: the data points are offset vertically for clarity…….. 126

Figure 4.25 : Wide scattering profile of blinded samples, including non-lesional, malignant; non-EMT (labelled as EMT negative) and EMT (labelled as EMT positive) breast tissues. 129

Figure 4.26 : Mean axial d-spacing (nm) for the 51 blinded groups of tissues; non-lesional and malignant; non-EMT (labelled as EMT negative) and EMT (labelled as EMT positive). The number of samples are shown in the brackets. Note: error bars represent standard deviation. Note: the data points are offset vertically for clarity…… 137

Figure 4.27 : PCA plot of all studied samples; non-lesional, benign, EMT (labelled as EMT positive) and non-EMT (labelled as EMT negative) tissues…… 140

University of Malaya

LIST OF TABLES

Table 2.1 : Summary of EMT markers and roles…….. 13

Table 2.2: : A range of published studies on EMT in breast cancers for the past 10 years. 16

Table 2.3 : Survey of Raman spectroscopy breast cancer research work carried out over the past decade. 25

Table 2.4 : Published EMT cancer progression studies over the past decade using Raman spectroscopic investigations. 32

Table 2.5 : A listing of advantages and disadvantages of Raman spectroscopy in breast cancer diagnostic and EMT studies..... 34

Table 2.6 : The SAXS features and their relation to the structural components in breast tissue. 40

Table 2.7 : Summary by the features studied by previous work using SAXS..... 40

Table 3.1 : The breast cancer tissue samples classified on the basis of pathology for Raman spectroscopy measurement. 48

Table 3.2 : The breast cancer tissue samples classified on the basis of pathology for SAXS measurement. 48

Table 3.3 : EMT scores and molecular classification of breast cancer samples 54

Table 3.4 : Pathological characteristics of the study samples……... 57

Table 3.5 : General of specifications of Renishaw inVia Raman spectroscopy……. 62

Table 3.6 : Components that make-up the Beamline 1.3W: SAXS and its functions…….. 69

Table 3.7 : Properties of the SAXS machine at Beamline 1.3W: SAXS. 70

Table 4.1 : The Principal Component (PC) numbers of non-lesional, EMT and non-EMT FFPE dewaxed samples. 86

Table 4.2 : Raman assignments of the prominent peaks from the non-lesional, EMT and non-EMT breast tissues upon paraffin removal. The error of the intensity of the non-lesional, EMT

University of Malaya

and non-EMT tissues from the range of 600-1800 cm^{-1} is 0.96, 0.95 and 0.87 respectively ……... 91

Table 4.3 : The Principal Component (PC) numbers of EMT and non-EMT FFPE waxed samples. 93

Table 4.4 : Raman assignments of the prominent peaks from the FFPE waxed of EMT and non-EMT breast tissues upon paraffin removal. The error of the intensity of the EMT and non-EMT tissues from the range of 600-1800 cm^{-1} is 0.79 and 0.81 respectively..... 97

Table 4.5 : Summary peak parameters for 16 classified breast tissues. ... 125

Table 4.6 : SAXS peak parameters for the 51 blinded individual breast tissue cases, each in FFPE (dewaxed) condition. Pathology reporting was performed after conduct of the experiment. 133

Table 4.7 : Summary peak parameters for the 51 blinded breast tissues, each in FFPE (dewaxed) condition…… 136

Table 4.8 : Kruskal-Wallis test for differences in non-lesional, benign, EMT and non-EMT tissues according to measurements. Note: (**) highly significant, (*) significant, (##) provisional significance……. 138

Table 4.9 : Dunn's post hoc test to assess the significant differences in measurements between non-lesional, benign, EMT and non-EMT tissues. Note: (**) highly significant, (*) significant, (##) provisional significance. 138

University of Malaya

LIST OF SYMBOLS AND ABBREVIATIONS

λ	:	Wavelength of the incident X- rays
θ	:	Angle between the incident and scattered X- ray beams
a-Si	:	Amorphous Silicon
ALDH1	:	Aldehyde Dehydrogenase 1
AUC	:	Area under the curve
c	:	Sample concentration
CCD	:	Charge-Coupled Device
CSC	:	Cancer Stem Cells
CTC	:	Circulating Tumour Cells
d	:	Periodicity of the order within the scattering medium
DCIS	:	Ductal Carcinoma In-situ
DMM	:	Double Multilayer Monochromator
ECM	:	Extracellular Matrix
EGF	:	Epidermal Growth Factor
EMT	:	Epithelial-to-Mesenchymal Transition
FFPE	:	Formalin Fixed Paraffin Embedded
FWHM	:	Full Width at Half Maximum
GA-QDA	:	Genetic Algorithm-Quadratic Discriminant Analysis
$\dot{\imath}_{bg,PD}$	:	Beam intensity measured by a photodiode at the beam stop
$\bar{I}_{dark}$	:	Dark current pattern
$\bar{I}_{sam}$	:	Sample pattern
ICA	:	Independent Component Analysis

University of Malaya

ICH-GCP	;	International Conference on Harmonization – Guidelines for Good Clinical Practice
IDC	:	Invasive Ductal Carcinoma
IHC	:	Immunohistochemistry
ILC	:	Invasive Lobular Carcinoma
ISH	:	*In Situ* Hybridization
$\vec{k}_i$	:	Incoming photon,
$\vec{k}_i$	:	Scattered photon
L	:	Sample-Detector Distance
LCIS	:	Lobular Carcinoma In-situ
MMPs	:	Metalloproteinases
MRI	:	Magnetic Resonance Imaging
n	:	Order of the diffraction maxima (integer)
NNLS	:	Non-Negative Least Square
O-P-O	:	Phosphodiester Backbones
PB	:	Peripheral Blood
PBC	:	Primary Breast Cancer
PCA	:	Principle Component Analysis
PC-LDA	:	Principal Component-Linear Discriminant Analysis (
PET	:	Positron Emission Tomography
q	:	Scattering vector
$\vec{q}$	:	Modulus of scattering vector
ROC	:	Receiver Operating Characteristic
SANS	:	Small Angle Neutron Scattering
SAXS	:	Small Angle X-ray Scattering

University of Malaya

SEM	:	Scanning Electron Microscope
SLRI	:	Synchrotron Light Research Institute
SPS	:	Siam Photon Source
t_{dark}	:	Accumulation time
t_{sam}	:	Exposure time
T_{bg}	:	Background Transition
ZEB	:	Zinc-finger E-box binding

University of Malaya

LIST OF APPENDICES

Appendix A: Characteristic Peak of Raman Spectra in Breast Cancer Studies. 179

University of Malaya

CHAPTER 1: INTRODUCTION

1.1 Research Background

Breast cancers, one of the most prevalent forms of primary malignancy in females, represent a major threat to the health and longevity of women worldwide (Dietze et al., 2015; Geen et al., 2016). The majority of deaths among breast cancer patients are due to invasion and metastasis, features related to a process known as the epithelial to mesenchymal transition (EMT) (Fedele et al., 2017; Felipe Lima et al., 2016). During the EMT process, epithelial cells lose their intercellular adhesion and apical basal polarisation, promoting migratory capacity through the basement membrane and into connective tissue, with associated metastatic spread and elevated resistance to apoptosis (Kalluri & Neilson, 2003; Kalluri & Weinberg, 2009). In particular, EMT is associated with reduced expression of the epithelial marker E-cadherin (cadherin signifying calcium-dependent adhesion of a class of type-1 trans-membrane proteins) and enrichment of transcription factors, Snail, Twist, Slug and ZEB1 mRNA, important repressors of E-cadherin, as well as mesenchymal markers such as N-cadherin, vimentin and fibronectin.

Verifying biomarkers related to the EMT process are expected to be highly important in targeting EMT-associated breast cancer progression, also in developing treatment effective in inhibiting metastasis. Notwithstanding the considerable diagnostic efficacy of current technologies, including clinical breast examination, screening mammography and tissue sampling (e.g. fine needle aspiration, core biopsy, and surgical excisional biopsy), the EMT process in cancer progression remains obscure, metastatic cancer involving multistep processes. In so doing it is hoped to identify new targets for the prevention of metastasis and ultimately to improve breast cancer patient survival rates. Acknowledging that the onset of many diseases is preceded by biochemical change (Robbins, 1974), metabolomic diagnostic techniques are now used to study, profile and fingerprint the metabolites involved in the cellular processes. Acknowledging that the onset of many

University of Malaya

diseases is preceded by biochemical change (Robbins, 1974), metabolomic diagnostic techniques are now used to study, profile and fingerprint the metabolites involved in the cellular processes. Metabolomics is a fast growing field of research in terms of transcriptomics and proteomics, involving multicomponent analysis of all metabolites in a cell, tissue, organism and biological fluid (Bino et al., 2004; Nambiar et al., 2010; Nicholson & Wilson, 1989; Raamsdonk et al., 2001; Yizhak et al., 2010). The elucidation of biochemical pathways via investigation of metabolite levels has long depended on the use of Mass Spectrometry (MS)(Mischak et al., 2009; Woo et al., 2009; Zürbig & Mischak, 2008) and Nuclear Magnetic Resonance (NMR) (Gebregiworgis & Powers, 2012; J. Zhang et al., 2012; Zira et al., 2010). However, while MS offers greater sensitivity in detection of metabolites, major limitations include greater expense in sample preparation and low analytical reproducibility (Lucarelli et al., 2015). Furthermore, the derivative processes used to improve analytical capabilities can also result in metabolite degradation. Moreover, with metabolomics often dealing with distinguishing features between two different biological states, there is additional need for improvement in the analysis of target compounds in difficult biological matrices. Such factors nothwithstanding, the complexity and high dynamic range of metabolite concentrations pose additional challenges to the efficacy of qualitative and quantitative analyses.

Sensing of the gross biochemical perturbations that are occurring in pathological conditions can be achieved through optical spectroscopic methods in unison with multivariate statistical tools data analysis, with changes reflected in the optical properties, including absorption, scattering and fluorescence (González-Solís et al., 2011; Pichardo-Molina et al., 2007). Recent years have witnessed the increasingly popular use of Raman spectroscopy, identifying unknown components in materials and biological samples, further probing primary, secondary, tertiary, and quaternary structures of large biological molecules (Sabtu et al., 2020). In respect of medical diagnostics, several advantageous

University of Malaya

features are on offer, including high chemical specificity and an ability to obtain molecular information in the absence of staining or labelling. The acquisition of spectra can also be performed *in vitro*, *ex vivo*, or *in vivo*, in each case avoiding disruption of the cellular environment. This is a major advantage given that most biological assays utilise chemical biomarkers, often requiring conditions non-native to the biological environment. Detailed characterizations are obtained through molecular-specific energy shifts of inelastically scattered light, present interests in associations with disease progression included, with molecular changes reflected in individual bands in the Raman spectra (Rehman et al., 2013). Although Raman spectroscopy has not yet been established in the clinic, this technique has shown great potential in characterizing and discriminating between non-cancerous and cancerous tissues (Dukor et al., 1998; Haka et al., 2002, 2009; Stone, 2001), also between different pathologic grades of breast and other epithelial cancers (Kamemoto et al., 2011; Robichaux-Viehoever et al., 2007; Stone et al., 2002). The use of Raman spectroscopy in breast cancer investigations over the past 10 years and more, including in study of EMT, has been evaluated most predominantly using cell lines and/or animal models (Aigner et al., 2007; Chaturvedi et al., 2016a; Damonte et al., 2009; Kim et al., 2016; Marro et al., 2014; Nieva et al., 2012; Yan et al., 2005; L. Zhang & Min, 2017).

In concomitance to this, collagens are major proteins that exist in the extracellular matrix (ECM), the degeneration and penetration of the ECM seeming to be significant processes relating to the morbidity and mortality of cancer (Campbell, 2018). Specifically, association of invasive carcinoma cells with the ECM has been associated with an increase in proteins which then disrupt the ECM, permitting the invasion of carcinoma cells, proliferating steadily to surrounding tissue. Several studies have demonstrated disturbed collagen structures, in particular the degree of deterioration corresponding with invasive carcinoma development in breast (D'Arrigo & Fentiman,

University of Malaya

2004; Pucci Minafra et al., 1986; Tanjore & Kalluri, 2006). Fibrillar collagen types I and III are dominant in ECM (Kauppila et al., 1998), molecular alterations in these being potentially linked with invasive carcinoma. In electron micrograph studies by others, a particular focus has concerned fibrillar collagen in tumour-bearing tissues, degradation in this being observed. The suggestion has been that this could arise from a failure of the altered collagen to aggregate in such a way that would allow formation of the bundles typical of healthy tissue (Pucci-Minafra et al., 1987; Pucci Minafra et al., 1986). Further to this has been the suggestion that impaired supramolecular organization occurring during cancer progression could provide for pathways engendering invasion of neoplastic cells (Kidane et al., 1999; M.flug & P.Kopf-Maier, 1995) . The organization of the collagen molecules within the tissue can be analyzed through diffraction peaks that reflect the regular packing of the collagen. As one of the prime components of breast tissue, collagen has been studied through a plethora of techniques inclusive of small angle x-ray scattering (SAXS) (Fernández et al., 2002; R. A. Lewis et al., 2000; Sidhu et al., 2008). SAXS patterns have previously been analyzed in investigation of the condition of collagen fibrils in breast tissue in attempts to find distinguishable characteristics relatable to breast disease, all containing collagen structures affected by invasive cancer (Arnedos et al., 2009; Falzon et al., 2006; Fernández et al., 2002). In present work, we have specifically utilized x-ray synchrotron radiation of SAXS to gather information on the axial spacing and intensity of scattering patterns, reflective of internal collagenous structure, the scattering profiles being indicative of structural detail at the supramolecular level.

In regard to the potential of Raman spectroscopy and SAXS measurements for breast tissue diagnostic, studies have been carried out to investigate a range of *ex-vivo* human breast tissue specimens, characterising the chemical morphological composition in corresponding macromolecules (Anwar et al., 2015; Howlader et al., 2014; S. Lee et al.,

University of Malaya

2014; C. H. Liu et al., 1992), also to predict the breast tissue disease state during EMT development and discriminate between different types of breast tissue, healthy, benign or malignant, focusing in particular on those that are non-EMT and those that are EMT, respectively. In particular, attention is paid for the former to the changes in lipid, protein and nucleic acid since these are the major structures responsible for biochemical changes in the process of cancer development (Rehman et al., 2013). In this respect, while numbers of studies have investigated use of Raman and SAXS for tissue recognition, none to our knowledge have investigated spectroscopic diagnostic (e.g. by chemical composition and structural component analysis, respectively) of EMT breast cancer tissues, a matter of interest in present work.

1.2 Research Objectives

To our knowledge, in utilising Raman spectroscopy and Small Angle X-ray Scattering (SAXS) to examine EMT tissues, this is the first time that such a larger scale investigation will have been conducted anywhere, going well beyond the previous studies carried out elsewhere, and certainly the first such study in Malaysia. We are seeking to deconvolve and track the molecular content, to determine the molecular structure of cancer cells during the EMT process, respectively. Each of these processes involves biochemical changes linked with the early metastatic phenotype of cancer cell. The important objectives of this project are:

1. To characterise the Raman spectra of non-lesional and malignant breast tissue of EMT, on the basis of their lipid, protein and nucleic acid which correlate features with biochemistry and tissue morphology.
2. To determine the axial spacing and intensity of scattering patterns, reflective of internal collagenous structure where the scattering profiles being indicative of structural detail at the supramolecular level using SAXS.

University of Malaya

3. To understand the biochemical changes associated with cancer progression especially on the content of collagens, lipids, proteins and nucleic acids in the tissues, in particular the molecular mechanism driving the EMT process.

1.3 Thesis Structure

The content of this thesis is structured into five chapters to achieve the main objectives as described in the previous section, providing a detail background on the physics, engineering, analytical methods and biophysics involved in this research. A further structure of the thesis is as follows:

1. Chapter 1

Chapter 1 provides an introduction to the subject area, outlining the motivation and goals of the work, going on to detail the structure of the thesis

2. Chapter 2

The next section of Chapter 2 provides the general knowledge of breast cancer, including the types of breast cancer, the risk factors of breast cancer and diagnosis techniques of breast cancer. The explanation on the biological process of EMT, the role of EMT and collagen in breast cancer are also described in this chapter. The physics behind the Raman spectroscopy along with the review on the applications of Raman spectroscopy in breast cancer and EMT for the past 10 years have been discussed too. Besides Raman, the detail of SAXS on the physics and the summary of SAXS applications in breast cancer studies by previous researchers have also been described.

3. Chapter 3

Chapter 3 presents the methodology for both Raman and SAXS measurements. This chapter begins with the classification of samples used in both methods. The basic components and instrumentation setup for both Raman spectroscopy and SAXS were

University of Malaya

presented in this chapter. Additionally, the analyses involved in the measurements were described here which include Principle Component Analysis (PCA), Independent Component Analysis (ICA), Non-Negative Least Square (NNLS), Kruskal-Wallis test and Dunn's post hoc test.

4. Chapter 4

Chapter 4 presents the results obtained in both Raman and SAXS measurements. In the Raman study, the classification of breast tissues involves pre-processing techniques to reduce the noise and eliminate the unwanted background as well as the multivariate analysis employed to the pre-processed Raman spectra for paraffin extraction and biomolecular characterisation in all types of tissues. Our focus only on the changes of lipid, protein and nucleic acid in the samples. As for the SAXS measurements, the parameters for the classifications of breast tissues were presented here which include d-spacing, the amplitude of the third, fifth, ninth and tenth-order axial Bragg peaks, the magnitude of integrated intensity and the full width at half-maximum (FWHM) of the triglyceride (i.e. fat) peak.

5. Chapter 5

Chapter 5 concludes the overall results and ideas on the application of Raman spectroscopy and SAXS in monitoring the molecular of EMT in human breast cancer tissue. Finally, an overview on the future works in the area of clinical diagnostic where the needs to develop alternative methods in which rapid, high spatial resolution, high sensitivity and specificity and accuracy are crucial for diagnostic application in order to understand the metabolic fingerprint and scaffolding structure of breast tissue. For the particular situations under investigation, use of such facilities is allowing information to be obtained at an unprecedented level of detail, yielding new understanding of the affected tissues and the progression of disease. It can potentially be exploited as diagnostic

University of Malaya

indicators of clinical conditions and may enhance better treatment and understanding of the pathogenesis of the underlying mechanisms involved in support of histology.

University of Malaya

CHAPTER 2: LITERATURE REVIEW

2.1 Breast Cancer

Breast cancer is the most frequent cancer and becomes a major threat to women today. This disease impacts over 1.5 million women each year and causes the greatest number of deaths among women today. In 2015, 570 000 women died from breast cancer, that is 15% of all cancer deaths among women (*Breast cancer*, 2017).

The most common types of breast cancer include Invasive Ductal Carcinoma (IDC), Invasive Lobular Carcinoma (ILC), Ductal Carcinoma In-situ (DCIS) and Lobular Carcinoma In-situ (LCIS) (*Types of Breast Cancer*, 2016).

Breast cancer is a cancer that develops from breast tissue which starts when cells in the breast begin to grow in an uncontrolled way. Usually breast cancer start either in the lobule's cells, which are the glands that make breast milk (called as lobular cancer) or the ducts that carry milk from the lobules to the nipple (called as ductal cancer) as shown in Figure 2.1.

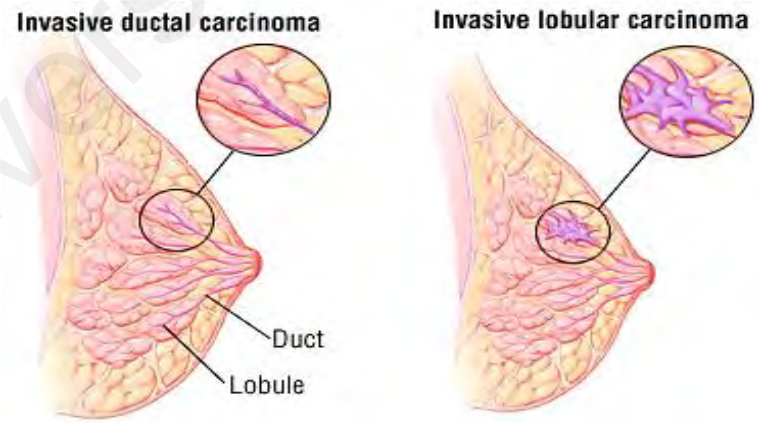


Figure 2.1: Ductal and lobular cancer. The circles' area are showed the invasive carcinoma cells ("BREAST CANCER: IDC (INVASIVE DUCTAL CARCINOMA) – CAUSES, SYMPTOMS AND CURE," 2017).

University of Malaya

There are several signs of breast cancer including, lump in the breast, change in breast shape, skin irritation or dimpling, breast or nipple pain, nipple retraction (turning inward), nipple discharge (other than breast milk) and redness, scaliness or thickening of the nipple or skin around the breast area. The risk factors of breast cancer include as described below (*Lifestyle-related Breast Cancer Risk Factors*, 2016):

- Obesity.
- Drinking alcohol.
- Lack of physical exercise.
- Having children late.
- Birth control – Studies found that women using oral contraceptives have slightly higher risk of breast cancer than women who have never used it. A few studies have looked at the effect of the birth control shot (Depo-Provera, an injectable form progesterone used for birth control shot) on the risk of breast cancer. Women currently using the birth control shot seem to have an increased risk in breast cancer.
- Hormone replacement therapy during menopause.
- Lack of breastfeeding – some studies suggest that breastfeeding may slightly lower the risk of breast cancer especially if it's continued for 1½ to 2 years.
- Genetics.

Many factors are involved in the initiation of breast cancer, but mortality is largely due to metastatic disease. The heterogeneity of breast cancers adds to a less than comprehensive description of progression of the biological features of the disease at first diagnosis, although the different biological behaviours and metastatic patterns observed among the distinct breast cancer phenotypes may suggest different mechanisms of invasion and metastasis. Thus said, specific genomic alterations have been utilised by other researchers

University of Malaya

(Eroles et al., 2012; Naderi et al., 2007; Wirapati et al., 2008; Yersal & Barutca, 2014) to define subtypes of breast cancer. However, the large number of genetic alterations present in tumour cells complicates the discrimination between genes that are critical for preserving the disease state and those that are merely coincidental. However, if breast cancer is detected and treated in its initial phase, statistics indicate some 90% of the patients can be cured. (Howlader et al., 2014)

Currently, a number of techniques are being used for early diagnosis and subsequent treatment of breast cancer. Early diagnosis of the disease certainly be expected to reduce the morbidity and mortility (McKinney et al., 2020; Myers et al., 2015). Standard techniques that are employed for breast cancer diagnosis include X-ray mammography, ultrasonography, positron emission tomography (PET) and magnetic resonance imaging (MRI). However, these techniques provide poor resolution, sensitivity and specificity (Shanmugam et al., 2016). It is also expensive and time-consuming. To gain a better understanding of malignancy, the histopathologist performed surgical procedures on cancer lesions for characterisation of the tissue but this procedure is known to cause trauma in patients (Shanmugam et al., 2016). For example, screening mammography aims to identify breast cancer at earlier stages of the disease, when treatment can be more successful, however the intrepretion of the the images obtained by mammography remains challenging including the false-positive results, which lead to patient anxiety, unnecessary follow-up and invasive diagnostic procedures (McKinney et al., 2020; Myers et al., 2015).

2.1.1 Epithelial-to-Mesenchymal Transition (EMT)

The epithelial-to-mesenchymal transition (EMT) is a biologic process by which the epithelial cells lose their cell polarity and cell-cell adhesion and acquire migratory and invasive properties characteristic of the mesenchymal cells (see Figure 2.2).

University of Malaya

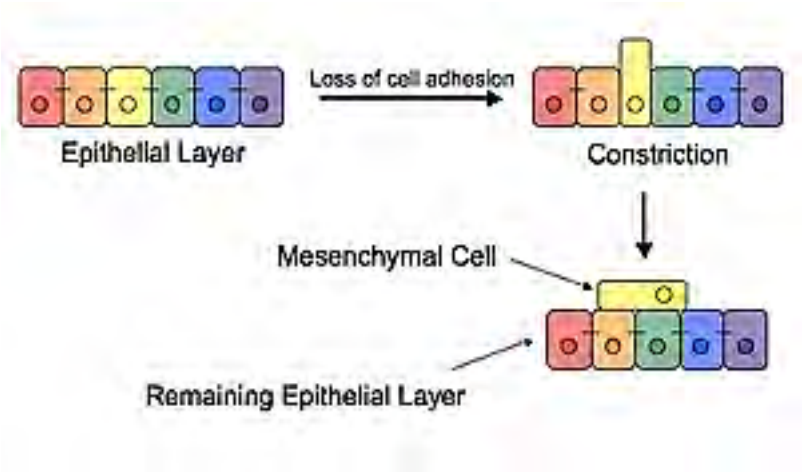


Figure 2.2: The EMT process ("Epithelial – mesenchymal transition," 2017).

EMT has been classified into three types as follow (Qureshi et al., 2015):

1. Type 1 EMT – occurs during embryogenesis and organ development.
2. Type 2 EMT – involved in wound healing, tissue regeneration and organ fibrosis.
3. Type 3 EMT – involved in cancer progression.

Several signalling pathways of EMT have been discovered in tumour progression. The process of EMT involves a loss of epithelial markers such as E-cadherin, claudin, occludin, plakophilin, cytokeratin and desmoplakins, gaining mesenchymal markers such as Vim1 (Vimentin), SNA1 (sometimes referred to as Snail), N-cadherin, Zeb 1 and Zeb 2. (Lamouille et al., 2014; Qureshi et al., 2015) A summary of EMT markers and roles is shown in Table 2.1.

University of Malaya

Table 2.1: Summary of EMT markers and roles.

Epithelial		Mesenchymal	
Marker	**Role**	**Marker**	**Role**
E-cadherin	Helps to attach epithelial cell sheets and maintaining epithelial cell polarity. (Hanahan & Weinberg, 2011; M. Y. Lee & Shen, 2012; Wang & Zhou, 2011)	**Vimentin**	Up-regulated in migratory cells. (Kokkinos et al., 2007)
Claudin	Transmembrane protein. (Kwon, 2013; Morin, 2005) Components of tight junctions, which are central for the regulation of paracellular permeability and the maintenance of epithelial cell polarity. (Kwon, 2013; Morin, 2005)	**N-cadherin**	Acts as a weak intercellular adhesion system. (Morin, 2005)
Occludin	Transmembrane proteins localized at tight junctions. (Ikenouchi et al., 2003) Responsible for establishing and maintaining epithelial cell polarity. (Ikenouchi et al., 2003)	**SNA1(Snail)**	Regulates E-cadherin. (Kokkinos et al., 2007)
Plakophilin	Group of proteins integral to the inner desmosomal plaque area. (Aigner et al., 2007) Act as scaffolding proteins required for proper desmosome assembly. (Aigner et al., 2007) Involve in desmosome-dependent cell adhesion and is specific to epithelial cells. (Valladares-Ayerbes et al., 2010)	**Zeb 1 & Zeb 2**	Zinc-finger E-box binding (ZEB) Initiate the loss of E-cadherin. (Hanrahan et al., 2017)
Desmoplakin	Essential for the maintenance of desmosomes in epithelial cells. (SATO et al., 2016)		
Cytokeratin	Intermediate filament proteins of epithelial cells which are expressed in a tissue- and differentiation state-specific manner. (Van der Velden et al., 1997)		

University of Malaya

Over the past few decades, intensive studies have been made to understand how cancer metastasizes and tremendous work focused on the role of EMT in cancer progression. Accumulating evidence suggests that EMT has become a critical role in cancer progression, through which tissue epithelial cancers invade and metastasize (Voulgari & Pintzas, 2009).

Nearly 80% of malignant tumours are developed from epithelial tissues. In several types of cancer cells, EMT plays a decisive role during metastatic progression (Qureshi et al., 2015). The loss of epithelial features and gain mesenchymal characteristic enables a cancer cell to spread and increase invasive properties. During the metastatic progression of carcinoma, polarised epithelial tumour cells acquired invasive and migratory features, leave the initial site, invade the basement membrane beneath, intravasate into blood or lymph vessels, move through the blood circulation, extravasate from the circulation, move into the secondary site and grow at the metastatic foci (M. Y. Lee & Shen, 2012).

The initiation of metastatic progression requires invasion which is enabled by EMT. Carcinoma cells in a primary tumour lose cell-cell adhesion and invade the basement membrane with increased migratory and invasive properties. The carcinoma cells later intravasate into blood or lymph vessels and move through the circulation. Later, these circulating tumour cells (CTC) exit from the blood circulation to form micrometastases. The EMT program enables the carcinoma cells to metastasis and then once they invade and reach their new home, they undergo MET, to revert back to epithelial form at these metastatic sites. Figure 2.3 showed the EMT and MET process in metastatic progression.

University of Malaya

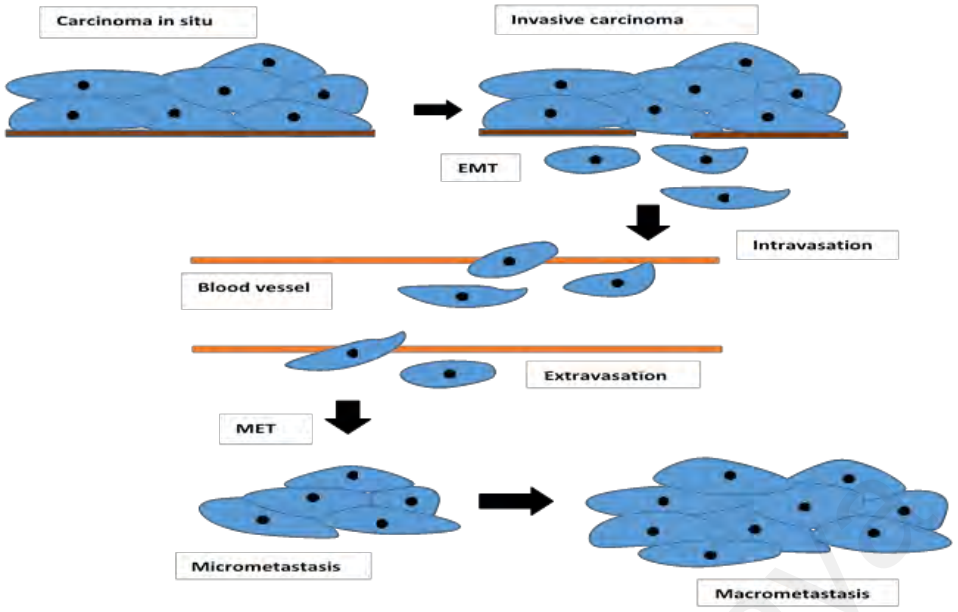


Figure 2.3: The EMT and MET in tumour invasion.

2.1.2 EMT in Breast Cancer

To date, the EMT program has been known in the metastases for epithelial cancers including breast cancer. One of the main challenges of conducting clinical studies on EMT is that it is difficult to detect individual cancer cells that have "transitioned or are transitioning" in real time in human tissue. Usually, at the time of histopathological assessment, the cell has either not yet undergone EMT or has already undergone EMT or MET. Therefore, the phase of transition cannot easily evaluate at a specific time or growth stage (Wu et al., 2016). The role of EMT in breast cancer has been demonstrated via a number of *in vitro* and *in vivo* studies by various researchers all around the world to provide the evidence on the presence of EMT in breast cancer.

As shown in Table 2.2, the majority of evidence of EMT in breast cancer progression is evaluated from studies in cancer cell lines and/or animal models. Hence, its relevance in human cancer is still under debate and investigation. Despite recent advances in technologies, the EMT programme in cancer progression much remains unclear as cancer is a complex and multi-process and EMT represents only part of the metastasis and tumour

invasion. Understanding of EMT programme in cancer progression will help in targeting EMT-associated cancer progression and give cancer patients hope to provide the best selection of treatment protocols.

Table 2.2: A range of published studies on EMT in breast cancers for the past 10 years.

Year	Research group	Cancer source	Major findings
2003	(Xue et al., 2003)	Mouse models	The study was of expression of FSP1/S100A4 in the EMT in PyV-mt mice. The results show the direct link of FSP1 expression to the development of metastatic potential in primary tumour.
2003	(Ackland et al., 2003)	Cell line (PMC42-LA)	Stimulation of PMC42-LA cells with epidermal growth factor (EGF) leads to epithelio-mesenchymal transition–like changes including up-regulation of vimentin and down-regulation of E-cadherin.
2007	(Aigner et al., 2007)	Cell line (MDA-MB-231)	Results showed that ZEB1 affects expression of several genes that are critically involved in epithelial polarity, supporting the regulatory role of ZEB1in EMT. This study provided strong evidence for a key function of ZEB1 in late-stage cancer progression.
2008	(Trimboli et al., 2008)	Mouse models	The group developed stromal- and epithelial specific cre-transgenic mice to directly visualize EMT during cancer progression *in-vivo*. The study claimed first direct evidence for EMT in breast cancer, suggesting that EMT in breast cancer can be specifically promoted by the myc oncogene.
2009	(Damonte et al., 2009)	Mouse models	Immunohistochemistry was demonstrated to provide reliable and specific confirmation of EMT-phenotype in mouse mammary tumours. This study provided experimental and immunohistochemical evidence of EMT showing that dual epithelial and mesenchymal staining of tumour spindle cells identifies some, but not all, EMT-type tumours in the mouse. Based on their observations, the group suggested that EMT is not a single tumour 'type' but is a process including a number of phenotypes and arise from a variety of mechanisms.
2009	(Aktas et al., 2009)	Blood samples	A total of 226 blood samples from 39 metastatic breast cancer patients were evaluated during a follow-up of palliative chemo-, antibody – or hormonal therapy for the expression of the stem cell marker aldehyde dehydrogenase 1 (ALDH1) and markers for EMT; TWIST, Akt2, and PI3Kα, the findings being correlated with the presence of circulating tumour cells (CTC) and response to therapy. The data obtained by the group indicate that a major proportion of CTC of metastatic breast cancer patients shows EMT and tumour stem cell characteristics.

University of Malaya

Table 2.2: Continued.

Year	Research group	Cancer source	Major findings
2011	(Kallergi et al., 2011)	Blood samples	Investigation on EMT markers expressed in CTC for early and metastatic breast cancer using peripheral blood samples. The results obtained clearly indicate that EMT markers of Twist and vimentin are expressed in CTCs of patients with early stage and metastatic breast cancer.
2012	(Mego et al., 2012)	Blood samples	Investigations made of mRNA transcripts of EMT-inducing transcription factors (TFs) in tumour cells from the peripheral blood (PB) of primary breast cancer (PBC) patients. The results obtained demonstrate that CTCs with EMT phenotype may occur in the peripheral circulation of PBC patients and that neoadjuvant therapies are unable to eliminate CTCs undergoing EMT.
2013	(Yu et al., 2013)	Blood samples	The study demonstrated a dual colorimetric in RNA *in situ* hybridization (ISH) analysis to characterise EMT in circulating tumour cells (CTCs) in breast cancer. The data obtained support the evidence of EMT occurring in relation to blood-borne dissemination of human breast cancer.
2016	(Kim et al., 2016)	Cell lines (MCF-7 and MDA-MB-231)	The group found epidermal growth factor (EGF)-stimulated activation of Smad2/3, upregulated several key EMT markers, inhibited E-cadherin expression, promoted EMT, enhanced migration and invasion in MCF-7 and MDA-MB-231 breast cancer cells.
2017	(Lakhtakia et al., 2017)	Human breast tissues	This study sought to identify the phenotype of EMT and cancer stem cells (CSC) in metastatic and non-metastatic breast cancer in Omani women and to establish their association with conventional clinico-pathological predictors of breast cancer. They found that there was a strong association of metastasis with younger women ≤ 40 years, lymphovascular invasion and expression of EMT. The expression of Vimentin, fibronectin and MMP-11 in stroma were definitive markers which are useful for EMT identification while CSC can be identified by the $CD44^{+}/CD24^{-/low}$ phenotype. EMT has a strong association with Ki67 PI while CSC has a strong association with HER2-like cancers.
2018	(Lin et al., 2018)	1. Cell lines (MC-7, T47D, SK-BR3, MDA-MB-231 & BT549) 2. Human breast tissues 3. Mouse models	The group reported that Notch 3 expression positively correlates with GATA binding protein 3 (GATA-3) and both are associated with estrogen receptor-α (ERα) expression in breast cancer cells. They have demonstrated *in vivo* and *in vitro* that Notch 3 inhibits EMT and metastasis of breast cancer by upregulation of GATA-3 expression. This study showed strong evidence demonstrating that the activation of Notch 3 maintains the EMT and suppresses metastasis through the GATA-3 transcription regulation in breast cancer.

University of Malaya

2.1.3 Collagen

Collagens are the main proteins that exist in the extracellular matrix (ECM) (Schonemark *et al.,* 1997). The ECM is made up of an intricate network of macromolecules which take up most of the space between cells. Collagen fibrils creates a matrix that aids in structural support in multiple directions in the connective tissue. The fibrils constitute groups of microfibrillar bundles, which are packed laterally in an approximately hexagonal close- packed arrangement. The collagen molecules have hydrogen bonds that connect each other in a staggered arrangement longitudinally and this structure is repeated periodically along the fibril axis.

A triple-stranded helical molecule of about 300 nm long make up the basic unit of collagen (see Figure 2.4). The axial arrangement of collagen molecules in their fibrils intrinsically exhibit an overlapping pattern, in which each molecule alternates a distance d from its neighbours, where d is the axial periodicity of the collagen. In a healthy breast tissue, collagen fibrils have a radius of about 90nm, and an axial periodicity of d = 65 nm (Fernandez *et al.,* 2004).

Whilst the exact role of the ECM in cancer remains an unsolved puzzle, it is reported that its structure is critically disturbed in malignant breast lesions (Poste G and Fidler I, 1990). In order to detect breast cancer, we are able to analyse the structural alterations in the collagen of ECM. Diffraction pattern of the breast tissue sample, which shows variation could possibly suggest damage to the collagen structure in the ECM, which would be associated with cancer. Collagen degradation and dissemination of the collagen bundles in invasive carcinogenic tissues have been observed with a scanning electron microscope (SEM), whereby a plentiful of thin fibrils disorderly moving in all directions throughout the stroma was seen (Pucci- Minafra *et al.,* 1998).

The disintegrated appearance of fibrils may denote that the new collagen is unable to aggregate properly to reassemble the typical bundles found in healthy tissues. In the study

University of Malaya

by Pucci- Minafra *et al.,* they deduced that the collection of altered collagen molecules during cancer progression damages the true supramolecular organization of the collagenous stroma and may provide a permissive pathway for the invasive growth of neoplastic cells.

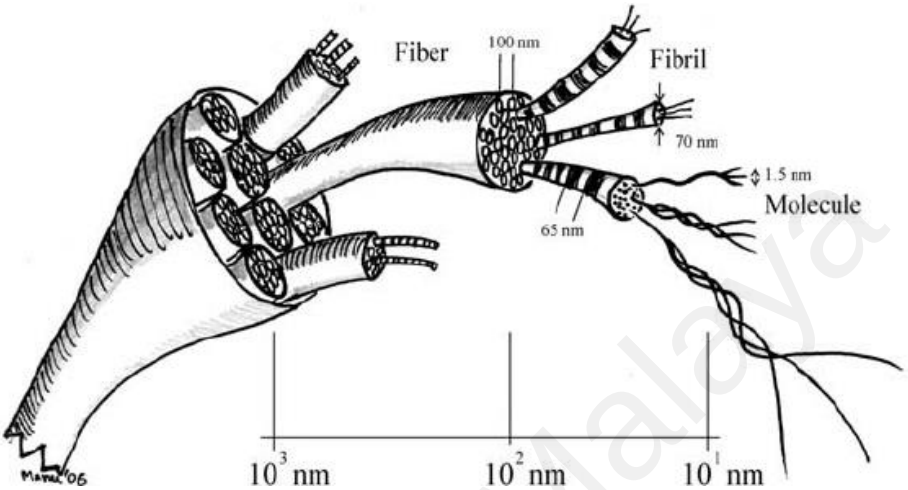


Figure 2.4: The order of arrangement of fibrillar collagen in breast tissue. Image extracted from Sabeena Sidhu, 2009.

2.2 Raman Spectroscopy

The first spectroscopic experiments were reported by Sir Isaac Newton in which he used a glass prism for spectral dispersion and the term spectrum was introduced in his classic volume *Opticks* in the 17th century. There were major achievements in spectroscopy in the early 19th century with the invention of the optical grating by Joseph Fraunhofer who developed the first grating spectroscope. Advanced developments in spectroscopic studies were achieved in the late 19th and 20th century by many researchers, and these achievements have contributed enormously to the development of spectroscopy.

Spectroscopy is an analysis technique to investigate the interaction of electromagnetic radiation with atoms and molecules. The main spectroscopies used to detect vibrations in molecules based on the mechanism of infrared absorption and Raman scattering. They are

widely used to provide information on chemical structures and physical form to identify substances from the characteristics of spectral patterns (fingerprinting).

Raman spectroscopy is a vibrational spectroscopy based on inelastic light scattering. Light scattering involving two-photon process, the incident and scattered photons. Mostly, scattered photons are of the same energy to that of incident photons, which is called elastic or Rayleigh scattering, but scattered photons occasionally gain or lose energy resulting in frequency shift and this scattering is known as inelastic or Raman scattering. The frequency difference between the incident and scattered photons is known as Raman shift, expressed in wavenumber which have units of inverse length.

Wave equation

$$\lambda = \frac{c}{v} \tag{2.1}$$

Whera λ is wavelength, c is the speed of light and v is the frequency of the radiation.

Photon energy from Planck's equation:

$$\Delta E = hv \tag{2.2}$$

Where h is Planck's constant.

Wave vector:

$$\bar{v} = \frac{v}{c} = \frac{1}{\lambda} \tag{2.3}$$

Raman shift:

$$\Delta\bar{\boldsymbol{v}} = \frac{\mathbf{1}}{\boldsymbol{\lambda_o}} - \frac{\mathbf{1}}{\boldsymbol{\lambda_1}} \tag{2.4}$$

where $\Delta\bar{v}$ is the Raman shift expressed in wavenumber, λ_o is the excitation wavelength, and λ_1 is the Raman spectrum wavelength

University of Malaya

Raman scattering is a weak phenomenon in that only approximately one in every 10^6 – 10^8 photons which scatter is Raman scattered. Figure 2.5 shows the basic processes of Rayleigh and Raman scattering.

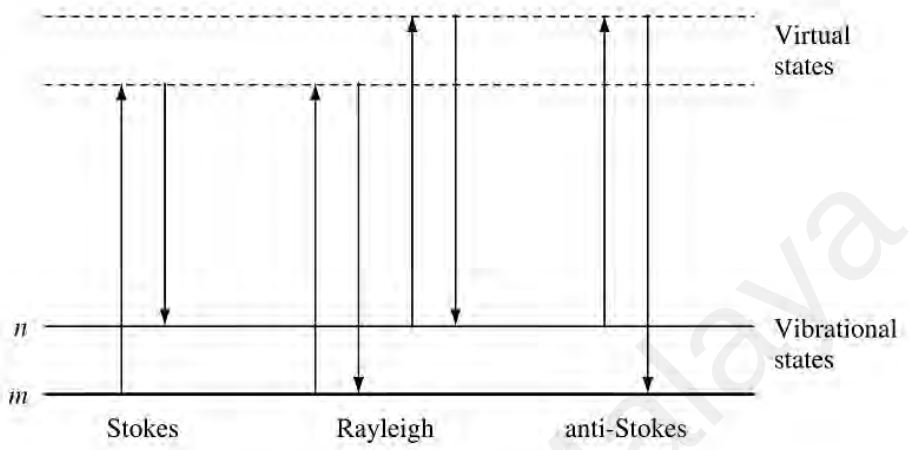


Figure 2.5: The Rayleigh and Raman scattering processes.

- Rayleigh Scattering
  - Most photons scatter this way.
  - It does not involve energy change.
  - The photons/lights return to the same energy state.
- Raman Scattering
  - The photons from the ground vibrational state, m, lead to absorption of energy by the molecule and the molecule gains vibrational energy, n (promotional to a higher energy vibrational state)
  - This is called Stokes scattering, the scattered photon has less energy than the incident photon.
  - However, due to thermal energy, some molecules may be present in an excited state, n. The molecules lose energy and the photon scatters from n state to the ground state, m. This is called anti-Stokes scattering, which is the scattered photon has higher energy than the incident photon.

University of Malaya

2.2.1 The Applications of Raman Spectroscopy in Cancer Detection

In recent years, the application of Raman spectroscopy has been remarkably increased in the field of medicine especially in the detection of a variety of cancers. Despite better diagnosis and improved treatment expanding patient's lifespan, the prognosis for many cancer diseases remains dire. Poor diagnosis of tumour can lead to incomplete excision of cancerous tissue or unnecessary removal of healthy tissue (Ember et al., 2017). Instead of removing tissue for pathological analysis, it is reasonable to use spectroscopy technique such as Raman spectroscopy to obtain diagnostic information about suspicious lesions. Raman spectroscopy can potentially identify the functional groups and chemical bonds that are present in the tissues or/and cells. Hence, it is possible to evaluate the structure of the protein, lipids, nucleic acids, and carbohydrates in a molecule and also the changes in their chemical structure due to the process of cancer development. Thus, this technique possibly monitors the progression of cancer and allows prediction of the chemical pathway of the disease progression. As cancer disease cause the changes in the molecular composition of the affected tissues, these changes can be reflected in the spectra. If the spectral changes are specific enough for a particular disease state, hence, these spectral can be used as a phenotypic marker of the disease.

