COMPARATIVE PROTEOMIC, GLYCOPROTEOMIC AND PHOSPHOPROTEOMIC ANALYSES OF BREAST CANCER TISSUES

ABDULLAH SALEH SALEH ABDULLAH AL-WAJEEH

UNIVERSITI SAINS MALAYSIA

2019

COMPARATIVE PROTEOMIC, GLYCOPROTEOMIC AND PHOSPHOPROTEOMIC ANALYSES OF BREAST CANCER TISSUES

by

ABDULLAH SALEH SALEH ABDULLAH AL-WAJEEH

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

October 2019

This thesis is dedicated to …

My beloved mother and my late father,

My beloved wife,

My brothers and sisters

ACKNOWLEDGEMENT

بسم الله الرحمن الرحيم
{نَرْفَعُ دَرَجَاتٍ مَّن نَّشَآءُ وَفَوْقَ كُلِّ ذِي عِلْمٍ عَلِيمٌ} - (76- سورة يوسف)

All praise and thanks are due to ALLAH SUBHANH WA TAALA, the Lord of the world, for giving me health, strength, knowledge, and patience to complete this work. His Majesty said (And remember! your Lord caused to be declared publicly): "If you are grateful, I will add more (favours) unto you" (Ibrahim: 7).

As the Prophet MOHAMMED, Peace be Upon Him, said: "Whoever does not thank people (for their favours) has not thanked Allah (properly)." Therefore, my appreciation and sincere gratitude go to my main supervisor, Dr. Mohd Nazri Bin Ismail, for his diversified help, support, and encouragement. I wish to express my deepest thanks for his thoughtful comments and commitments, precious advice, and guidance. Furthermore, I would like to express my deep gratitude to my co-supervisor, Prof. Dr. Aishah Binti Abdul Latiff, for her support, patience, and guidance during my study. She has widened my horizons in conducting research. Her contributions were invaluable and extraordinary. Her way of directing a postgraduate student was unique. I am privileged to work under their supervision during my PhD study.

I would like to thank Dr. Salizawati Binti Muhamad Salhimi for her support and guidance. In addition, I would like to express my deep thankfulness to Dr. Imran Abdulkhleq for his extensive help and support during this study and to all members of Seberang Jaya hospital and administrative staff.

My thanks and gratitude also go to Dr. Mohammed Alsairafy, the director of Anti-Doping, Lab Qatar (ADLQ), Mrs. Samira, the deputy director, and to all ADLQ centre members, my colleagues, technicians, and administrative staff.

I would like to express my thanks and gratitude to the director of the Analytical Biochemistry Research Centre (ABrC), USM, Penang Prof. Dr. Zafarina Zainuddin, and to all ABrC members, my colleagues, technicians, and administrative staff. My acknowledgement also goes to the Institute of Postgraduate Studies and USM library for their help and support. Special thanks to all my thesis committee members and viva panel whom I always remain grateful for their guidance and valuable comments through the whole thesis process.

My sincere thanks goes to my family members who are always in my heart: my father for his endless and continuous encouragement and constant support, and my mother for her continuous prayers and inspiration. My sincere gratitude goes to my dearest brothers for their continuous support and encouragement and my sisters for always keeping a smile on my face and motivating me all the time.

Moreover, I would like to express great thanks to all colleagues and faculty members at Ibb University, Yemen, for their distinguished cooperation, support, and help during my study. Finally, my appreciation goes to my friends Dr. Majed Ahmed Al-Mansoub and Dr. Redhwan Alnakhalny for their support and encouragement during my study.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

α-FP	Alpha-fetoprotein
ACS	American cancer society
AGC	Automatic gain control
AJCC	American joint committee on cancer
ASR	Age-standardized rate
CA15.3	Cancer antigen 15.3
CA27-29	Cancer antigen 27-29
CA19-9	Cancer antigen 19-9
CAD	Collision activated dissociation
CEA	Carcinoembryonic antigen
CHCA	Hydroxycinnamic acid
CID	Collision induced dissociation
DAVID	Database for Annotation, Visualization and Integrated Discovery
DCC	Doping control centre
DCIS	Ductal carcinoma in situ
DDA	Data-dependent analysis
DHB	Dihydroxy benzoic acid
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ECD	Electron capture dissociation
EI	Electron ionisation
ESI	Electrospray ionisation
ESI MS	Electrospray ionisation mass spectrometry
ESI-QTOF	Electrospray-quadruple-time-of-light
ETD	Electron transfer dissociation
FDA	Food and drug administration
FDP	Fibrin degradation products
FDR	False discovery rates
FNA	Fine needle aspiration
FTICR	Fourier-transform ion cyclotron resonance
GC	Gas chromatography

GlcNAc	N-acetylglucosamine
HCD	Higher-energy collision dissociation
HER2/NEU	Human epidermal growth factor receptor 2
HR	Hormone receptor
HRT	Hormone replacement therapy
IAA	Iodoacetic acid
IBC	Inflammatory breast cancer
IDC	Invasive ductal carcinoma
ILC	Invasive lobular carcinoma
KBH	Kepala Betas Hospital
KIT	Proto-oncogene c-Kit or tyrosine-protein kinase
LCIS	Lobular carcinoma *in situ*
LIT	Linear ion trap
LTQ	Linear trap quadrupole
MALDI	Matrix-assisted laser desorption ionisation
MAN	Monosaccharides like mannose
MOAC	Metal oxide affinity chromatography
MRI	Magnetic resonance imaging
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MUC-1	Mucin-like membrane (glycoprotein)
NCR	National cancer registry
PA	Plasminogen activator
PET	Positron emission tomography
PSA	Prostate-specific antigen
PTM	Post-translational modification
QIT	Quadrupole ion trap
QMS	Quadrupole mass spectrometer
Q-TOF	Quadrupole time-of-flight
RB	Running buffer
RNA	Ribonucleic acid
SA	Sinapinic acid
SJH	Seberang Jaya Hospital
TF	Thomsen-friedenreich

TFA	Trifluoroacetic acid
TiO2	Titanium dioxide
TIS	Timed ions selector
TOF	Time-of-flight
TPA	Tissue polypeptide antigen
TPS	Tissue polypeptide specific antigen
UniProtKB	The uniprot knowledgebase

ANALISIS PERBANDINGAN PROTEOMIK, GLIKOPROTEOMIK DAN FOSFOPROTEOMIK KE ATAS TISU KANSER PAYUDARA

ABSTRAK

Kanser payudara adalah punca kematian kedua tertinggi dalam kalangan wanita di seluruh dunia. Dalam populasi Malaysia, kanser payudara adalah salah satu jenis kanser yang paling biasa dalam kalangan wanita. Kajian ini bertujuan untuk mengenalpasti protein-potein tumor dan laluan isyarat yang berpotensi menjadi ciri unik karsinogenesis kanser payudara melalui teknik proteomik termaju, dengan melibatkan perbandingan antara profil protein tisu yang sihat dan kanser. Dalam kajian ini, profil protein kanser payudara dan tisu normal yang berdekatan diambil dari wanita Malaysia diperolehi dengan menggunakan teknik pemisahan menggunakan GELFREE, pengenalpastian menggunakan LTQ-Orbitrap LC-MS/MS dan analisis bioinformatik. Data daripada pemprofilan protein dalam kajian ini telah menunjukkan bahawa terdapat banyak protein yang memmpunyai modifikasi pasca-translasi (PTM) seperti fosforilasi dan glikosilasi. Kajian ini mendapati sejumlah 137 protein unik