For the past few decades, the potential of Raman spectroscopy in cancer detection has been actively explored. Many researchers have studied the potential of Raman spectroscopy to differentiate between cancerous, non-cancerous and pre-cancerous cells (dysplastic). This technique also has been used to understanding the progression of the disease at the molecular level. Such work has been done using live cells, pellet, formalin-fixed paraffin-embedded (FFPE) cells, scaffolds as well as whole tissues and body fluids.

A number of researchers have reported on the application of Raman spectroscopy on breast cancers as it is the most frequent cancer among women worldwide. The conventional

University of Malaya

techniques used for the breast cancers diagnosis face some limitations which are slow, invasive and expensive. In addition, these techniques offer limited information regarding the screened region of affected tissues and they do not provide sufficient information of the grade of malignancy of the tumour and the progression of the cancer. For instance, in mammography, 10-14% of clinically diagnosable cancers will not be detected (Nyström et al., 1993). Because of all these limitations, the applications of Raman spectroscopy on breast cancers have increased a great deal. This section will review on previous papers in the attempt to develop diagnostic techniques for breast cancers based on in vivo, ex vivo, in vitro and formalin-fixed paraffin by using Raman spectroscopy for the past 10 years.

2.2.2 The Application on Raman Spectroscopy on Breast Cancer

Several studies have shown that Raman spectroscopy is capable of differentiating cancerous and benign breast lesions from normal tissue with high sensitivity and specificities. Perhaps the most extensive work on the application of Raman spectroscopy for cancer detection has been for breast cancers as this cancer is the most common cancer among women worldwide. Although routine screening using the conventional technique can aid early detection of malignancy, lesions identified with this technique must be biopsied and evaluated histopathologically to determine the presence of malignancy. Contrasting with this are a number of current screening techniques for breast cancer detection, including mammography, ultrasound and magnetic resonance imaging (MRI). In evaluation of the performance of such modalities, each of the techniques has associated drawbacks and inaccuracies. (Bassett et al., 1991; Carter et al., 1989; Lennarth Nyström et al., 1993; Rehman et al., 2013) For example, while mammography is the most common technique for breast cancer diagnosis, 10-14% of clinical diagnosis of cancers are not detectable using this approach. (Lennarth Nyström et al., 1993; Rehman et al., 2013) Additionally, mammography can result in high false positive or negative outcomes, the prevailing sensitivity and specificity producing uncertainty. (Depciuch et al., 2016; Krishna

University of Malaya

et al., 2008) Ultrasound imaging is not as sensitive or specific as mammography (Rehman et al., 2013) and shares with MRI the problem of a low imaging resolution (3 mm in the case of MRI, 40 μm for digital mammogram techniques, (Karellas & Vedantham, 2009) 70 μm for the a-Si (amorphous silicon) mammography, (Sharyl J. Nass, I. Craig Henderson, 2001) and 1.0 to 1.5 mm for breast ultrasound imaging), (Malik et al., 2018) MRI also being expensive. (Depciuch et al., 2016) The limitations of each of the modalities together with the potential albeit low radiation risk from mammography encourages development of a rapid, cost-effective, highly sensitive technique, note being made of Raman spectroscopy in breast cancer studies, also its potential as a diagnostic tool.

Table 2.3 shows a survey of Raman spectroscopy breast cancer research work carried out over the past decade, demonstrating the potential of Raman spectroscopy to distinguish normal tissues in comparison with pathological tissues which achieved high sensitivity and specificity. The works also revealed that the normal breast tissue has higher content of lipids compared to benign and malignant tissues. In addition, the protein level was found higher in both benign and malignant tissues compared to normal tissues. The Raman spectra can be interrogated using principal component, linear-discriminant analysis (PCA/LDA), (Ramanujam et al., 1996) also in environments such as MATLAB® that compresses the information held by the spectra without losing any relevant data; (Jackson, 1991) LDA accentuates the differences in spectra measured from the different pathological groups, minimizing these spectral differences within each group (Otto, 1999).

A detailed database of Raman peaks is crucial to understand the chemical fingerprint of the samples. Most researchers largely depending on prior studies that focused on defining the band assignments of different peaks. Appendix A provides such information, detailing the major Raman peaks within the spectral range of interest in studies of breast cancer.

University of Malaya

Table 2.3: Survey of Raman spectroscopy breast cancer research work carried out over the past decade.

Year	Research Group	Samples	Raman system	Analysis method	Major findings
2008	(H. Abramczyk et al., 2008)	Human breast tissues	- Ramanor U1000 Raman spectrometer (Jobin Yvon) and spectra physics 2017-04S argon ion laser - 514nm and 488 nm with output power of 25mW and 100mW respectively - 0.5s integration time; - 8cm^{-1} spectral resolution	- Principal Component Analysis (PCA) - Matlab least-squares fitting algorithm	– The study showed excellent discrimination of normal and abnormal breast tissue and the ability of Raman spectroscopy to diagnose breast cancer accurately – Also shown was that normal breast tissue has a characteristic set of Raman bands: C-C and C=C stretching bands of carotenoids and C-H symmetric and asymmetric bands of lipids (fats). These were not visible in the benign tumour tissue or in the malignant tissue.
2009	(Haka et al., 2009)	Human breast tissues	- Raman system and Raman optical fiber probe - 830nm diode laser with excitation power between 100 and 150 mW - 10 to 30s integration time	- Spectral diagnostic algorithms - Non-negativity constrained least-squares. - X^2 analysis (used to calculate the goodness of fit and the error associated with model fitting)	– First prospective study of Raman spectroscopy in diagnosing normal, benign and malignant breast tissues in freshly excised surgical specimens. – The study justified use of a Raman spectroscopic algorithm for breast cancer diagnosis, developed *in vitro* on a large *ex vivo* clinical data set that closely mimics the *in vivo* environment. – Prospective testing of the algorithm on the clinical data set resulted in 83% sensitivity, 93% specificity, 36% positive predictive value and 99% negative predictive value in distinguishing cancerous from normal and benign tissues.

University of Malaya

Table 2.3: continued.

Year	Research Group	Samples	Raman system	Analysis method	Major findings
2010	(Yangzhe Wu et al., 2010)	Cell lines (MDA-MB-435 (435) and MDA-MB-435/BRMS1 (435/BRMS1))	- Renishaw inVia Raman spectrometer - 785 nm laser - 63x objective NA = 0.90 - accumulation at 10 s - wavenumber range of 500–1800 cm^{-1}	-Pre-processing: Baseline correction, Savitzky-Golay - Cosmic Ray Removal (CRR) and signal-to-baseline intensity	– The group demonstrated the feasibility of AFM and Raman techniques applications for *in situ* measurement of cellular biomechanics and biochemical components of breast carcinoma. – The study showed that the BRMS1 expression altered the physiochemical and biomechanical properties of the parental cell membrane surface and induced an increase in cell adhesion, cell elasticity and stiffness. This suggests that 435/BRMS1 cells are not metastatic and have low metastasis ability compared to 435 cells which are metastatic.
2011	(Halina Abramczyk et al., 2011a)	Human breast tissue	- Ramanor U1000 (Jobin Yvon, JY) Raman spectrometer - ion Ar laser (488,514 nm) - laser beam doubled SHG of the Nd:YAG laser (532 nm) - Average laser intensity was 10 mW - 50× objective 0.50 NA to the spot of 200 nm	Not stated	– The group reported Raman 'optical biopsy' images of human cancerous breast tissue. Demonstrated label-free Raman imaging capability in accurately characterising breast cancer tissue and distinguishing between normal and malignant types. The Raman images revealed an inhomogeneous distribution of different compounds, mainly carotenoids and that fatty acids in breast cancerous tissues differ significantly from that of the surrounding normal, non-cancerous breast tissue. – Results showed the composition of carotenoids and lipids in the cancerous breast tissues differing significantly from that of the surrounding normal breast tissues. Results also showed that label-free Raman imaging has the ability to accurately characterize breast cancer tissue and to distinguish between normal, and malignant types. The study discussed possibilities for Raman imaging replacing conventional biopsy.

University of Malaya

Table 2.3: continued.

Year	Research Group	Samples	Raman system	Analysis method	Major findings
2012	(Brozek-Pluska et al., 2012a)	Human breast tissues	- Ramanor U1000 Raman spectrometer (Jobin Yvon) - ion Ar laser 514nm - laser beam doubled SHG of the Nd:YAG laser (532 nm) - average laser intensity was 10 mW. - 50x objective to the spot of 650nm. - WITEC Project Plus software.	Basis analysis or the cluster method	– Data showed the most significant differences between non-cancerous and cancerous tissues to be found in spectral regions characteristic of the vibrations of carotenoids, lipids and proteins. – Spectra of normal breast tissues found to be dominated by carotenoids and lipids whereas spectra of cancerous tissues are characteristic of protein with low lipid and carotenoid content.
2013	(Damayanti et al., 2013)	Cell lines (MCF10A, MCF10AneoT, MCF10CA1h, and MCF10CA1a)	- Senterra Confocal Raman Spectrometer - 633 nm He-Ne laser with 20mWpower - 50× objective - 45 s integration time	- PCA - PLS-DA - Elastic net regularized regression	– First application of Raman spectroscopy in diagnosing non-malignant, premalignant, malignant and metastatic stages of breast cancer in a three-dimensional (3-D) cell culture model that closely mimics an *in vivo* environment. – The results obtained show for both cell models, 2-D and 3-D, that Raman spectroscopy has sufficient resolution to discriminate the four stages of breast cancer progression with near 100% accuracy. The spectra showed polysaccharide, lipid, nucleic acid, aromatic amino acid and extracellular matrix components involved in breast cancer staging.

University of Malaya

Table 2.3: continued.

Year	Research Group	Samples	Raman system	Analysis method	Major findings
2013	(Hu et al., 2013)	Human breast tissues	- Portable Raman spectroscopy system -785 nm laser with excitation power 5mW. - 60 integration time - Spectral region of 800 – 1800 cm^{-1}	-PCA - Support vector machine-recursive feature elimination (SVM-RFE)	– Diagnosis of normal, benign and malignant human breast tissues. – Data showed that the spectra of normal tissues are assigned to vibrational modes of lipids. In comparison, both benign and malignant tissues indicate the presence of higher protein content and lower lipids. The malignant tissues showed relatively more lipids and nuclei compare to benign tissues. – SVM-RFE produced better discrimination, sensitivity and specificity and is more suitable for diagnosing breast cancer disease than PCA.
2014	(Potcoava et al., 2014a)	Cell lines (T47D, MDA-MB-231, LNCaP and PC3)	- Raman microscope (Olympus IX70 inverted research microscope). - 785 nm diode laser with 45mW maximum power. - 60× objective NA=1.2 - 20s integration time - Coherent anti-Stokes Raman scattering (CARS) spectroscopy.	- Matlab least-squares fitting - PCA - Linear discriminant analysis (LDA)	– Investigated changes in intracellular lipids in breast cancer and prostate cancer cells upon exposure to the hormone medroxyprogesterone acetate (MPA) and synthetic androgen R1881. – The study demonstrated Raman-based methods can play an important role in determining the role of hormones content affecting the growth potential of cancers. – The results showed hormone-treated cancer cells T47D and LNCaP have greater number and size of intracellular lipid droplets and higher degree of saturation than untreated cells. The cancer cells of MDA-MB-231 and PC3 showed no significant changes upon treatment.

University of Malaya

Table 2.3: continued.

Year	Research Group	Samples	Raman system	Analysis method	Major findings
2015	(Halina Abramczyk et al., 2015)	Cell lines (MCF10A and MDA-MB-231)	- Raman microscope (Witec alpha 300 RSA) - 40x objective - 532nm laser with excitation power 38 mW - 0.5 s integration time (Raman image) - 20s integration time (Raman spectra) - Spectral ranges of 100 – 1780 cm^{-1} and 2600 – 3940 cm^{-1}.	- PCA - Matlab least-squares fitting	– Diagnosed the role of lipid droplets in cell line cultures that closely mimic an *in vivo* environment of various stages in human breast cancer tissue. – The number of cytoplasmic lipid droplets in live non-malignant MCF10A breast tissue cells is some 2 × lower than in mildly malignant breast cells (MCF7) and some 4 × lower than in highly malignant breast cells (MDA-MB-231). Hence, the higher amount of lipid droplets correlates with increased aggressiveness of the cancer.
2016	(Depciuch et al., 2016)	Human breast tissues	- SmartRaman DXR (Thermo Scientific) Raman spectrometer - 780nm laser and laser power 14mW - spectral range of 500–4000cm^{-1}	Not stated	– Comparative study of spectroscopic measurements on patients with the following: normal non-cancerous breast tissue; breast cancer tissues before chemotherapy; breast cancer tissues after chemotherapy; and normal breast tissues received around the cancerous breast region. – The spectra obtained from breast cancer patients showed changes in carotenoid and fat amount. Changes in carbohydrate and protein levels (e.g lack of amino acids, changes in amino acid concentration, structural changes) in comparison with normal breast tissues also observed. – Significant increase in the amount of lipids (2888 cm^{-1}, 2926 cm^{-1}) in the tissue after previous chemotherapy observed. – This study showed that Raman spectroscopy can be used in monitoring changes in breast cancer as well as in monitoring the effectiveness of chemotherapy.

University of Malaya

Table 2.3: continued.

Year	Research Group	Samples	Raman system	Analysis method	– Major findings
2017	(Q. Li et al., 2017)	Human breast tissues	- Ocean Optics QE65Pro miniature fibre optic Raman spectrometer - 785 nm laser and laser power 30mW - 30s integration time. - Spectral range from 700 cm^{-1} to 2000 cm^{-1}.	- Algorithm of adaptive net analyte signal weight K-local hyperplane (ANWKH) - K-local hyperplane distance nearest-neighbour (HKNN) - adaptive weight k-local hyperplane (AWKH) - support vector machine (SVM)	– This study demonstrated detection of breast cancer using a laser 785 nm miniature Raman spectrometer to develop a portable, low-cost and *in vivo* cancer diagnostic method. – ANKWH algorithm was proposed to increase the classification accuracy. – Results showed the proposed classification algorithm to be effective and can obtain high classification accuracy (93.1%), sensitivity (99.2%), and specificity (79.7%).
2018	(Jafarzadeh et al., 2018)	Cell line (SKBR3)	- Raman microscope (Thermo Nicolet Almega, USA). - 532 nm Nd-Yag laser (Almega) with laser power density at the sample was of ~100mW/µm3. - Acquisition time per cell is 64 seconds. - Spectral window: 600-3000cm^{-1}. - Resolution: 4cm^{-1} - 40x objective	- PCA - Genetic algorithm-quadratic discriminant analysis (GA-QDA)	– Detecting distinct radiobiological responses in human breast cancer cells using Raman spectroscopy.The group reported effects of radiation dose and post-irradiation time on molecular changes in human breast cancer. – The results showed the molecular changes were mostly reversible for sample with 2 Gy gamma irradiation). Doses > 4Gy showed irreversible changes in sample cells and these changes mainly seen at wavenumbers of 2700 – 2800 cm^{-1}. – This study opens up new possibilities in understanding some remarkable molecular and structural changes induced in breast cancer cells after radiotherapy.

University of Malaya

2.2.3 The Application on Raman Spectroscopy on Detection of EMT in Breast Cancer

As epithelial-to-mesenchymal transition (EMT) plays an important role in cancer progression, the study on EMT has arisen as one of the important studies in clinical research. Despite that the EMT process associated with tumour progression is now becoming well accepted, the molecular and cellular events that initiate the EMT have not been clearly elucidated (Ackland et al., 2003). The metabolism changes of EMT have been studied through analysis of mRNA and protein levels of key metabolic enzymes. Even though previous work on gene expression provides important understanding on metabolic changes during EMT in cancer, direct visualization of the relevant metabolites is lacking especially at the single-cell level (L. Zhang & Min, 2017). In study of EMT metabolic changes, important in developing cancer therapy strategies and drug resistance, (Marro et al., 2014; Rysman et al., 2010; Vazquez-Martin et al., 2008) it is crucial to develop alternative analytical techniques, Raman spectroscopy being well-positioned in this respect. Table 2.4 reviews key published literature focusing on Raman spectroscopy of EMT in cancer, each demonstrating utility of the method. Generally, the studies listed in Table 2.4 showed the capability of Raman spectroscopy in monitoring the biomolecular changes in EMT and it was shown that the lipid content in the cells was correlated with the EMT process. The Raman spectroscopy studies of EMT in breast cancer, while limited in number, all point to the considerable advantages of Raman spectroscopy applied in breast cancer diagnostics and EMT studies (Tables 2.5). The limitations of the Raman spectroscppy for molecular diagnostics also listed in Table 2.5. It is nevertheless apparent that further studies on the relationships between EMT markers and breast cancer are needed, helping to identify early stage breast cancer metastasis biomarkers.

University of Malaya

Table 2.4: Published EMT cancer progression studies over the past decade using Raman spectroscopic investigations.

Year	Research Group	Samples	Raman system	Analysis method	Descriptions
2012	(Nieva et al., 2012)	Cell lines (MDA-MB-435, MDA-MB-468, MDA-MB-231, SKBR3, MCF7 and MCF10A)	- Renishaw inVia Raman spectrometer - 514nm laser with 10mW excitation power. - 60x objective.	- PCA - Partial Least Square Discriminant Analyses (PLS-DA)	– Raman micro-spectroscopy used to explore the lipid phenotype associated with breast cancer malignancy. – Results obtained by the group suggested the lipid phenotype of breast cancer cells is a signal of the proclivity to mesenchymal transition related to the metastatic spread.
2014	(Marro et al., 2014)	Cell lines (MCF10A and MDA-MB-435)	- Renishaw inVia Raman spectrometer - 514nm laser with 5mW excitation power - 60x objective (Leica, 0.75 NA).	- Multivariate Curve Resolution (MCR) - PLS-DA	– First application of Multivariate Curve Resolution (MCR) to Raman spectra to assess the metabolic composition of breast cancer cells undergoing the epithelial to mesenchymal transition. – The analysis of MCF10A cells showed initial changes in the lipid content of cells associated with their EMT phenotype. The MCR analysis showed that the EMT affects the lipid profile of cells, increasing tryptophan but maintaining a low fatty acid content in comparison with highly metastatic cells.
2016	(Chaturvedi et al., 2016a)	Cell lines (HMECs, HMLE, HMLE-Twist, T47D, HMLE-Ras, BT-474 and MDAMB231)	- Renishaw inVia Raman spectrometer - 785nm with ~28 mW excitation power - 50x L (NA=0.50) long working distance objective. - 25s acquisitions for the lower wavenumber window (700-1800 cm^{-1}) and for the high wavenumber window (2800-3000 cm^{-1}).	Principal Component-Linear Discriminant Analysis (PC-LDA)	– Monitored biomolecular changes associated with the transformation of a normal cell into an invasive cell associated with breast cancer. – Based on the results obtained, the biomolecular changes in invasive cell lines could be distinguished from the non-invasive cell lines. – The study showed that lipid richness was correlated with the cell transformation (EMT) as well as cancer progression

University of Malaya

Table 2.4: continued.

Year	Research Group	Samples	Raman system	Analysis method	Descriptions
2017	(L. Zhang & Min, 2017)	Cell lines (MCF-7)	Raman scattering (SRS) microscopy	- Raman image comparison	– The group used metabolic labelling and SRS microscopy, a strategy of biorthogonal chemical imaging, to directly visualize the changes in anabolic metabolism during EMT in cancer at the single-cell level, showing that metabolic labelling coupled with SRS microscopy can be a straightforward method in imaging cancer metabolism. – Found that after EMT, there was less active anabolism of amino acids, glucose, and choline, reflecting slower protein and membrane synthesis in mesenchymal cells.

University of Malaya

Table 2.5 : A listing of advantages and disadvantages of Raman spectroscopy in breast cancer diagnostic and EMT studies.

The advantages of Raman spectroscopy	Non-invasive and non-destructive (Butler et al., 2016; Krishna & Rubina, 2015; Nieva et al., 2012; J Surmacki et al., 2013).
	Rapid (Butler et al., 2016; Krishna et al., 2006; Krishna & Rubina, 2015; Nieva et al., 2012).
	Cost effective (Kumar et al., 2007).
	Minimum sample preparation (Butler et al., 2016; Tu & Chang, 2012).
	High sensitivity and specificity (Marro et al., 2014).
	Provide direct biochemical information of the samples (Halina Abramczyk et al., 2011a).
	Provide detail chemical information of the tissue samples at molecular level (Haka et al., 2005; Vargas-Obieta et al., 2016)
	Can be used on variety types of samples (Ember et al., 2017).
	Provide high spatial resolution (Halina Abramczyk et al., 2011a; Butler et al., 2016).
The disadvantages of Raman spectroscopy	Long acquisition time due to the weak Raman signal (Eberhardt et al., 2015)
	Prone to interference from other molecules such as paraffin (Meksiarun et al., 2017)
	Susceptible to noise interference (low signal to noise ratio), fluorescence and turbidity (Gonzales et al., 2019)

2.3 Small Angle Scattering (SAS)

Small angle scattering is a technique based on deflection of collimated radiation away from the straight trajectory after it interacts with structures that are much larger than the wavelength of the radiation. The deflection is small about 0.1° - 10°, hence the name small-angle. There are two primary sources for small angle scattering experiments which are small angle x-ray scattering (SAXS) and small angle neutron scattering (SANS). The most important feature of the SAS method is its potential for analysing the inner structure of disordered systems and the application of this method is a unique way to obtain direct structural information on systems with random arrangement of density inhomogeneities.

2.3.1 Small Angle X-ray Scattering (SAXS)

Present work utilises Small Angle X-ray Scattering (SAXS) available at Synchrotron Light Research Institute (SLRI), Thailand. SAXS uses the coherent scattering obtained at

small angle to classify tissues (Changizi et al., 2005; Fernández et al., 2004). This technique can identify structures as small as 0.1 nm. Furthermore, SAXS has evolved into becoming a commonly used technique to characterize macromolecules in low resolution. The most important advantage of SAXS compared to other techniques is that it can be executed under an extensive variety of solution conditions, including near physiological state and for a wide range of molecular sizes. In the case of biological macromolecules, the advantage of SAXS over crystallography is that a crystalline sample is not required.

The organization of the collagen molecules within the tissue sample can be analyzed through the diffraction peaks, which result from the regular packing of the collagen. Diffraction effects, which are usually observed as constructive or destructive interference can be generated as small angle x-ray scattering is principally coherent. Through these interference patterns, which carry information regarding the molecular structure of the tissues, we are able to identify changes that are present due to cancer. This type of measurement displays the interference, which corresponds to the collagen structures within the tissue examined within the order of 10 to 100 nm (Fernández et al., 2002; R. A. Lewis et al., 2000). Diffraction patterns arise from the inhomogeneities in the distribution of electrons in the material. The radial distribution of the observed scattering pattern can also be used to infer the particle's density and surface area of the sample constituent. Studies have proved convincing contrast between healthy and diseased tissues (Conceião et al., 2009; Fernández et al., 2004; Sidhu et al., 2009). These studies have all been carried out using synchrotron radiation sources, utilizing the high photon coherence over long distances.

Coherent scattering follows Bragg's Law as follow:

$$n\lambda = 2d \sin \theta \tag{2.5}$$

University of Malaya

where n is the order of the diffraction maxima (integer), λ is the wavelength of the incident X- rays, d is the periodicity of the order within the scattering medium and θ is the angle between the incident and scattered X- ray beams. Figure 2.6 illustrates the scattering of SAXS, where the elastic scattering of the incident photons is conserved and interference in the wave fronts of the scattered photons can be seen on the detector. The rings form by the scattering pattern adhere to the Bragg's law creating a well-defined annulus of high and low intensities whose radii can be measured.

Scattering data can be obtained by linking the observed intensity to the modulus of scattering vector $\vec{\boldsymbol{q}}$

$$\vec{q} = \vec{k}_f - \vec{k}_i \tag{2.6}$$

$$|\vec{k}_i| = \frac{2\pi}{\lambda} = |\vec{k}_f| \tag{2.7}$$

Where the incoming photon, $\vec{\boldsymbol{k}}_{\boldsymbol{i}}$ and the scattered photon, $\vec{\boldsymbol{k}}_{\boldsymbol{f}}$ define $\vec{\boldsymbol{q}}$. From the geometry defined in Figure 2.7, the scattering vector can be described as:

$$q = \frac{n\lambda}{2\sin\theta} = \frac{2\pi n}{q} \tag{2.8}$$

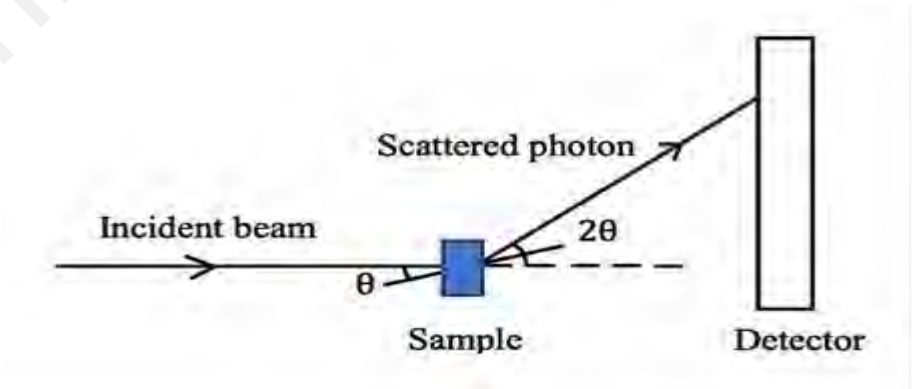


Figure 2.6: Small angle x-ray scattering geometry.

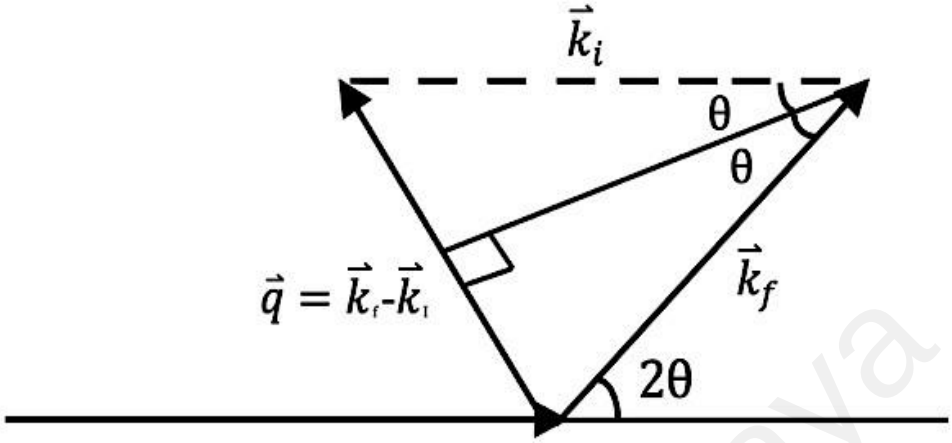


Figure 2.7: Construction of the scattering vector, $\vec{q}$ with incoming photon, $\vec{k}_i$ and the scattered photon, $\vec{k}_f$.

2.3.2 Bragg's Law

Bragg diffraction is commonly known as the interference of reflected electromagnetic waves by a solid. The orientationally symmetry of the azimuthal angle of the diffracted X-rays is caused by the random orientation of the sample's structure with respect to each other. Concentric cones of radiation that is being displayed as ring patterns on the detector is due to the circular symmetry in the azimuthal angle. The radii and intensity of the rings are unique to the scattering sample.

The scattering pattern that arises give information related to the structure of the object based on Bragg's Law. The movement of the electronic cloud in the atom upon collision with X-rays, an oscillating electric dipole is formed, which re-radiate waves with the same wavelength as the incident photon but in a slightly different direction. This phenomenon is known as Rayleigh scattering where the energy of an incident X-ray photon is small compared to the ionization energy of the target which occurs in the energy below 10 keV.

University of Malaya

The wavelength produced is equal to the incident X-ray or energy. The only difference is the trajectory direction. These re-emitted wave fields interfere with each other either constructively or destructively producing a diffraction pattern on a detector film (see Figure 2.8). The resulting wave interference pattern is the basis of diffraction analysis.

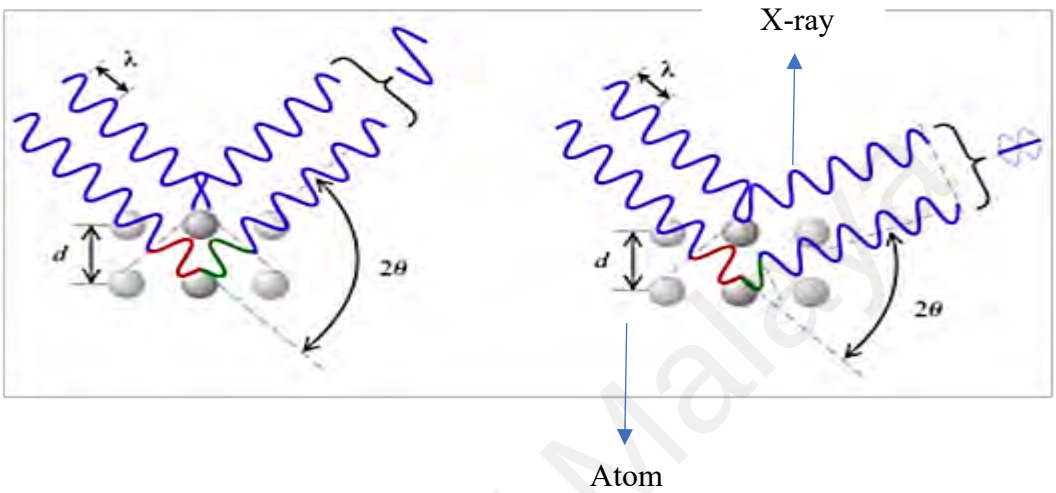


Figure 2.8: Diagram shows the phase shift which causes constructive (left) and destructive (right) interferences. Image extracted from https://commons.wikimedia.org/wiki/File:Loi_de_bragg.png.

According to Bragg's law, for constructive interference to be achieved, it requires the path length difference between X-rays scattered by crystals from different planes to be equal to an integer number, n of wavelength, λ. When the scattered waves interfere constructively, they remain in phase since the difference between the path lengths of two waves undergoing interference is given by 2d sin θ (see Figure 2.9).

University of Malaya

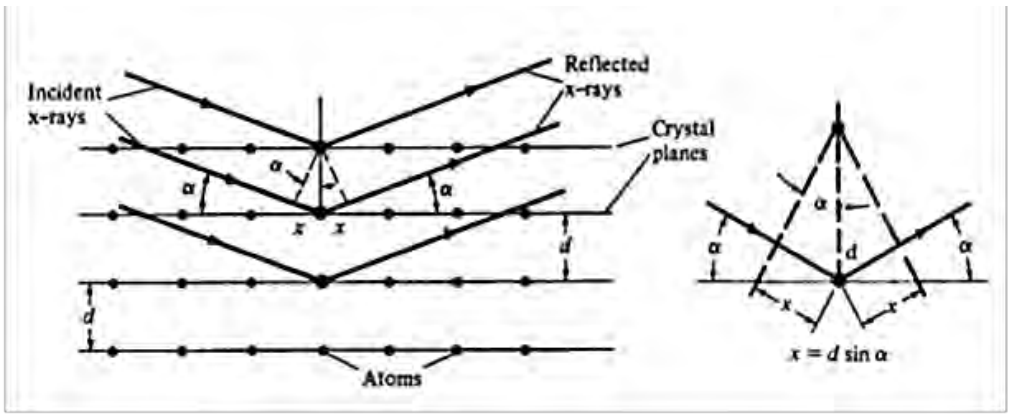


Figure 2.9: Diagram shows the x-ray diffraction. Image extracted from https://en.wikipedia.org/wiki/Xray_crystallography#/media/File:Bragg_diffraction_2.svg.

2.3.3 SAXS in Breast Tissue

SAXS has previously used to examine breast, all of which contain collagen structures affected by invasive cancer (Falzon et al., 2006; Fernández et al., 2002; R. A. Lewis et al., 2000). Table 2.6 shows the parameters used by previous studies that gave significant results to analyse the structural components of breast tissue. Table 2.7 summarizes the work done by previous researchers that studied breast cancer tissues using SAXS. All of the SAXS breast tissue studies previously have utilized synchrotron x-ray source in order to obtain fast and accurate data as in the present work. However, no study in our knowledge have worked on the epithelial-mesenchymal breast cancer tissue and determine its structural components in response to metastasis.

University of Malaya

Table 2.6: The SAXS features and their relation to the structural components in breast tissue.

SAXS feature	Physically represents
Collagen axial d- spacing	Spacing between the gap-overlap section of the collagen molecules
Collagen axial peak amplitude	Degree of long-range order in the collagen fibrils
Collagen axial peak width (full width half maximum- FWHM)	Wide peak correspond to a large variability in the d- spacing. Now peaks correspond to low variability in the d-spacing suggesting highly ordered fibrillar collagen overlap spacings.
Amorphous scatter	Proportional to the total area per unit mass. Tells is about variation in the electron density of the entire heterogeneous sample.
Amorphous scatter between the third and fifth order peak	Integrated intensity between q = 0.2-0.6 nm^{-1}. Used to describe the specific surface area of the scatterers.

Table 2.7: Summary by the features studied by previous work using SAXS.

Group	Features observed	Conclusions
Lewis et al., (2000)	**Intensity in axial and Bessel peaks**	Scattered intensity in the axial and Bessel peaks was much less in tumour tissue than in normal tissue and benign lesions.
	Spacing of 3rd axial peaks: d- spacing	Larger for benign lesions compared to malignant and normal tissue
Fernandez et al., (2004)	**Axial period of collagen structure**	Axial period of collagen is slightly larger in the fibrils surrounding invasive carcinoma than in healthy areas.
	Average intensity of scattering	The average intensity of scattering from cancerous regions is higher than the intensity from healthy regions.
Round et al., (2005)	**3rd order axial peaks**	Differences were observed in normal samples up to 6 cm away from tumour site
	Equatorial peak area	Obtained difference between normal and benign tissues but no differences observed in cancer-invaded tissues.

University of Malaya

Table 2.7: continued.

Group	Features observed	Conclusions
Conceicao et al., (2009)	**Diffuse scattering**	The diffuse scattering from tumour samples is higher than from benign and normal samples.
	Third- order d- spacing	The third-order axial spacing is larger in malignant lesion than in normal tissue or benign lesion.

2.3.4 Axial (Meridional Peak Features)

Collagen is a long periodic structure, which produces maxima in scattered intensity on SAXS images, where the sharpened and width of the peaks reflects the regularity in electron density. For well-oriented systems such as collagen, the d-spacing can be easily read from the axial Bragg peaks. The SAXS patterns from oriented collagen fibrils have distinct features in the meridional direction due to this axial periodicity (Suhonen et al., 2005).

Axial peaks from the axially periodic collagen structures are wider for malignant than for healthy tissue, which has been interpreted as a decrease of ordering with malignancy. Figure 2.10 shows an example of the scattering vector versus the intensity of a healthy breast tissue sample imaged using SAXS at the SLRI. The third and fifth order axial peaks can be seen at $q = 0.49\ nm^{-1}$ And $q=0.56\ nm^{-1}$ respectively. Most SAXS studies dealing with biological tissues such as the breast (Changizi et al., 2005; Falzon et al., 2006; Fernández et al., 2004; R. A. Lewis et al., 2000; Robert A. Lewis et al., 1999) emphasize on the odd ordered Bragg Peaks to obtain a sense of periodicity within the collagen present in the sample. Specifically, in the current study, special attention was made to the 3rd and 5th order Bragg peaks.

University of Malaya

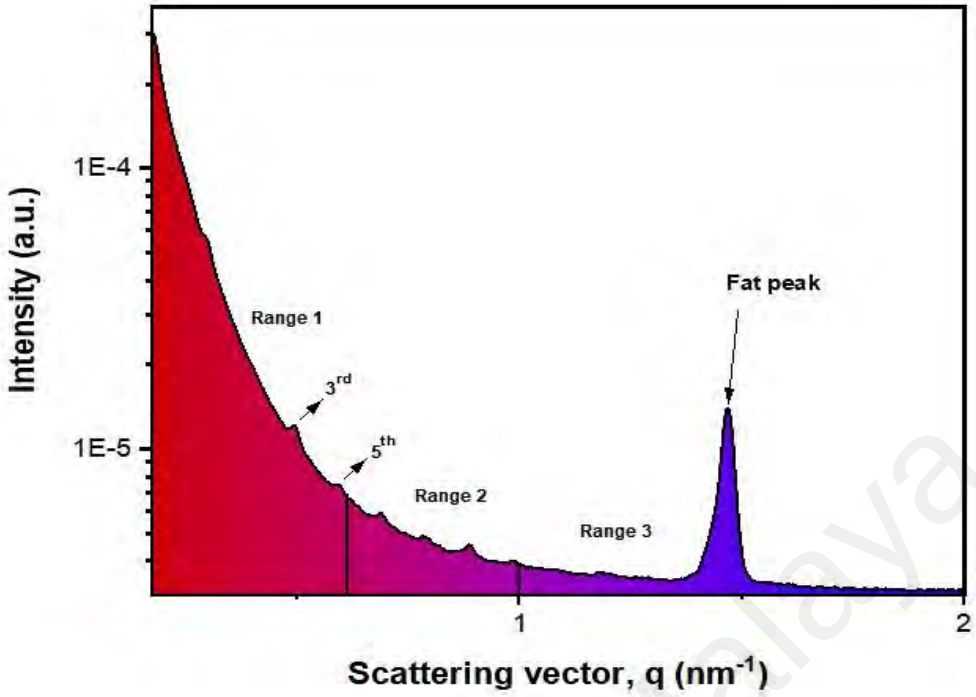


Figure 2.10: Typical one-dimensional profile extracted from a SAXS image of healthy tissue of present work. A fat peak is found in range 3 from 1.0 - 2.0 nm^{-1}. The 3rd and 5th order peaks can be found in range 1 from 0.2 - 0.6 nm^{-1}.

Similar experimental methods with the utilization of synchrotron radiation were practised by two independent groups, where SAXS was being implemented to investigate the condition of collagen fibrils in the breast tissue in attempts to find distinguishable characteristics that can be related to breast disease. Lewis et al., 2000 acquired SAXS images by averaging the whole tissue sample whilst SAXS images were acquired by Fernandez et al., 2004 by selecting several points on the tissue sample (1- 0.25 mm intervals) and considered each image as a discrete observation relative to the sample pathology (Fernández et al., 2004; R. A. Lewis et al., 2000). Both studies signified that there appears to be an increase in the d-spacing in invaded tissues compared to healthy or benign tissues. The axial d-spacing of collagen in invaded adipose tissues discovered by Fernandez et al. is 65.3 ± 0.2 mm and 65.0 ± 0.1 mm in benign lesions and healthy tissues. On the other hand, the d-spacing of collagen found by Lewis et al., 2000 for invasive carcinoma and normal tissues is within the range of 64.7 - 65.22 nm, whereas a smaller

d-spacing was obtained for benign tissues between 64.4 and 64.7 nm (R. A. Lewis et al., 2000).

Furthermore, Lewis et al. noted that a lower scattered intensity of the axial peaks can be found in tumour tissue compared to either normal tissue or benign lesion aligned with other work (Tanjore & Kalluri, 2006; Sidhu et al., 2014). Moreover, a higher total scattering intensity was signified by Fernandez et al., 2004 for tissues disrupted by cancer than in areas of healthy tissue which is attributable to the specific area of the scatters in tumours being enlarged (Fernández et al., 2004). Scattering vector, q on the range of q = 0.20 nm^{-1} and 0.56 nm^{-1} was considered and it was found that lower intensity corresponds to healthy tissue, whereas slightly higher intensity in benign lesions and producing highest intensity in malignant lesions. Both studies suggest that the axial d- spacing is larger in invaded tissues than in healthy or benign, but the small differences in the absolute d- spacing values between the two studies can perhaps be attributed to differences in experimental conditions or calibrations.

In addition, the peak amplitude and peak full width half maximum features have shown differences between diseased and healthy breast tissues. The peak amplitude indicates the degree of long-range order in the collagen fibrils which arises from the fact that x-ray scattering is a process of constructive interference. Round et al., 2005 found that the amplitude of the 3rd order axial peak was higher for benign tissues compared to malignant tissues (Round et al., 2005). The reduction in the peak amplitude for malignant tissues can be attributed to the action of Metalloproteinases (MMPs) on the malignant tissue.

The FWHM of the peak describes the distribution of the particle sizes within the sample. In the case of collagen, the FWHM related to the distribution of axial d-spacings

in the fibrillar collagen. A narrow peak related to a low variability in the axial d-spacings which indicates that all of the fibrils are very similar in terms of their d-spacings and structure and vice versa for wide peaks.

University of Malaya

CHAPTER 3: METHODOLOGY

3.1 Breast tissue samples

3.1.1 Raman Spectroscopy Measurement

Present study investigates non-lesional, EMT and non-EMT human breast cancer tissues, from 23 patients. The tissues were acquired following fully informed consent from patients who were subject to surgical treatment, with approval for use in research granted by the Medical Research Ethics Committee, University Malaya Medical Centre in accordance with the International Conference on Harmonization – Guidelines for Good Clinical Practice (ICH-GCP) and Declaration of Helsinki. The samples divided were into two groups; the first group comprises formalin-fixed and paraffin embedded (FFPE) dewaxed samples and the second group comprise FFPE samples waxed (Table 3.1). The rationale for comparing dewaxed and waxed FFPE samples was to consider the extent to which wax (paraffin) impregnation of cancer tissues affected the Raman spectroscopy findings. The study was piloted with waxed samples (10 samples – Group 2) without inclusion of any non-lesional tissue. However, this was discontinued in favour of dewaxed samples (27 samples – Group 1) to eliminate the confounding effect of wax on the Raman analysis. Group 1 was the main study cohort and included non-lesional material whenever this was available for the cohort cases in the pathology archive. Although not comparable in sample size, both Group 1 and Group 2 findings are shared in this report for completeness of data. There was a duplication of four cancers (2 EMT positive and 2 EMT negative) where both waxed and dewaxed samples were analysed. These are marked in Table 3.1. Since the study involves human samples available in the diagnostic pathology archives, it was limited by availability of such tissues. Only 10 breast cancers (4 EMT positive and 6 EMT negative) had non-lesional tissue from the same excision (mastectomy) available in the archives.

University of Malaya

Table 3.1: The breast cancer tissue samples classified on the basis of pathology for Raman spectroscopy measurement.

Groups	Sample number	Sample pathology		Material source
Group 1	EMT-9	Breast cancer-EMT	Same patient	FFPE (dewaxed)
	EMT-10	Non-lesional		FFPE (dewaxed)
	EMT-11	Breast cancer- EMT	Same patient	FFPE (dewaxed)
	EMT-12	Non-lesional		FFPE (dewaxed)
	EMT-13	Breast cancer-EMT	Same patient	FFPE (dewaxed)
	EMT-14	Non-lesional		FFPE (dewaxed)
	EMT-15	Breast cancer-EMT	Same patient	FFPE (dewaxed)
	EMT-16	Non-lesional		FFPE (dewaxed)
	EMT-17	Breast cancer-EMT		FFPE (dewaxed)
	EMT-18	Breast cancer-Non-EMT		FFPE (dewaxed)
	EMT-19	Breast cancer-EMT		FFPE (dewaxed)
	EMT-20	Breast cancer-EMT		FFPE (dewaxed)
	EMT-21	Breast cancer-EMT		FFPE (dewaxed)
	EMT-22	Breast cancer-EMT		FFPE (dewaxed)
	EMT-23	Breast cancer-non-EMT	Same patient	FFPE (dewaxed)
	EMT-24	Non-lesional		FFPE (dewaxed)
	EMT-25	Breast cancer-non-EMT	Same patient	FFPE (dewaxed)
	EMT-26	Non-lesional		FFPE (dewaxed)
	EMT-27	Breast cancer-non-EMT	Same patient	FFPE (dewaxed)
	EMT-28	Non-lesional		FFPE (dewaxed)
	EMT-29	Breast cancer-non-EMT	Same patient	FFPE (dewaxed)
	EMT-30	Non-lesional		FFPE (dewaxed)
	EMT-31	Breast cancer-non-EMT	Same patient	FFPE (dewaxed)
	EMT-32	Non-lesional		FFPE (dewaxed)
	EMT-33	Breast cancer-non-EMT	Same patient	FFPE (dewaxed)
	EMT-34	Non-lesional		FFPE (dewaxed)
	EMT-35	Breast cancer-non-EMT		FFPE (dewaxed)
Group 2	EMT-13	Breast cancer-EMT		FFPE (waxed)
	EMT-36			FFPE (waxed)
	EMT-37			FFPE (waxed)
	EMT-38			FFPE (waxed)
	EMT-9			FFPE (waxed)
	EMT-39	Breast cancer-non-EMT		FFPE (waxed)
	EMT-40			FFPE (waxed)
	EMT-33			FFPE (waxed)
	EMT-41			FFPE (waxed)
	EMT-18			FFPE (waxed)

University of Malaya

3.1.2 Small Angle X-ray Scattering (SAXS) Measurement

This work examined breast tissue samples acquired by the Department of Pathology, University of Malaya, obtained from surgical excision or mastectomies. The tissues were acquired following fully informed consent from patients who were subject to surgical treatment, with approval for use in research granted by the Medical Research Ethics Committee, University Malaya Medical Centre in accordance with the International Conference on Harmonization – Guidelines for Good Clinical Practice (ICH-GCP) and Declaration of Helsinki. For the first category of investigations, primary analysis was carried out by the collaborating pathologists on all eight of the tissue samples that were acquired from five patients; these were histologically categorized as EMT-1 to EMT-8 (see Table 3.2). Following this, SAXS measurements were also carried out on 26 paraffin-wax embedded (subsequently dewaxed) breast tissues from 17 patients (indicated as EMT-9 to EMT-35, as shown in Table 3.2), the pathological state of these not being revealed to the SAXS investigators (deemed blinded samples). For all cases, the samples were dissected out, obtaining nominal dimensions 5 mm (length) × 5 mm (breadth) × 1 mm (thickness), with two sub-samples taken of each selected case, fixed in formalin solution and kept in Eppendorf tubes (at 10% formaldehyde in water). Prior to SAXS measurements, excess formalin was with- drawn from the tissues and the tissues taped onto a metal frame (10 × 10 mm) on both sides using Kapton tape, a process performed at the synchrotron site.

University of Malaya

Table 3.2: The breast cancer tissue samples classified on the basis of pathology for SAXS measurement.

Research Number	Note	Sample Pathology	Material source
EMT-1	EMT-1 & EMT-2 are tissues from the same patient	EMT positive cancer	FFPE (dewaxed)
EMT-2		Non-lesional	
EMT-3		Fibroadenoma (Benign Tumour)	Formalin-fixed
EMT-4		Fibroadenoma (Benign Tumour)	
EMT-5	EMT- 5 & EMT-6 are tissues from the same patient	Non-lesional	
EMT-6		EMT negative cancer	
EMT-7	EMT- 7 & EMT-8 are tissues from the same patient	Non-lesional	
EMT-8		EMT negative cancer	
EMT-9	EMT-9 & EMT-10 are tissues from the same patient	Blinded (Pathological condition not revealed to SAXS investigator)	FFPE (dewaxed)
EMT-10			
EMT-12			
EMT-13	EMT-13 & EMT-14 are tissues from the same patient		
EMT-14			
EMT-15	EMT-15 & EMT-16 are tissues from the same patient		
EMT-16			
EMT-17			
EMT-18			
EMT-19			
EMT-20			
EMT-21			
EMT-22			
EMT-23	EMT-23 & EMT-24 are tissues from the same patient		
EMT-24			
EMT-25	EMT-25 & EMT-26 are tissues from the same patient		
EMT-26			
EMT-27	EMT-27 & EMT-28 are tissues from the same patient		
EMT-28			
EMT-29	EMT-29 & EMT-30 are tissues from the same patient		
EMT-30			
EMT-31	EMT-31 & EMT-32 are tissues from the same patient		
EMT-32			
EMT-33	EMT-33 & EMT-34 are tissues from the same patient		
EMT-34			
EMT-35			

University of Malaya

3.2 Breast tissue samples preparation

3.2.1 Preparation of breast cancer tissue into FFPE blocks

In regard to sampling, fresh mastectomy or tumour excision specimens of breast cancer, received by the histopathology laboratory, have been sliced to some 1 cm thickness, subsequently immersed in 10% neutral buffered formalin* to allow tissue fixation (preservation). The ratio of volume of formalin to tissue was about 10:1 while the duration of fixation was a minimum of 6 hours, larger specimens (such as mastectomies) typically being fixed overnight. After fixation, the specimens were sampled for tumour, surgical margins and other areas of interest. Non-lesional samples were from mastectomies performed for removal of tumour and were not separate biopsies. The sample would accordingly not be immediately adjacent to the tumour and was usually several cm away as the purpose of such a sample was to examine the normal portion of the breast. Each sample, of some 3 to 5 mm in thickness and not more than 2 cm x 2.5 cm in length and breadth, were placed into a uniquely labelled plastic cassettes and kept in 10% neutral buffered formalin to await subsequent processing into formalin-fixed paraffin embedded (FFPE) blocks. The processing of the samples, from formalin-fixed to a paraffinised block was carried out sequentially, with dehydration by alcohol followed by use of xylene for clearing of alcohol and subsequent replacement of xylene with paraffin wax (via impregnation). Conducted by the Department of Pathology, University of Malaya, this process was completely automated using a Leica TP 1020 tissue processor. The processing schedule was as follows:

i) two changes of 10% formalin for 81 minutes each

ii) two changes of 95% alcohol for 81 minutes each

iii) three changes of 100% alcohol for 81 minutes each

iv) two changes of xylene for 81 minutes each

University of Malaya

v) three changes of wax for 81 minutes each

After impregnation with paraffin, the tissue was embedded in a desired orientation in the cassette with a metal mould as the backing upon a hot plate. The cassette was then filled with liquid paraffin and placed on a cold plate to solidify the paraffin. The paraffin block was then popped out of the mould, creating a formalin fixed paraffin embedded (FFPE) block with the plastic cassette which was then ready for microtome sectioning. Microtomed sections were used for histopathology examination**.

**10% neutral buffered formalin was manually prepared from monobasic Sodium phosphate 175g, dibasic Sodium phosphate 818.5g, formaldehyde 37% 5 litre and water 45 litre, with pH of 6.5. Hence the actual amount of dissolved formaldehyde in the 10% formalin was 3.7-4.0%*

***Histological diagnosis of breast carcinoma - In the routine histopathology laboratory of the Department of Pathology, University of Malaya, 4 micron-thick sections were microtomed from the FFPE, dewaxed and stained with Haematoxylin and Eosin (H&E) for examination under the microscope for histopathology diagnosis.*

3.2.2 Dewaxing of the FFPE Blocks

The cases recruited for this study were histologically confirmed breast carcinoma and two cases of benign tumour (fibroadenoma). The EMT status of the breast carcinomas have been determined by the collaborating pathologists based on y scoring of E-cadherin (epithelial) and vimentin (mesenchymal) expression by immunohistochemistry as described in section 2.1.4. For the breast carcinoma cases, selection of study material included additional samples of normal (healthy, non-lesional) tissues from cancer free regions of mastectomies. Exclusion was made of cases with scanty tumour material in the FFPE blocks, the concern arising that future pathology review could be compromised. Dewaxing of the FFPE material for SAXS was performed manually using the following steps:

University of Malaya

i) The chosen FFPE blocks were melted down at 65 to 70°C for about 3 hours

ii) The sample was retrieved from the melt and dewaxed with xylene for 3-4 hours

iii) The dewaxed sample was subjected to 100% alcohol cleansing for 30 to 60 minutes

iv) The sample was then washed under running water until clean.

Dewaxed and rehydrated samples were then subsampled for Raman spectroscopy and SAXS measurements. Two subsamples were taken of each selected case, comprising slices of 4µm thickness for Raman measurement and 1 mm thickness for SAXS measurement. Later, the 1 mm thick sliced tissues were then kept in 10% neutral buffered formalin within Eppendorf tubes to await their SAXS examination.

3.2.3 Determination of EMT status by immunohistochemistry (IHC) and scoring

For this study, 12 breast cancers exhibiting EMT and 11 breast cancers not exhibiting EMT were selected from a concurrent study on EMT prevalence in human breast cancer tissues, archived at the Department of Pathology, University of Malaya. In brief, 4 micron-thick sections microtomed from FFPE blocks of breast cancers were run on a Ventana Benchmark XT automated system for IHC staining using a mouse-monoclonal antibody to E-cadherin (1:50; Dako: Clone NCH-38) and Vimentin (1:500 Dako:Clone V9). Positive IHC expression is depicted by brown colour staining of the cells. As is standard practice in the UMMC laboratory, a positive control is incorporated with each test slide. The positive control for E-cadherin is a known E-cadherin positive breast cancer while normal breast epithelium serves as an internal positive control. The positive control for Vimentin is lymphoid cells in human tonsil, while smooth muscle cells of blood vessels also serve as internal positive control. Cytoplasmic membranous expressions by cancer cells for E-cadherin and Vimentin were assessed microscopically and semiquantified for: (i) percentage of malignant cells expressing E-cadherin or

University of Malaya

Vimentin, as 0 (0-<1%), 1 (1-10%), 2 (11-50%), 3 (more than 50%) and; (ii) intensity of staining: 0 (negative), 1 (weak), 2 (moderate) and 3 (strong). For both E-cadherin and vimentin independently, the percentage and intensity scores were multiplied to obtain a final score. Final scores of 0-4 were classified as low expression and scores of 6-9 as high expression. EMT was considered to be present if the cancer showed a low E-cadherin final score and/or a high Vimentin final score. In addition, EMT positivity was further classified as "complete" (EMT-CV) when there was both a low E-cadherin and high Vimentin score and as "intermediate" with there was either a low E-cadherin or high Vimentin score. Intermediate EMT was further categorised as EMT-C where EMT positivity was based only on a low E-cadherin score and as EMT-V where EMT positivity was based only on a high Vimentin score (Yang et al., 2020). Scoring for E-cadherin and Vimentin was achieved through consensus scoring by four practising histopathologists (including two of the authors) in a larger concurrent study on EMT, of which the current study is a subset. Only expressions in visualised cancer cells were scored and not non-cancer elements (like stromal cells, blood vessels or inflammatory cells). Figure 3.1 provides images of an assortment of samples illustrating these categories, of which EMT-20 also illustrates the difference in morphology between vimentin-positive cancer cells and vimentin-positive stromal (benign) cells. Table 3.3 details the EMT scores of the various cancers tabulated against molecular predictive status.

University of Malaya

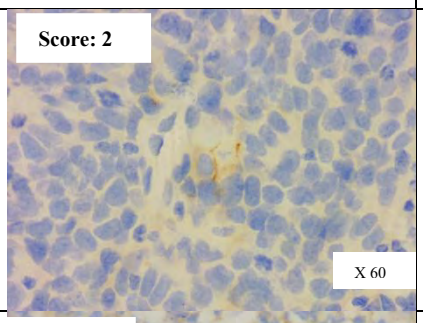
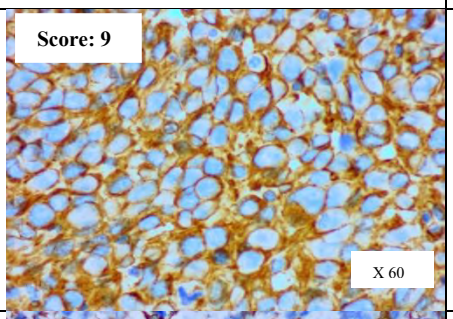
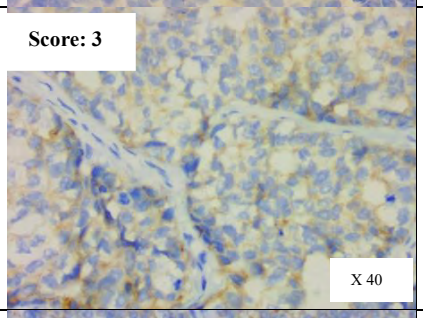
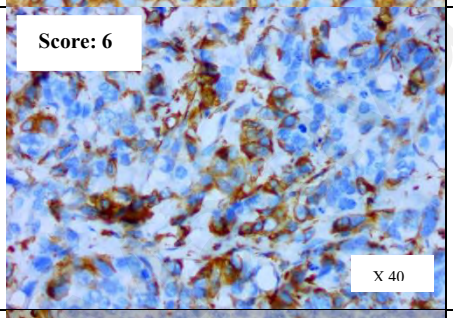
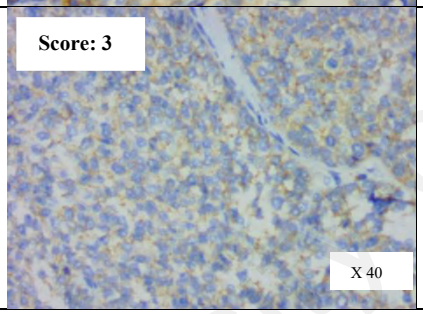
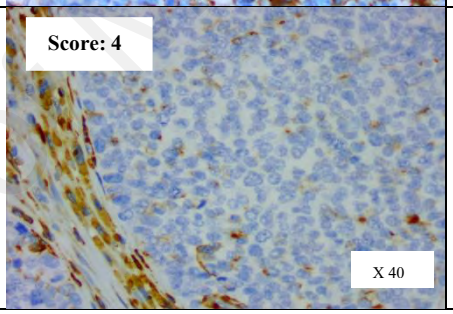
Sample number	E cadherin expression	Vimentin expression	EMT status
EMT-9			Positive, EMT-CV
EMT-13			Positive, EMT-CV
EMT-11			Positive, EMT-C

Figure 3.1: Photomicrographs of five breast cancer samples illustrating their E-cadherin and Vimentin scores, captured at comparable magnifications. Positive IHC expression is depicted by brown colour staining of the cells. Images of samples EMT-11, EMT-20 and EMT-23 illustrate the strong Vimentin staining of stromal cells which are not considered for scoring of cancer cells. EMT-20 illustrates the difference in morphology of vimentin-positive tumour cells at the right lower triangle of the picture (which show the cancer characteristics of cellular pleomorphism and enlarged nuclei) from vimentin-positive stromal (benign) cells (which are elongated and have small bland nuclei). The column at the right indicates the designated EMT status of the illustrated samples.

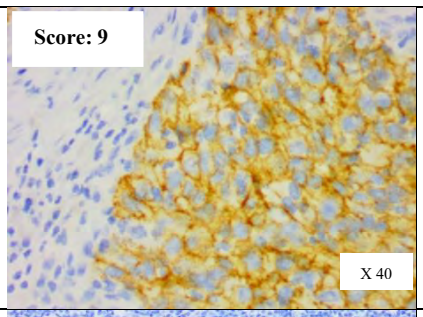
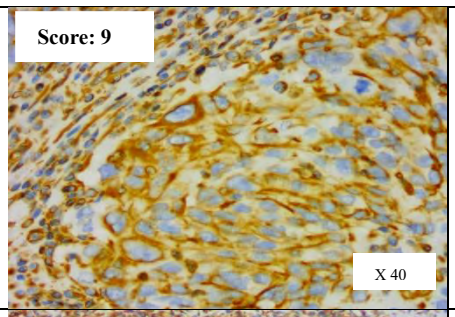
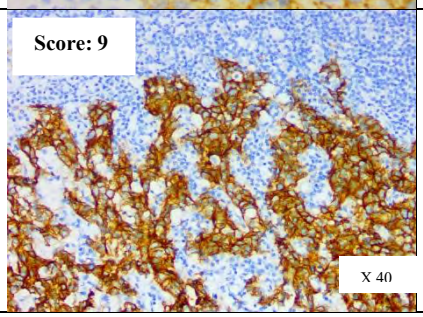
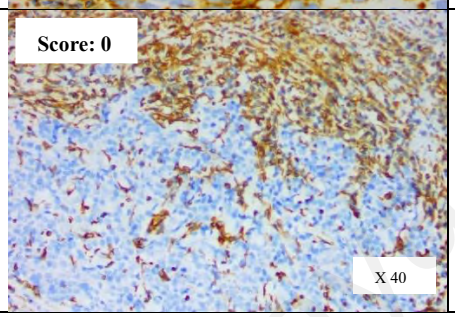
EMT-20			Positive, EMT-V
EMT-23			Negative

Figure 3.2: continued.

University of Malaya

Table 3.3: EMT scores and molecular classification of breast cancer samples.

Sample Number	Vimentin			E-cadherin			EMT status	Molecular classification (HR/HER2/TN)
	% score	Intensity score	Final score	% score	Intensity score	Final score		
EMT-9	3	3	9	2	1	2	Positive, EMT-CV	TN
EMT-11	2	2	4	3	1	3	Positive, EMT-C	HR
EMT-13	2	3	6	3	1	3	Positive, EMT-CV	TN
EMT-15	2	3	6	3	3	9	Positive, EMT-V	TN
EMT-17	3	3	9	3	3	9	Positive, EMT-V	TN
EMT-18	0	0	0	3	2	6	Negative	TN
EMT-19	3	3	9	3	1	3	Positive, EMT-C	TN
EMT20	3	3	9	3	3	9	Positive, EMT-V	TN
EMT-21	2	3	6	3	2	6	Positive, EMT-V	TN
EMT-22	0	0	0	3	1	3	Positive, EMT-V	HR
EMT-23	0	0	0	3	3	9	Negative	HR
EMT-25	0	0	0	3	3	9	Negative	HR
EMT-27	0	0	0	3	3	9	Negative	HR
EMT-29	0	0	0	3	3	9	Negative	HER2
EMT-31	1	3	3	3	3	9	Negative	HR
EMT-33	0	0	0	3	2	6	Negative	HER2
EMT-35	1	3	3	3	2	6	Negative	HER2
EMT-36	2	3	6	0	1	0	Positive, EMT-V	TN
EMT-37	0	0	0	2	2	4	Positive, EMT-C	HER2
EMT-38	3	3	9	3	3	9	Positive, EMT-V	HR
EMT-39	0	0	0	3	2	6	Negative	TN
EMT-40	1	1	1	3	2	6	Negative	TN
EMT-41	0	0	1	3	2	6	Negative	HR

University of Malaya

3.2.4 Determination of tumour size, histological grade, estrogen receptor (ER), progesterone receptor (PR) and HER2 status

It is the routine practice in UMMC for the maximum dimensions (cm) of all excised breast cancers to be recorded in the histopathology reports. The H&E slides of cancers for this study were retrieved and re-evaluated for histological grade based on the modified Bloom and Richardson system which incorporates an evaluation of nuclear pleomorphism, tubule formation and mitotic activity (Elston & Ellis, 1991). The presence of calcifications was also noted.

All cancers for this study had been routinely stained for estrogen receptor (ER), progesterone receptor (PR) and HER2 status by IHC using the Ventana BenchMark automated system (see Table 3.4). The IHC slides were retrieved and expressions for ER, PR and HER2 evaluated according to guidelines of the American Society of Clinical Oncology / College of American Pathologists (ASCO / CAP) (Hammond et al., 2010; Wolff et al., 2014). Cancers which were ER and PR positive irregardless of HER2 expression were categorized as hormone receptor (HR) positive. Cancers which were negative for ER and PR and, were HER2 positive (i.e. HER2 IHC score 3 or IHC score 2 but confirmed amplified by *in-situ* hybridization) were categorized as HER2 enriched. Cancers which were negative for ER, PR and not HER2 enriched were categorized as triple negative (TN).

University of Malaya

Table 3.4: Pathological characteristics of the study samples.

Research number	Pathology	EMT status	Vimentin score	eCadherin score	Size (cm)	Histological grade	Calcifications (Yes/No)	Molecular class (based on ER/PR/HER2 status)	Material for SAXS
EMT-1	Breast cancer	EMT	6	3	3.8	3	No	HER2	FFPE (dewaxed) x2
EMT-2	Normal breast of EMT-1						No		FFPE (dewaxed) x2
EMT-3	Fibroadenoma (benign tumour)						No		Formalin fixed x2
EMT-4	Fibroadenoma (benign tumour)						No		Formalin fixed x2
EMT-5	Normal breast of EMT-6						No		Formalin fixed x2
EMT-6	Breast cancer	Non-EMT	0	6	3.4	3	No	HR	Formalin fixed x2
EMT-7	Normal breast of EMT-8						No		Formalin fixed x2
EMT-8	Breast cancer	Non-EMT	0	6	3	3	No	HR	Formalin fixed x2
EMT-9	Breast cancer	EMT	9	2	6	3	No	TN	FFPE (dewaxed) x2
EMT-10	Normal breast of EMT-9						No		FFPE (dewaxed) x2
EMT-11	Breast cancer	Non-EMT	4	3	5.5	2	No	HR	FFPE (dewaxed) x2
EMT-12	Normal breast of EMT-11						No		FFPE (dewaxed) x2
EMT-13	Breast cancer	EMT	6	3	7	3	No	TN	FFPE (dewaxed) x2

University of Malaya

Table 3.4: continued.

Research number	Pathology	EMT status	Vimentin score	eCadherin score	Size (cm)	Histological grade	Calcifications (Yes/No)	Molecular class (based on ER/PR/HER2 status)	Material for SAXS
EMT-14	Normal breast of EMT-13						No		FFPE (dewaxed) x2
EMT-15	Breast cancer	EMT	6	3	8.3	3	No	TN	FFPE (dewaxed) x2
EMT-16	Normal breast of EMT-15						No		FFPE (dewaxed) x2
EMT-17	Breast cancer	EMT	9	9	2.5	3	No	TN	FFPE (dewaxed) x2
EMT-18	Breast cancer	Non-EMT	0	6	1.5	2	Yes	TN	FFPE (dewaxed) x2
EMT-19	Breast cancer	EMT	9	3	3.5	3	No	TN	FFPE (dewaxed) x2
EMT-20	Breast cancer	EMT	9	9	3.5	3	No	TN	FFPE (dewaxed) x2
EMT-21	Breast cancer	EMT	6	6	3.5	3	No	TN	FFPE (dewaxed) x2
EMT-22	Breast cancer	EMT	0	3	5	3	Yes	HR	FFPE (dewaxed) x2
EMT-23	Breast cancer	Non-EMT	0	9	2	3	No	HR	FFPE (dewaxed) x2
EMT-24	Normal breast of EMT-23						No		FFPE (dewaxed) x2
EMT-25	Breast cancer	Non-EMT	0	9	4	2	No	HR	FFPE (dewaxed) x2
EMT-26	Normal breast of EMT-25						No		FFPE (dewaxed) x2

University of Malaya

Table 3.4: continued.

Research number	Pathology	EMT status	Vimentin score	eCadherin score	Size (cm)	Histological grade	Calcifications (Yes/No)	Molecular class (based on ER/PR/HER2 status)	Material for SAXS
EMT-27	Breast cancer	Non-EMT	0	9	2.7	3	No	HR	FFPE (dewaxed) x2
EMT-28	Normal breast of EMT-27						No		FFPE (dewaxed) x2
EMT-29	Breast cancer	Non-EMT	0	9	4	3	Yes	HER2	FFPE (dewaxed) x2
EMT-30	Normal breast of EMT-30						No		FFPE (dewaxed) x2
EMT-31	Breast cancer	Non-EMT	3	9	1.5	1	Yes	HR	FFPE (dewaxed) x2
EMT-32	Normal breast of EMT-32						No		FFPE (dewaxed) x2
EMT-33	Breast cancer	Non-EMT	0	6	2.5	3	Yes	HER2	FFPE (dewaxed) x2
EMT-34	Normal breast of EMT-33						No		FFPE (dewaxed) x2
EMT-35	Breast cancer	Non-EMT	3	6	5	3	N	HER2	FFPE (dewaxed) x2

Key:
EMT status: EMT = vimentin scores 6 to 9 and / or ecadherin score 0 to 4; non-EMT = vimentin score 0 to 4 and /or ecadherin score 6 to 9.
Histological grade is according to modified Bloom and Richardson system.
Molecular class: HR = Positive for ER and PR irregardless of HER2 expression; HER2 = HER2 enriched (IHC score 3 or confirmed amplifed by ISH) and negative for ER and PR; TN = Triple negative i.e. negative for ER, PR and not HER2 enriched.

University of Malaya

3.3 Experimental Method: Raman Spectroscopy

3.3.1 Instrumentation of Raman Spectroscopy

Figure 3.2 shows a schematic and functional diagram of the main components of a typical Raman spectrometer, the assembled system consisting of an excitation source (monochromatic and most typically that offered by a laser), the optical path system, the Rayleigh (elastic scattering) rejection filter, a monochromator/dispersing element, (spectrograph) and a detector. Through use of the excitation source, the sample is irradiated with monochromatic light that impacts on and thereby interacts with sample molecules, most essentially with the electron cloud and bonds (setting them into vibration), resulting in scattered light. The spectrometer processes this scattered light, eliminating the unwanted portion (as in for instance the elastically scattered component) and dispersing the remainder into its constituent wavelengths. Given that the intensity of the Raman (inelastic) scattering component is very weak compared to that of the elastic component, a notch filter or high-performance long pass filter is used to block the Rayleigh scattered light since otherwise this would easily overpower the weak signal of the Raman scattered light. The Raman scattered light is then detected, most typically through use of a deep depletion charge-coupled device (CCD) detector. Finally, through use of a computer, a Raman spectrum in digital form is displayed, generally as a function of energy (frequency) difference between incident and scattered photons.

The wavenumber difference between incident light and Raman scattered light is known as the Raman shift expressed in wavenumber (cm^{-1}), being unique for individual molecules of a particular type. The molecules in the material have specific vibrational modes allowing the Raman spectrum of each material to be used for material identification and analysis. Figure 3.3 shows the layout of the Renishaw inVia Raman spectroscopy equipment, with all the key components highlighted and the general specification is tabulated in Table 3.5.

University of Malaya

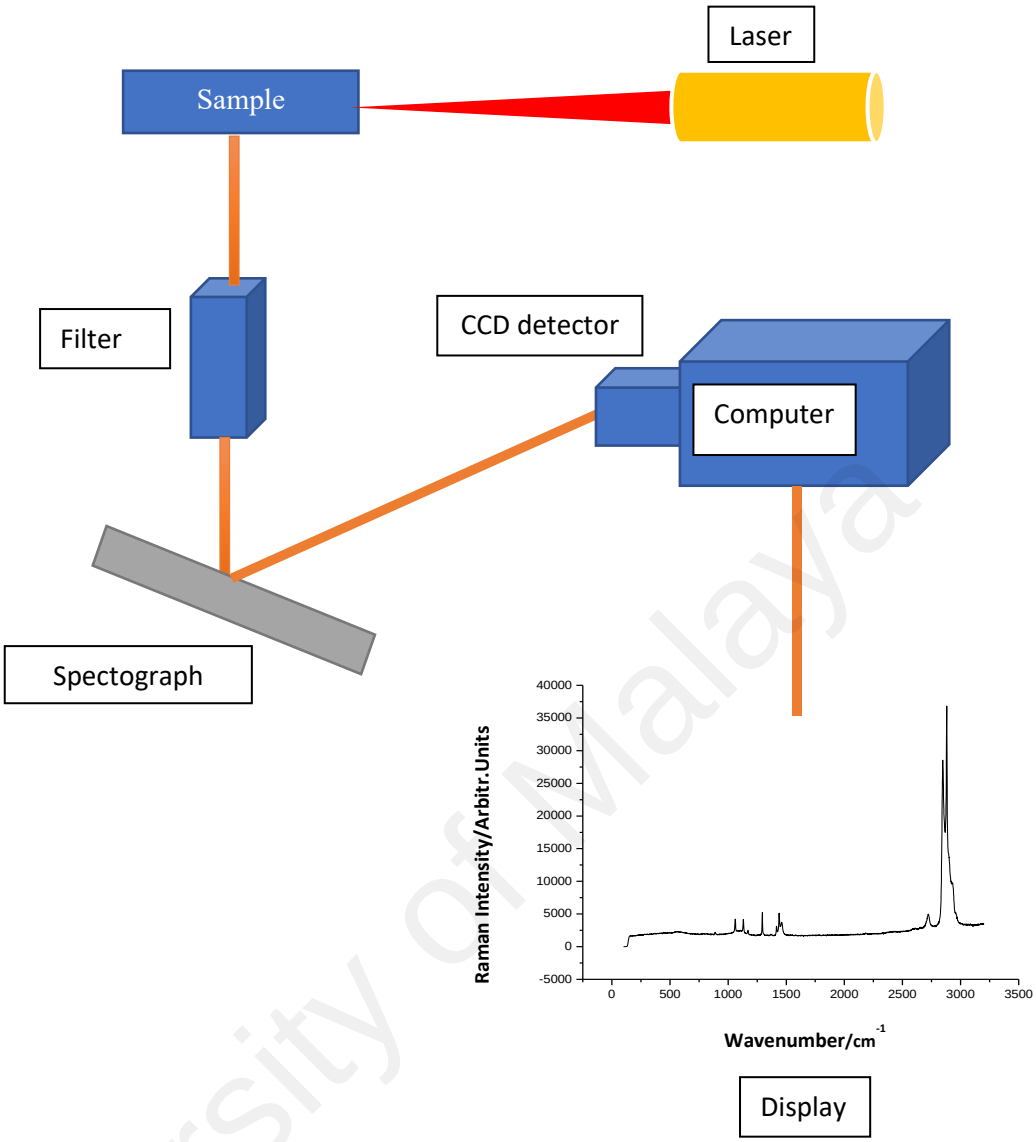


Figure 3.3: Schematic diagram of Raman spectrometer and a resulting Raman spectrum.

University of Malaya

Figure 3.4: The components of Renishaw inVia Raman spectroscopy. Image extracted from https://www.ccmr.cornell.edu/wpcontent/uploads/sites/2/2016/08/Wire4_Training_Modules_Compilation-Part-1.pdf.

Table 3.5: General of specifications of Renishaw inVia Raman spectroscopy.

Wavelength range	200 nm to 2200 nm
Laser supported	From 229 nm to 1064 nm
Spectral resolution	0.3 cm^{-1} (FWHM) Highest typically necessary: 1 cm^{-1}
Stability	< ± 0.01cm^{-1}
Low wavenumber cut-off	5 cm^{-1} Lowest typically necessary:100 cm^{-1}
High wavenumber cut-off	30,000 cm^{-1} Standard: 4,000 cm^{-1}
Spatial resolution (lateral)	0.25 µm Standard: 1 µm
Spatial resolution (axial)	< 1 µm Standard:<2 µm. Dependent on objective and laser
Detector size	1024 pixel × 256 pixel
Supply voltage	110 V AC to 240 V AC +10% - 15%
Supply frequency	50 Hz or 60 Hz
Typical power consumption (spectrometer)	150 W

University of Malaya

WiRE 3.3 (Windows-based Raman Environment) is a software developed by Renishaw to allow system control and configuration as well as analysis and processing of acquired data. WiRE software are created to:

1. Control Renishaw inVia Raman spectroscopy
2. View the sample
3. Control data collection parameters
4. View data
5. Provide processing function to improve data obtained
6. Provide analysis options to extract information from Raman data
7. Enable data and results to be printed and exported for analysis or reporting.

3.3.2 Raman Spectroscopy Measurement

In present work, the Raman spectra of the cancerous breast tissues for group 1 and 2 (refer Table 3.1) were acquired by the author and the Department of Physics, University of Malaya using a Renishaw Raman system (Apply Innovation, Gloucestershire, UK) (Figure 3.4). The spectroscopy consists of a laser, objective lens, and charge-coupled device (CCD) detector. To excite the sample, a laser excitation beam of 532 nm was focused onto the sample through a 50x objective microscope. Each sample was exposed to the laser excitation for 20 seconds, detecting in the spectral range 100 – 3200 cm^{-1}. Raman spectra were displayed through use of a computer using Renishaw Wire software. Three to five Raman spectra were acquired from each tissue sample.

In regard to phospholipids, a major constituent of the plasma membrane forming the outermost layer of animate cells, they are composed of two fatty acids that help form a diacylglycerol. The latter activates proteins that have a role in various signalling cascades. The Raman signal sensitivity of the fatty acids has been described in detail in Levchenko

University of Malaya

and Qu (2018), the authors recording that the ratio of the bands at 1665 cm^{-1} (this being the stretching mode, proportional to the amount of unsaturated C=C bonds) and 1440 cm^{-1} (this being the C–H_2 bending mode, proportional to the amount of saturated C–C bonds) being indicative of the degree of saturation of the fatty acids. In present analysis, the respective bands are observed to centre around 1659 cm^{-1} and 1441 cm^{-1}, a minor variation to that reported in Levchenko and Qu (2018) (Levchenko & Qu, 2018). For the nucleic acids, the phosphodiester backbones (O-P-O) exhibits many of the spectral complexities of DNA and RNA have stretching motion that give rise to peaks in the 800 to 900 cm^{-1} region (Chen et al., 2010). The nucleosides are sensitive to the torsion of the phosphodiester bond network of the nucleic acid backbone structure and do not depend on the identity of the attached nucleotic base (Benevides et al., 2005). Additionally, Raman spectroscopy can distinguish between unique backbone conformations of A-DNA (807 ± 3 cm^{-1}), B-DNA (835 ± 7 cm^{-1}) and Z-DNA (745 ± 2 cm^{-1}).

3.3.3 Analysis of Raman Spectra

3.3.3.1 Pre-processing data

Data pre-processing is an important step in improving the accuracy of subsequent classification analyses (Lasch, 2012), also to prevent mix-up between the main information of Raman assignment bands and the noise (Rinnan, 2014). As Raman spectra are prone to a variety of contaminants (e.g. detected cosmic ray events, fluorescence background, noise etc.), pre-processing procedures is necessary to be undertaken in the Raman raw data, use being made of OriginPro version 2019b software (OriginLab Corporation, Northampton, Massachusetts, USA). The whole Raman spectra analysis was done by the author with the help of statistician from Institute of Mathematical Science, University of Malaya. Prior to multivariate analysis, the steps for the pre-processed raw data of Raman spectra is illustrated in Figure 3.4.

University of Malaya

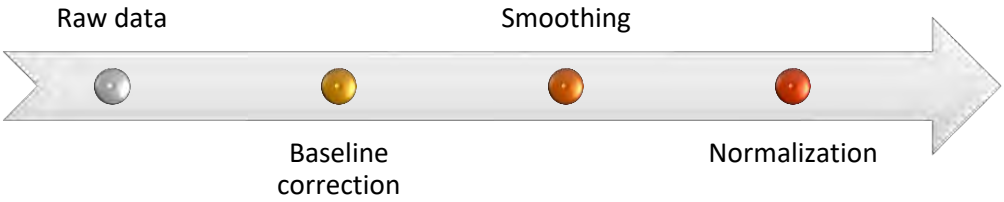


Figure 3.5: The procedure of pre-processing of Raman spectra.

Eight points of baseline correction was applied to each spectrum. Raman spectra were smoothed by using a Savitzky-Golay filter (5th order, 13 points) to reduce the noise (Ibrahim et al., 2017). Subsequently, the intensity of the Raman spectra were corrected by vector normalization (Ibrahim et al., 2017), the pre-processed data then being ready for further analysis. Figure 3.5 shows the Raman spectra of raw data and the after pre-processing data.

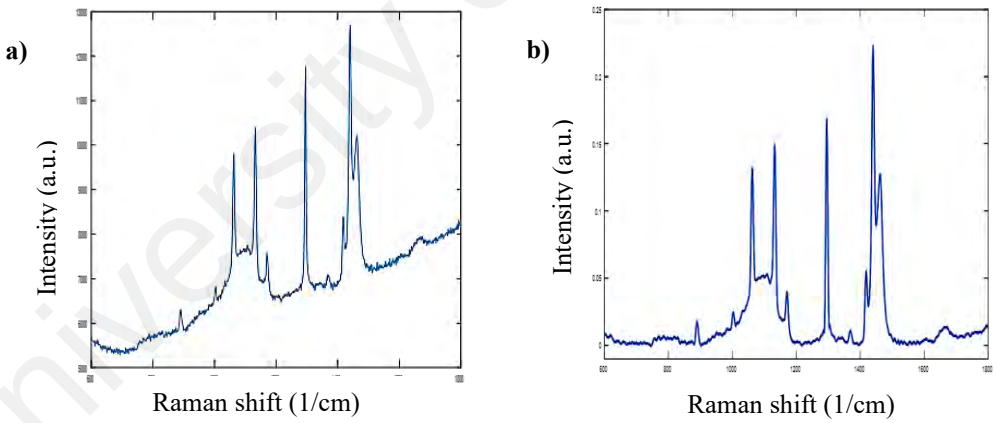


Figure 3.6: The (a) raw and (b) pre-processed data of Raman spectra.

3.3.3.2 Multivariate analysis

Multivariate analysis was performed on each pre-processed Raman spectrum, seeking differences between the various cancerous human breast tissue samples under study. The analysis consists of several steps as shown in Figure 3.6 below.

University of Malaya

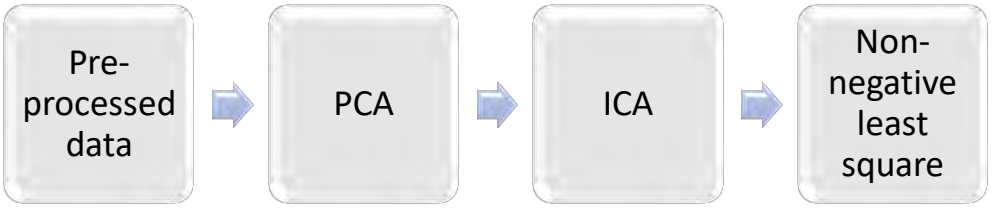


Figure 3.7: The multivariate analysis for both groups of Raman spectra data.

(a) Principal Component Analysis (PCA)

Principal Component Analysis (PCA), an unsupervised data transformation technique, reduces the dimensionality of the data while retaining the most significant information for further data analysis (Gautam et al., 2015). In this, the original coordinate was transformed to a new coordinate system referred to as the principal component (PC). The first PC contains the greatest variance from the data, the second the next greatest and so on. In present work the first three PCs were used; these later being used to performe Independent Component Analysis (ICA) for paraffin extraction.

(b) Independent Component Analysis (ICA)

With the sample tissues embedded in paraffin, it is crucial to extract and eliminate the paraffin signal, prominent wax peaks tending to interfere with Raman spectral data (Faoláin et al., 2005; Meksiarun et al., 2017). Commonly and as previously mentioned, the FFPE tissue samples are dewaxed using xylene. However, previous studies have shown the dewaxing procedure to be a time consuming process, with the structures of the samples potentially altered and a residual layer of paraffin potentially remaining in the sample tissues (Ali et al., 2012; Faoláin et al., 2005). In this work, the Raman spectra of Group 2 samples were digitally dewaxed using ICA and Non-Negative Least Squares (NNLS) analysis as proposed by Gobinet et al. and Meksiarun et al. (Gobinet et al., 2008; Meksiarun et al., 2017).

University of Malaya

ICA, again an unsupervised analysis technique and an extension to PCA, was used to separate the linearly mixed data. The ICA is applied to estimate the paraffin component to extract pure component information from the spectra of FFPE tissue samples. In present study, ICA was performed using the FastICA algorithm for estimation of the paraffin component in the breast cancer tissue spectra. ICA is determined from the Raman spectra matrix, X(n x m) where n are the observations and m are the wavenumbers. The X matrix can be reconstructed into a linear model by use of the following equation:

$$X = A\,.S \tag{3.1}$$

with A a (n × k) mixing matrix and S a (k × m) matrix representing independent components, k being the number of components. To solve relation (1), the unmixing matrix W (k × n) needs to be calculated, based on observations of X, to obtain the independent component $\hat{S}$ (k × m) where matrix W should be the inverse of A:

$$\hat{S} = W\,.X \tag{3.2}$$

(c) Non-negative least square (NNLS)

The estimated components from ICA were later used for subtraction from the Raman spectra by using Non-Negative Least Squares (NNLS) fitting. The NNLS problem is given by the following:

$$x' = argmin\,\|d - Cx\|_2 \text{ , subject to } x \geq 0 \tag{3.3}$$

with d the Raman spectra, x the paraffin concentration profile, c the estimated component of paraffin and x' the non-negative concentration profile of x.

The paraffin removed spectra were then calculated by subtracting the Raman spectra with the estimated component of paraffin as shown in the following relation:

University of Malaya

$$d' = d - Cx' \tag{3.4}$$

where d' is the paraffin removed spectra. The NNLS fittings were performed using MATLAB software (Mathworks Inc., MATLAB version R2017b).

3.4 Experimental Method: Small Angle X-ray Scattering (SAXS)

3.4.1 Instrumentation of Synchrotron radiation

Synchrotron radiation is the electromagnetic radiation emitted when charged particles are accelerated radially. It is produced in synchrotrons using bending magnets, undulators and wigglers. The properties of synchrotron radiation ranges from having a broad spectrum (covers from microwaves to hard x-rays) to having high flux and high brilliance. High intensity photon beam allows rapid experiments or use of weakly scattering samples. Furthermore, highly collimated photon beam is generated by a small divergence using a collimator. Figure 3.7 shows the set-up of the Beamline 1.3W: SAXS at Synchrotron Light Research Institute (SLRI), Thailand. Table 3.6 summarizes the main components that make up the beamline in SLRI which consists of the electron gun, linear accelerator, booster synchrotron and storage ring.

Figure 3.8: Set-up of the Beamline 1.3W: SAXS at Synchrotron Light Research Institute (SLRI) in Thailand.

University of Malaya

Table 3.6: Components that make-up the Beamline 1.3W: SAXS and its functions.

No	Components	Function
1	**Electron gun**	Enormous number of electrons is produced by applying electricity to heat up the cathode of the electron gun to release electrons. High positive voltage is applied to attract the electrons into the linear accelerator.
2	**Linear accelerator: LINAC**	To achieve 40 MeV, electrons are being accelerated by microwaves and later sent to the booster synchrotron.
3	**Booster synchrotron**	The energy of the electrons is increased through circular motion in radio waves until it achieves an energy level of 1GeV in l6 seconds or nearing the speed of light.
4	**Storage ring**	The storage ring functions to increase the electron energy to 1.2 GeV. The ring contains many types of magnets in order to bend the movement of electron circularly. The energy being released by the electrons is called the "synchrotron light". Due to relativistic effects, the radiation being emitted is highly collimated.

3.4.2 Experimental Setup: Synchrotron source

The SAXS experiment in present work was performed at Beamline 1.3W: SAXS Synchrotron Light Research Institute (SLRI), Thailand. The main properties of the X-ray scattering beamline at SLRI is listed in Table 3.7. A schematic of the beamline is shown in Figure 3.8 with a fixed wavelength (λ = 1.379Å., Energy = 9 kev). The beamline optics were able to produce focal spot, collimated with the use of slits at the sample position. The tissue sample was inserted carefully into the sample holder and aligned with the radiation beam and imaged using SAXS to obtain a diffraction pattern. Sample is attached to a sample stage and the measurement is performed in air because they were easier to handle and to maintain the hydration of the samples during the experiment, rather than in a vacuum environment. The tissue samples were placed in the sample holder of 1.83 nm away from the detector. The detector used to collect the X-ray scattering data is a two-dimensional detector (Rayonix SX 165 CCD) of 2048 x 2048 pixels that record the scattering pattern. To lessen air scattering pattern and absorption attenuations, it is

University of Malaya

important to keep the detector inside a vacuum chamber. The sample-detector distance allowed recording of the momentum transfer at range 0.06 $nm^{-1} \leq q \leq 2.50$ nm^{-1}. In order to avoid the detector being saturated due to the primary intensity, a beam stop of 4.5 mm diameter was placed centralized and adjacent to the detector. Silver Behenate was used as a standard sample to calibrate and normalize the SAXS patterns in order to obtain the precise reciprocal space scale of each image (Huang et al., 1993). The measured time was fixed at 5 minutes to obtain a maximum photon count without saturating the detector. Once imaged, the tissue was removed from the metal frame and was kept into the receptacles again for preservation. A MATLAB-based data processing program called SAXSIT which is developed by Synchrotron Light Research Institute Thailand is used to process the measurement 2-D scattering patterns for further analysis.

Table 3.7: Properties of the SAXS machine at Beamline 1.3W: SAXS.

Technical Information	
Application	Nano structural investigation of materials
Source of radiation	Multipole wiggle
Photon flux	~ 2 x 10^9 phs/s
Photon energy range	6 - 9 keV
Beam size	2 x 1 mm (HxV)
Beam collimation	three slits and one guard slit
Type of monochromator	Double Multilayer Monochromator (W/B4C)
Energy bandwidth	1 %
Focusing element	Toroidal mirror
Detectors	CCD (Rayonix SX165, LX170HS)
Sample-detector distance	140- 4500mm
Minimum q	0.06nm
Measurable characteristic length	~ 1 -100nm
Type of sample	Powder/ solid/ liquid/ fiber
Sample environment	In air

University of Malaya

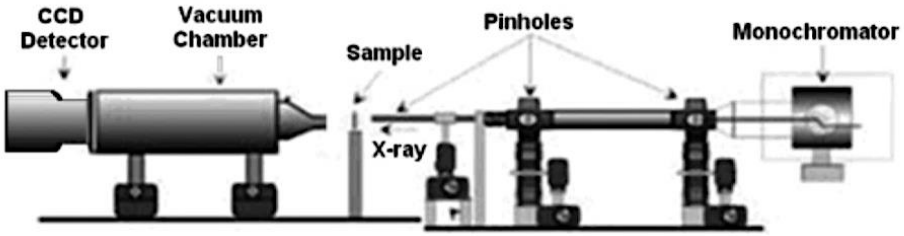


Figure 3.9: Experimental setup of the SAXS experiment. Image extracted from Sabeena Sidhu, 2009.

3.4.2.1 Photon source and beamline optics

(a) Photon source

The Siam Photon Source (SPS) is a 1.2 GeV synchrotron light source of Thailand. It is operated at the maximum beam current of 150 mA. The SAXS beamline utilizes the synchrotron radiation from a bending magnet of the SPS. At photon energy between 6-9 keV, the calculated photon flux density of above 10^{11} photons/sec/0.1% BW/ mrad^2/ mm^2 is obtained. Figure 3.9 shows the schematic diagram of the SAXS beamline.

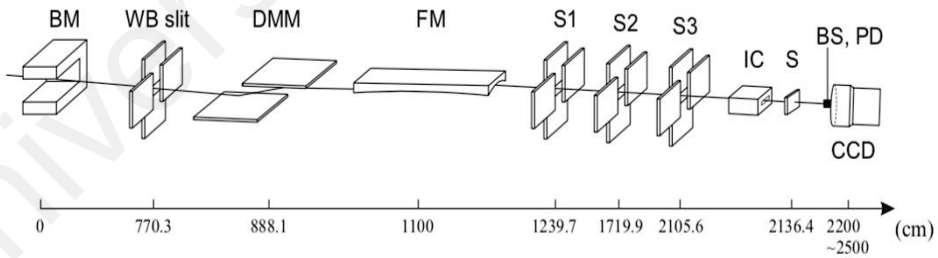


Figure 3.10: Schematic diagram of the SAXS beamline. BM: bending magnet, WB slit: white beam slit, DMM: double multilayer monochromator, FM: focusing mirror, S1- S3: slits, IC: Ionization chamber, S: sample, BS: beam stop, PD: photodiode, CCD: charge-coupled device detector. Image extracted from Siriwart Soontaranon, 2011.

University of Malaya

(b) Charge Coupled Device (CCD) detector

A Charge Coupled Device (CCD) is a type of indirect solid-state detector. A scintillator, usually CsI:TI, is used to convert the X-rays into visible photons. Optically focusing devices, such as lenses or fiber optics, are typically used for demagnification in order to match the scintillator size with the CCD size and finally a silicon CCD chip is used to detect the light photons, converting them into electronic signal and digitalizing them.

A CCD detector (Mar CCD) of diameter 165 mm is used in present investigation for capturing the scattering pattern. To monitor the beam intensity before and after exposure, an ionization chamber was installed in front of the sample holder and a photodiode was placed in front of a beam stop. The sample-detector distance between 0.5-4 meter can be varied by removing or adding the vacuum tubes between the sample and detector.

(c) Double multilayer monochromator and focusing mirror

The Double Multilayer Monochromator (DMM) is one of the most important components to increase the efficiency of experiments in SAXS beamline. The main optics contains a DMM and a toroidal focusing mirror to focus the x-rays to the sample position. The photon flux at the sample position is estimated to be about 2 x 10^9 Photons/sec/mm^2. The requirements for the monochromator include continuous scan with an axis of driving, 1-2° of Bragg angle range and reduced parallelism error of the mirrors in the range.

It consists of 150 periodic layers of W/B4C, having d-spacing of 2.5 nm, coated on a Si substrate (the mirrors are fabricated by Rigaku Innovative Technology Inc.). Calibration of the DMM has been carried out by measuring the K-absorption edge of Ni, Co and Fe metal foils. The photon energy was varied by rotating the DMM angle while an ionization chamber with the foil attached in the front was used to monitor the beam intensity. A Si-111 crystal was employed as an energy analyser and an ionization chamber

University of Malaya

used for monitoring the beam intensity is placed behind the crystal. The measured energy bandwidth for the photon energy of 6, 7, 8 and 9 keV were found to be 0.92%, 1.01%, 0.93% and 0.95%, respectively. The DMM therefore provides the monochromatic X-ray with the energy resolution of approximately 1% in the energy range between 6-9 keV. The advantage of using DMM is that higher photon flux can be achieved.

In order to increase photon flux at the sample position, a toroidal mirror is used to focus the beam in both horizontal and vertical direction. The mirror was made from ultra-low expansion glass substrate coated with Rh and Pt. It has the tangential and the sagittal radii of curvature of 213, 334 cm and 5.33 cm, respectively. In order to optimize the beam acceptance of the mirror, the operation is designed at 5 mrad grazing angle. The beam size at the sample position obtained by calculating the FWHM of the Gaussian function fitted into the histogram has been found to be 1 x 0.2 mm^2.

3.4.3 Analysis of SAXS Spectra

The SAXSIT program has been developed by BL1.3W: SAXS team to process the 2-dimensional scattering patterns, manipulate the data, fitting and modelling. The code is written in MATLAB, which is executable. It is able to do pattern alignment, sample-detector distance calibration, subtraction of background patterns, circular averaging to acquire 1-dimensional scattering profiles as a function of the scattering vector, q from 2-dimensional scattering patterns, radial integration to obtain 1D profile as a function of azimuthal angle and normalization of data sets by area under the curve.

(a) Analyse standard

This module is used to calculate vertical and horizontal pixel shifts to be used for pattern alignment so that its actual center is at the primary beam position. The module uses a pattern of some standard material which gives circular scattering pattern, e.g. silver behenate. The scattering profiles are calculated from the pattern at three points and fitted

University of Malaya

with Gaussian to find peak positions. The program will then estimate the radius and scan along the circular path. The peak positions are then used to find the center of the circle. The centre of the circular pattern is assumed to be the primary beam position. After the centre is found, the vertical and horizontal pixel distances will be calculated between this centre and a specified origin of coordinate.

(b) Read parameter measurements

This module reads in measurements parameters from an excel file. The parameters must be input in a specific order in the second column of the excel file. Users are recommended to use a template file MeasPara.xls given with the SAXSIT package to make the input file. The text files (filenames with .txt above) contain values of integrated intensity readings from ionization chamber and photodiode. These values are used to normalize the scattering intensities.

(c) Load pattern/background subtraction

This module is intended to be used for background subtraction. The program will load patterns specified in the parameter file. The background subtracted pattern is calculated from

$$I_{subbg} = I_{sam} - (1 - c)\tilde{I}_{bg} \tag{3.5}$$

Where,

$$\tilde{I}_{sam} = \frac{\bar{I}_{sam} - \bar{I}_{dark}}{T_{sam} \cdot i_{sam,IC}} \tag{3.6}$$

$$\tilde{I}_{bg} = \frac{\bar{I}_{bg} - \bar{I}_{dark}}{T_{bg} \cdot i_{bg,IC}} \tag{3.7}$$

University of Malaya

$$\bar{I}_{sam} = \frac{I_{sam}}{t_{sam}} = sample\ pattern\ divide\ by\ sample\ exposure\ time, t_{sam}$$

(3.8)

$$\bar{I}_{bg} = \frac{I_{bg}}{t_{bg}} = background\ pattern\ divided\ by\ exposure\ time, t_{bg}$$

(3.9)

$$\bar{I}_{dark} = \frac{I_{dark}}{t_{dark}}$$

$$= dark\ current\ pattern\ divided\ by\ accumulation\ time, t_{dark}$$

(3.10)

$i_{sam,IC}$ and $i_{bg,IC}$ are x-ray intensity, measured by the ion chamber, integrated over the period of the measurement of sample and background, respectively.

c = sample concentration ($0 \leq c \leq 1$). The default value is 0.0 meaning the background will not be scaled.

$T_{sam} = \frac{i_{sam,PD}/i_{sam,IC}}{i_{cell,PD}/i_{cell,IC}}$ = sample transition, where $i_{sam,PD}$ is beam intensity measured by a photodiode at the beam stop, integrated over the measurement period of the sample and $i_{cell,PD}$ is that for the empty sample cell.

$T_{bg} = \frac{i_{bg,PD}/i_{bg,IC}}{i_{cell,PD}/i_{cell,IC}}$ = background transition, where $i_{bg,PD}$ is beam intensity measured by a photodiode at the beam stop, integrated over the measurement period of the solvent. The SAXS imaged obtained at the end is seen in Figure 3.10 where it consists of arc with radius that increase in response to q.

University of Malaya

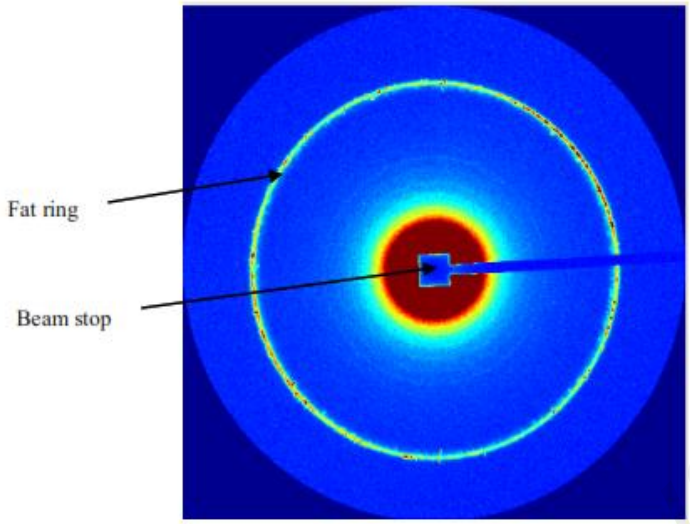


Figure 3.11: Example of a SAXS pattern obtained from the 2-Dimensional detector.

(d) Calculating scattering profile

A scattering profile is calculated from the scattering pattern along a radial path starting from the specified origin. The intensity values are obtained from each pixel lying in the path. The number of points which the intensity is obtained is approximately the number of pixels in that path. The radial distance from the centre to a certain point on the detector plane is translated into the scattering angle θ and then to a scattering vector q in the unit nm^{-1} using the input 'sample to detector distance' (in mm), the pixel size (in mm) and the x-ray energy (in keV). The scattering angle is given by

$$\boldsymbol{2\theta = tan^{-1}(\frac{D}{L})} \tag{3.11}$$

Where D represents the distance from the origin to the point on the detector plane, calculated using the pixel size, and L is the sample-detector distance. The scattering vector is given by

$$\boldsymbol{q = \frac{4\pi}{\lambda} sin\theta} \tag{3.12}$$

University of Malaya

Where λ is the x-ray wavelength calculated from the specified x-ray energy. In the 'choose direction to calculate profile' the use is required to enter the azimuthal angle φ in which the profile is calculated. The azimuthal angle value is referenced from the left hand side horizontal midplane of the pattern (the direction from the centre horizontally to the left is – degree) and increases in the clockwise direction.

If circular averaging is to be performed the user gives the angular range and angular step size for the averaging. If the 'direction to calculate profile' is 90°, the 'averaging range' is 10°. The 'step size' is 0.25°, the profile will then be calculated from 80° to 100°with 0.25° increment. The averaging is performed by integration of the profile:

$$\boldsymbol{I_{ave}(q)} = \frac{\int_{\varphi_{min}}^{\varphi_{max}} \boldsymbol{I(q,\varphi)d\varphi}}{\int_{\varphi_{min}}^{\varphi_{max}} \boldsymbol{d\varphi}} \tag{3.13}$$

A valid q range is also calculated just for information when the user provides the beam stop (4.5mm) and detector radii. This is the q-range which falls within the distance between the edge of the beam stop and the detector edge. The outputs are given in an excel file having the filename the same as the pattern filename with a suffix '_SP_Xdeg.xls', where X is the profile central angle. The output excel file contains 4 worksheets:

1. Averaged I (2theta), containing the averaged intensity profile as a function of 2θ,
2. Non-averaged I (2theta), containing the profile calculated at the central angle as a function of 2θ,
3. Averaged I(q), containing circularly averaged intensity profile as a function of the scattering vector q and
4. Averaged I(q), in valued q range containing the same data as the worksheet Averaged I(q) but only those in the valid q range.

University of Malaya

3.4.3.1 Image Acquisition and Pre-processing

Each image acquired at the synchrotron source was saved as the header file containing the time the image was taken, the image name and other unique identifications. There were three processing steps that were performed in order to extract the information from the SAXS images, as described below:

1. Calibration of the images for q- spacing,
2. Normalization of the intensity. This was necessary for comparison of images between experimental sessions and x-ray sources,
3. Radial integration over the 2D images to obtain 1D data of the intensity as a function of the scattering vector, q from which feature parameters were extracted.

(a) Calibration of a q-spacing

Silver behenate is used as calibration standard for SAXS measurements. The images from the synchrotron source were calibrated for q-spacing using silver behenate. The exposure time for acquiring the silver behenate images was 5 minutes, the same exposure time used for imaging the breast tissue samples with the synchrotron source. The silver behenate images were calibrated between each experimental session to ensure that the d-spacings were exactly the same.

(b) Normalization of intensity

Normalization of the intensity in the SAXS images was required to correct the variation in the incident beam intensity and sample attenuation. Data was normalized to compensate for the variation in the intensity of the incident beam flux and to take into account the exposure time and the decay of the synchrotron beam current.

University of Malaya

(c) Radial integration

The maxima due to scattering for well-ordered systems are observed as 'rings' about the centre of the scattering image. By radially integrating the image, a 1-dimensional profile of the intensity over the entire image can be obtained. Fitting procedures such as the Gaussian fitting of the intensity peaks using OriginPro 8.1 are easier to perform on 1D data rather than with a 2D image. Before the 2D image data was radially integrated, the beam stop was masked out. The intensities were averaged as a function of q, starting from the edge of the masked beam stop and outward to the defined edge of the circle region of interest. The resulting 1D data was exported to an Excel file.

3.4.4 Statistical analysis

After the accurate sample scattering profiles were obtained, a few parameters that illustrates the structural information were generated by fitting the functions using OriginPro2019. Fitting of the peaks to Gaussians have allowed generation of the parameters: FWHM, amplitudes, centre position and area of the individual peaks. Axial periodicities (d-spacings), extracted from Gaussian peak centres, were obtained using equation (4). To determine those variables differentiating between groups, discriminant analysis was performed, R software (Version 3.6.1) being used to perform the Kruskal-Wallis test, with subsequent performance of Dunn's post hoc test to identify the pairs of tissues which differed significantly for all measurements considered, p values substantiating parameter extractions of practical consequence in constructing a diagnostic model. The analysis have be done by our research group with the help of statitistician from the Institute of Mathematical Science, University of Malaya.

University of Malaya

CHAPTER 4: RESULTS AND DISCUSSIONS

4.1 Raman Spectroscopy Data Analysis

We understand the work herein to represent the first known Raman spectroscopic study on EMT in human breast tissue, the data being supported by multivariate analysis. Previous studies by others have used cell lines in investigating molecular exchanges occurring in the EMT process (Chaturvedi et al., 2016b; Marro et al., 2014; Nieva et al., 2012), comparisons being made between EMT and non-EMT tissue, including extraction of general metabolic features of neoplastic cells. This has led to the identification of changes in nucleic acid, protein and lipid (Talari et al., 2015), reflected in the lower wavenumber (LWN) range, 600 to 1800 cm^{-1} (Chaturvedi et al., 2016b). In regard to the small number of samples investigated herein, it needs to be stressed that the EMT process relates to a relatively brief period of existence, by definition occurring at the onset of malignancy. Associated with this is the clear diagnostic challenge within the imaging department of detecting early stage changes, be it in visualised soft tissue mass, spiculation, calcifications etc.

4.1.1 Pre-processing data

Example pre-processed Raman spectra data from non-lesional, EMT and non-EMT breast cancer tissue obtained from the average of all the samples are shown in Figure 4.1, with apparent appreciable paraffin contribution bands within the low wavenumber region for both dewaxed and waxed samples. The pure paraffin spectra were measured as the reference shown in Figure 4.2. These are strongly overlapped with important tissue components throughout the entire spectrum that have been obtained for each sample. Multivariate analysis allows delineation of the biological components of EMT in breast cancer tissues, the paraffin signal from the various Raman spectra effectively being

University of Malaya

removed. Example Raman spectra following paraffin removal within the fingerprint region (600 – 1800 cm^{-1}), for all types of breast tissue samples, are shown in Figure 4.3. From these two figures, it is apparent that the EMT and non-EMT breast cancer tissues exhibit differences, albeit at a rather subtle level, in terms of intensity, position, shape and peak assignments in the Raman spectra.

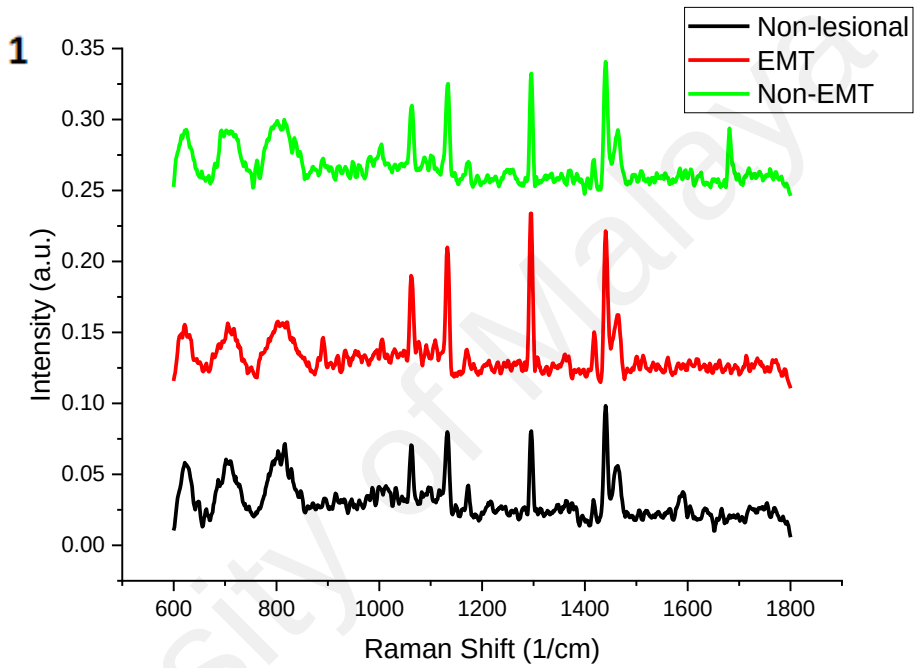


Figure 4.1: The pre-processed Raman spectra from all the samples obtained from the average of all the samples. Figure label number 1 is for FFPE dewaxed samples and figure label number 2 is for FFPE without dewaxed samples. (Black line: Non-lesional; Red line: EMT; Green line: Non-EMT) . Note: the x-axis represents the Raman shift and y-axis is Raman intensity. The lines are offset for clarity.

University of Malaya

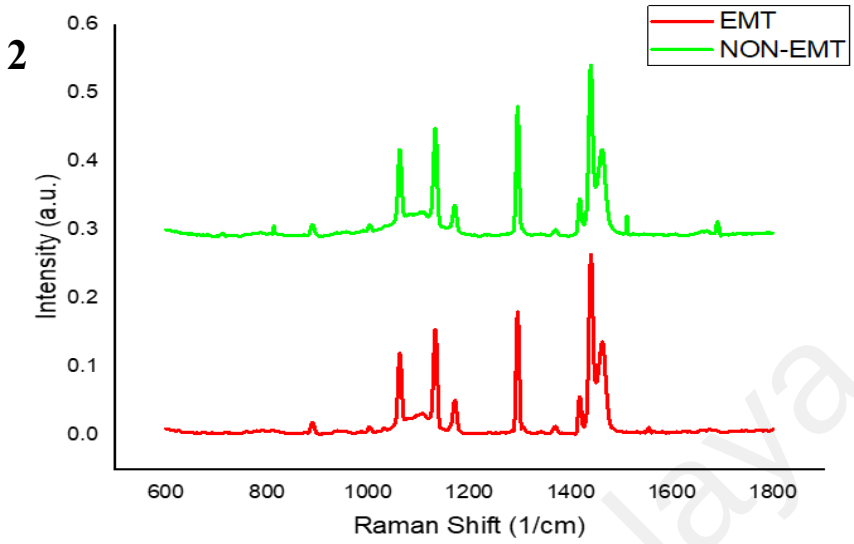


Figure 4.1: continued.

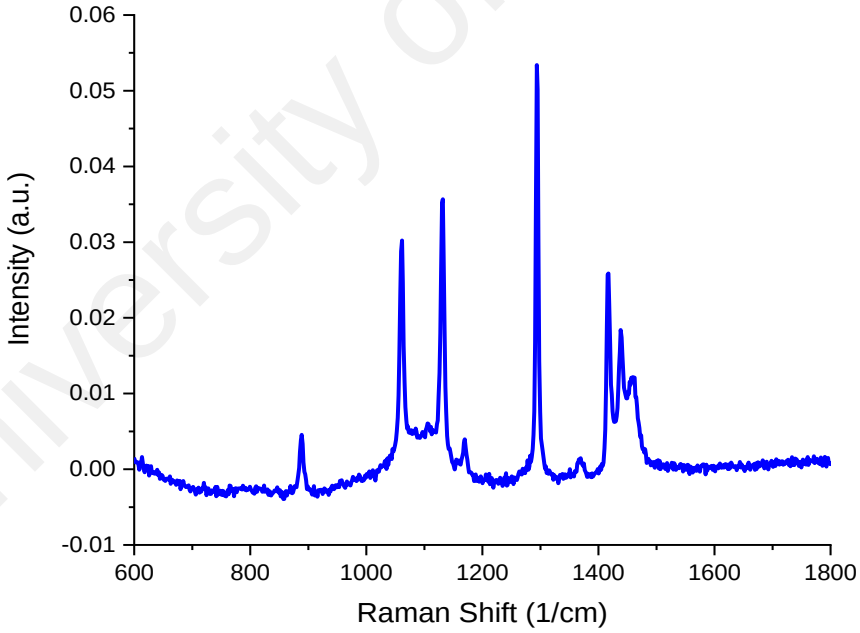


Figure 4.2: The spectra of pure paraffin.

University of Malaya

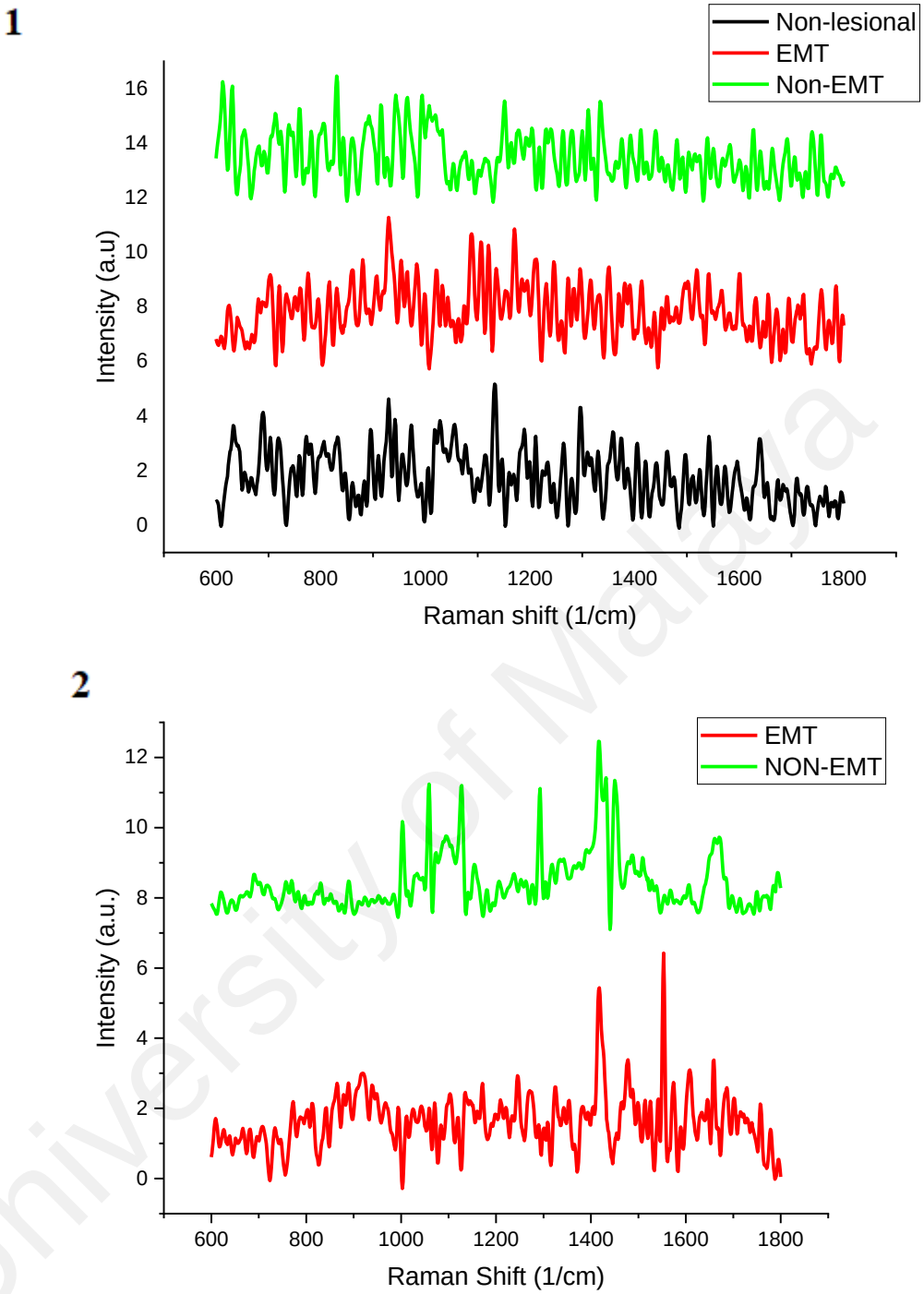


Figure 4.3: The Raman spectra of all types of samples following extraction of the paraffin wax signal. Figure label number 1 is for FFPE dewaxed samples and figure label number 2 is for FFPE without dewaxed samples. (Black line: Non-lesional; Red line: EMT; Green line: Non-EMT). Note: the x-axis represents the Raman shift and y-axis is Raman intensity. The lines are offset for clarity.

University of Malaya

4.1.2 Multivariate analysis: Paraffin removal process

Present investigation has been made on formalin fixed paraffin embedded (FFPE) breast tissue samples, requiring the removal of paraffin due to highly intense signal that will alter the underlying tissue (Gobinet et al., 2008). The paraffin removal process has been conducted on both FFPE conditions; dewaxed and waxed breast tissue samples. Despite the fact that the FFPE dewaxed tissue has undergo the dewaxing process chemically, the paraffin component in Raman spectra is still present as shown in Figure 4.4. The multivariate analysis is therefore crucial for both cases, including FFPE dewaxed and waxed prior to remove the paraffin component in the breast tissue samples. The multivariate analysis technique has been discussed intensively in Chapter 3 (section 3.3.3.2), consisting of PCA, ICA and NNLS. For further discussion of this, see the section immediately below in which the FFPE dewaxed and waxed breast tissue samples are compared, showing the removal of paraffin contaminant prior to obtain the significant Raman assignment bands within 600 to 1800 cm^{-1}.

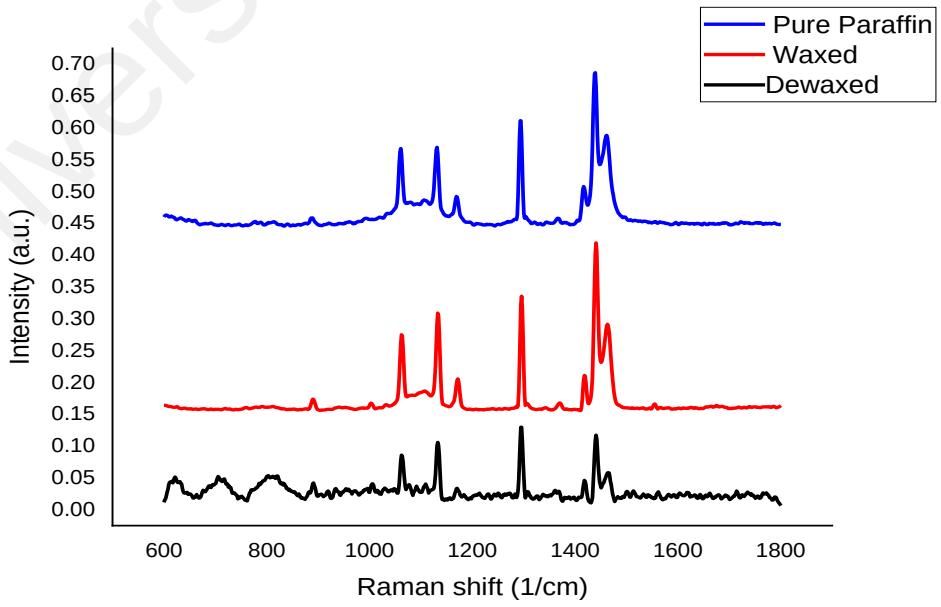


Figure 4.4: The comparison between pure paraffin, FFPE waxed and FFPE dewaxed samples. The lines are offset for clarity.

4.1.2.1 FFPE dewaxed samples

The PCA analysis of FFPE dewaxed samples have been investigated on all tissue types, including non-lesional, EMT and non-EMT breast tissues, comprising of 10, 9 and 8 of number of samples, respectively. Table 4.1 represents the Principal Component (PC) numbers together with the eigenvalue, percentage of variance and cumulative for non-lesional, EMT and non-EMT breast tissues. The first three PC numbers have been selected; PC1, PC2 and PC3 for all such cases (see Figure 4.5), which will be used for the ICA analysis. The implementation of ICA for paraffin extraction in FFPE samples has been reported previously by Vrabie et al, Gobinet et al., and Meksiarun et al. (Gobinet et al., 2008; Meksiarun et al., 2017; Vrabie et al., 2007). The ICA method is performed on the PC loadings in order to obtain the estimated components by using OriginPro 2019. Subsequently, the estimated components will be subtracted from the FFPE dewaxed spectra for all tissue types using NNLS, utilising Matlab R2017b software. The mean Raman spectra of FFPE waxed of non-lesional, EMT and non-EMT breast tissues after paraffin removal process is shown in Figure 4.6. The significant Raman assignment bands within 600 to 1800 cm^{-1} from the prominent peaks obtained in Figure 4.6 are listed in Table 4.2.

University of Malaya

Table 4.1: The Principal Component (PC) numbers of non-lesional, EMT and non-EMT FFPE dewaxed samples.

Principal Component Number	Eigenvalue	Percentage of Variance (%)	Cumulative (%)
NON-LESIONAL			
1	0.027	24.358	24.358
2	0.018	16.536	40.894
3	0.013	11.427	52.321
4	0.011	10.182	62.502
5	0.010	9.057	71.560
6	0.009	8.143	79.703
7	0.008	7.621	87.324
8	0.008	7.087	94.411
9	0.006	5.589	100.000
10	1.564×10^{-32}	1.416×10^{-29}	100.000
Principal Component Number	**Eigenvalue**	**Percentage of Variance (%)**	**Cumulative (%)**
EMT			
1	0.081	47.629	47.629
2	0.026	15.368	62.997
3	0.014	8.358	71.355
4	0.012	7.220	78.575
5	0.011	6.558	85.133
6	0.010	5.950	91.083
7	0.009	5.200	96.283
8	0.006	3.717	100.000
9	1.657×10^{-32}	9.729×10^{-30}	100.000
Principal Component Number	**Eigenvalue**	**Percentage of Variance (%)**	**Cumulative (%)**
NON-EMT			
1	0.102	50.817	50.817
2	0.039	19.375	70.193
3	0.018	8.970	79.162
4	0.012	6.189	85.352
5	0.011	5.425	90.776
6	0.020	4.855	95.631
7	0.009	4.369	100.000
8	2.398×10^{-32}	1.196×10^{-29}	100.000

University of Malaya

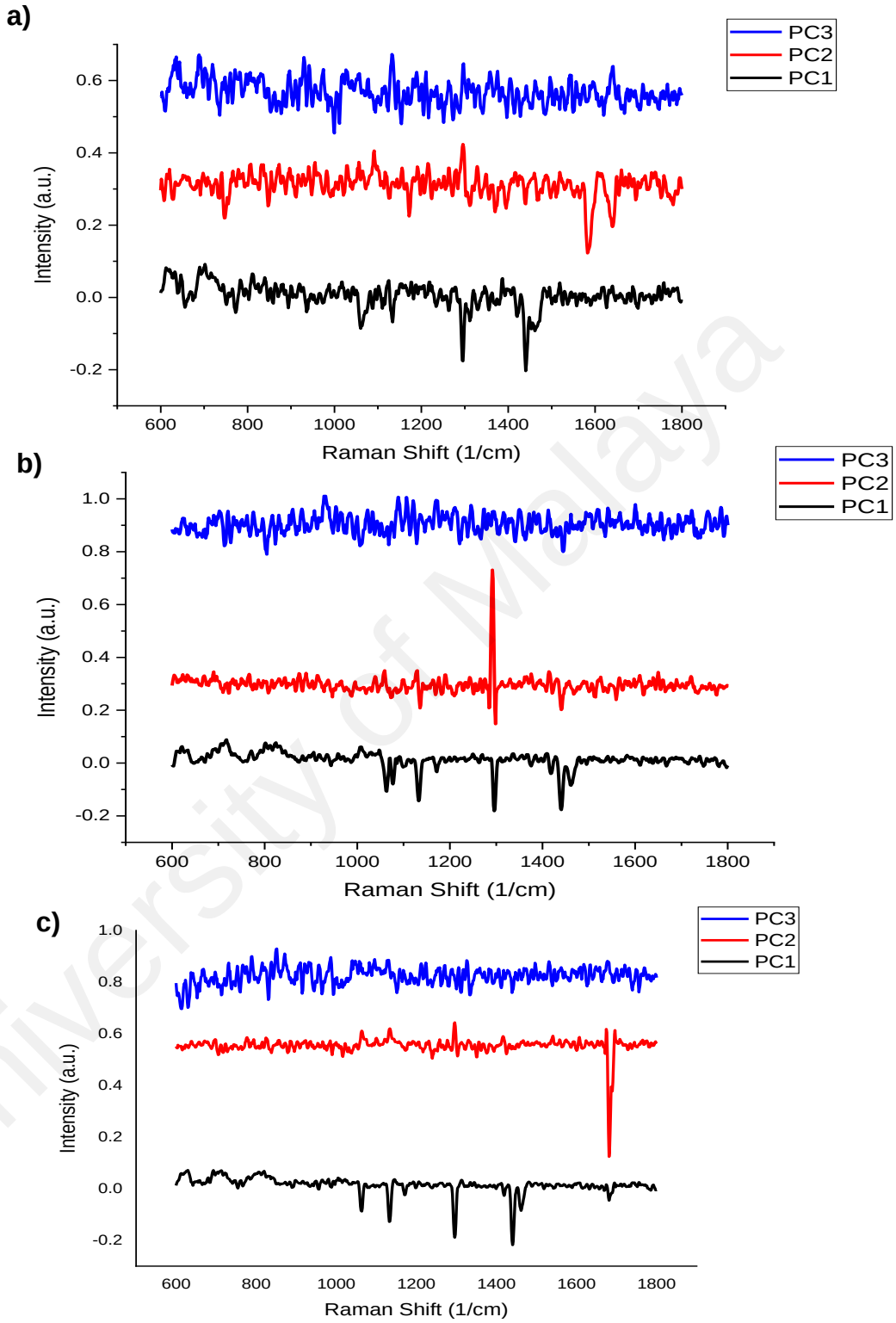


Figure 4.5: Principal Component (PC) loadings for (a) non-lesional, (b) EMT and (c) non-EMT of FFPE dewaxed samples. The lines are offset for clarity.

University of Malaya

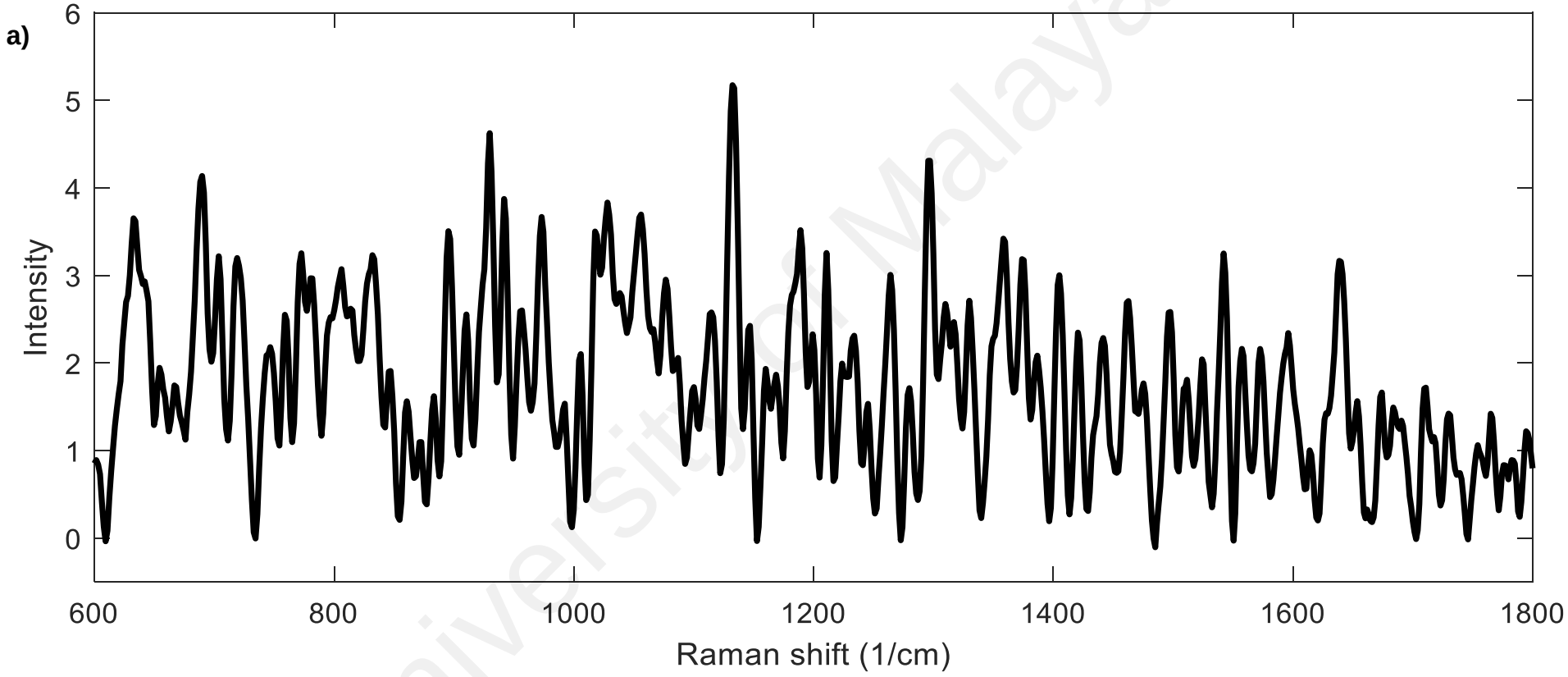


Figure 4.6: The mean of Raman spectra of (a) non-lesional, (b) EMT and (c) non-EMT breast tissue after paraffin removal process.

University of Malaya

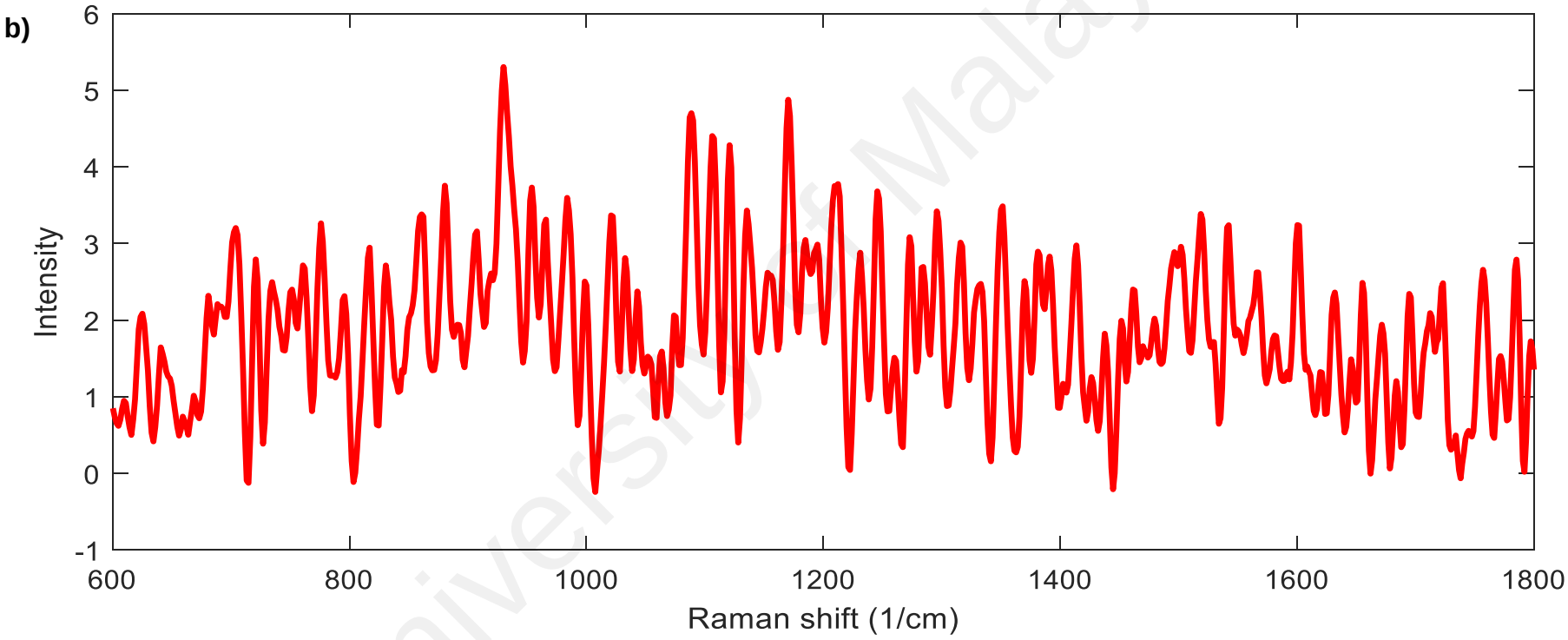


Figure 4.6: continued.

University of Malaya

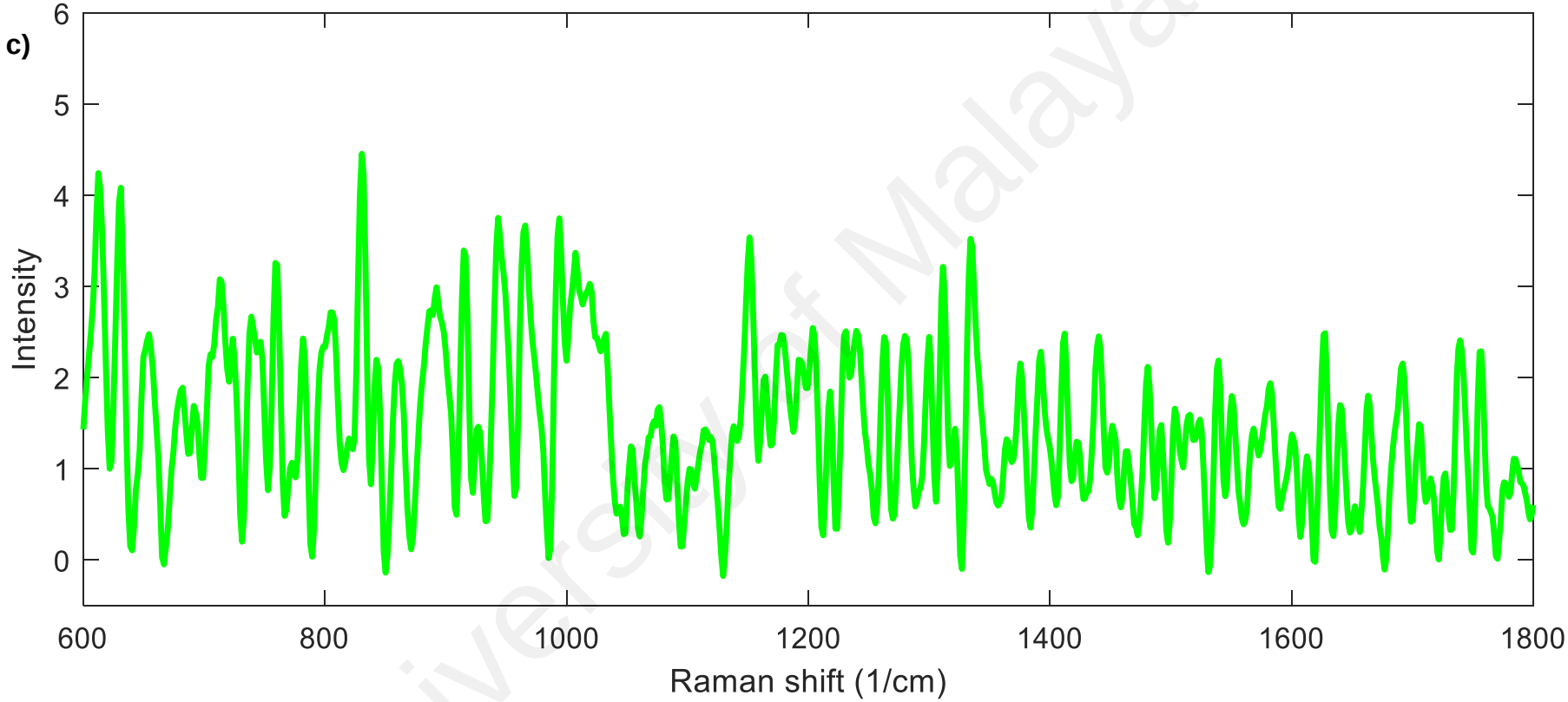


Figure 4.6:continued.

University of Malaya

Table 4.2: Raman assignments of the prominent peaks from the non-lesional, EMT and non-EMT breast tissues upon paraffin removal. The error of the intensity of the non-lesional, EMT and non-EMT tissues from the range of 600-1800 cm^{-1} is 0.96, 0.95 and 0.87 respectively.

WAVENUMBER (cm^{-1})	ASSIGNMENT	INTENSITY
NON-LESIONAL		
720	Adenine	3.094
782	Cytosine and thymine	2.888
815	O–P–O stretch (nucleic acid)	2.621
832	Out of plane ring breathing tyrosine/O-P-O stretch DNA	3.320
930	C-C stretch of proline ring and valine and protein backbone (a-helix conformation)/glycogen	4.626
1005	Carotenoids	2.145
1077	C-C stretch lipids	2.951
1211	Tryptophan and phenylalanine V mode	3.256
1223	Amide III	1.945
1264	Bending mode (= CH) of lipids	3.010
1280	CH twist and bend (protein)	1.711
1318	CH_2 twist and bend (nucleic acids, proteins, lipids)	2.463
1388	C-H stretching (protein)	2.027
1441	Phospholipids	2.290
1639	Amide I in α-helix	3.195
1654	Amide I (protein)	1.540
1710	Lipids (Ester carbonyl stretching mode)	1.716
WAVENUMBER (cm^{-1})	**ASSIGNMENT**	**INTENSITY**
EMT		
751	Symmetric breathing of tryptophan	2.405
831	Out of plane ring breathing tyrosine/O-P-O stretch DNA	2.736
880	DNA	3.756
1033	C-H in plane bending mode of phenylalanine	2.818
1043	Proline	2.394
1053	C-C stretching (lipid)	1.538
1088	C-N stretching of proteins and lipids, stretching mode (C-C) and stretching mode (C-O) of phospholipids	4.702
1121	Stretching mode (C-C) of lipids from trans-segments	4.280
1155	- Beta carotene (CH/C–C) in lipids - Carotenoids	2.694
1230	Amide III	2.891

University of Malaya

Table 4.2: continued.

WAVENUMBER (cm^{-1})	ASSIGNMENT	INTENSITY
EMT		
1315	Amide III (α-helix) of proteins	3.013
1370	Myoglobin	2.518
1451	CH_2 bending mode (lipid)	1.95564
1461	CH_2 and CH_3 symmetric deformation of proteins	2.417
WAVENUMBER (cm^{-1})	**ASSIGNMENT**	**INTENSITY**
NON-EMT		
712	C-N stretching (lipids)	3.067
782	Cytosine and thymine	2.405
831	Out of plane ring breathing tyrosine/O-P-O stretch DNA	4.452
1033	C-H in plane bending mode of phenylalanine	2.408
1053	C-C stretching (lipid)	1.256
1150	C-C stretching vibration of carotenoids	3.540
1230	Amide III	2.464
1263	- Amide III of proteins - =C-H in plane bending (lipid)	2.447
1280	CH twist and bend (protein)	2.455
1300	CH_2 deformation (lipid)/adenine, cytosine	2.450
1453	CH_2 deformation (protein)	1.466
1582	Adenine, guanine (nucleic acid)	1.933

4.1.2.2 FFPE waxed samples

For the case of FFPE waxed samples, the PCA analysis have been undertaken on 5 for each of EMT and non-EMT breast tissue samples. The details on the Principal Component (PC) numbers can be found in Table 4.3. In this case, five PC loadings has been utilised for both sample types; EMT and non-EMT as shown in Figure 4.8. Similarly, as described in 4.1.2.1 (FFPE dewaxed samples) the generated PC loadings will be implemented in ICA analysis in order to obtain the estimated paraffin component which will be subsequently subtracted from the Raman spectra using NNLS method. The mean of Raman spectra of FFPE waxed EMT and non-EMT breast tissues after paraffin removal

University of Malaya

is demonstrated in Figure 4.8. The significant Raman assignment bands within 600 to 1800 cm^{-1} from the prominent peaks obtained in Figure 4.9 is listed in Table 4.4.

Table 4.3: The Principal Component (PC) numbers of EMT and non-EMT FFPE waxed samples.

Principal Component Number	Eigenvalue	Percentage of Variance (%)	Cumulative (%)
EMT			
1	0.063	75.650	75.650
2	0.012	14.051	89.701
3	0.007	8.344	98.045
4	0.002	1.955	100.000
5	6.927×10^{-33}	8.339×10^{-30}	100.000
Principal Component Number	**Eigenvalue**	**Percentage of Variance (%)**	**Cumulative (%)**
NON-EMT			
1	0.069	76.816	76.816
2	0.014	15.501	92.317
3	0.005	5.850	98.166
4	0.002	1.834	100.000
5	2.686×10^{-32}	2.980×10^{-29}	100.000

University of Malaya

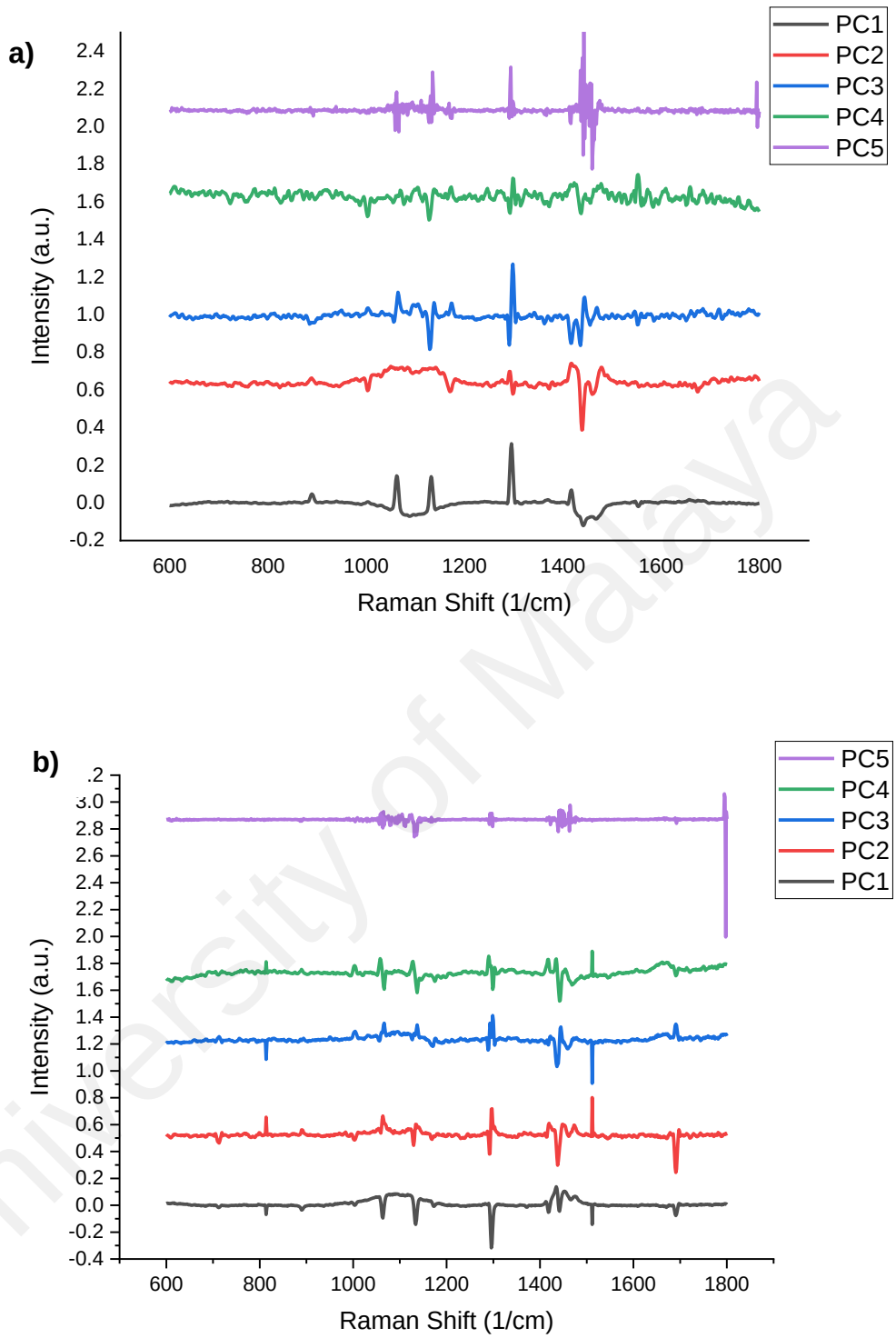


Figure 4.7: PC loadings for (a) EMT and (b) non-EMT tissues of FFPE waxed samples. The lines are offset for clarity.

University of Malaya

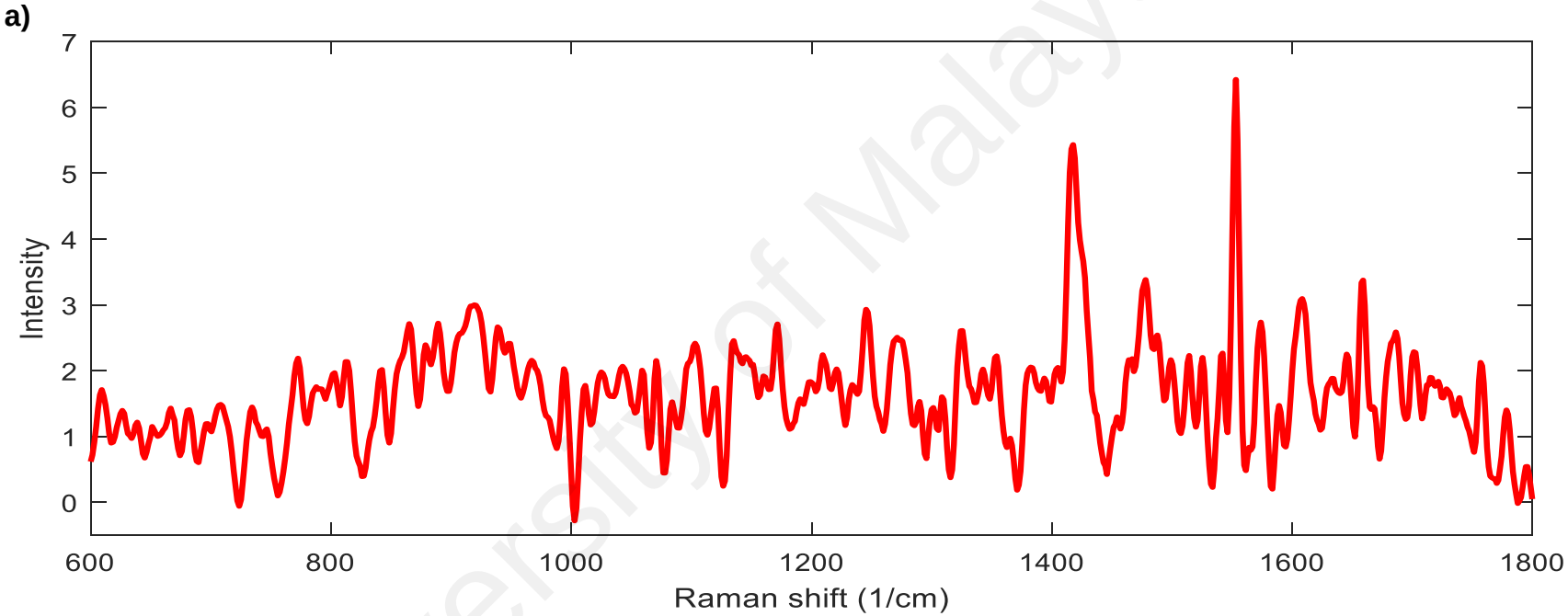


Figure 4.8: The mean of Raman spectra of FFPE waxed of (a) EMT and (b) non-EMT breast tissue after paraffin removal process.

University of Malaya

b)

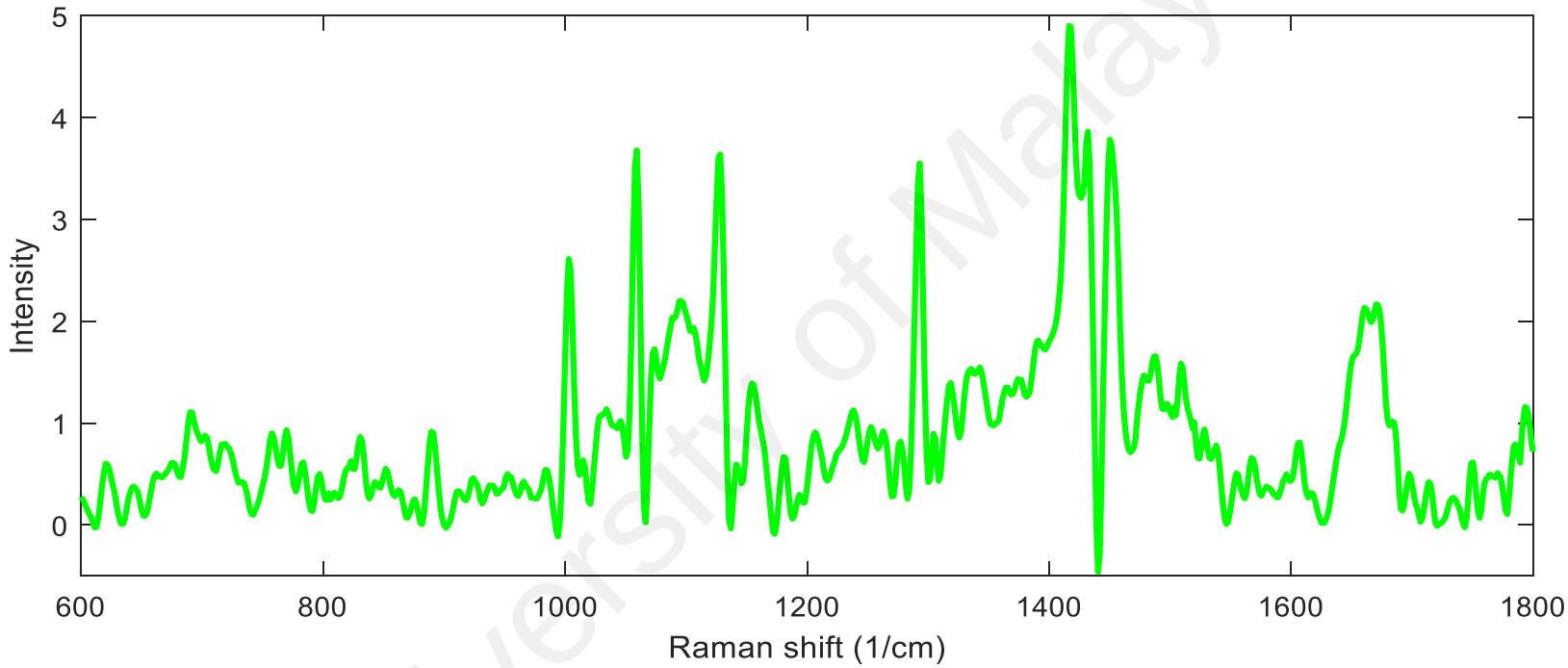


Figure 4.8: continued.

University of Malaya

Table 4.4: Raman assignments of the prominent peaks from the FFPE waxed of EMT and non-EMT breast tissues upon paraffin removal. The error of the intensity of the EMT and non-EMT tissues from the range of 600-1800 cm^{-1} is 0.79 and 0.81 respectively.

WAVENUMBER (cm^{-1})	ASSIGNMENT	INTENSITY
	EMT	
878	DNA	2.379
938	CH_2 lipid	2.674
1043	Proline	2.063
1083	Phospholipid	1.516
1120	Stretching mode (C-C) of lipids from trans-segments	1.771
1171	C-C stretch amino acid	2.707
1210	Tryptophan and phenylalanine V mode	2.253
1272	Amide III (α-helix), collagen	2.516
1301	Lipids	1.442
1343	CH_3, CH_2 wag (collagen)	2.022
1366	Guanine, TRP (protein), lipids in tumour bed	0.893
1453	CH_2 deformation (protein)	1.288
1526	C=C stretching mode of β-carotene	2.183
1574	Tryptophan, nucleic acids (guanine, adenine) TRP protein	2.736
1659	Lipid	3.390
WAVENUMBER (cm^{-1})	**ASSIGNMENT**	**INTENSITY**
	NON-EMT	
643	C-C twisting mode of tyrosine	0.379
676	Stretch (bending (CCN)), Vinyl & Porphyrin modes	0.613
720	- Adenine	0.790
820	Tyrosine	0.577
832	Out of plane ring breathing tyrosine/O-P-O stretch DNA	0.848
853	Ring breathing of tyrosine and C-C stretch of proline ring	0.553
876	-Tryptophan - C-C stretch, hydroxyproline in carcinoma	0.254
890	Structural protein modes of tumours	0.928
925	C-C stretch of proline ring/glucose/lactic acid	0.463
1003	Symmetric ring breathing mode of phenylalanine	2.620
1031	C–H in phenylalanine	1.188
1046	C-O stretching of carbohydrates in tumour	0.997
1128	C–N stretching in protein or C–O stretching in carbohydrates	3.697
1206	Hydroxyproline, tyrosine (collagen assignment)	0.910
1264	Bending mode (= CH) of lipids	0.871

University of Malaya

Table 4.4: continued.

WAVENUMBER (cm^{-1})	ASSIGNMENT	INTENSITY
NON-EMT		
1277	Amide III (α-helix), collagen	0.819
1305	Amide III, α-helix, C-C stretch & C-H (protein)	0.915
1318	CH_2 twist and bend (nucleic acids, proteins, lipids)	1.389
1344	A and G of DNA/RNA and CH deformation of proteins in tumour bed	1.526
1451	CH_2 bending mode (lipid) in malignant tissue	3.907
1659	Lipid	2.059

University of Malaya

4.1.3 Metabolite Characterisation of non-lesional, EMT and non-EMT tissues

All the prominent bands are obtained from the Raman spectra of paraffin wax extracted. We only focus on the changes of lipid, protein and nucleic acid in the samples. Firstly, we compared three breast tissues; non-lesional, EMT, and non-EMT of the FFPE dewaxed samples. Then we compared two tissues: EMT and non-EMT of FFPE without dewaxed samples. The peaks were obtained from the average of all samples. All the assignments are based on the review by Sabtu et., al (2019).

4.1.3.1 Lipid

(a) Dewaxed sample

Raman spectra bands consistent with the presence of lipid are found in all types of samples; non-lesional, EMT and non-EMT tissues. The prominent peak lipid bands are shown in Figure 4.9 complete with peak assignments. The peaks were obtained from the average of all samples. As seen in Figure 5, the Raman CH group signal in EMT tissues, centered around 1451 cm^{-1}, are of lower intensity than that of the corresponding CH group signals in both non-lesional and non-EMT tissue, centered around 1264 cm^{-1} and 1263 cm^{-1}, respectively. We have found that the CH lipid bands in non-lesional tissues, 42% and 21% greater intensities in comparison to that EMT and non-EMT tissues respectively. Phospholipid was found in non-lesional tissues at peaks 1441 cm^{-1} and in both EMT and non-EMT tissues at peaks 1088 cm-1 with greater intensity by 60% and 110% in EMT tissue samples compared to that non-lesional and non-EMT tissue respectively. At position 1121 cm^{-1} and 1053 cm^{-1}, C-C stretching mode in lipid is found to exist in EMT and non-EMT tissues respectively while in non-lesional tissue the C-C stretch is found at peak 1077 cm^{-1}. The greatest intensity of lipid C-C stretching was found in the EMT tissues; 109% higher intensity compared to that non-EMT tissues and 37% higher

intensity compared to that non-lesional tissue. A comparison of the relative intensity of non-lesional, EMT and non-EMT tissue also included in Figure 4.10 to Figure 4.12.

(b) Waxed sample

The Raman spectra of lipid found in the waxed samples are depicted in Figure 4.10 together with the peaks assignment. The Raman band observed at peak 938 cm^{-1} in EMT tissues and 1451 cm^{-1} in the non-EMT tissues are marker peaks of CH group of lipid, showing 37% greater intensity in the non-EMT tissues compared to the EMT tissues. Phospholipid was found in both the EMT and non-EMT tissues with slighty higher intensity in EMT tissues by 2% compared to non-EMT tissues. The C-C stretching of lipid was found at peak 1120 cm^{-1} for EMT tissues and 1078 cm^{-1} for non-EMT tissues with greater intensity by 21% in EMT tissues compared to that of non-EMT tissues. A comparison of the relative intensity of both EMT and non-EMT tissue is shown in Figure 4.13 to Figure 4.15.

Numerous studies have indicated that abnormal levels of lipids in tissues correlate with metastasis of various types of cancer, including breast cancer (Marro et al., 2014; Nieva et al., 2012). An early feature of carcinogenesis is the activation of lipid metabolism, being the hallmark of many types of cancer (Hilvo et al., 2011; Marro et al., 2014). It has been reported that lipids have an important role in controlling the adhesion of cell and migration processes (Chaturvedi et al., 2016b; Murai, 2012), a matter supported by clinical data showing lipid-rich breast cancer to be responsible for half of the deaths of patients with cancer (Machalekova et al., 2012; Murai, 2012). With EMT a process which acquires a mesenchymal phenotype with increased migratory features, present observation of high intensities of Raman bands corresponding to lipids in EMT tissues are seen to be in accordance with this.

University of Malaya

The results from Raman investigations of phospholipid in breast tissue obtained herein is in line with previous work by Marro et al., 2014 (Marro et al., 2014) and Chaturvedi et al., 2016 (Chaturvedi et al., 2016b), both indicated that the increases of phospholipid to be associated with the EMT process. Present results have shown that the phospholipid intensity in EMT tissue to be 69% greater compared to that of non-lesional tissues. For CH group lipids the signal intensity has been observed to be significantly greater in non-lesional tissues in comparison to that of EMT and non-EMT tissues, agreeing with the findings of Brozek-Pluska et al., 2012 (Brozek-Pluska et al., 2012b) and Li et al., 2017 (Q. Li et al., 2017).

University of Malaya

CH group lipids (*)

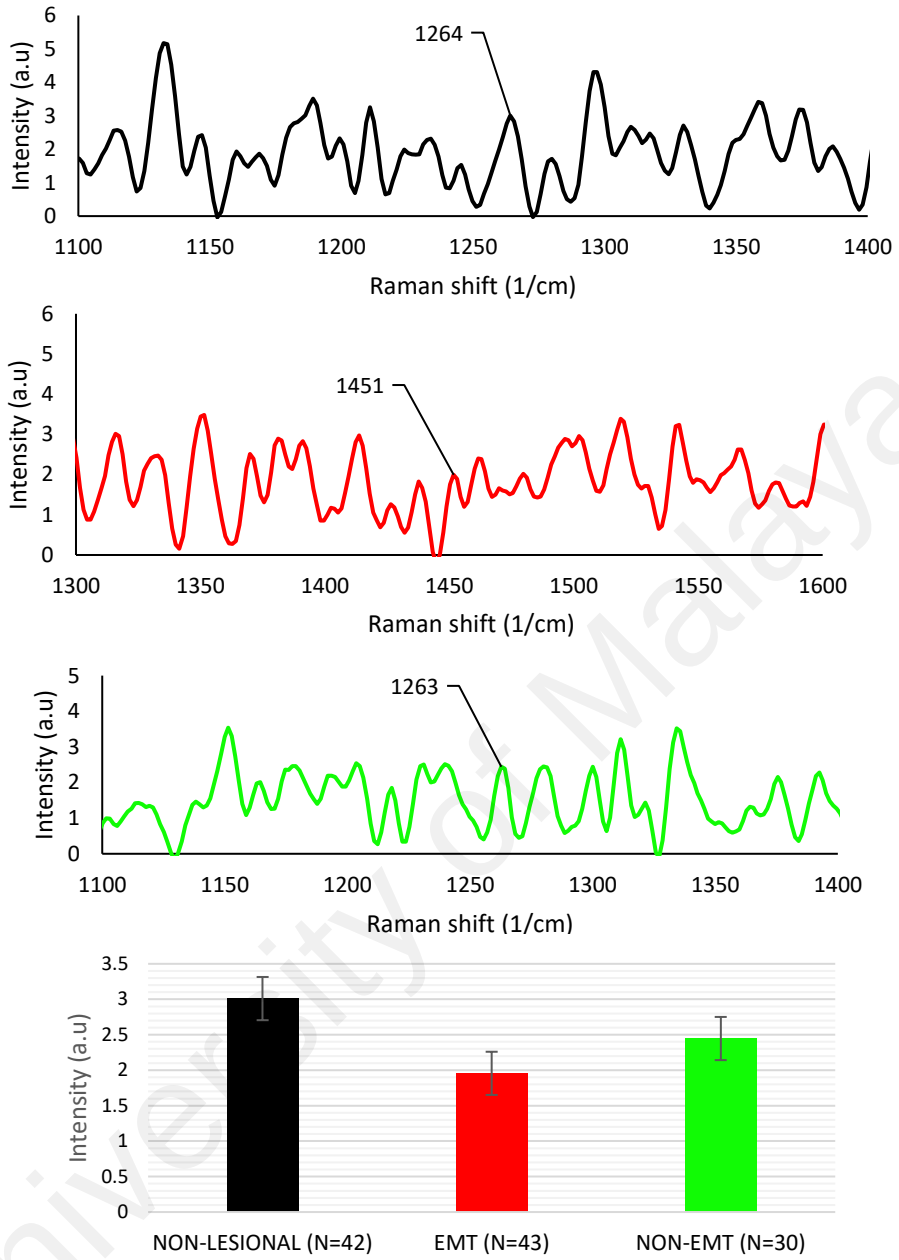


Figure 4.9: Summary of CH group of lipid vibrational bands observed for dewaxed samples (Group 1) in non-lesional, EMT and non-EMT breast cancer tissue. The black colour trace represents the non-lesional tissues, red colour represents EMT tissues and green colour represents non-EMT tissues. Comparison of the relative intensity of CH group lipid Raman bands in dewaxed samples between non-lesional, EMT and non-EMT breast tissue samples are depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. (*) denotes a statistically significant difference between the groups reported by the ANOVA test.

University of Malaya

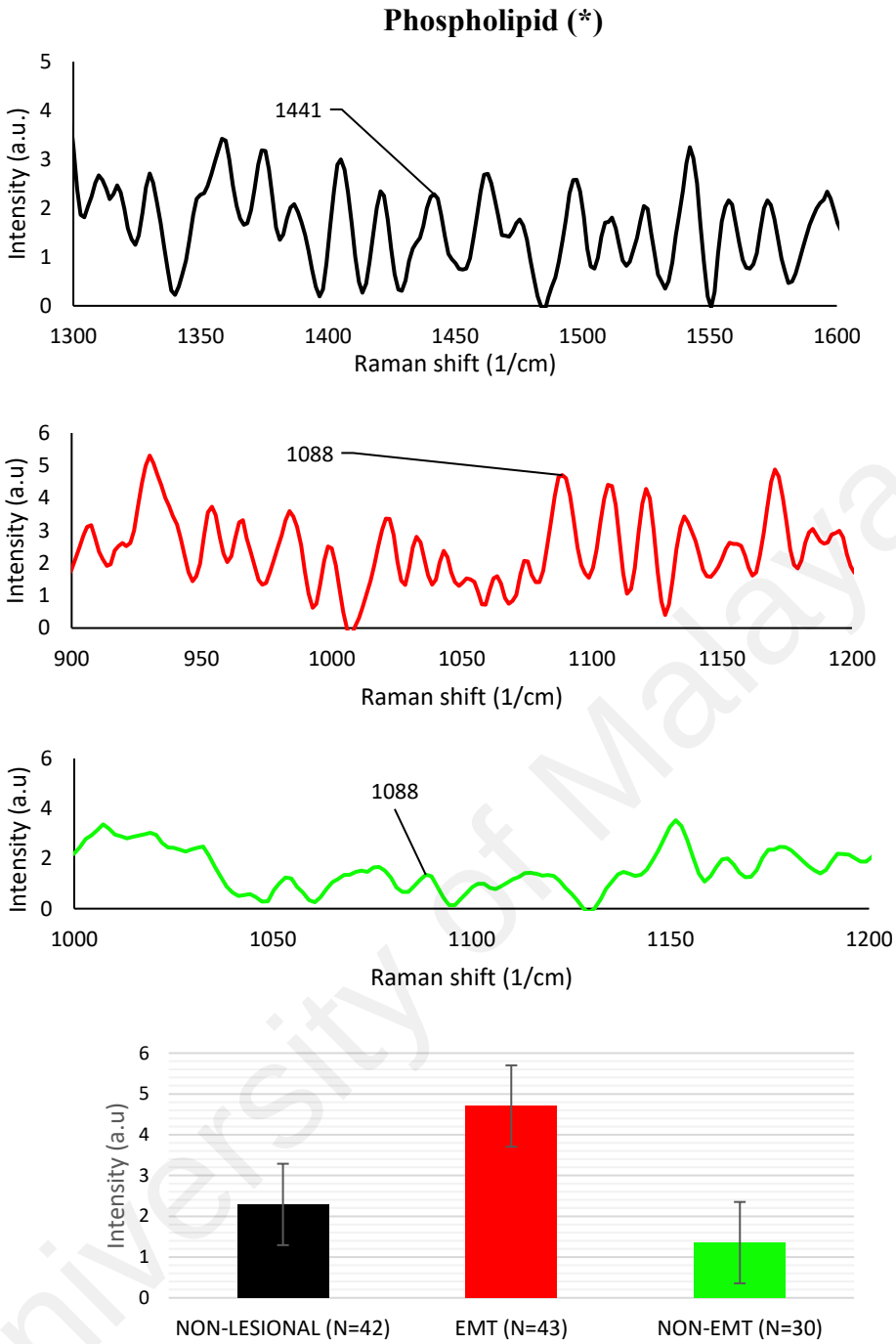


Figure 4.10: Summary of phospholipid vibrational bands observed for dewaxed samples (Group 1) in non-lesional, EMT and non-EMT breast cancer tissue. The black colour trace represents the non-lesional tissues, red colour represents EMT tissues and green colour represents non-EMT tissues. Comparison of the relative intensity of phospholipid Raman bands in dewaxed samples between non-lesional, EMT and non-EMT breast tissue samples are depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. (*) denotes a statistically significant difference between the groups reported by the ANOVA test.

University of Malaya

C-C stretching

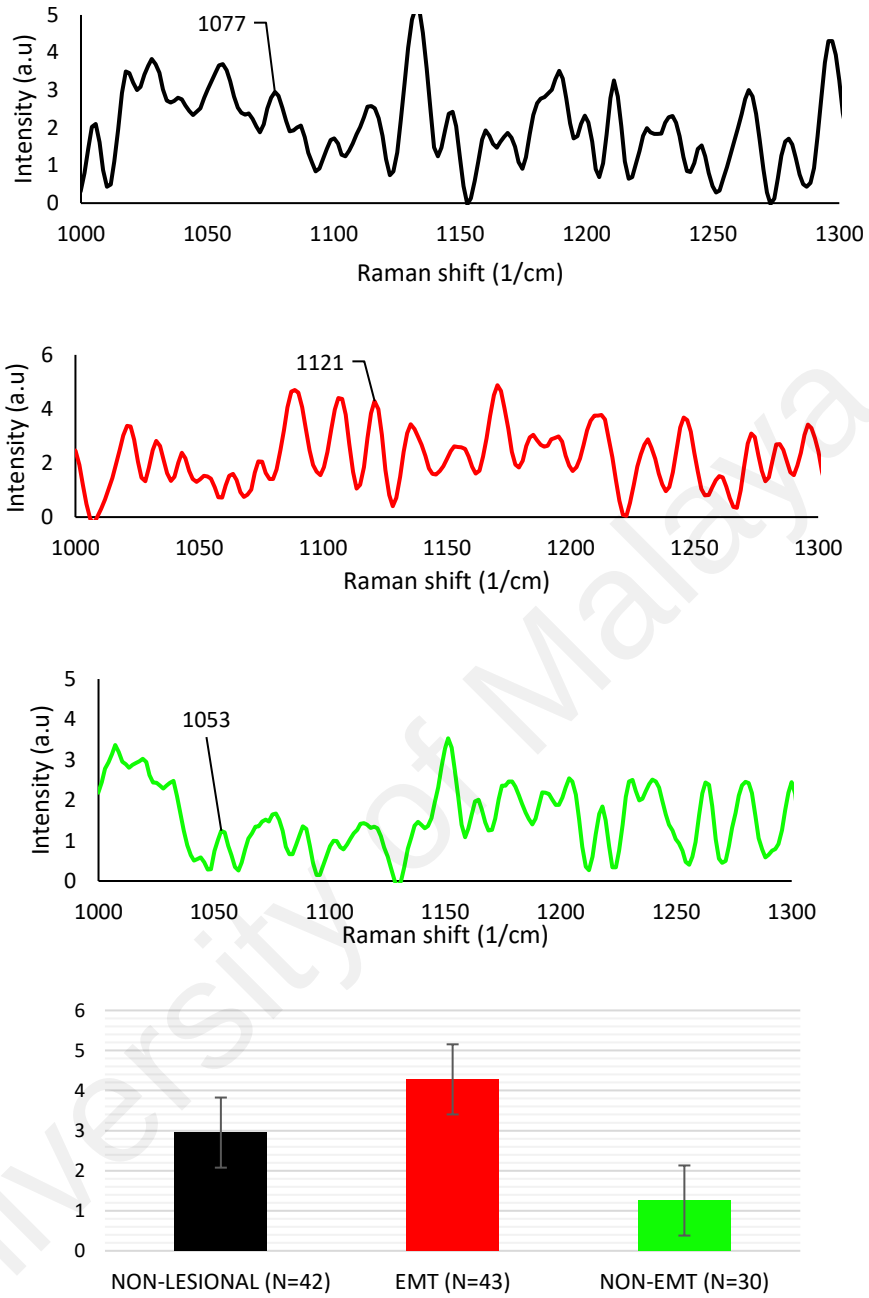


Figure 4.11: Summary of C-C stretching of lipid vibrational bands observed for dewaxed samples (Group 1) in non-lesional, EMT and non-EMT breast cancer tissue. The black colour trace represents the non-lesional tissues, red colour represents EMT tissues and green colour represents non-EMT tissues. Comparison of the relative intensity of C-C stretching lipid Raman bands in dewaxed samples between non-lesional, EMT and non-EMT breast tissue samples are depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. No statistically significant differences between the groups are reported by the ANOVA test.

University of Malaya

CH group lipids (*)

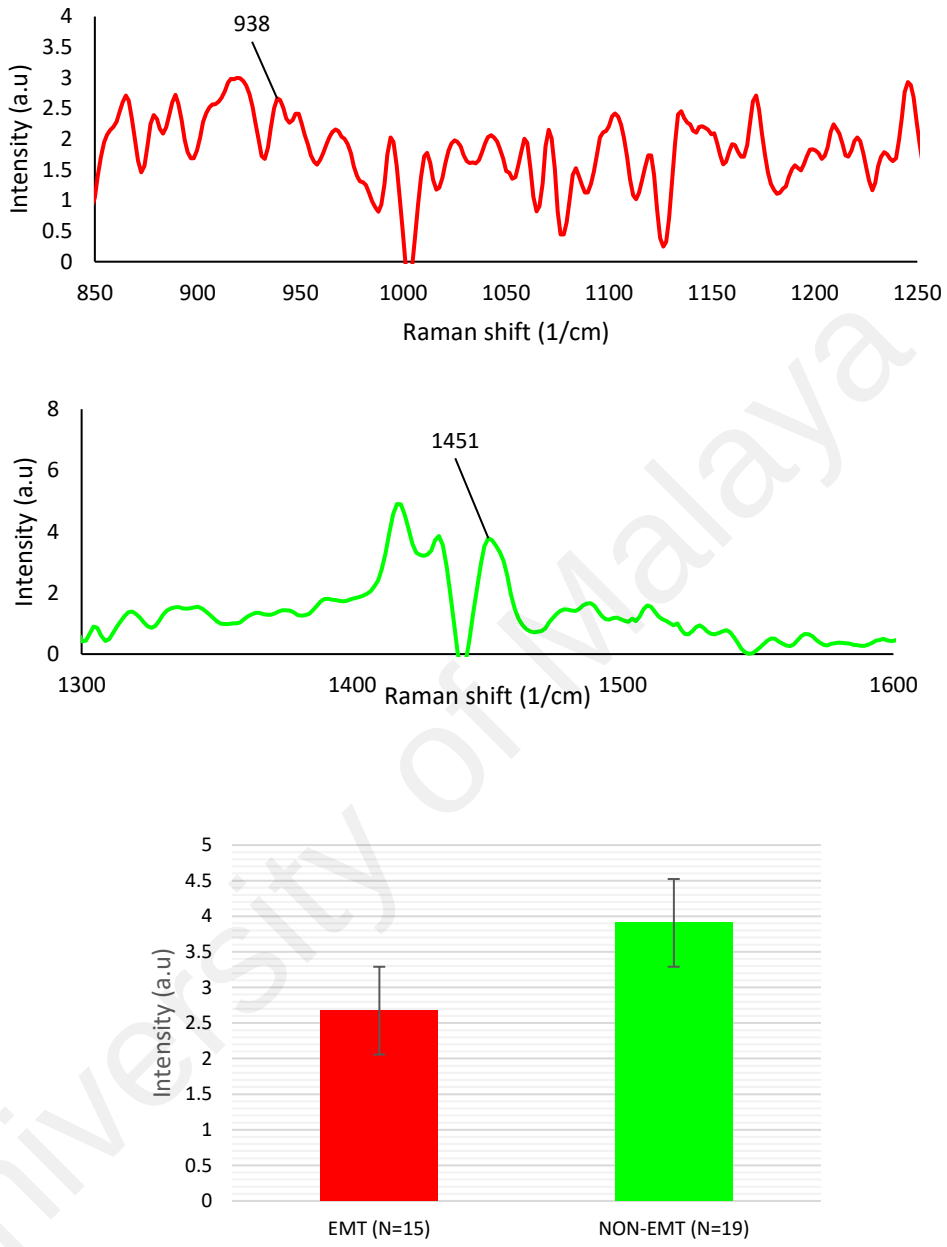


Figure 4.12: Summary of CH group lipid vibrational bands observed for waxed samples (Group 2) in EMT and non-EMT breast cancer tissue. The red colour represents EMT tissues and green represents non-EMT tissues. Comparison of the relative intensity of CH group lipid Raman bands in waxed samples between EMT and non-EMT breast tissue samples is depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. (*) denotes a statistically significant difference between the groups reported by the t-test.

University of Malaya

Phospholipid (*)

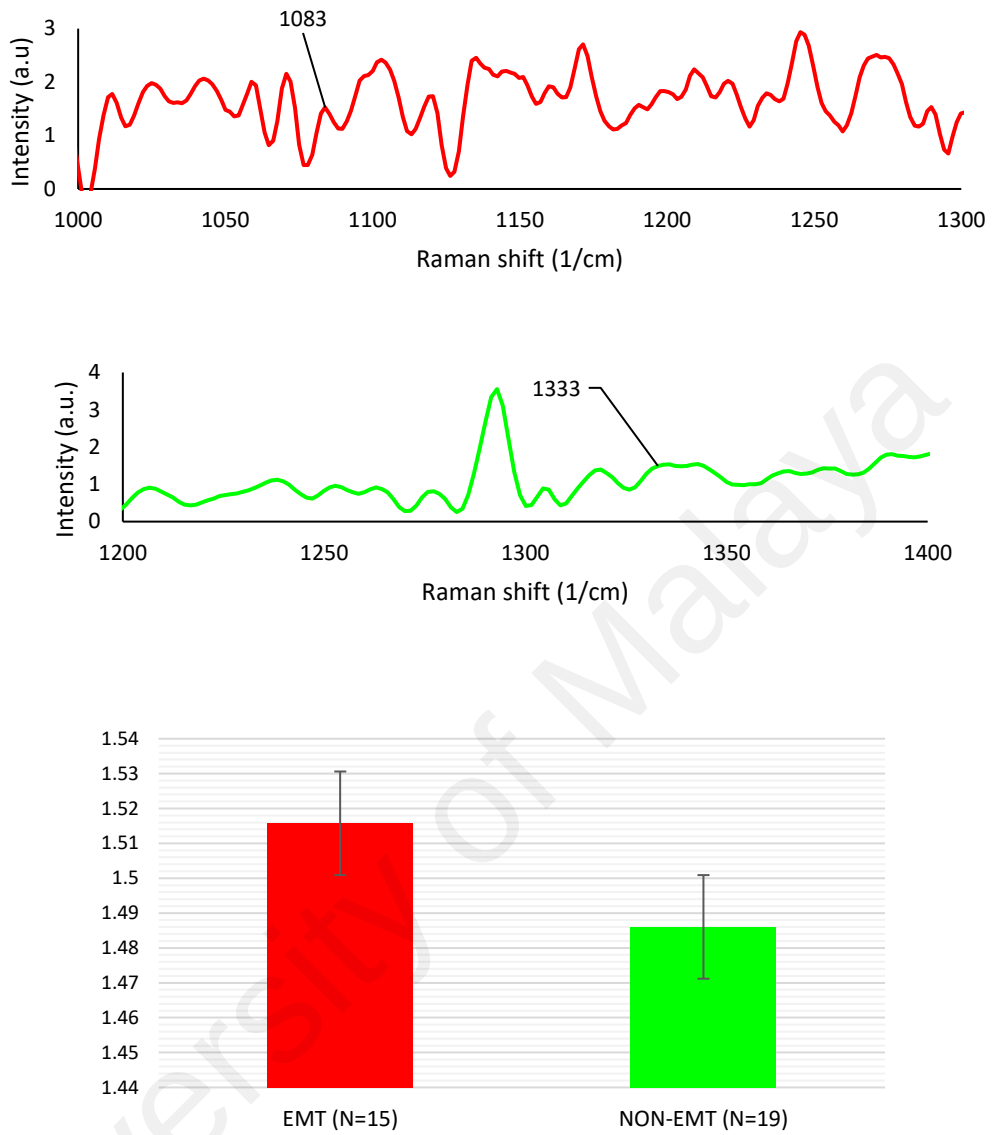


Figure 4.13: Summary of phospholipid vibrational bands observed for waxed samples (Group 2) in EMT and non-EMT breast cancer tissue. The red colour represents EMT tissues and green represents non-EMT tissues. Comparison of the relative intensity of phospholipid Raman bands in waxed samples between EMT and non-EMT breast tissue samples is depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. (*) denotes a statistically significant difference between the groups reported by the t-test.

University of Malaya

C-C stretching

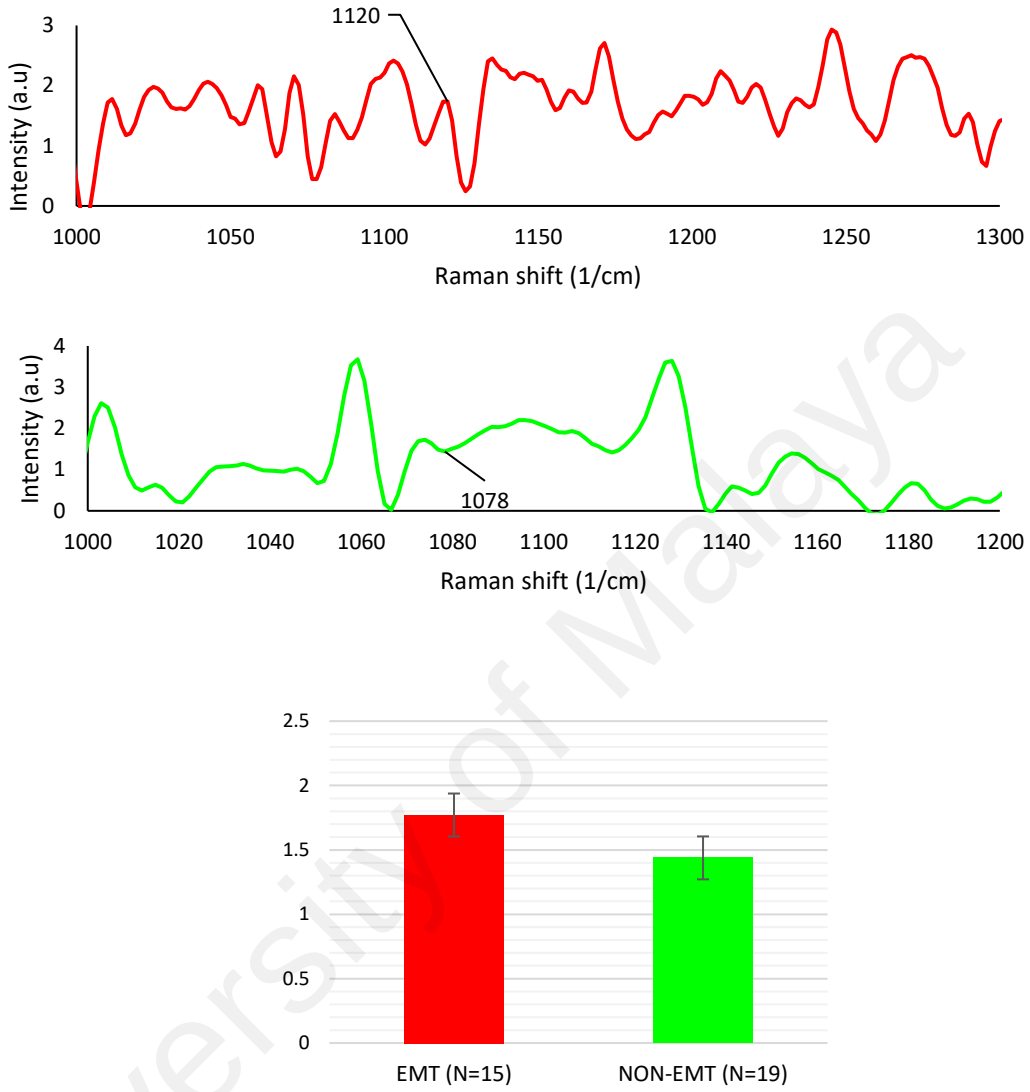


Figure 4.14: Summary of C-C stretching vibrational bands observed for waxed samples (Group 2) in EMT and non-EMT breast cancer tissue. The red colour represents EMT tissues and green represents non-EMT tissues. Comparison of the relative intensity of C-C stretching lipid Raman bands in waxed samples between EMT and non-EMT breast tissue samples is depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. No statistically significant difference between the groups reported by the t-test.

University of Malaya

4.1.3.2 Protein

(a) Dewaxed sample

In present work, greater intensity protein bands were observed in EMT compared to non-lesional and non-EMT breast cancer tissue. In non-lesional tissue Amide III was centered around 1223 cm^{-1} while in EMT tissue it was found at 1230 cm^{-1} and 1315 cm^{-1}; in non-EMT tissues the Amide III was found at 1230 cm^{-1} and 1263 cm^{-1}. CH group protein peaks were found for all three tissue types, at 1388 cm^{-1} for non-lesional tissues, 1461 cm^{-1} for EMT tissues and 1453 cm^{-1} for non-EMT tissues. The peak assigned to Amide I, a collagen assignment, was found at 1639 cm^{-1} in non-lesional tissues while it was absent in both EMT and non-EMT breast cancer tissues. Figure 4.11 presents mean Raman spectra for peaks attributed to proteins in non-lesional, EMT and non-EMT breast tissues samples. Comparison is made of the intensities of the protein bands in all samples, as shown in Figure 4.16 to Figure 4.18.

(b) Waxed sample

In the waxed samples, Amide III and CH group proteins were detected in both EMT and non-EMT tissues, with Amide III found at 1272 cm^{-1} in EMT tissues and 1277 cm^{-1} in non-EMT tissues. The CH group of protein are observed at 1343 cm^{-1} in both EMT and non-EMT tissues. Figure 8 presents mean Raman spectra for peaks attributed to proteins in both EMT and non-EMT breast tissues samples. Comparison is made of the intensities of the protein bands for all samples (Figure 4.19 to Figure 4.20).

University of Malaya

Amide III

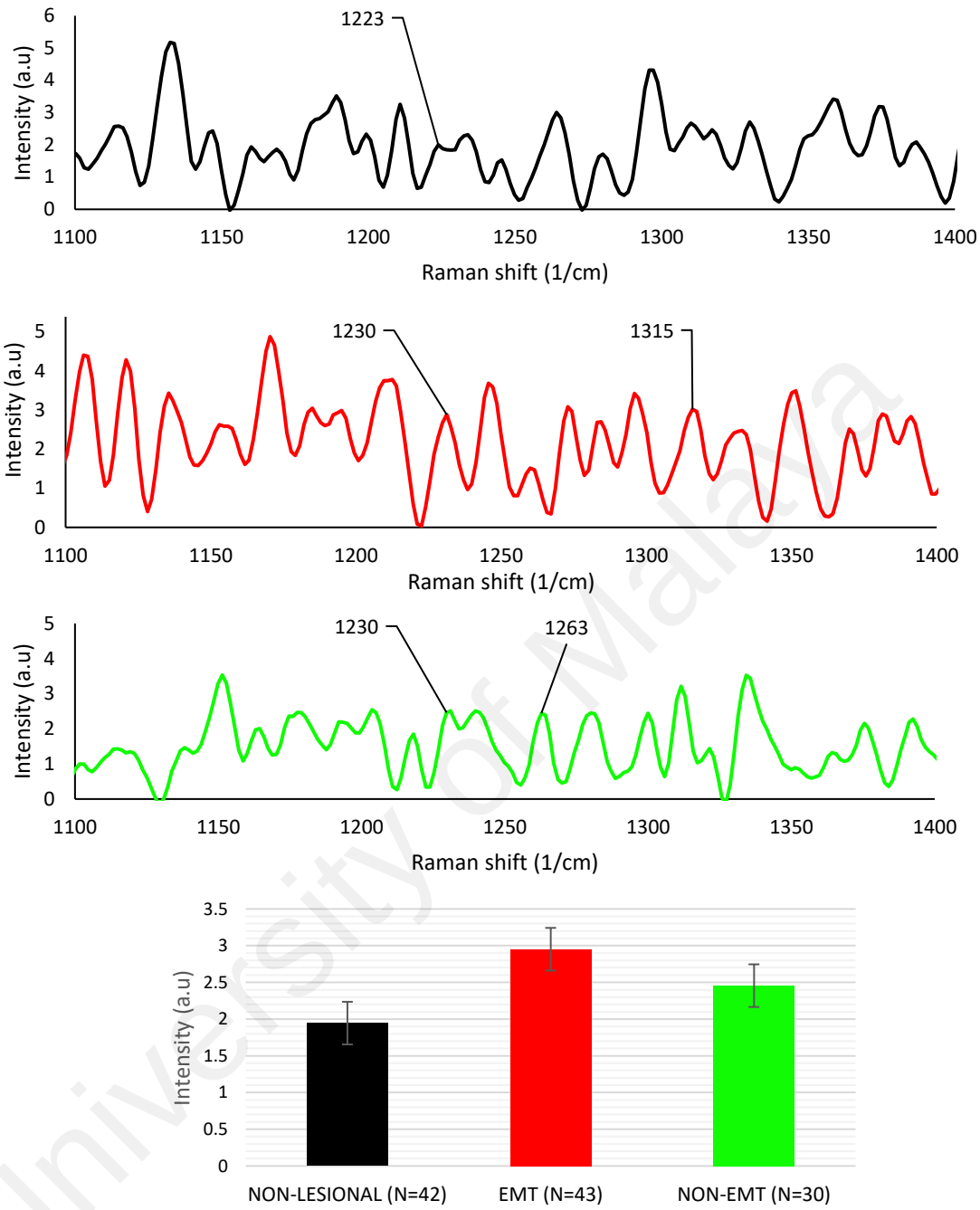


Figure 4.15: Summary of Amide III vibrational bands observed for dewaxed samples (Group 1) of non-lesional, EMT and non-EMT breast cancer tissue. The black colour represents the non-lesional tissues, while red represents EMT tissues and green represents non-EMT tissues. Comparison of relative intensity of Amide III Raman bands for dewaxed samples of non-lesional, EMT and non-EMT breast tissue is depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. Note: the x-axis represents the tissue type and y-axis is Raman intensity. No statistically significant difference between the groups reported by the ANOVA test.

University of Malaya

CH group proteins (*)

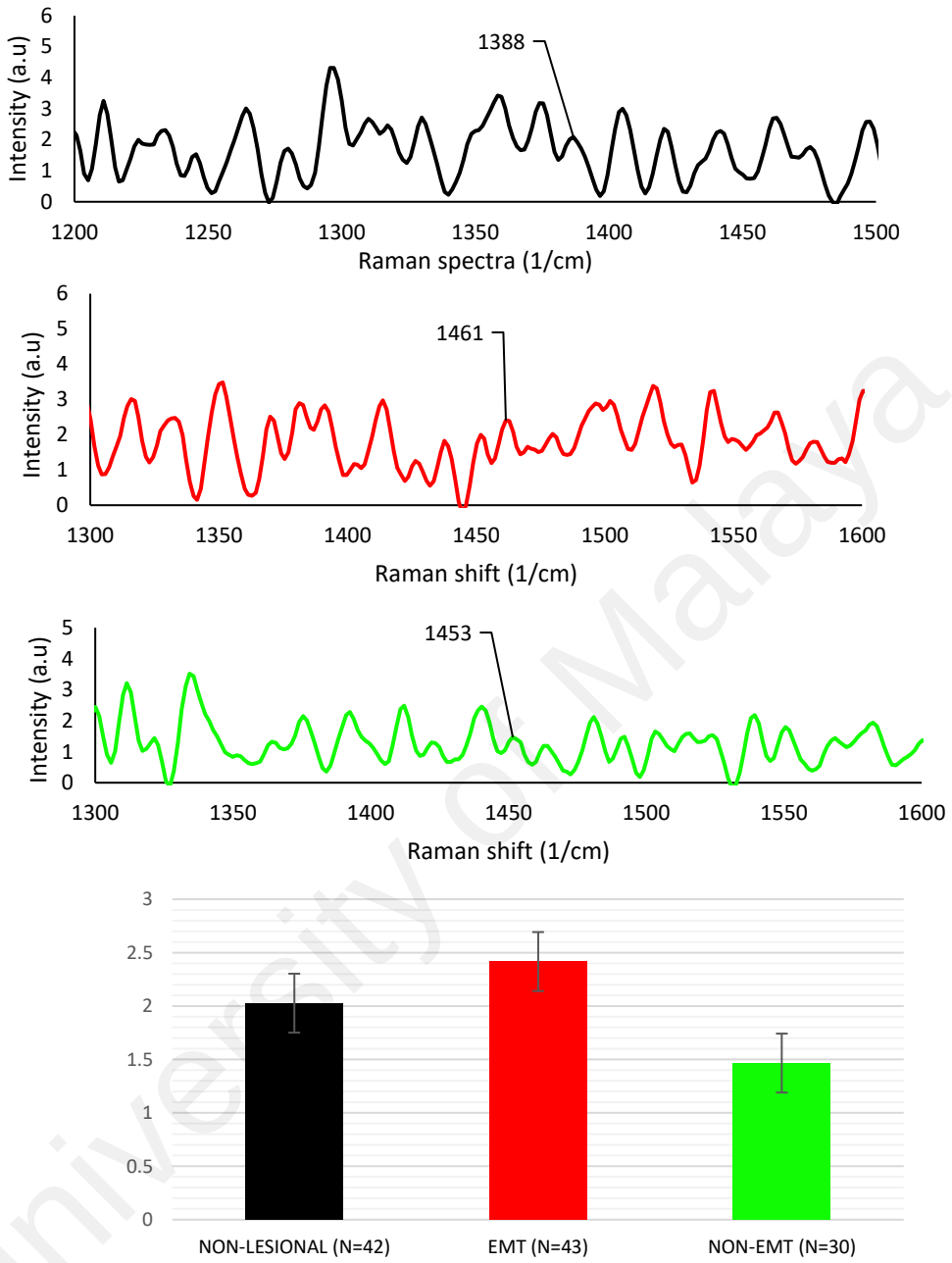


Figure 4.16: Summary of CH group protein vibrational bands observed for dewaxed samples (Group 1) of non-lesional, EMT and non-EMT breast cancer tissue. The black colour represents the non-lesional tissues, while red represents EMT tissues and green represents non-EMT tissues. Comparison of relative intensity of protein Raman bands for dewaxed samples of non-lesional, EMT and non-EMT breast tissue is depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. Note: the x-axis represents the tissue type and y-axis is Raman intensity. (*) denotes a statistically significant difference between the groups reported by the ANOVA test.

University of Malaya

Amide I

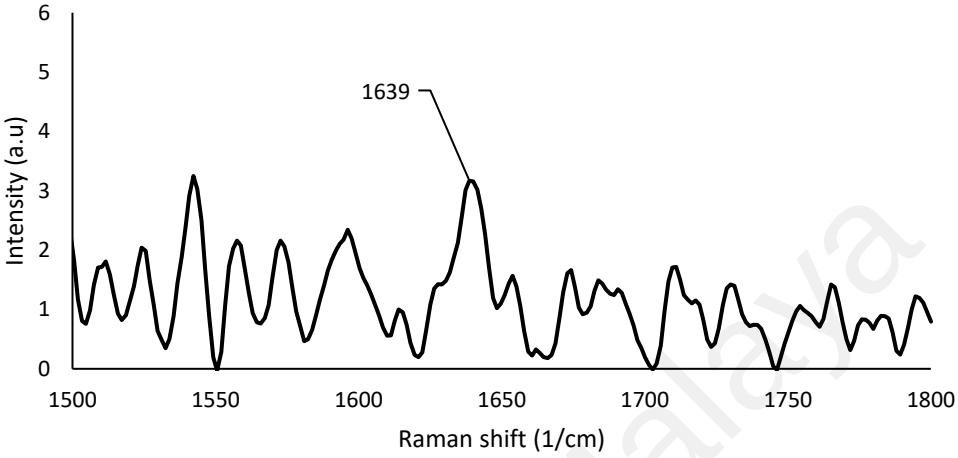


Figure 4.17: Amide I of non-lesional tissue.

University of Malaya

Amide III

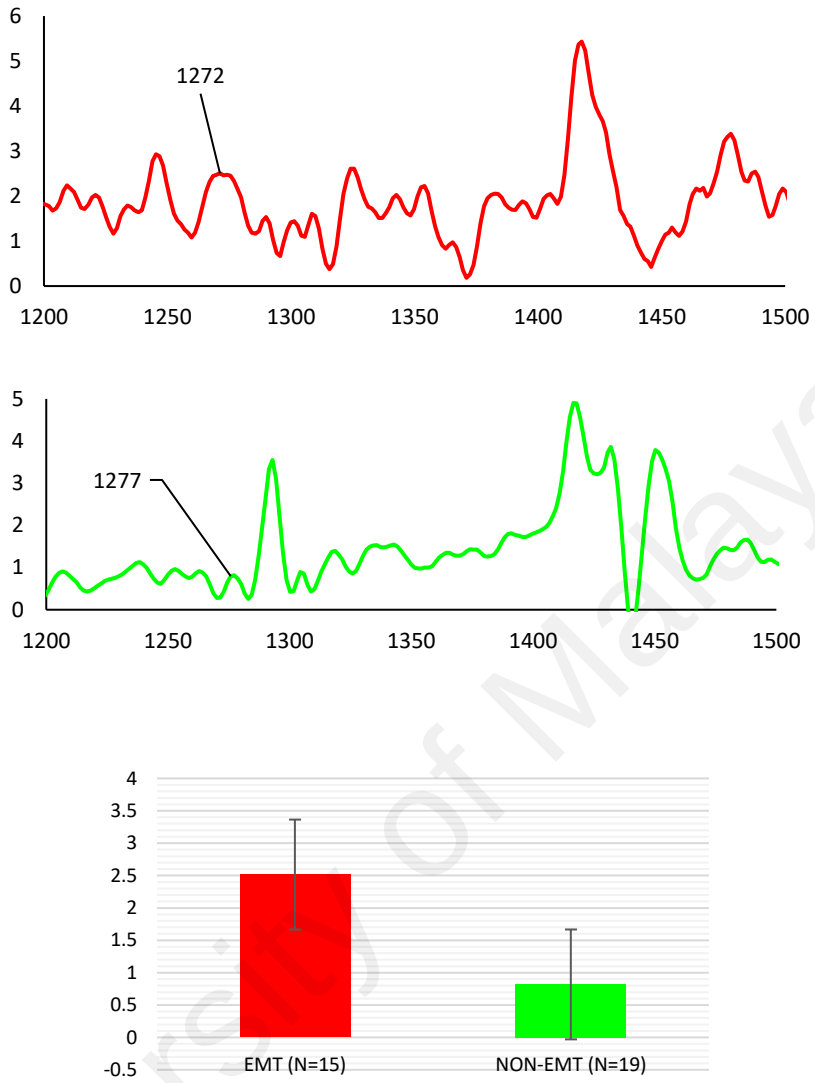


Figure 4.18: Summary of Amide III vibrational bands observed for waxed samples (Group 2) of EMT and non-EMT breast cancer tissue. The colour red represents EMT tissues and green represents non-EMT tissues. Comparison of the relative intensity of protein Raman bands for waxed samples of EMT and non-EMT breast tissue us depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. Note: the x-axis represents the tissue type and y-axis is Raman intensity. No statistically significant differences between the groups are reported by the t-test.

University of Malaya

CH group proteins

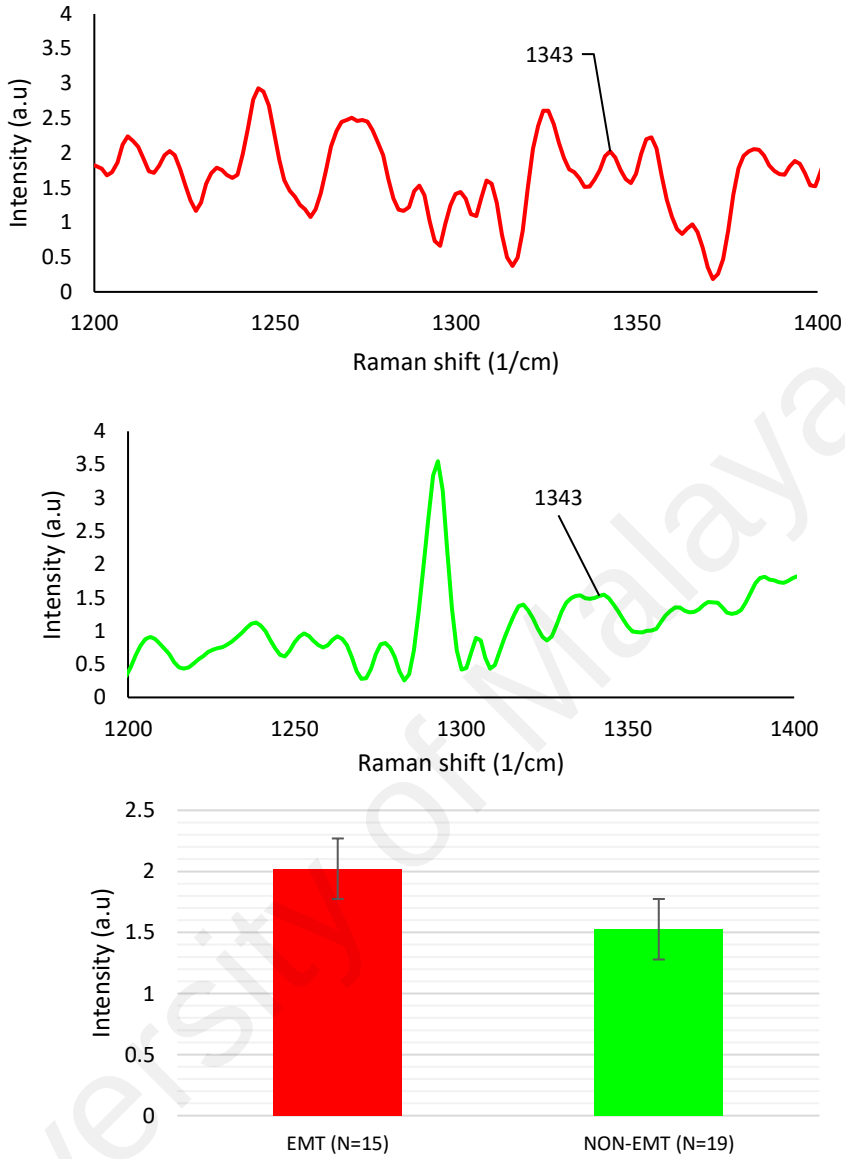


Figure 4.19: Summary of CH group protein vibrational bands observed for waxed samples (Group 2) of EMT and non-EMT breast cancer tissue. The colour red represents EMT tissues and green represents non-EMT tissues. Comparison of the relative intensity of protein Raman bands for waxed samples of EMT and non-EMT breast tissue us depicted in the bar charts. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. Note: the x-axis represents the tissue type and y-axis is Raman intensity. No statistically significant differences between the groups are reported by the t-test.

University of Malaya

There is some evidence (see below) of associations between proteins and cancer (Conklin et al., 2011; Gilles et al., 1999; F. Liu et al., 2016; Morin, 2005), in regard to present interest playing an important role in the EMT process, including overexpression of certain proteins. Among these, in cells that have been observed to be in progression towards invasive forms of breast cancer are: vimentin, N-cadherin, snail family zinc finger 1 (SNAIL), snail family zinc finger 2 (SLUG), twist family bHLH transcription factor 1 (TWIST), zinc finger E-box binding homeobox 1 (ZEB1) and zinc finger E-box binding homeobox 2 (ZEB2) (Grzegrzolka et al., 2015; Tomaskovic-Crook et al., 2009). It has been reported that high protein levels are linked to the division of cells, migration and gain cell proliferation in tumours (Brozek-Pluska et al., 2012a). Previous studies by others, using Raman spectroscopy on breast cancer tissues (but not on EMT specifically), have shown cancerous breast tissues contain more protein (Q. Li et al., 2017b; Manoharan et al., 2003). The protein bands found in this study are dominated by the amide groups.

The increased intensities of protein bands in Raman spectra are a result of increased proliferation in the EMT tissues, the Snail family having been shown to be involved in proliferation control during EMT processes (Savagner, 2010). The latter is described by a distinct pathway that explored in direct association with cancer progression (i.e. to trigger EMT-associated process), involving the protein tyrosine phosphatase Pez (Wyatt et al., 2007). The phosphorylation of proteins on tyrosine residues by tyrosine kinases and the reversal of the process by protein tyrosine phosphatases (PTPs) regulate many diverse cellular functions from proliferation to differentiation. The PTP Pez is an intracellular PTP that is located at the adherents junctions (AJ) in endothelial and epithelial cells (C. Wadham et al., 2000; Carol Wadham et al., 2003).

University of Malaya

Present work has found that the intensity of CH group protein in the dewaxed samples is greater in EMT tissues compared to that of non-lesional and non-EMT tissues by 18% and 49%, respectively. Meanwhile, in the waxed samples, the intensity of CH group protein is found to be 28% greater in EMT tissues compared to non-EMT tissues. The CH group protein results obtained herein are similar to that found by previous researcher; Hu et al., 2013 (Hu et al., 2013). Li et al., 2017 (Q. Li et al., 2017b) observed Amide III in cancerous tissue and close to absent in normal tissue, where present results has found that the intensity of Amide III in EMT tissues is greater to non-lesional tissues and non-EMT tissues by 41% and 18%, respectively. For the case of waxed samples, the intensity of Amide III is 102% greater in the EMT tissues compared to the non-EMT tissues.

4.1.3.3 Nucleic Acids

(a) Dewaxed sample

Raman spectra of nucleic acid bands were found in FFPE dewaxed tissue non-lesional, EMT and non-EMT samples. The peak assignments are showed in Figure 4.21. The relative intensities of nucleic acid band in non-lesional, EMT and non-EMT tissues have been compared, as shown in Figure 4.13. The O-P-O stretch of DNA has been found in non-lesional tissue at 832 cm-1 and at 831 cm-1 in both EMT and non-EMT breast cancer tissues.

(b) Waxed sample

For the case of FFPE waxed tissue samples, Raman spectra of the O-P-O stretch DNA bands are in both EMT and non-EMT tissues at 827 cm-1 and 832 cm-1 respectively as shown in Figure 4.22. The comparison of relative intensities for both EMT tissues also included in Figure 4.2.

University of Malaya

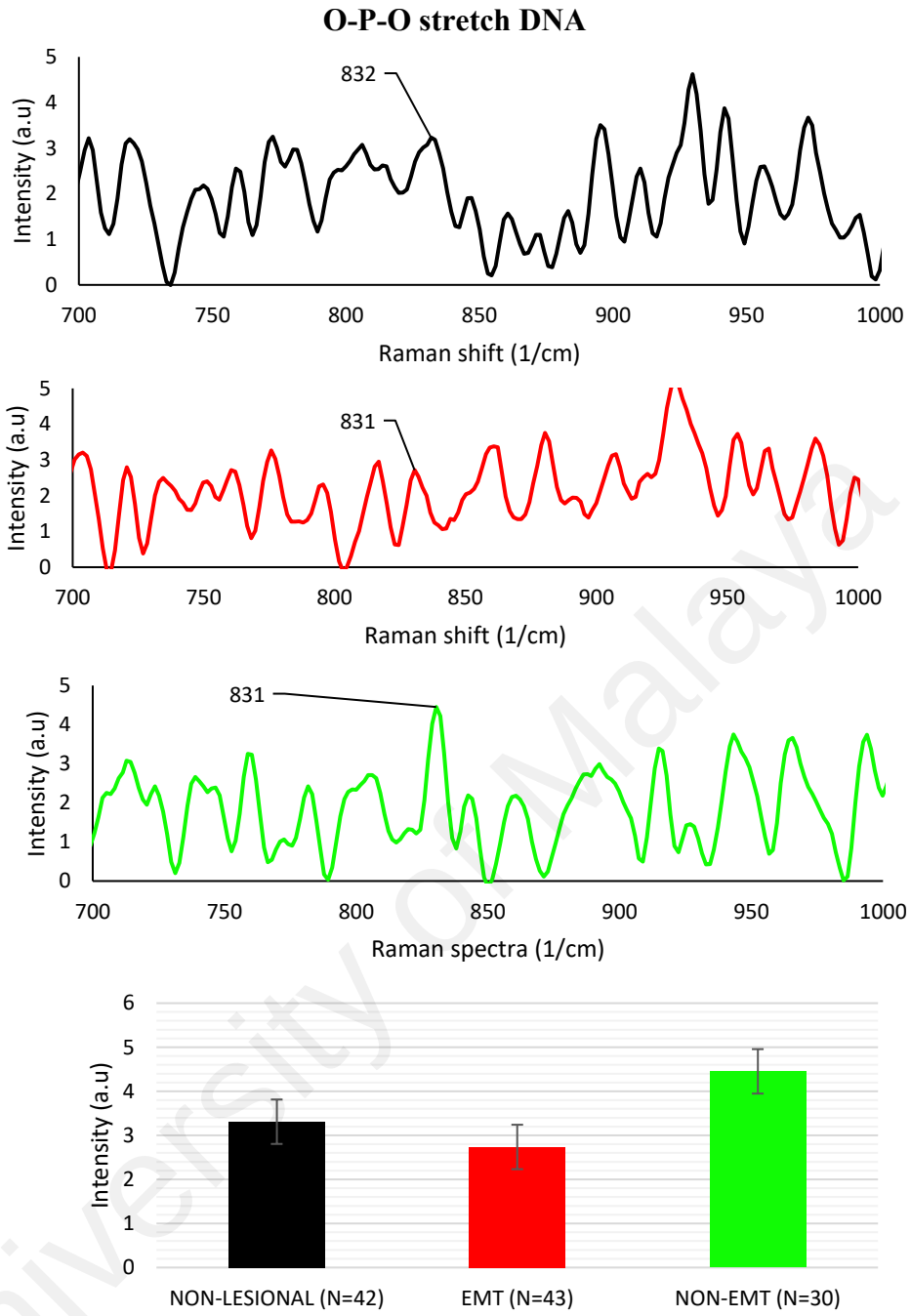


Figure 4.20: Summary of nucleic acid vibrational bands observed for dewaxed samples (Group 1) in non-lesional, EMT and non-EMT breast cancer tissue. The colour black represents the non-lesional tissues, while red represents EMT tissues and green represents non-EMT tissues. Comparison of the relative intensity of nucleic acid Raman bands for dewaxed samples in non-lesional, EMT and non-EMT breast cancer tissue is depicted in the bar chart. Each column represents the average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. Note: the x-axis represents the tissue type and y-axis is Raman intensity. No statistically significant differences between the groups are reported by the ANOVA test.

University of Malaya

O-P-O stretch DNA

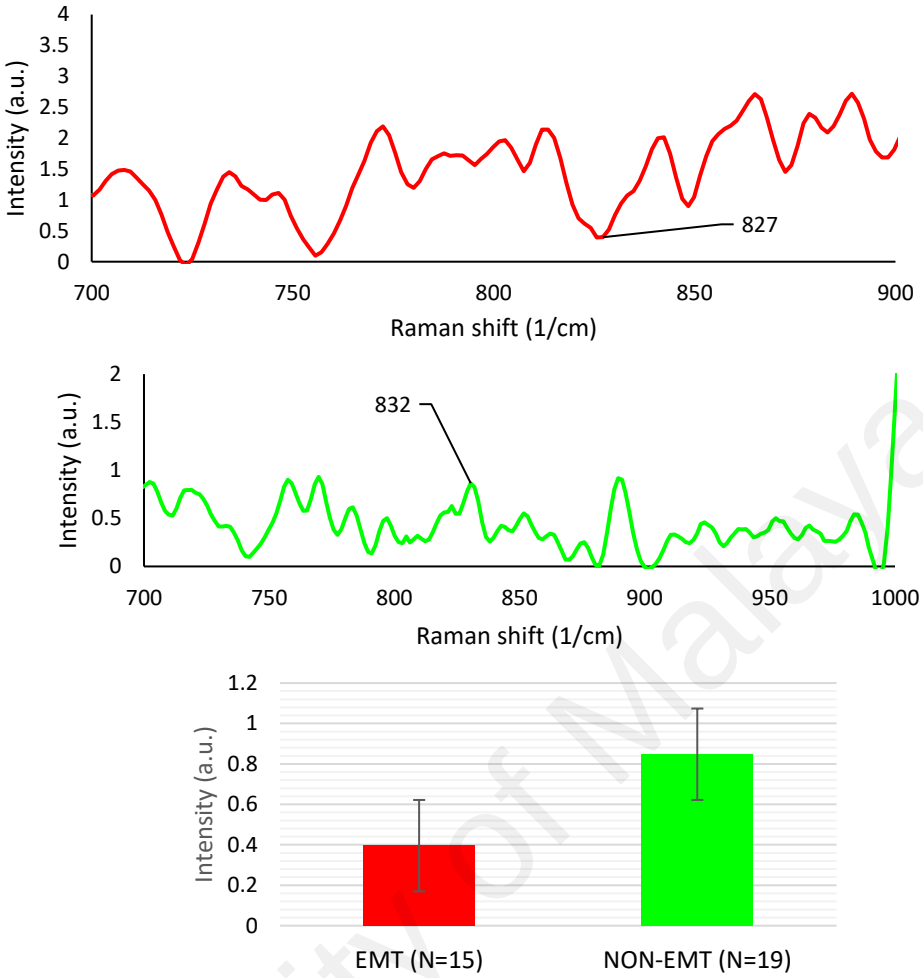


Figure 4.21: The nucleic acid vibrational bands of O-P-O stretch observed for waxed samples (Group 2) in non-EMT breast cancer tissue. Each column represents average of spectra for each type of tissue. The value of N represents the number of Raman spectra acquired for each sample, shown within the brackets. The error bar represents standard deviation. The x-axis represents the wavenumber while the y-axis represents the intensity. No statistically significant differences between the groups are reported by the t-test.

University of Malaya

Nucleic acids are basic molecules within which all the information needed to control and build cells are stored. Therefore, the nucleic acids may be excellent biomarkers for screening of breast cancer. In the study by Li et al. in 2015 on discrimination of breast cancer tissues from normal tissues the O-P-O stretch DNA signal was observed to be attenuated in cancer tissues (Q. B. Li et al., 2015). In present study, an attenuated nucleic acid signal is also observed in EMT tissues, reflecting the changing configurations and components of nucleic acids during EMT processes in cancer progression. For the dewaxed samples, the intensity of O-P-O stretch DNA is found to be 19% greater in non-lesional tissues compared to EMT breast cancer tissues and 48% higher in non-EMT tissues compared to that EMT tissues. Whereas in the waxed samples, the O-P-O stretch DNA of non-EMT tissues is 73% higher compare to the EMT tissues.

4.1.4 Statistical Analysis

4.1.4.1 Principal Component Analysis (PCA)

The general trends in the data from these measurements were studied using PCA. Two analyses of PCA were performed; the first PCA was performed on the dewaxed samples and the second PCA was performed on the waxed samples. The method used for the analyses are described below:

1. Dewaxed samples:
    a) Non-lesional (N=10)
    b) EMT (N= 9)
    c) Non-EMT (N=8)

University of Malaya

2. Waxed samples

a) EMT (N=5)

b) Non-EMT (N=5)

For the dewaxed samples, the data obtained from PCA are shown in figures 4.23A and 4.23B. For Figure 4.23A, the PCA results show that majority of the non-lesional samples are clustered in the negative area around the PC3 axis with values of between 0.00 to -1.00. Figure 4.23B which is part of Figure 4.23A, focuses on the EMT and non-EMT samples. From the figure, it shows that the majority of EMT samples are clustered in the negative area from 0.00 to -1.00 along the PC3 axis while the majority of non-EMT samples are clustered in the positive area from 0.00 to 2.00 along the PC3 axis. For the waxed samples, the score plots of the waxed samples in Figure 4.23C show the presence of two groups with one outlier from the EMT group. PC1 and PC3 account to 62.6% of the total variance. The ANOVA results for intensity show that significant differences exist, and the Tukey's multiple comparisons test revealed statistical significance while considering the EMT – non-EMT pair (p-value < 10-4) as well EMT– non-lesional pair (p-value < 10-7). On the other hand, the ANOVA test reports that there is a significant difference between waxed and dewaxed samples (p-value < 10-16) when intensity is of concern.

University of Malaya

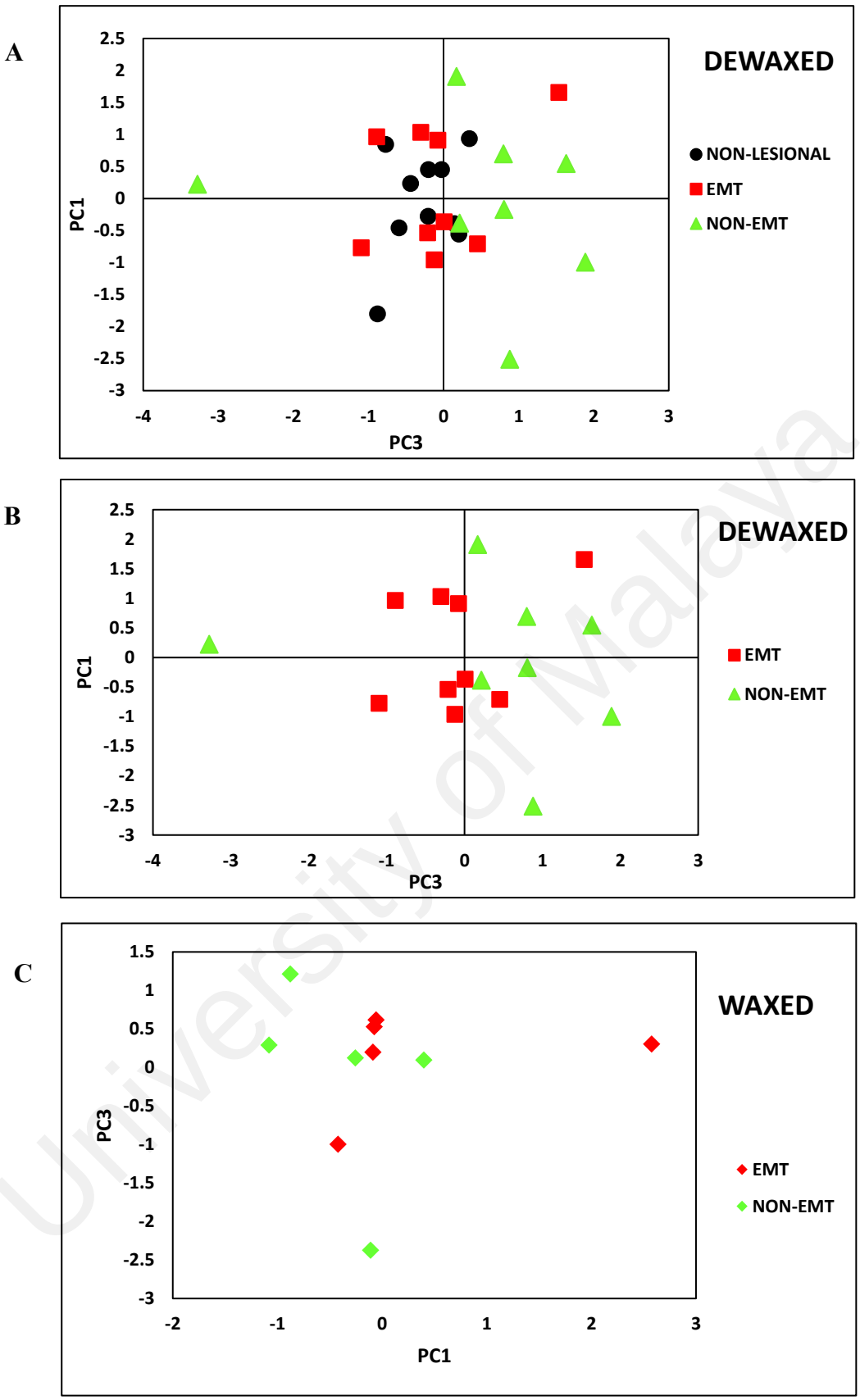


Figure 4.22: PCA plots of all studied samples; non-lesional, EMT and non-EMT breast tissues of both dewaxed (A and B) and waxed (C) samples.

University of Malaya

Finally, it is important to examine biochemical and medical biochemistry projections that may have traction in considering the present combination of Raman spectroscopy and histology findings. In respect of cell transformation in the breast, present work shows severely increased phospholipid content and C-C stretching during epithelial to stromal mesenchymal trans-differentiation. Attention is drawn to potential links to several original contributions within which depleting metabolic markers have been demonstrated clinically. These involve lung (Hao et al., 2016), rare childhood cancers (Gros et al., 2019), renal cell cancers (Pandey et al., 2020) and colorectal cancers (Arima et al., 2020), until now lacking evidence in breast cancer. The growing field of what is referred to as deutenomics has consistently shown the increased significance of deuterium depletion via natural cellular ketogenic substrate oxidation. The underlying medical biochemistry mechanisms, described by Boros et al. (Boros et al., 2017; László G Borosa et al., 2016), have considered how defective mitochondria with diminished low deuterium ketogenic fatty acid substrate oxidation can hamper recycling of deuterium depleted metabolic water. This is seen to be obtained via tricarboxylic acid cycle (TCA cycle) hydratase reactions (Boros & Somlyai, 2014). Such mechanism could preserve normal epithelial cellular mesenchymal phenotype in breast to prevent or reverse cancer formation.

University of Malaya

4.2 Small Angle X-ray Scattering (SAXS) Data Analysis

4.2.1 SAXS analysis of 16 classified breast tissues

Figure 4.24 depicts the scattering profile of each of the tissue groups analyzed in this work, including non-lesional, benign and malignant, EMT and non-EMT. Based on the scattering profiles, a number of features were analyzed, as follows: (i) the 3rd, 5th, 9th and 10th order Bragg peaks located between 0.175 – 0.200 nm^{-1}, corresponding to the axial organization of the collagen fibrils; (ii) the peak at 1.44 nm-^{1} for EMT and non-EMT lesions; (iii) 1.46 nm^{-1} for non-lesional tissue, corresponding to the scattering of triglycerides acknowledging this to be one of the most significant of the constituents of lipids. Fat peaks are not present in any of the benign tissue, stroma not containing triglycerides. Finally, is a fourth category, a diffuse scattering component manifests a rapidly decreasing, resulting from disordered collagen fibril structures and other constituents of the extracellular matrix.

Observed is that the scattering intensity is greatest in EMT malignant tissue, followed by non-EMT malignant tissue and subsequently benign tissue, moderate in intensity, the lowest intensity being for non-lesional tissue. This arises from the propagation of cancer and invasion of normal tissues, leading to changes in collagen fibril structural organization, then causing the endogenous collagen to degenerate and new collagen of diminished structural organization to emerge (Pucci Minafra et al., 1986). Notably, EMT lesions manifest at the 9th and 10th order while other tissues manifest in 3rd and 5th order peaks, resolved to be due to the loss of E-cadherin expression in tissues suffering invasion and metastasis (Wang & Zhou, 2013). E-cadherin functions to yield mechanically strong adhesive links between cells in the tissue and is crucial in maintaining epithelial morphology (Harner-Foreman et al., 2017; Le Bras et al., 2012). It is then replaced by mesenchymal specific N-cadherin and integrins specific for a more transitory adhesion,

University of Malaya

activating the EMT, invasion and metastasis. In addition, matrix metalloproteinases (MMPs) emerge to degrade the basement membrane, leading to the migration of cancer into the surrounding stroma.

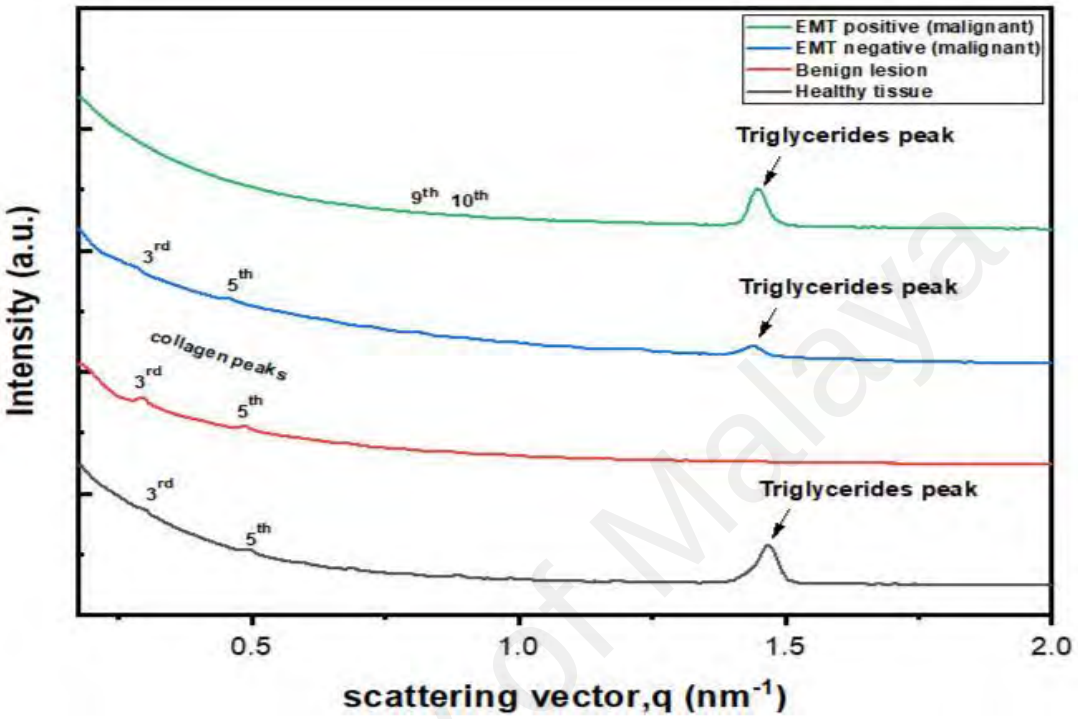


Figure 4.23: Typical scattering profiles of four groups of breast tissue analyzed in present work.

University of Malaya

Table 4.5 shows analysis of the axial spacing, peak area and FWHM (mean and range). Figure 4.25 demonstrates appreciably larger d-spacing for EMT malignant lesion compared to the other tissue types, a consequence of cancer propagation into the breast tissues and increase of N-cadherin expression, the tissue losing cell-cell adhesion as previously mentioned, causing the long-range intermolecular bonding rigidity of collagen to diminish. Further apparent are smaller values of axial peak areas in benign tissue and malignant lesions compared to normal tissue (Table 4.5 refers). The reduction of the structural order of collagen within lesions and consequent decrease in intensity can be associated with the activity of collagenases, a group of collagen degrading enzymes (Round et al., 2005). Present study has observed benign lesions to be associated with a larger FWHM in the third-order peak than that for non-EMT lesions, also for normal tissue and even EMT at the 9th and 10th order (see Table 4.5), a result similar to that found by previous researchers (Sidhu et al., 2008). Larger FWHM is expected on the grounds that greater variability is to be found in the fibril packing structure of benign lesions.

University of Malaya

Table 4.5: Summary peak parameters for 16 classified breast tissues.

Parameters	Group of Tissue	Material Source	Mean ± sd
Axial d-spacing (nm)	Non-lesional (n=4)	Formalin fixed	63.7 ± 0.1
	Non-lesional (n=2)	FFPE (dewaxed)	63.4 ± 0.3
	Benign (n=4)	Formalin fixed	64.0 ± 0.7
	Non-EMT (n=4)	Formalin fixed	64.3 ± 0.8
	EMT (n=2)	FFPE (dewaxed)	64.7 ± 0.8
Axial Peak Area (a.u.)	Non-lesional (n=4)	Formalin fixed	$(3.4 \pm 1.1) \times 10^{-8}$
	Non-lesional (n=2)	FFPE (dewaxed)	$(4.7 \pm 2.2) \times 10^{-8}$
	Benign (n=4)	Formalin fixed	$(2.9 \pm 4.4) \times 10^{-8}$
	Non-EMT (n=4)	Formalin fixed	$(4.1 \pm 0.9) \times 10^{-9}$
	EMT (n=2)	FFPE (dewaxed)	$(4.7 \pm 0.19) \times 10^{-10}$
Third- order FWHM (nm^{-1})	Non-lesional (n=4)	Formalin fixed	$(16.3 \pm 0.7) \times 10^{-3}$
	Non-lesional (n=2)	FFPE (dewaxed)	$(15.6 \pm 1.0) \times 10^{-3}$
	Benign (n=4)	Formalin fixed	$(18.8 \pm 0.1) \times 10^{-3}$
	Non-EMT (n=4)	Formalin fixed	$(13.7 \pm 2.9) \times 10^{-3}$
Ninth-order FWHM (nm^{-1})	Non-lesional (n=4)	FFPE (dewaxed)	$(23.6 \pm 1.9) \times 10^{-3}$
	Non-lesional (n=2)	Formalin fixed	$(18.9 \pm 3.0) \times 10^{-3}$
	Benign (n=4)	Formalin fixed	$(15.0 \pm 0.6) \times 10^{-3}$
	Non-EMT (n=4)	FFPE (dewaxed)	$(3.7 \pm 1.8) \times 10^{-3}$
	EMT (n=2)	FFPE (dewaxed)	$(4.5 \pm 2.7)\ 10^{-3}$
Tenth-order FWHM (nm^{-1})	Non-lesional (n=4)	FFPE (dewaxed)	$(24.1 \pm 2.4) \times 10^{-3}$
	Non-lesional (n=2)	Formalin fixed	$(14.8 \pm 1.5) \times 10^{-3}$
	Benign (n=4)	Formalin fixed	$(12.1 \pm 1.4) \times 10^{-3}$
	Non-EMT (n=4)	FFPE (dewaxed)	$(16.0 \pm 4.4) \times 10^{-3}$
	EMT (n=2)	FFPE (dewaxed)	$(13.8 \pm 5.2)10^{-3}$
Triglyceride FWHM (nm^{-1})	Non-lesional (n=4)	Formalin fixed	$(53.7 \pm 13.6) \times 10^{-3}$
	Non-lesional (n=2)	FFPE (dewaxed)	$(40.0 \pm 10.0) \times 10^{-3}$
	Benign (n=4)	Formalin fixed	No peak
	Non-EMT (n=4)	Formalin fixed	$(88.8 \pm 15.4) \times 10^{-3}$
	EMT n=2)	FFPE (dewaxed)	$(49.1 \pm 27.9) \times 10^{-3}$
Triglyceride axial peak area (a.u.)	Non-lesional (n=4)	Formalin fixed	$(0.02 \pm 0.01) \times 10^{-6}$
	Non-lesional (n=2)	FFPE (dewaxed)	$(0.4 \pm 0.02) \times 10^{-6}$
	Benign (n=4)	Formalin fixed	No peak
	Non-EMT (n=4)	Formalin fixed	$(0.039 \pm 0.006)10^{-6}$
	EMT (n=2)	FFPE (dewaxed)	$(0.23 \pm 0.008)10^{-6}$

University of Malaya

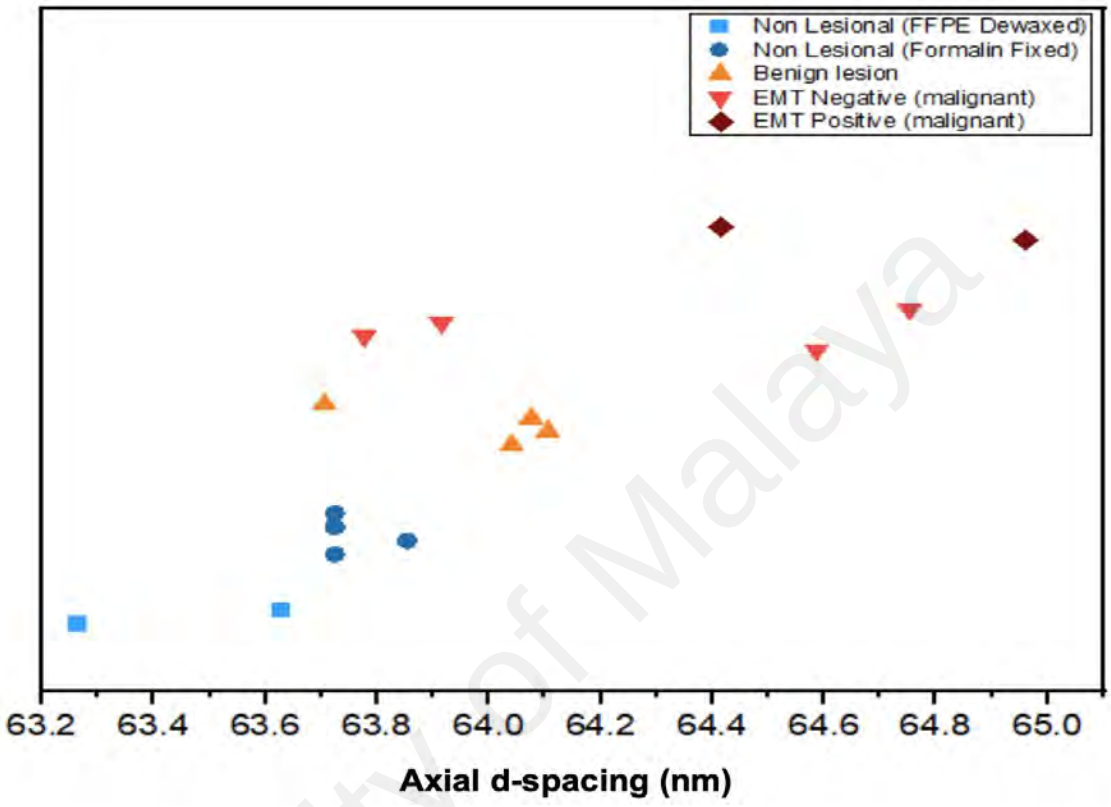


Figure 4.24: Individual axial d-spacing (nm) points for the 16 groups of classified tissues, including non-lesional, benign, non-EMT (labelled as EMT negative) and EMT (labelled as EMT positive). Note: the data points are offset vertically for clarity

University of Malaya

4.2.2 SAXS analysis of 51 blinded breast tissues

In regard to differences in axial d-spacing between non-lesional and malignant breast tissues, as seen in the summary results of the 16 classified breast tissues investigation (Figure 4.26), this is associated with the disruption of collagen fibrils structure. Note that collagen is a long-range periodic structure, repeated along the fibril axis, containing molecules packed laterally to each other with hydrogen bonds, binding to each other longitudinally in a staggered arrangement. Tissue remodeling is an essential step during the malignant process, transformation of epithelial cells being associated with metalloproteinases in collagen degradation by synthesis of fibrillary and non-fibrillary matrix proteins (Coussens & Werb, 1996). Extensive alteration of the ECM has been observed, collagen derangement being attributed to enzymatic degradation and altered neosynthesis. It is evident that at the supramolecular level changes in collagen in tissue invaded by cancer can be described quantitatively from the SAXS patterns, paving the way for tissue classification and improved understanding of mechanisms of cancer growth.

In this section, using a similar experimental SAXS setup to that above, the study has been repeated for 51 blinded breast tissues (FFPE (dewaxed)), seeking to distinguish differences between the tissues groups, non-lesional and malignant (EMT and non-EMT breast tissue). Biological tissue SAXS studies (Arnedos et al., 2009; Changizi et al., 2005; Falzon et al., 2006; Fernández et al., 2004) often place emphasis on odd-ordered Bragg peaks in obtaining the periodicity within the collagen present in the sample. It is apparent that the odd peaks are larger than the neighboring even peaks, revealing a qualitative feature of the structure. The nth order Bragg peak derives from the nth Fourier component of the electron density unit cell (Arthur, 1923), contributing to the odd cosine waves with

University of Malaya

feature that the amplitude is of the opposite sign at the centre of the unit interval than it is at the edges. This is a known feature of the structure of long-chain hydrocarbons exist in extracellular matrix (ECM).

Highly-ordered healthy collagen exhibit sharp and prominent peaks from the 3rd to the 10th order; conversely, axial ordering gradually disappears in invaded regions, most particularly for EMT tissues, as in Figure 4.26, demonstrating greater degradation in diseased tissues than surrounding healthier tissue. Variations in axial periodicity result from a change in crimping angle (James et al., 1991), also if the normal type I collagen molecule is replaced by the trimer in which the three chains are identical (Pucci Minafra et al., 1986) and high concentration of type III is ruled out (James et al., 1991). It has been suggested that the contact of epithelial cells with collagen I contributes to increased cell motility accompanying EMT, critical in disease progression (Liotta L & Kohn E, 2001). For invaded regions the scattering intensity has been observed to be significantly greater, the triglyceride peak also being broader, EMT breast cancer tissue in particular where collagen fibrils can suffer “peeling off”, corresponding to large increase in the specific surface area per unit volume of scatterers, agreeing with the findings of Fernandez et al. (Fernández et al., 2004).

University of Malaya

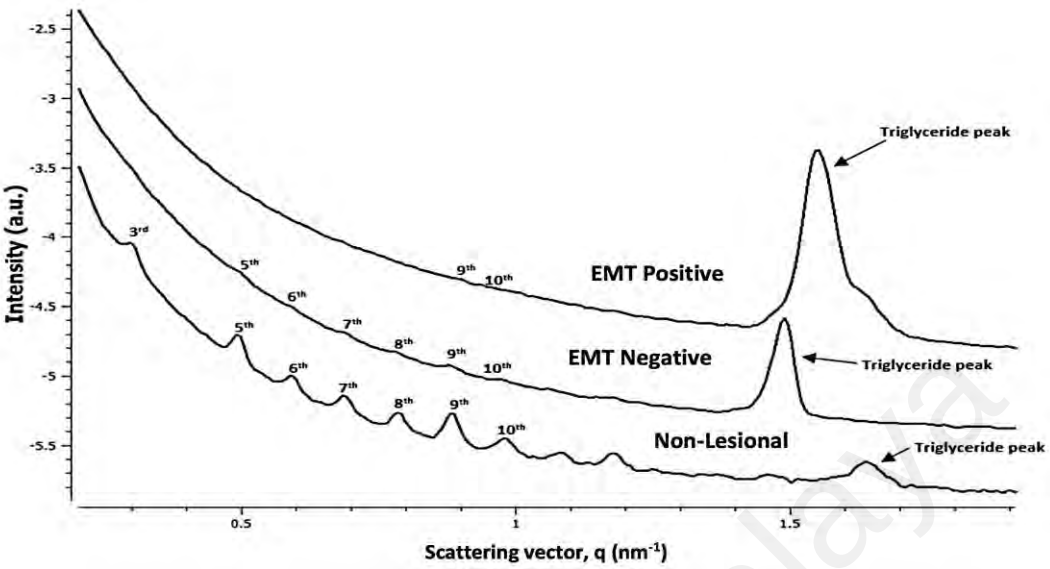


Figure 4.25: Wide scattering profile of blinded samples, including non-lesional, malignant; non-EMT (labelled as EMT negative) and EMT (labelled as EMT positive) breast tissues.

Fibrils orientation was determined from the maximum intensity of the collagen peaks; the calculated axial peak area and FWHM are detailed in Table 4.7. Small but systematic differences are seen in axial d-spacing, the period of collagen being slightly greater in fibrils surrounding invasive carcinoma cells than in non-lesional areas, interpreted as an increase in axial periodicity with malignancy. It has also been reported (Pertovaara, 2000) that breast tumours display aberrant collagen bundle organization in the tumour stroma. The FWHM and peak amplitudes show significant differences between diseased and non-lesional breast tissues. In the case of collagen, the FWHM is related to the distribution of axial d-spacings in the fibrillar collagen and the amount of variability in the distribution of particle sizes in the sample. The FWHM of the investigated samples show a general trend of decreasing FWHM with severity of disease, in particular EMT. A larger width would depict more variability in the distance between the collagen fibrils compared to a peak with a smaller width. Here, it is important to keep in mind that the peak area is the

University of Malaya

incorporation of the peak width and peak amplitude. The axial peak area describes the extent of the long-range order in the packing structure of the collagen, arising from the fact that x-ray scattering involves constructive interference, whereby the larger the area, the greater the degree of long-range order of the samples. In Table 4.7, it can be observed that the axial peak area decreases with malignancy. The reduction of the peak area is possibly caused by the action of collagen-degrading enzymes known as collagenases as found in other work (Round et al., 2005). In the case of a greater degree of ordering (i.e. non-lesional tissue) this will be portrayed as a combination of broader width and greater amplitude, thus of larger axial peak area, vice-versa in the less ordered malignant tissue. It should be noted, as stated in Table 2.6 (Chapter 2), that peak area depends on the abundance of material within sample, thickness and composition in particular. The dissected-out samples are of nominal dimensions 5 mm (length) x 5 mm (breadth) x 1 mm (thickness), assumed to be sufficiently equal across samples to allow use of the peak area as parameter for quantitative comparisons.

While 5th, 9th and 10th axial orders are present in both non-EMT and EMT, an observed effect of degradation in the poorly ordered collagen of EMT is reduction of FWHM and axial peak area (Table 4.7). Comparison can also be made with the earlier work on the classified 16 tissues, the d-spacing being within the same range. Special care needs to be taken to include all key parameters, indicators arising from the change in collagen fibril packing (axial peak d-spacing) and the amount of collagen in the tissue (integrated intensities and area), correlating with morphological features and cell types of the tissues. For instance, while EMT-17, EMT-18 and EMT-31 produce typical scattering spectra for non-lesional tissue, the axial area of EMT-17 for all d-spacing orders are much less than non-lesional and non-EMT tissues; additionally, the FWHM of all d-spacing orders for EMT-18 are much less than non-lesional tissue while EMT-31 shows greater

University of Malaya

d-spacing than non-lesional tissue. Accordingly, EMT-17 is classified as EMT while EMT-18 and EMT-31 are classified as non-EMT. As with other studies, present studies find collagen fibril d-spacing, axial peak amplitude and the area of the lateral peaks equatorial peak features to be distinguishing parameters.

Table 4.8 summarises the peak parameters of the several tissue types investigated. As expected, the FWHM of triglyceride peak is greater in EMT tissues compared to non-lesional tissues. Of interest are triglycerides, fat being a major source of energy for growth and membrane synthesis in solid tumours (Neuschwander-Tetri, 2010; Swinnen et al., 2006), the proportions of fat and fibrous tissue present in breast being associated with enhanced cell migration and aggressiveness, characteristic of metastatic cells, suggestive of activation of de novo lipogenesis in EMT induced cells (Medes et al., 1953).

Lewis et al. (2000) (R. A. Lewis et al., 2000) acquired SAXS images by averaging over the whole tissue sample while SAXS images acquired by Fernandez et al. (Fernández et al., 2004) resulted from the selecting of several points on the tissue sample (1 - 0.25 mm intervals), each image considered a discrete observation relative to the sample pathology. The axial d-spacing of collagen in benign and non-lesional breast tissues reported by Fernandez et al (2004) (Fernández et al., 2004) are 65.3 ± 0.2 nm and 65.0 ± 0.1 nm, respectively. On the other hand, the d-spacing of collagen found by Lewis et al. (2000) (R. A. Lewis et al., 2000) for invasive carcinoma and normal tissues is within the range 64.7 nm - 65.2 nm, whereas a smaller d-spacing was obtained for benign tissues between 64.4 and 64.7 nm. The present study found the d-spacing in invasive EMT, at 64.5 ± 0.2 nm, to be larger than in histopathologically invasive non-EMT at 63.8 ± 0.1 nm and non-lesional d = 63.6 ± 0.1 nm tissues, as illustrated in Figure 4.27 (the error bars in the graph represent the standard deviation). It should be noted that the formalin fixation samples of present investigation are found to change and reduce the absolute d-spacing

University of Malaya

of the collagen fibril in comparison to previous research, although it does not alter the relative differences between non-lesional and malignant tissues, being of great value since the transport and restoration of samples is greatly facilitated by formalin fixation. The larger standard deviation of the d-value in invaded regions (EMT tissues) can be explained by the weak Bragg reflections and higher background.

University of Malaya

Table 4.6: SAXS peak parameters for the 51 blinded individual breast tissue cases, each in FFPE (dewaxed) condition. Pathology reporting was performed after conduct of the experiment.

NON-LESIONAL BREAST TISSUE						
Research Number	**Order range**	**FWHM Order range (nm^{-1})**	**Axial peak area (a.u.) range**	**Axial d- spacing (nm) average**	**FWHM Triglyceride (nm^{-1})**	**Axial peak area Triglyceride (a.u.)**
10A	3,5,9,10	0.020 - 0.014	5.20×10^{-7} - 4.64×10^{-9}	63.5	0.05	6.69×10^{-6}
10B	3,5,9,10	0.021 - 0.030	6.81×10^{-7} - 3.73×10^{-8}	62.5	0.08	8.63×10^{-7}
12A	3,5,9,10	0.015 - 0.044	4.93×10^{-8} - 1.23×10^{-8}	63.8	0.06	1.03×10^{-6}
12B	3,5,9,10	0.018 - 0.027	1.69×10^{-7} - 7.35×10^{-8}	63.6	0.04	1.96×10^{-6}
14A	3,5,9,10	0.012 - 0.004	1.18×10^{-9} - 4.36×10^{-11}	62.8	0.03	5.30×10^{-6}
14B	5,9,10	0.021 - 0.013	7.69×10^{-9} - 2.10×10^{-9}	64.7	0.06	6.34×10^{-7}
16A	5,9,10	0.026 - 0.017	2.77×10^{-8} - 2.88×10^{-9}	63.5	0.04	2.68×10^{-7}
16B	5,9,10	0.022 - 0.012	5.76×10^{-9} - 7.65×10^{-10}	64.8	0.04	2.95×10^{-7}
24A	3,5,9,10	0.021 - 0.028	4.08×10^{-7} - 1.71×10^{-8}	63.3	0.05	3.18×10^{-8}
24B	5,9,10	0.024 - 0.027	5.61×10^{-8} - 8.86×10^{-9}	63.7	0.11	8.53×10^{-7}
26A	5,9,10	0.025 - 0.006	1.38×10^{-8} - 2.68×10^{-10}	64.4	0.03	3.74×10^{-8}
26B	5,9,10	0.019 - 0.016	3.89×10^{-8} - 2.24×10^{-9}	63.8	0.05	7.26×10^{-6}
28A	5,9,10	0.022 - 0.027	6.82×10^{-8} - 7.97×10^{-9}	63.7	0.06	7.88×10^{-7}
28B	5,9,10	0.020 - 0.026	3.44×10^{-8} - 5.07×10^{-9}	63.8	0.05	1.78×10^{-6}
30A	5,9,10	0.023 - 0.023	2.87×10^{-8} - 5.08×10^{-9}	64.2	0.06	1.20×10^{-6}
30B	3,5,9,10	0.020 - 0.026	2.94×10^{-7} - 1.60×10^{-8}	63.3	0.06	3.73×10^{-7}
32A	3,5,9,10	0.021 - 0.030	2.56×10^{-7} - 9.35×10^{-9}	63.0	0.04	2.34×10^{-6}
32B	5,9,10	0.023 - 0.012	2.46×10^{-8} - 7.69×10^{-10}	63.4	0.05	1.26×10^{-6}
34A	3,5,9,10	0.007 - 0.025	1.10×10^{-7} - 2.54×10^{-10}	63.6	0.02	1.75×10^{-9}
34B	5,9,10	0.014 - 0.029	1.51×10^{-8} - 1.09×10^{-9}	63.9	0.06	1.08×10^{-7}

University of Malaya

Table 4.6: continued

NON-EMT BREAST CANCER TISSUE						
Research Number	**Order range**	**FWHM Order range (nm^{-1})**	**Axial peak area (a.u.) range**	**Axial d- spacing (nm) average**	**FWHM Triglyceride (nm^{-1})**	**Axial peak area Triglyceride (a.u.)**
18A	3,5,9,10	0.014 - 0.006	7.31×10^{-8} - 4.35×10^{-10}	63.8	0.08	3.63×10^{-7}
18B	3,5,9,10	0.017 - 0.015	2.43×10^{-7} - 2.01×10^{-9}	63.6	0.04	4.99×10^{-7}
23A	3,5,9	0.014 - 0.029	1.13×10^{-7} - 4.05×10^{-8}	63.9	0.04	5.00×10^{-8}
23B	5,9,10	0.007 - 0.017	1.64×10^{-9} - 1.78×10^{-10}	63.4	0.03	1.38×10^{-8}
25A	5,9,10	0.011 - 0.015	1.23×10^{-9} - 9.32×10^{-10}	64.4	0.04	1.15×10^{-8}
25B	5,9,10	0.009 - 0.014	1.99×10^{-9} - 4.03×10^{-10}	63.8	0.05	3.73×10^{-7}
27A	9,10	0.023 - 0.014	8.40×10^{-9} - 2.86×10^{-9}	63.8	0.06	6.51×10^{-7}
27B	9,10	0.009 - 0.015	1.16×10^{-9} - 2.58×10^{-9}	64.5	0.08	2.63×10^{-7}
29A	5,9,10	0.016 - 0.029	2.73×10^{-8} - 9.25×10^{-9}	64.0	0.03	8.16×10^{-7}
29B	5,9,10	0.023 - 0.011	1.12×10^{-7} - 2.08×10^{-9}	63.3	0.04	4.96×10^{-6}
31A	3,5,9,10	0.020 - 0.024	3.82×10^{-7}- 7.40×10^{-9}	63.2	0.05	5.74×10^{-7}
31B	No peaks					
33A	9,10	0.010 - 0.010	1.22×10^{-9} - 1.12×10^{-9}	64.3	0.04	1.95×10^{-6}
33B	5,9,10	0.019 - 0.006	2.78×10^{-8} - 2.51×10^{-10}	64.3	0.05	1.67×10^{-7}
35A	5,9	0.020 - 0.018	6.75×10^{-8} - 1.08×10^{-8}	63.5	0.04	3.20×10^{-7}
35B	5,9	0.023 - 0.018	1.20×10^{-7}- 9.06×10^{-9}	64.1	0.06	7.18×10^{-7}
EMT BREAST CANCER TISSUE						
Research Number	**Order range**	**FWHM Order range (nm^{-1})**	**Axial peak area (a.u.) range**	**Axial d- spacing (nm) average**	**FWHM Triglyceride (nm^{-1})**	**Axial peak area Triglyceride (a.u.)**
9A	9,10	0.012- 0.019	2.61×10^{-9} - 2.18×10^{-9}	65.2	0.08	1.02×10^{-6}
9B	9,10	0.009 - 0.005	6.31×10^{-10} - 1.34×10^{-10}	64.7	0.09	2.45×10^{-7}

University of Malaya

Table 4.6: continued

EMT BREAST CANCER TISSUE						
Research Number	**Order range**	**FWHM Order range (nm^{-1})**	**Axial peak area (a.u.) range**	**Axial d- spacing (nm) average**	**FWHM Triglyceride (nm^{-1})**	**Axial peak area Triglyceride (a.u.)**
13A	9,10	0.01 - 0.021	1.31×10^{-9} - 2.01×10^{-9}	65.9	0.03	1.27×10^{-6}
13B	5,9	0.012 - 0.013	8.09×10^{-9} - 2.75×10^{-9}	64.4	0.04	3.47×10^{-6}
15A	9,10	0.004 - 0.005	1.12×10^{-10} - 1.78×10^{-10}	65.7	0.04	1.64×10^{-7}
15B	9,10	0.01 - 0.007	1.79×10^{-9} - 1.74×10^{-9}	65.0	0.04	3.43×10^{-6}
17A	5,9,10	0.02 - 0.013	2.22×10^{-8} - 1.02×10^{-9}	63.7	0.05	1.75×10^{-6}
17B	5,9,10	0.019 – 0.017	1.82×10^{-8} - 3.2×10^{-9}	63.0	0.09	1.83×10^{-6}
19A	9,10	0.013 - 0.008	1.71×10^{-9} - 5.67×10^{-10}	64.7	0.04	4.42×10^{-8}
19B	9,10	0.012 - 0.004	2.02×10^{-10} - 6.08×10^{-10}	64.7	0.06	1.48×10^{-7}
20A	9,10	0.006 - 0.006	5.21×10^{-10} - 3.58×10^{-10}	65.6	0.06	1.74×10^{-6}
20B	9,10	0.007 - 0.003	5.31×10^{-10} - 1.41×10^{-10}	65.5	0.04	1.63×10^{-6}
21A	5,9	0.018 - 0.017	7.96×10^{-9} - 4.68×10^{-9}	63.6	0.07	3.97×10^{-7}
21B	5,9,10	0.0272 - 0.004	2.45×10^{-8} - 3.01×10^{-10}	64.2	0.07	1.17×10^{-6}
22A	5,9,10	0.016 - 0.002	8.14×10^{-9} - 1.23×10^{-10}	63.7	0.09	4.98×10^{-7}
22B	5,9,10	0.014 - 0.007	9.10×10^{-9} - 4.08×10^{-10}	64.5	0.07	2.13×10^{-6}

University of Malaya

Table 4.7: Summary peak parameters for the 51 blinded breast tissues, each in FFPE (dewaxed) condition.

SAXS Parameter	Group of Tissue	Mean ± sd
Axial d-spacing (nm)	**Non-lesional**	63.6 ± 1.0
	Non-EMT	62.3 ± 9.9
	EMT	64.5 ± 1.5
Axial Peak Area (a.u.)	**Non-lesional**	$(6.8 \pm 12.5) \times 10^{-8}$
	Non-EMT	$(4.3 \pm 7.7) \times 10^{-8}$
	EMT	$(0.5 \pm 0.61) \times 10^{-8}$
Third-order FWHM (nm^{-1})	**Non-lesional**	$(19.6 \pm 2.3) \times 10^{-3}$
	Non-EMT	$(17.2 \pm 2.9) \times 10^{-3}$
	EMT	No peak
Fifth-order FWHM (nm^{-1})	**Non-lesional**	$(22.3 \pm 3.4) \times 10^{-3}$
	Non-EMT	$(18.5 \pm 4.9) \times 10^{-3}$
	EMT	$(16.7 \pm 3.4) \times 10^{-3}$
Ninth- order FWHM (nm^{-1})	**Non-lesional**	$(25.1 \pm 8.3) \times 10^{-3}$
	Non-EMT	$(18.1 \pm 7.4) \times 10^{-3}$
	EMT	$(13.6 \pm 8.2) \times 10^{-3}$
Tenth- order FWHM (nm^{-1})	**Non-lesional**	$(20.2 \pm 10.4) \times 10^{-3}$
	Non-EMT	$(13.7 \pm 7.0) \times 10^{-3}$
	EMT	$(8.7 \pm 6.1) \times 10^{-3}$
Triglyceride FWHM (nm^{-1})	**Non-lesional**	$(51 \pm 19.4) \times 10^{-3}$
	Non-EMT	$(49 \pm 15.5) \times 10^{-3}$
	EMT	$(60 \pm 20.7) \times 10^{-3}$
Triglyceride Axial Peak Area (a.u.)	**Non-lesional**	$(1.7 \pm 2.17) \times 10^{-6}$
	Non-EMT	$(0.8 \pm 1.25) \times 10^{-6}$
	EMT	$(1.3 \pm 1.08) \times 10^{-6}$

University of Malaya

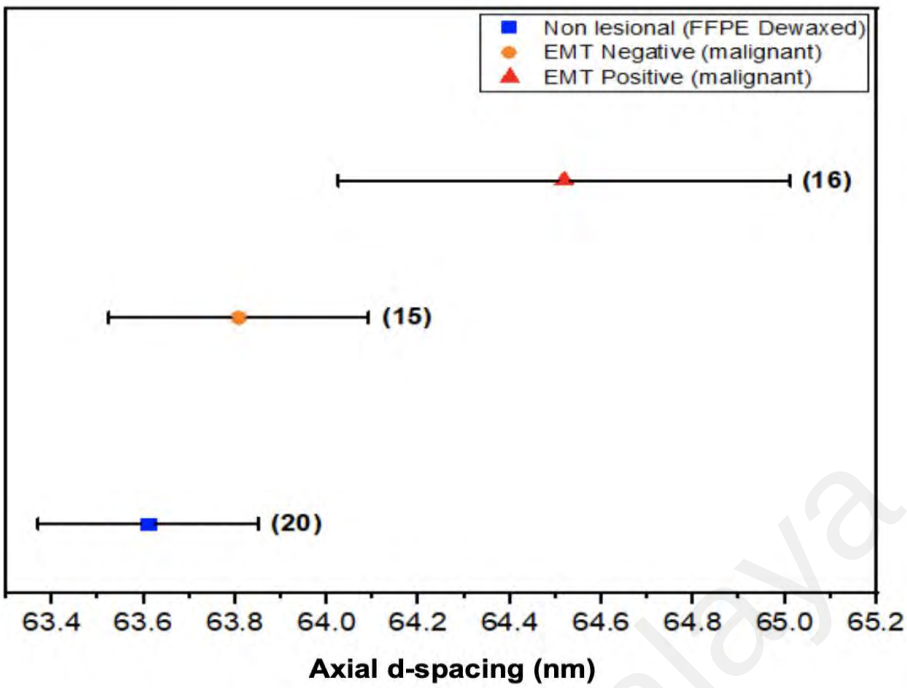


Figure 4.26: Mean axial d-spacing (nm) for the 51 blinded groups of tissues; non-lesional and malignant; non-EMT (labelled as EMT negative) and EMT (labelled as EMT positive). The number of samples are shown in the brackets. Note: error bars represent standard deviation. Note: the data points are offset vertically for clarity.

4.2.3 Statistical Analysis

Kruskal-Wallis test has been performed to examine if the four types of tissues; non-lesional, benign, EMT and non-EMT differed significantly by considering d-spacing, axial area, FWHM, FWHM triglyceride and axial area triglyceride (see Table 4.8). It has been found that differences exist between the four tissues for d-spacing, axial area and FWHM. Dunn's post hoc test was subsequently performed to identify the pairs of tissues which differed significantly for all five measurements considered (see Table 4.9). The p-values for axial area triglyceride and FWHM triglyceride showed insignificant differences for all four tissues. The EMT – non-lesional pair showed significant differences for d-spacing, axial area and FWHM while non-EMT – EMT appeared to be significantly different for axial area.

Table 4.8: Kruskal-Wallis test for differences in non-lesional, benign, EMT and non-EMT tissues according to measurements. Note: (**) highly significant, (*) significant, (##) provisional significance.

Measurements	p-value
d-spacing	0.014 **(**)**
Axial peak area	0.007 **(**)**
FWHM	0.002 **(**)**
Axial area triglyceride	0.049 **(*)**
FWHM triglyceride	0.065 **(##)**

Table 4.9: Dunn's post hoc test to assess the significant differences in measurements between non-lesional, benign, EMT and non-EMT tissues. Note: (**) highly significant, (*) significant, (##) provisional significance.

Samples	p-value				
	d-spacing	**Axial peak area**	**FWHM**	**Axial area triglyceride**	**FWHM triglyceride**
Benign – non-EMT	0.752	0.828	0.483	0.152	0.138
Benign – EMT	0.475	0.134	0.345	0.057**(##)**	0.053**(##)**
Benign – non-lesional	0.461	0.940	0.530	0.085	0.127
Non-EMT – EMT	0.144	0.043**(*)**	0.383	0.198	0.247
EMT – non lesional	0.752	0.600	0.028**(**)**	0.407	0.724
EMT– non-lesional	0.008**(**)**	0.005**(**)**	0.001(**)	0.450	0.338

Note: **(**)** highly significant, **(*)** significant, **(##)** provisional significance

University of Malaya

In present work the healthy and unhealthy tissues from the same patient were considered as independent. The Cook's distance was calculated, and two outliers were detected, belonging to patients 10 and 30 – both with non-lesional tissues in the FFPE (dewaxed) condition.

General trends in the data from measurements on the tissues were studied via principal component analysis (PCA) (see Figure 4.28) and the methods used, as described below:

1) Non-lesional - FFPE (dewaxed) (N = 9); Formalin fixed (N = 2)
2) Benign - Formalin fixed (N = 2)
3) Non-EMT - FFPE (dewaxed) (N = 8); Formalin fixed (N=2)
4) EMT - FFPE (dewaxed) (N = 9); Formalin fixed (N = 1)

Figure 4.28 indicates the presence of two clusters along the PC1 axis, one in the direction of negative PC1 (EMT) and another in the opposite direction (non-lesional). Two observations with non-lesional tissues belonging to patients 14 and 26 are seen in the EMT cluster. The PCA plot explains 60.2% of the total variation where 35.0% of the total variation is described by axis 1 while 25.2% is described by axis 2. An extreme observation belonging to patient 31 (non-EMT) whose D-spaceB measurement failed to show peak readings is seen in the PCA plot. Patient 24 (non-lesional) also appeared to be an outlier due to high readings of axial peak area and FWHM. These two observations were retained in the data as they did not affect the results of the Kruskal-Wallis test and the Dunn post hoc test.

University of Malaya

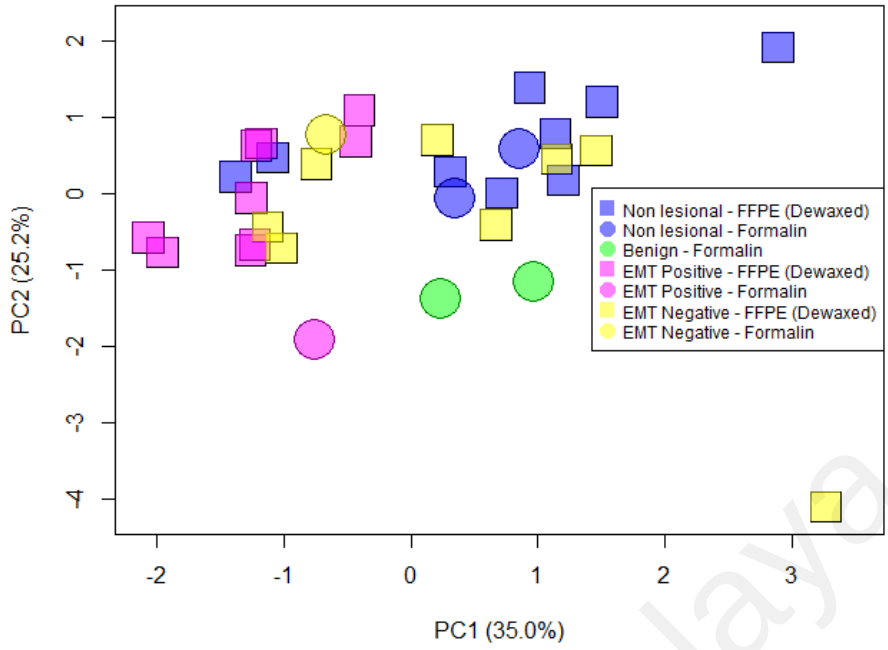


Figure 4.27: PCA plot of all studied samples; non-lesional, benign, EMT (labelled as EMT positive) and non-EMT (labelled as EMT negative) tissues.

University of Malaya

CHAPTER 5: CONCLUSION

5.1 Raman Spectroscopy

Inspection of biological tissue via Raman spectroscopy can yield a wealth of information regarding molecular bonds, potentially leading to the discovery and quantification of new intrinsic biomarkers associated with tissue classification. Particular instances may distinguish between 'diseased' and 'normal' tissue or between different stages of disease progression. In a laboratory setting, we demonstrated the Raman spectral analysis applied in unison with a combination of multivariate statistical techniques (PCA, ICA and NNLS) to discriminate cancer phenotypes, in particular EMT and non-EMT of breast tissue embedded in paraffin wax. The tissue samples have been investigated in dewaxed and waxed conditions. It was found possible to extract information concerning the relative quantities of various profiles of lipid, protein and nucleic acid composition of the breast tissues investigated herein. PCA indicated that the tissues of the EMT phenotype are potentially distinguishable, producing a spectral signature of the proclivity to mesenchymal transition, related to highly aggressive and metastatic spread. The ANOVA test followed by the Tukey's multiple comparison test showed that significant differences exist among non-lesional, EMT and non-EMT samples based on Raman intensity. Human breast tissues were used, to our knowledge being the first study tracking the molecular content of cancer cells in human tissue during the EMT process. The findings are summarised as follows:

1. For dewaxed samples, we found that CH lipid bands in non-lesional tissues are 42% and 21% greater in intensity in comparison to that EMT and non-EMT tissues, respectively. For the signal of waxed tissue samples, the CH group of lipid is 37% greater in non-EMT tissues compared to that of EMT tissues.

University of Malaya

2. Phospholipid bands greater in intensity by 60% and 110% were found in EMT tissue compared to non-lesional and non-EMT respectively. For the waxed samples, the phospholipid was found in both EMT and non-EMT tissues with slightly higher intensity in EMT tissues by 2% compared to that non-EMT tissues. Accordingly, the increase in lipid is associated with the process of EMT in breast cancer.

3. The greatest intensity of lipid C-C stretching was found in the EMT tissues; 109% higher compared to that non-EMT tissues and 37% higher compared to that of non-lesional tissue in the dewaxed samples. In the waxed tissues, C-C stretching of lipid was found to be greater in intensity by 21% in EMT tissues compared to that in non-EMT tissues.

4. In the dewaxed samples, the intensity of CH group protein was found to be 18% and 49% greater in EMT tissues compared to that of non-lesional and non-EMT tissues, respectively. In the waxed samples, the intensity of CH group protein was found to be 28% greater in EMT tissues compared to non-EMT tissues.

5. In the dewaxed samples, the intensity of Amide III in EMT tissues has been found to be 41% greater compared to non-lesional tissues and 18% greater in non-EMT tissues. In the waxed samples, the intensity of Amide III has been found to be 102% greater in the EMT tissues compared to the non-EMT tissues

6. Attenuated levels of O-P-O stretch DNA have been observed in EMT breast tissues, reflecting the changing configurations and components of nucleic acids during the EMT processes in cancer progression. For the dewaxed samples, the intensity of O-P-O stretch DNA is found to be 19% greater in non-lesional tissues compared to EMT breast cancer tissues and 48% higher in non-EMT tissues compared to that EMT tissues.

University of Malaya

In the case of waxed samples, the O-P-O stretch DNA of non-EMT tissues has been observed to be 73% higher compared to the EMT tissues.

7. Raman spectroscopy, supported by multivariate analysis (PCA, ICA and NNLS), is found to be a reliable instrument in assessing metabolic changes in EMT breast cancer tissues.

8. In terms of intensity, the results from ANOVA affirm that dewaxed and waxed samples differ significantly.

Present results point to a particularly important aspect concerning the utility of the Raman techniques that have been applied herein, namely the possibility of undertaking evaluation of the enormous archival tissue material available in clinical diagnostic laboratories. In this, one can see potential utility in further assisting in stratification of cells in terms of the metastasis initiation process phenotype.

University of Malaya

5.2 SAXS

It is shown that the SAXS technique is able to produce rapid high-quality interpretable diffraction (structural) data from breast excision specimens, employed herein to seek differentiation between normal, benign and malignant breast tissues, the latter consisting of non-EMT and EMT cancers. From the Kruskal-Wallis and Dunn post hoc tests, we have found that the difference between the d-spacing of the EMT lesion and non-lesional tissue is highly significant ($p < 0.008$). The respective d-spacings for the various tissues are: for EMT lesion, in the range 64.5 - 64.7 nm; for non-EMT lesions, FFPE (dewaxed) and formalin fixed, 63.8 nm and 64.3 nm, respectively; for benign lesions 64.0 nm; for non-lesional tissues FFPE (dewaxed) 63.7 nm and; for non-lesional tissue (formalin-fixed) in the range 63.4 - 63.6 nm. For EMT breast tissue, the disturbed collagen structure is most notable in the 9th and 10th order. Thus said, present study has found that axial d-spacing on its own does not offer the sufficiency needed in determining early breast disease. Instead, other features, such as, the amplitude of the third, fifth, nine and tenth-order axial Bragg peaks are needed, together with the magnitude of the integrated intensity and the full-width at half-maximum of the fat peak, offering significant differences between tissue types. As such, the presence of disease is thought to be best represented by a combination of factors, rather than any single specific trait. Of particular note in regard to formalin fixation of samples is that no alteration is observed to occur in the relative differences in collagen d-spacing between non-lesional and malignant tissues. This is a matter of great importance given that preserved sample and also retrospective study of samples is greatly facilitated by formalin fixation. Present findings indicate that molecular structure characteristics of breast tissue, obtained at the SAXS

University of Malaya

(supramolecular) level, can be used as markers of disease progression, SAXS analysis offering potential as a viable diagnostic procedure in support of breast clinics.

5.3 Future works

In this present work we demonstrated the capability of Raman spectroscopy and SAXS to study the metabolic changes in EMT breast cancer tissues. The content of lipid, protein, nucleic acids and collagen have been observed and it showed the differences between the non-lesional, EMT and non-EMT breast tissues.

Overall, we achieved all the objectives stated in Section 1.2. Based on the Raman results obtained in this work, the Raman spectroscopy is capable to characterise and distinguish between the non-lesional and malignant breast tissue of EMT on the basis of their lipid, protein and nucleic acid content. The SAXS analysis also showed the capability of SAXS to determine the axial spacing, intensity of scattering patterns and the scattering profiles which reflects the internal collagenous structure of the breast cancer tissues. Based on the present work, Raman spectroscopy and SAXS proved to be useful techniques to monitor the the molecular changes (e.g collagen, lipid, protein and nucleic acids) related to the EMT process in the cancer progression.

Despite current technologies, breast cancer remains one of the most common causes of death among women throughout the world. Histopathology and biomedical imaging are the current techniques that have been used as common practice for breast cancer diagnosis. However, these techniques reflect some limitation and drawbacks such as time consuming, expensive, poor resolution and leads to vague results prone to human interpretations. The demand for highly sensitive, accurate and rapid techniques urges the researchers to develop Raman spectroscopic and SAXS techniques in breast cancer

University of Malaya

diagnosis. Both Raman spectroscopy and SAXS have emerged as a diagnostic tool in breast cancer studies due to its rapid, and high specificity and sensitivity.

Many studies have been conducted to provide the evidence of EMT in breast cancer progression and metastasis in vitro (cell lines) and in vivo (mouse models). It needs to be stressed that the EMT process relates to a relatively brief period of existence, occurring at the onset of malignancy. In the absence of soft tissue mass and associated density changes, detecting early stage changes via mammographic imaging is a particular challenge, evidence of spiculations, calcifications and disturbed breast architecture being diagnostically equivocal. Further to be acknowledged are the several primary challenges in histopathological evaluation of EMT in human cancer tissue. Difficulties arise in real time tracking of individual cancer cells that have transitioned or are in transition. One issue concerns the staging of EMT, whether it has occurred, whether it has yet to undergo transition or whether the reverse process has occurred, via the mesenchymal-to-epithelial transition (MET) (Foroni et al., 2012; Iwatsuki et al., 2010; Mallini et al., 2014; Thiery et al., 2009). Nevertheless, the prototypical spindle-shaped morphology of the fibroblasts can be used to advantage to differentiate between epithelial cancer cells and fibroblasts. Thus said, epithelial cells that have undergone transition are poorly distinguishable from fibroblasts.

More work is still required before these methods can be used in clinical practice, requiring comprehensive research in translating Raman and SAXS data as the current technique of spectroscopic needs to employ several multivariate statistical tools. Therefore, it is crucial to develop a convenient methodology to standardise protocol for data acquisition and data analysis into the clinical setting being easily interpretable and producing rapid and accurate result.

University of Malaya

The above needs can be further augmented by expansion of the underpinning studies that have been undertaken to date, including real-time early stage cancer diagnosis by using optical fibres, coupled with a biopsy needle and incorporated into Raman spectrometer. A number of studies have already demonstrated the effectiveness of so called 'optical needle biopsies' (Frank et al., 1995; Haka et al., 2006; Pogue et al., 2004). In vivo Raman spectroscopic diagnosis capable to provide rapid diagnosis, reduce the number of biopsies and high-resolution information of the biomarkers responsible to the cancer progression. Hence, this *in vivo* method has the potential to be developed and fabricated for the clinical practice.

As with the structural component analysis, future study can be carried out to determine at what spacing the SAXS is sensitive to the presence of cancer and then raster scan point by point over the tissue section. All tissue is composed of collagen that provides structure to the tissue exactly analogous to the scaffolding on a building. Even without scaffolding, all buildings are built around a structure. In tissue, when there is infiltrating cancer then it pushes through the structure, altering the spacing, in other words causing structural alterations.

Verifying biomarkers related to the EMT process are expected to be highly important in targeting EMT-associated breast cancer progression, also in developing treatment effective in inhibiting metastasis. We do understand the difficulties to determine the existence of EMT in breast cancer tissues due to the EMT process relates to a relatively brief period of existence, by definition occurring at the onset of malignancy. This is the main reason of the limited numbers of samples investigated in this research. However, more samples are needed in future works to study the character of EMT in breast cancer tissue comprehensively.

University of Malaya

REFERENCES

Abramczyk, H., Placek, I., Brozek-Płuska, B., Kurczewski, K., Morawiec, Z., & Tazbir, M. (2008). Human breast tissue cancer diagnosis by Raman spectroscopy. *Spectroscopy*, *22*(2–3), 113–121.

Abramczyk, H., Surmacki, J., Brozek-Płuska, B., Morawiec, Z., & Tazbir, M. (2009). The hallmarks of breast cancer by Raman spectroscopy. *Journal of Molecular Structure*, *924–926*(C), 175–182.

Abramczyk, Halina, Brozek-Pluska, B., Surmacki, J., Jablonska, J., & Kordek, R. (2011a). The label-free Raman imaging of human breast cancer. *Journal of Molecular Liquids*, *164*(1–2), 123–131.

Abramczyk, Halina, Surmacki, J., Kopeć, M., Olejnik, A. K., Lubecka-Pietruszewska, K., & Fabianowska-Majewska, K. (2015). The role of lipid droplets and adipocytes in cancer. Raman imaging of cell cultures: MCF10A, MCF7, and MDA-MB-231 compared to adipocytes in cancerous human breast tissue. *Analyst*, *140*(7), 2224–2235.

Ackland, M. L., Newgreen, D. F., Fridman, M., Waltham, M. C., Arvanitis, A., Minichiello, J., … Thompson, E. W. (2003). Epidermal growth factor-induced epithelio-mesenchymal transition in human breast carcinoma cells. *Laboratory Investigation*, *83*(3), 435–448.

Aigner, K., Dampier, B., Descovich, L., Mikula, M., Sultan, A., Schreiber, M., … Schweifer, N. (2007). The transcription factor ZEB1 (δEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity. *Oncogene*, *26*(49), 6979–6988.

Aktas, B., Tewes, M., Fehm, T., Hauch, S., Kimmig, R., & Kasimir-Bauer, S. (2009). Stem cell and epithelial-mesenchymal transition markers are frequently overexpressed in circulating tumor cells of metastatic breast cancer patients. *Breast Cancer Research*, *11*(4), 1–9.

Ali, S. M., Bonnier, F., Tfayli, A., Lambkin, H., Flynn, K., McDonagh, V., … Byrne, H. J. (2012). Raman spectroscopic analysis of human skin tissue sections ex-vivo : evaluation of the effects of tissue processing and dewaxing. *Journal of Biomedical Optics*, *18*(6), 1–12.

Anwar, S., Firdous, S., Rehman, A., & Nawaz, M. (2015). Optical diagnostic of breast cancer using Raman, polarimetric and fluorescence spectroscopy. *Laser Physics Letters*, *12*(4), 1–6.

Arima, K., Lau, M. C., Zhao, M., Haruki, K., Kosumi, K., Mima, K., … Ogino, S. (2020). Metabolic profiling of formalin-fixed paraffin-embedded tissues discriminates normal colon from colorectal cancer. *Molecular Cancer Research*, *18*(6), 883–890.

University of Malaya

Arnedos, M., Nerurkar, A., Osin, P., A'Hern, R., Smith, I. E., & Dowsett, M. (2009). Discordance between core needle biopsy (CNB) and excisional biopsy (EB) for estrogen receptor (ER), progesterone receptor (PgR) and HER2 status in early breast cancer (EBC). *Annals of Oncology*, *20*(12), 1948–1952.

Bassett, L. W., Liu, T. H., Giuliano, A. E., & Gold, R. H. (1991). The prevalence of carcinoma in palpable vs impalpable, mammographically detected lesions. *American Journal of Roentgenology*, *157*(1), 21–24.

Benevides, J. M., Overman, S. A., & Thomas, G. J. (2005). Raman, polarized Raman and ultraviolet resonance Raman spectroscopy of nucleic acids and their complexes. *Journal of Raman Spectroscopy*, *36*(4), 279–299.

Bhushan, B., & Pradhan, A. (2013). Detection of breast & cervical cancer using Raman. *International Journal of Innovative Research in Science, Engineering and Technology*, *2*(8), 3420–3425.

Bino, R. J., Hall, R. D., Fiehn, O., Kopka, J., Saito, K., Draper, J., … Sumner, L. W. (2004). Potential of metabolomics as a functional genomics tool. *Trends in Plant Science*, *9*(9), 418–425.

Bitar, R. A., Martinho, S., Paulo, S., & Tierra-criollo, C. J. (2006). Biochemical analysis of human breast tissues using Fourier-transform Raman spectroscopy. *Journal of Biomedical Optics*, *11*(5), 1–8.

Boros, L. G., Collins, T. Q., & Somlyai, G. (2017). What to eat or what not to eat - That is still the question. *Neuro-Oncology*, *19*(4), 595–596.

Boros, L. G., & Somlyai, G. (2014). NADPH Synthesis, Intramolecular Deuterium Disequilibrium and Water Pools of Mammalian Cells. *Molecular Cell*, *55*(2), 253–263.

Breast cancer. (2017). Retrieved September 7, 2017, from World Health Organization website: http://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/

Brozek-Pluska, B., Musial, J., Kordek, R., Bailo, E., Dieing, T., & Abramczyk, H. (2012b). Raman spectroscopy and imaging: Applications in human breast cancer diagnosis. *Analyst*, *137*(16), 3773–3780.

Butler, H. J., Ashton, L., Bird, B., Cinque, G., Curtis, K., Dorney, J., … Martin, F. L. (2016). Using Raman spectroscopy to characterize biological materials. *Nature Protocols*, *11*(4), 664–687.

Campbell, K. (2018). Contribution of epithelial-mesenchymal transitions to organogenesis and cancer metastasis. *Current Opinion in Cell Biology*, *55*, 30–35.

University of Malaya

Carter, C. L., Allen, C., & Henson, D. E. (1989). Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer*, *63*(1), 181–187.

Changizi, V., Oghabian, M. A., Sarkar, S., Speller, R. D., & Kheradmand, A. A. (2005). Application of small angle X-ray scattering (SAXS) for differentiation between normal and cancerous breast tissues. *Iranian Journal of Radiation Research*, *2*(4), 205–210.

Chaturvedi, D., Balaji, S. A., Bn, V. K., Ariese, F., Umapathy, S., & Rangarajan, A. (2016b). Different phases of breast cancer cells: Raman study of immortalized, transformed, and invasive cells. *Biosensors*, *6*(4), 4–7.

Chen, Y., Eldho, N. V., Dayie, T. K., & Carey, P. R. (2010). Probing adenine rings and backbone linkages using base specific isotope-edited Raman spectroscopy: application to group II Intron Ribozyme domain V. *Biochemistry*, *49*(16), 3427–3435.

Conceião, A. L. C., Antoniassi, M., & Poletti, M. E. (2009). Analysis of breast cancer by small angle X-ray scattering (SAXS). *Analyst*, *134*(6), 1077–1082.

D'Arrigo, C., & Fentiman, I. S. (2004). Pathology of breast carcinoma. *International Journal of Clinical Practice*, *58*(1), 29–34.

Damayanti, N. P., Fang, Y., Parikh, M. R., Craig, A. P., Kirshner, J., & Irudayaraj, J. (2013). Differentiation of cancer cells in two-dimensional and three-dimensional breast cancer models by Raman spectroscopy. *Journal of Biomedical Optics*, *18*(11), Article#117008.

Damonte, P., Gregg, J. P., Borowsky, A. D., Keister, B. A., & Cardiff, R. D. (2009). EMT tumorigenesis in the mouse mammary gland. *Laboratory Investigation*, *87*(12), 1218–1226.

Depciuch, J., Kaznowska, E., Zawlik, I., Wojnarowska, R., Cholewa, M., Heraud, P., & Cebulski, J. (2016). Application of Raman spectroscopy and infrared spectroscopy in the identification of breast cancer. *Applied Spectroscopy*, *70*(2), 251–263.

Dietze, E., Sistrunk, C., Miranda-Carboni, G., O'Regan, R., & Seewaldt, V. (2015). Triple-negative breast cancer in African-American women: disparities versus biology. *Nature Reviews Cancer*, *4*, 248–254.

Dukor, R. K., Liebman, M. N., & Johnson, B. L. (1998). A new, non-destructive method for analysis of clinical samples with FT-IR microspectroscopy. Breast cancer tissue as an example. *Cellular and Molecular Biology*, *44*(1), 211–217.

Eberhardt, K., Stiebing, C., Matthaüs, C., Schmitt, M., & Popp, J. (2015). Advantages and limitations of Raman spectroscopy for molecular diagnostics: an update. *Expert Review of Molecular Diagnostics*, *15*(6), 773–787.

University of Malaya

Elston, C. W., & Ellis, O. (1991). Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*, *19*, 403–410.

Ember, K. J. I., Hoeve, M. A., McAughtrie, S. L., Bergholt, M. S., Dwyer, B. J., Stevens, M. M., … Campbell, C. J. (2017). Raman spectroscopy and regenerative medicine: a review. *Regenerative Medicine*, *2*(12), 1–10.

Eroles, P., Bosch, A., Alejandro Pérez-Fidalgo, J., & Lluch, A. (2012). Molecular biology in breast cancer: Intrinsic subtypes and signaling pathways. *Cancer Treatment Reviews*, *38*(6), 698–707.

Falzon, G., Pearson, S., Murison, R., Hall, C., Siu, K., Evans, A., … Lewis, R. (2006). Wavelet-based feature extraction applied to small-angle x-ray scattering patterns from breast tissue: A tool for differentiating between tissue types. *Physics in Medicine and Biology*, *51*(10), 2465–2477.

Faoláin, E. Ó., Hunter, M. B., Byrne, J. M., Kelehan, P., Lambkin, H. A., Byrne, H. J., & Lyng, F. M. (2005). Raman spectroscopic evaluation of efficacy of current paraffin wax section dewaxing agents. *Journal of Histochemistry and Cytochemistry*, *53*(1), 121–129.

Fedele, M., Cerchia, L., & Chiappetta, G. (2017). The epithelial-to-mesenchymal transition in breast cancer: Focus on basal-like carcinomas. *Cancers*, *9*(10), 1–19.

Felipe Lima, J., Nofech-Mozes, S., Bayani, J., & Bartlett, J. (2016). EMT in Breast Carcinoma—A Review. *Journal of Clinical Medicine*, *5*(7), Article#65.

Fernández, M., Keyriläinen, J., Karjalainen-Lindsberg, M. L., Leidenius, M., Von Smitten, K., Fiedler, S., & Suortti, P. (2004). Human breast tissue characterisation with small-angle X-ray scattering. *Spectroscopy*, *18*(2), 167–176.

Fernández, M., Keyriläinen, J., Serimaa, R., Torkkeli, M., Karjalainen-Lindsberg, M. L., Tenhunen, M., … Suortti, P. (2002). Small-angle x-ray scattering studies of human breast tissue samples. *Physics in Medicine and Biology*, *47*(4), 577–592.

Foroni, C., Broggini, M., Generali, D., & Damia, G. (2012). Epithelial-mesenchymal transition and breast cancer: Role, molecular mechanisms and clinical impact. *Cancer Treatment Reviews*, *38*(6), 689–697.

Frank, C. J., McCreery, R. L., & Redd, D. C. (1995). Raman spectroscopy of normal and diseased human breast tissues. *Analytical Chemistry*, *67*(5), 777–783.

Gautam, R., Vanga, S., Ariese, F., & Umapathy, S. (2015). Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Techniques and Instrumentation*, *2*, Article#8(2015).

Gebregiworgis, T., & Powers, R. (2012). Application of NMR metabolomics to search for human disease biomarkers. *Combinatorial Chemistry & High Throughput Screening*, *15*(8), 595–610.

University of Malaya

Geen, Shanmugam, K., Kathiresan, G., Subrahmanyam, S., & Durairaj, K. (2016). Raman fingerprints in detection of breast cancer. *Journal of Biosensors, Biomarkers and Diagnostics*, *1*(1), 1–11.

Gobinet, C., Sebiskveradze, D., Vrabie, V., Tfayli, A., Piot, O., & Manfait, M. (2008). Digital dewaxing of Raman spectral images of paraffin-embedded human skin biopsies based on ICA and NCLS. *European Signal Processing Conference*, 2008, 1–5.

Gonzales, W. V., Mobashsher, A. T., & Abbosh, A. (2019). The progress of glucose monitoring—A review of invasive to minimally and non-invasive techniques, devices and sensors. *Sensors, 19(4),* Article#800.

González-Solís, J. L., Aguiñaga-Serrano, B. I., Martínez-Espinosa, J. C., & Oceguera-Villanueva, A. (2011). Stage determination of breast cancer biopsy using raman spectroscopy and multivariate analysis. *AIP Conference Proceedings*, *1364*, 33–40.

Gros, S. J., Holland-Cunz, S. G., Supuran, C. T., & Braissant, O. (2019). Personalized treatment response assessment for rare childhood tumors using microcalorimetry–exemplified by use of carbonic anhydrase IX and aquaporin 1 inhibitors. *International Journal of Molecular Sciences*, *20*(20), 1–8.

H. Abramczyk, I. Placek, B.Bro˙zek-Płuska, K. Kurczewski, Z.Morawiec, and M. T. (2008). Human breast tissue cancer diagnosis by Raman spectroscopy. *Spectroscopy*, *22*, 113–121.

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*, *102*(35), 12371–12376.

Haka, A. S., Shafer-Peltier, K. E., Fitzmaurice, M., Crowe, J., Dasari, R. R., & Feld, M. S. (2002). Identifying microcalcifications in benign and malignant breast lesions by probing differences in their chemical composition using raman spectroscopy. *Cancer Research*, *62*(18), 5375–5380.

Haka, A. S., Volynskaya, Z., Gardecki, J. A., Nazemi, J., Lyons, J., Hicks, D., … Feld, M. S. (2006). In vivo margin assessment during partial mastectomy breast surgery using Raman spectroscopy. *Cancer Research*, *66*(6), 3317–3322.

Haka, A. S., Volynskaya, Z., Gardecki, J. A., Nazemi, J., Shenk, R., Wang, N., … Feld, M. S. (2009). Diagnosing breast cancer using Raman spectroscopy: prospective analysis. *Journal of Biomedical Optics*, *14*(5), Article#054023.

Hammond, M. E. H., Hayes, D. F., Dowsett, M., Allred, D. C., Hagerty, K. L., Badve, S., … Wolff, A. C. (2010). American society of clinical oncology/college of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *Journal of Clinical Oncology*, *28*(16), 2784–2795.

University of Malaya

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of Cancer : The Next Generation. *Cell*, *144*(5), 646–674.

Hanrahan, K., O'Neill, A., Prencipe, M., Bugler, J., Murphy, L., Fabre, A., … Watson, R. W. (2017). The role of epithelial-mesenchymal transition drivers ZEB1 and ZEB2 in mediating docetaxel-resistant prostate cancer. *Molecular Oncology*, *11*(3), 251–265.

Hao, D., Sarfaraz, M. O., Farshidfar, F., Bebb, D. G., Lee, C. Y., Card, C. M., … Weljie, A. M. (2016). Temporal characterization of serum metabolite signatures in lung cancer patients undergoing treatment. *Metabolomics*, *12*(3), 1–9.

Harner-Foreman, N., Vadakekolathu, J., Laversin, S. A., Mathieu, M. G., Reeder, S., Pockley, A. G., … Boocock, D. J. (2017). A novel spontaneous model of epithelial-mesenchymal transition (EMT) using a primary prostate cancer derived cell line demonstrating distinct stem-like characteristics. *Scientific Reports*, *7*, 1–18.

Hilvo, M., Denkert, C., Lehtinen, L., Müller, B., Brockmöller, S., Seppänen-Laakso, T., … Orešič, M. (2011). Novel theranostic opportunities offered by characterization of altered membrane lipid metabolism in breast cancer progression. *Cancer Research*, *71*(9), 3236–3245.

Hossain, M. K., Cho, H.-Y., Kim, K.-J., & Choi, J.-W. (2014a). Silver nanostar patterned substrate for label-free characterization of breast cancer cells based on surface-enhanced Raman spectroscopy. *Science of Advanced Materials*, *6*(11), 2491–2495.

Howlader, N., Noone, A., & Krapcho, M. (2014). SEER Cancer Statistics Review 1975-2011. *National Cancer Institute*. Retrieved from https://seer.cancer.gov/archive/csr/1975_2011/

Hu, C., Wang, J., Zheng, C., Xu, S., Zhang, H., Liang, Y., … Xu, W. (2013). Raman spectra exploring breast tissues : Comparison of principal component analysis and support vector machine-recursive feature. *Medical Physics*, *40*(6), 1–7.

Huang, T. C., Toraya, H., Blanton, T. N., & Wu, Y. (1993). X-ray powder diffraction analysis of silver behenate, a possible low-angle diffraction standard. *Journal of Applied Crystallography*, *26*(pt 2), 180–184.

Ibrahim, O., Maguire, A., Meade, A. D., Flint, S., Toner, M., Byrne, H. J., & Lyng, F. M. (2017). Improved protocols for pre-processing Raman spectra of formalin fixed paraffin preserved tissue sections. *Analytical Methods*, *9*(32), 4709–4717.

Ikenouchi, J., Matsuda, M., Furuse, M., & Tsukita, S. (2003). Regulation of tight junctions during the epithelium-mesenchyme transition: direct repression of the gene expression of claudins/occludin by Snail. *Journal of Cell Science*, *116*(10), 1959–1967.

Iwatsuki, M., Mimori, K., Yokobori, T., Ishi, H., Beppu, T., Nakamori, S., … Mori, M. (2010). Epithelial-mesenchymal transition in cancer development and its clinical significance. *Cancer Science*, *101*(2), 293–299.

University of Malaya

Jackson, J. E. (1991). *A User's Guide to Principal Components*. New York: John Wiley & Sons.

Jafarzadeh, N., Mani-Varnosfaderani, A., Gilany, K., Eynali, S., Ghaznavi, H., & Shakeri-Zadeh, A. (2018). The molecular cues for the biological effects of ionizing radiation dose and post-irradiation time on human breast cancer SKBR3 cell line: A Raman spectroscopy study. *Journal of Photochemistry and Photobiology B: Biology*, *180*, 1–8.

Kallergi, G., Papadaki, M. A., Politaki, E., Mavroudis, D., Georgoulias, V., & Agelaki, S. (2011). Epithelial to mesenchymal transition markers expressed in circulating tumour cells of early and metastatic breast cancer patients. *Breast Cancer Research*, *13*(3), Article#R59.

Kalluri, R., & Neilson, E. G. (2003). Epithelial-mesenchymal transition and its implications for fibrosis. *Journal of Clinical Investigation*, *112*(12), 1776–1784.

Kalluri, R., & Weinberg, R. a. (2009). The basics of epithelial-mesenchymal transition. *Journal of Clinical Investigation*, *119*(6), 1420–1428.

Kamemoto, L. E., Misra, A. K., Sharma, S. K., Goodman, M. T., Luk, H., Ava C. Dykes, & Acosta, T. (2011). Near-infrared micro-Raman spectroscopy for in vitro detection of cervical cancer. *Applied Spectroscopy*, *64*(3), 255–261.

Karellas, A., & Vedantham, S. (2009). Detectors for digital mammography. In G. Whitman & T. Haygood (Eds.), *Digital Mammography: A Practical Approach*, (pp. 1-17). Cambridge: Cambridge University Press.

Kast, R. E., Serhatkulu, G. K., Cao, A., Pandya, A. K., Dai, H., Thakur, J. S., … Rabah, R. (2007). Raman spectroscopy can differentiate malignant tumors from normal breast tissue and detect early neoplastic changes in a mouse model. *Biopolymers*, *89*(3), 235–241.

Kauppila, S., Stenbäck, F., Risteli, J., Jukkola, A., & Risteli, L. (1998). Aberrant type I and type III collagen gene expression in human breast cancer in vivo. *Journal of Pathology*, *186*(3), 262–268.

Kidane, G., Speller, R. D., Royle, G. J., & Hanby, A. M. (1999). X-ray scatter signatures for normal and neoplastic breast tissues. *Physics in Medicine and Biology*, *44*(7), 1791–1802.

Kim, J., Kong, J., Chang, H., Kim, H., & Kim, A. (2016). EGF induces epithelial-mesenchymal transition through phospho-Smad2/3-Snail signaling pathway in breast cancer cells. *Oncotarget*, *7*(51), 85021–85032.

Kokkinos, M. I., Wafai, R., Wong, M. K., Newgreen, D. F., Thompson, E. W., & Waltham, M. (2007). Vimentin and epithelial-mesenchymal transition in human breast cancer - Observations in vitro and in vivo. *Cells Tissues Organs*, *185*(1–3), 191–203.

University of Malaya

Krishna, C. M., Kurien, J., Mathew, S., Rao, L., Maheedhar, K., Kumar, K. K., & Chowdary, M. V. P. (2008). Raman spectroscopy of breast tissues. *Expert Review of Molecular Diagnostics*, *8*(2), 149–166.

Krishna, C. M., Prathima, N. B., Malini, R., Vadhiraja, B. M., Bhatt, R. A., Fernandes, D. J., … Kartha, V. B. (2006). Raman spectroscopy studies for diagnosis of cancers in human uterine cervix. *Vibrational Spectroscopy*, *41*, 136–141.

Krishna, C. M., & Rubina, S. (2015). Raman spectroscopy in cervical cancers: an update. *Journal of Cancer Research and Therapeutics*, *11*(1), 10-17.

Kumar, K. K., Chowdary, M. V. P., Mathew, S., Rao, L., Krishna, C. M., & Kurien, J. (2007). Raman spectroscopic diagnosis of breast cancers: evaluation of models. *Journal of Raman Spectroscopy*, *38*, 1538–1553.

Kwon, M. J. (2013). Emerging roles of claudins in human cancer. *International Journal of Molecular Sciences*, *14*(9), 18148–18180.

Lakhtakia, R., Aljarrah, A., Furrukh, M., & Ganguly, S. S. (2017). Epithelial mesenchymal transition (EMT) in metastatic breast cancer in Omani women. *Cancer Microenvironment*. *10*(1–3), 25–37.

Lamouille, S., Xu, J., & Derynck, R. (2014). Molecular mechanisms of epithelial-mesenchymal transition. *Nature Reviews Molecular Cell Biology*, *15*(3), 178–196.

Lasch, P. (2012). Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging. *Chemometrics and Intelligent Laboratory Systems*, *117*, 100–114.

László G Borosa, D'Agostino, D. P., Katze, H. E., Roth, J. P., Meuilletg, E. J., & Somlyai, G. (2016). Submolecular regulation of cell transformation by deuterium depleting water exchange reactions in the tricarboxylic acid substrate cycle. *Medical Hypotheses*, *87*, 69–74.

Le Bras, G. F., Taubenslag, K. J., & Andl, C. D. (2012). The regulation of cell-cell adhesion during epithelial-mesenchymal transition, motility and tumor progression. *Cell Adhesion and Migration*, *6*(4), 365–373.

Lee, M. Y., & Shen, M. R. (2012). Epithelial-mesenchymal transition in cervical carcinoma. *American Journal of Translational Research*, *4*(1), 1–13.

Lee, S., Chon, H., Lee, J., Ko, J., Chung, B. H., Lim, D. W., & Choo, J. (2014). Rapid and sensitive phenotypic marker detection on breast cancer cells using surface-enhanced Raman scattering (SERS) imaging. *Biosensors and Bioelectronics*, *51*, 238–243.

Levchenko, S. M., & Qu, J. (2018). Biomolecular component analysis of phospholipids composition in live HeLa cells. *Biosensors*, *8*(123), 1–8.

University of Malaya

Lewis, R. A., Rogers, K. D., Hall, C. J., Towns-Andrews, E., Slawson, S., Evans, A., … Dance, D. R. (2000). Breast cancer diagnosis using scattered X-rays. *Journal of Synchrotron Radiation*, *7*(5), 348–352.

Lewis, Robert A., Rogers, K. D., Hall, C. J., Towns-Andrews, E., Slawson, S., Evans, A., … Dance, D. R. (1999). Preliminary observations of breast tumor collagen using synchrotron radiation. *Medical Applications of Penetrating Radiation*, *3770*(July), 32-37.

Li, Q. B., Wang, W., Liu, C. H., & Zhang, G. J. (2015). Discrimination of breast cancer from normal tissue with Raman spectroscopy and chemometrics. *Journal of Applied Spectroscopy*, *82*(3), 450–455.

Li, Q., Hao, C., & Xu, Z. (2017). Diagnosis of breast cancer tissues using 785 nm miniature Raman spectrometer and pattern regression. *Sensors*, *17*(3), Article#627.

Lifestyle-related Breast Cancer Risk Factors. (2016). Retrieved September 8, 2017, from American Cancer Society website: https://www.cancer.org/cancer/breast-cancer/risk-and-prevention/lifestyle-related-breast-cancer-risk-factors.html

Lin, H., Liang, Y., Dou, X., Chen, C., Wei, X., Zeng, D., … Zhang, G. (2018). Notch3 inhibits epithelial – mesenchymal transition in breast cancer via a novel mechanism , upregulation of GATA-3 expression. *Oncogenesis*, *7*, Article# 59.

Liu, C.-H., Zhou, Y., Sun, Y., Li, J. Y., Zhou, L. X., Boydston-White, S., … Alfano, R. R. (2013). Resonance Raman and Raman spectroscopy for breast cancer detection. *Technology in Cancer Research & Treatment*, *12*(4), 371–382.

Liu, C. H., Das, B. B., Glassman, W. L. S., Tang, G. C., Yoo, K. M., Zhu, H. R., … Alfano, R. R. (1992). Raman, fluorescence, and time-resolved light scattering as optical diagnostic techniques to separate diseased and normal biomedical media. *Journal of Photochemistry and Photobiology, B: Biology*, *16*(2), 187–209.

Lucarelli, G., Rutigliano, M., Galleggiante, V., Giglio, A., Palazzo, S., Ferro, M., … Ditonno, P. (2015). Metabolomic profiling for the identification of novel diagnostic markers in prostate cancer. *Expert Review of Molecular Diagnostics*, *15*(9), 1211–1224.

M.flug, & P.Kopf-Maier. (1995). The basement membrane, its involvement in carcinoma cell invasion. *Acta Anat*, *152*, 69–84.

Machalekova, K., Kajo, K., & Bencat, M. (2012). Unusual occurrence of rare lipid-rich carcinoma and conventional invasive ductal carcinoma in the one breast: case report. *Case Reports in Pathology*, *2012*, 1–4.

Mahadevan-Jansen, A., & Richards-Kortum, R. (1996). Raman spectroscopy for the detection of cancers and precancers. *Journal Of Biomedical Optics*, *1*(1), 31–70.

Malik, B., Terry, R., Wiskin, J., & Lenox, M. (2018). Quantitative transmission ultrasound tomography: Imaging and performance characteristics. *Medical Physics*, *45*(7), 3063–3075.

University of Malaya

Mallini, P., Lennard, T., Kirby, J., & Meeson, A. (2014). Epithelial-to-mesenchymal transition: What is the impact on breast cancer stem cells and drug resistance. *Cancer Treatment Reviews*, *40*(3), 341–348.

Marro, M., Nieva, C., Sanz-pamplona, R., & Sierra, A. (2014). Molecular monitoring of epithelial-to-mesenchymal transition in breast cancer cells by means of Raman spectroscopy. *Biochimica et Biophysica Acta*, *1843*, 1785–1795.

McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., … Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, *577*(7788), 89–94.

Mego, M., Mani, S. A., Lee, B., Li, C., Evans, K. W., Cohen, E. N., … Reuben, J. M. (2012). Expression of epithelial-mesenchymal transition-inducing transcription factors in primary breast cancer: the effect of neoadjuvant therapy. *International Journal of Cancer*, *130*(4), 808–816.

Meksiarun, P., Ishigaki, M., Huck-Pezzei, V. A. C., Huck, C. W., Wongravee, K., Sato, H., & Ozaki, Y. (2017). Comparison of multivariate analysis methods for extracting the paraffin component from the paraffin-embedded cancer tissue spectra for Raman imaging. *Scientific Reports*, *7*, Article#44890.

Mischak, H., Coon, J. J., Novak, J., Weissinger, E. M., Schanstra, J., & Dominiczak, A. F. (2009). Capillary electrophoresis–mass spectrometry as a powerful tool in biomarker discovery and clinical diagnosis: an update of recent developments. *Mass Spectrometry Reviews*, *28*(5), 703–704.

Morin, P. J. (2005). Claudin proteins in human cancer: promising new targets for diagnosis and therapy. *Cancer Research*, *65*(21), 9603–9606.

Murai, T. (2012). The role of lipid rafts in cancer cell adhesion and migration. *International Journal of Cell Biology*, *2012,* Article# 763283.

Myers, E. R., Moorman, P., Gierisch, J. M., Havrilesky, L. J., Grimm, L. J., Ghate, S., … Sanders, G. D. (2015). Benefits and harms of breast cancer screening: a systematic review. *JAMA - Journal of the American Medical Association*, *314*(15), 1615–1634.

Naderi, A., Teschendorff, A. E., Barbosa-Morais, N. L., Pinder, S. E., Green, A. R., Powe, D. G., … Caldas, C. (2007). A gene-expression signature to predict survival in breast cancer across independent data sets. *Oncogene*, *26*(10), 1507–1516.

Nambiar, P. R., Gupta, R. R., & Misra, V. (2010). An "Omics" based survey of human colon cancer. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, *693*(1–2), 3–18.

Nicholson, J. K., & Wilson, I. D. (1989). High resolution proton magnetic resonance spectroscopy of biological fluids. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *21*(4–5), 449–501.

University of Malaya

Nieva, C., Marro, M., Santana-codina, N., Rao, S., Petrov, D., & Sierra, A. (2012). The lipid phenotype of breast cancer cells characterized by Raman microspectroscopy : towards a stratification of malignancy. *PLOS One*, *7*(10), Article# e46456.

Nyström, L., Rutqvist, L. E., Wall, S., Lindgren, A., Lindqvist, M., Rydén, S., Andersson,…Frisell, J. (1993). Breast cancer screening with mammography : overview of Swedish randomised trials. Lancet, 341(8851), 973–978.

Otto, M. (1999). *Chemometrics: Statistics and computer application in analytical chemistry* (3rd ed.). Weinheim, Germany: *Wiley-VCH*.

Pandey, N., Lanke, V., & Vinod, P. K. (2020). Network-based metabolic characterization of renal cell carcinoma. *Scientific Reports*, *10*(1), 1–13.

Pertovaara, A. (2000). Plasticity in descending pain modulatory systems. *Progress in Brain Research*, *129*(40), 231–242.

Pichardo-Molina, J. L., Frausto-Reyes, C., Barbosa-García, O., Huerta-Franco, R., González-Trujillo, J. L., Ramírez-Alvarado, C. A., … Medina-Gutiérrez, C. (2007). Raman spectroscopy and multivariate analysis of serum samples from breast cancer patients. *Lasers in Medical Science*, *22*, 229–236.

Pogue, B. W., Jiang, S., Dehghani, H., Kogel, C., Soho, S., Srinivasan, S., … Paulsen, K. D. (2004). Characterization of hemoglobin, water, and NIR scattering in breast tissue: analysis of intersubject variability and menstrual cycle changes. *Journal of Biomedical Optics*, *9*(3), 541–552.

Potcoava, M. C., Futia, G. L., Aughenbaugh, J., Schlaepfer, I. R., & Gibson, E. A. (2014a). Raman and coherent anti-Stokes Raman scattering microscopy studies of changes in lipid content and composition in hormone-treated breast and prostate cancer cells. *Journal of Biomedical Optics*, *19*(11), Article#111605.

Pucci-Minafra, I., Luparello, C., Schillaci, R., & Sciarrino, S. (1987). Ultrastructural evidence of collagenolytic activity in ductal infiltrating carcinoma of the human breast. *International Journal of Cancer*, *39*(5), 599–603.

Pucci Minafra, I., Minafra, S., Tomasino, R. M., Sciarrino, S., & Tinervia, R. (1986). Collagen changes in the ductal infiltrating (scirrhous) carcinoma of the human breast. A possible role played by type I trimer collagen on the invasive growth. *Journal of Submicroscopic Cytology and Pathology*, *18*(4), 795–805.

Qureshi, R., Arora, H., & Rizvi, M. A. (2015). EMT in cervical cancer: Its role in tumour progression and response to therapy. *Cancer Letters*, *356*(2), 321–331.

Raamsdonk, L. M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M. C., … Oliver, S. G. (2001). A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnology*, *19*(1), 45–50.

University of Malaya

Ramanujam, N., Mahadevan, A., Richards-Kortum, R., Mitchell, M. F., Thomsen, S., Malpica, A., … Wright, T. (1996). Development of a multivariate statistical algorithm to analyze human cervical tissue fluorescence spectra acquired in vivo. *Lasers in Surgery and Medicine*, *19*(1), 46–62.

Rehman, I. ur, Movasaghi, Z., & Shazza Rehman. (2013). *Vibrational Spectroscopy for Tissue Analysis*. Florida, United States: CRC Press.

Reveendran, A., Varghese, S., Senthil Kumar, V., Ranganathan, V., & Karthikeyan, T. M. (2018). Micro raman spectroscopic analysis of lobular carcinoma tissues. *Asian Journal of Pharmaceutical and Clinical Research*, *11*(9), 172–175.

Rinnan, Å. (2014). Pre-processing in vibrational spectroscopy-when, why and how. *Analytical Methods*, *6*(18), 7124–7129.

Robbins, S. L. (1974). *Pathologic basis of disease* (5th ed.). PA, USA: WB Saunders Co.

Robichaux-Viehoever, A., Kanter, E., Shappell, H., Billheimer, D., Jones, H., & Mahadevan-Jansen, A. (2007). Characterization of Raman spectra measured in vivo for the detection of cervical dysplasia. *Applied Spectroscopy*, *61*(9), 986–993.

Round, A. R., Wilkinson, S. J., Hall, C. J., Rogers, K. D., Glatter, O., Wess, T., & Ellis, I. O. (2005). A preliminary study of breast cancer diagnosis using laboratory based small angle x-ray scattering. *Physics in Medicine and Biology*, *50*(17), 4159–4168.

Rysman, E., Brusselmans, K., Scheys, K., Timmermans, L., Derua, R., Munck, S., … Swinnen, J. V. (2010). De novo lipogenesis protects cancer cells from free radicals and chemotherapeutics by promoting membrane lipid saturation. *Cancer Research*, *70*(20), 8117–8126.

Sabtu, S. N., Abdul Sani, S. F., Bradley, D. A., Looi, L. M., & Osman, Z. (2020). A review of the applications of Raman spectroscopy for breast cancer tissue diagnostic and their histopathological classification of epithelial to mesenchymal transition. *Journal of Raman Spectroscopy*, *51*(3), 380–389.

Sato, R., Semba, T., Saya, H., & Arima, Y. (2016). Concise review : stem cells and epithelial- mesenchymal transition in cancer : biological implications and therapeutic targets. *Stem Cells, 34(8),* 1997–2007.

Sharyl J. Nass, I. Craig Henderson, J. C. L. (Ed.). (2001). *Mammography and Beyond: Developing Technologies for the Early Detection of Breast Cancer.* Washington (DC): National Academy of Sciences.

Sidhu, S., Siu, K. K. W., Falzon, G., Hart, S. A., Fox, J. G., & Lewis, R. A. (2009). Mapping structural changes in breast tissue disease using x-ray scattering. *Medical Physics*, *36*(7), 3211–3217.

Sidhu, S., Siu, K. K. W., Falzon, G., Nazaretian, S., Hart, S. A., Fox, J. G., … Lewis, R. A. (2008). X-ray scattering for classifying tissue types associated with breast disease. *Medical Physics*, *35*(10), 4660–4670.

University of Malaya

Stone, N. (2001). *Raman Spectroscopy of Biological Tissue for Application in Optical Diagnosis of Malignancy*. (Doctoral dissertation, Cranfield University). Retrieved from http://hdl.handle.net/1826/4015

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*(7), 564–573.

Suhonen, H., Fernández, M., Serimaa, R., & Suortti, P. (2005). Simulation of small-angle x-ray scattering from collagen fibrils and comparison with experimental patterns. *Physics in Medicine and Biology*, *50*(22), 5401–5416.

Surmacki, J, Musial, J., Kordek, R., & Abramczyk, H. (2013). Raman imaging at biological interfaces: applications in breast cancer diagnosis. *Molecular Cancer*, *12*(48), 12.

Surmacki, Jakub, Brozek-Pluska, B., Kordek, R., & Abramczyk, H. (2015). The lipid-reactive oxygen species phenotype of breast cancer. Raman spectroscopy and mapping, PCA and PLSDA for invasive ductal carcinoma and invasive lobular carcinoma. Molecular tumorigenic mechanisms beyond Warburg effect. *Analyst*, *140*(7), 2121–2133.

Talari, A. C. S., Movasaghi, Z., Rehman, S., & Rehman, I. U. (2015). Raman spectroscopy of biological tissues. *Applied Spectroscopy Reviews*, *50*(1), 46–111.

Tanjore, H., & Kalluri, R. (2006). The role of type IV collagen and basement membranes in cancer progression and metastasis. *American Journal of Pathology*, *168*(3), 715–717.

Thiery, J. P., Acloque, H., Huang, R. Y. J., & Nieto, M. A. (2009). Epithelial-Mesenchymal Transitions in Development and Disease. *Cell*, *139*(5), 871–890.

Trimboli, A. J., Fukino, K., Bruin, A. De, Wei, G., Shen, L., Tanner, S. M., … Leone, G. (2008). Direct Evidence for Epithelial-Mesenchymal Transitions in Breast Cancer. *Cancer Research*, *68*(3), 937–945.

Tu, Q., & Chang, C. (2012). Diagnostic applications of Raman spectroscopy. *Nanomedicine: Nanotechnology, Biology, and Medicine*, *8*(5), 545–558.

Types of Breast Cancer. (2016). Retrieved September 8, 2017, from American Cancer Society website: https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/types-of-breast-cancer.html

Utzinger, U. R. S., An, D. L. H., Ahadevan-jansen, A. M., Alpica, A. M., Follen, M. I., & Richards-kortum, R. (2001). Near-Infrared Raman Spectroscopy for in Vivo Detection of Cervical Precancers. *Applied Spectroscopy*, *55*(8), 955–959.

Valladares-Ayerbes, M., Díaz-Prado, S., Reboredo, M., Medina, V., Lorenzo-Patiño, M. J., Iglesias-Díaz, P., … Antón-Aparicio, L. M. (2010). Evaluation of plakophilin-3 mRNA as a biomarker for detection of circulating tumor cells in gastrointestinal cancer patients. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of*

University of Malaya

the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology, *19*(6), 1432–1440.

Van der Velden, L. A., Schaafsma, H. E., Manni, J. J., Ruiter, D. J., Ramaekers, F. C. S., & Kuijpers, W. (1997). Cytokeratin and vimentin expression in normal epithelium and squamous cell carcinomas of the larynx. *European Archives of Oto-Rhino-Laryngology*, *254*(8), 376–383.

Vargas-Obieta, E., Martínez-Espinosa, J. C., Martínez-Zerega, B. E., Jave-Suárez, L. F., Aguilar-Lemarroy, A., & González-Solís, J. L. (2016). Breast cancer detection based on serum sample surface enhanced Raman spectroscopy. *Lasers in Medical Science*, *31*(7), 1317–1324.

Vazquez-Martin, A., Colomer, R., Brunet, J., Lupu, R., & Menendez, J. A. (2008). Overexpression of fatty acid synthase gene activates HER1/HER2 tyrosine kinase receptors in human breast epithelial cells. *Cell Proliferation*, *41*(1), 59–85.

Voulgari, A., & Pintzas, A. (2009). Epithelial-mesenchymal transition in cancer metastasis: Mechanisms, markers and strategies to overcome drug resistance in the clinic. *Biochimica et Biophysica Acta - Reviews on Cancer*, *1796*(2), 75–90.

Vrabie, V., Gobinet, C., Piot, O., Tfayli, A., Bernard, P., Huez, R., & Manfait, M. (2007). Independent component analysis of Raman spectra: application on paraffin-embedded skin biopsies. *Biomedical Signal Processing and Control*, *2*(1), 40–50.

Wang, Y., & Zhou, B. P. (2011). Epithelial-mesenchymal transition in breast cancer progression and metastasis. *Chinese Journal of Cancer*, *30*(9), 603–611.

Wang, Y., & Zhou, B. P. (2013). Epithelial-mesenchymal transition—a hallmark of breast cancer metastasis. *Cancer Hallmarks*, *1(1)*, 38–49.

Winnard Jr, P. T., Zhang, C., Vesuna, F., Kang, J. W., Garry, J., Dasari, R. R., … Raman, V. (2017). Organ-specific isogenic metastatic breast cancer cell lines exhibit distinct Raman spectral signatures and metabolomes. *Oncotarget*, *8*(12), 20266–20287.

Wirapati, P., Sotiriou, C., & Kunkel, S. (2008). Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures, *Breast Cancer Researh*, *10*(4), Article#R65.

Wolff, A. C., Hammond, M. E. H., Hicks, D. G., Dowsett, M., McShane, L. M., Allison, K. H., … Hayes, D. F. (2014). Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/college of American pathologists clinical practice guideline update. *Archives of Pathology and Laboratory Medicine*, *138*(2), 241–256.

Woo, H. M., Kim, K. M., Choi, M. H., Jung, B. H., Lee, J., Kong, G., … Chung, B. C. (2009). Mass spectrometry based metabolomic approaches in urinary biomarker study of women's cancers. *Clinica Chimica Acta*, *400*(1–2), 63–69.

University of Malaya

Wu, Yangzhe, McEwen, G. D., Harihar, S., Baker, S. M., DeWald, D. B., & Zhou, A. (2010). BRMS1 expression alters the ultrastructural, biomechanical and biochemical properties of MDA-MB-435 human breast carcinoma cells: An AFM and Raman microspectroscopy study. *Cancer Letters*, *293*(1), 82–91.

Wu, Yanyuan, Sarkissyan, M., & Vadgama, J. (2016). Epithelial-mesenchymal transition and breast cancer. *Journal of Clinical Medicine*, *5*(13).

Xue, C., Plieth, D., Venkov, C., Xu, C., & Neilson, E. G. (2003). The gatekeeper effect of epithelial-mesenchymal transition regulates the frequency of breast cancer metastasis 1. *Cancer Research*, *63*, 3386–3394.

Yan, X. L., Dong, R. X., Wang, Q. G., Chen, S. F., Zhang, Z. W., & Zhang, X. J. (2005). Raman spectra of cell from breast cancer patients. *Guang Pu Xue Yu Guang Pu Fen Xi*, *25*(1), 58–61.

Yang, J., Antin, P., Berx, G., Blanpain, C., Brabletz, T., Bronner, M., … Sheng, G. (2020). Guidelines and definitions for research on epithelial–mesenchymal transition. *Nature Reviews Molecular Cell Biology*, *21*(6), 341–352.

Yersal, O., & Barutca, S. (2014). Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World Journal of Clinical Oncology*, *5*(3), 412–424.

Yizhak, K., Benyamini, T., Liebermeister, W., Ruppin, E., & Shlomi, T. (2010). Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics*, *26*(12), 255–260.

Yu, M., Bardia, A., Wittner, B. S., Stott, S. L., Smas, M. E., Ting, D. T., … Haber, D. A. (2013). Circulating Breast Tumor Cells Exhibit Dynamic Changes in Epithelial and Mesenchymal Composition. *Science*, *339*(6119), 580–584.

Zhang, J., Wei, S., Liu, L., Nagana Gowda, G. A., Bonney, P., Stewart, J., … Raftery, D. (2012). NMR-based metabolomics study of canine bladder cancer. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, *1822*(11), 1807–1814.

Zhang, L., & Min, W. (2017). Bioorthogonal chemical imaging of metabolic changes during epithelial-mesenchymal transition of cancer cells by stimulated Raman scattering microscopy. *Journal of Biomedical Optics*, *22*(10), 1–7.

Zheng, C., Liang, L., Xu, S., Zhang, H., Hu, C., Bi, L., … Xu, W. (2014). The use of Au@SiO2 shell-isolated nanoparticle-enhanced Raman spectroscopy for human breast cancer detection. *Analytical and Bioanalytical Chemistry*, *406*(22), 5425–5432.

Zhu, J., Zhou, J., Guo, J., Cai, W., Liu, B., Wang, Z., & Sun, Z. (2013). Surface-enhanced Raman spectroscopy investigation on human breast cancer cells. *Chemistry Central Journal*, *7*(1), Article#37.

Zira, A. N., Theocharis, S. E., Mitropoulos, D., Migdalis, V., & Mikros, E. (2010). 1H NMR metabonomic analysis in renal cell carcinoma: A possible diagnostic tool. *Journal of Proteome Research*, *9*(8), 4038–4044.

Zúñiga, W. C., Jones, V., Anderson, S. M., Echevarria, A., Miller, N. L., Stashko, C., … Storrie-Lombardi, M. C. (2019). Raman spectroscopy for rapid evaluation of surgical margins during breast cancer lumpectomy. *Scientific Reports*, *9*(1), 1–16.

Zürbig, P., & Mischak, H. (2008). Capillary electrophoresis coupled to mass spectrometry for biomarker discovery and diagnosis of kidney diseases. *Contributions to Nephrology*, *160*, 107–126.

University of Malaya

LIST OF PUBLICATIONS AND PAPERS PRESENTED

1. **Sabtu, S.N.**, Abdul Sani, S.F, Looi, L.M., Chiew, S.F., Pathmanathan, D., Bradley D.A., & Osman, Z. (2021). Indication of high lipid content in epithelial-mesenchymal transitions of breast tissues. *Scientific Reports, 11(1),* Article#3250.

2. Mohd Sobri, S. N., Abdul Sani, S. F., **Sabtu, S. N**., Looi, L. M., Chiew, S. F., Pathmanathan, D., … Bradley, D. A. (2020). Structural Studies of Epithelial Mesenchymal Transition Breast Tissues. *Scientific Reports*, *10(1)*, Article#1997.

3. **Sabtu, S.N.**, Abdul Sani, S.F., Bradley D.A, Looi, L.M., & Osman, Z. (2019). A review of the applications of Raman spectroscopy for breast cancer tissue diagnostic and their histopathological classification of epithelial to mesenchymal transition. *Journal of Raman Spectroscopy*, *51(3),*380-389.

University of Malaya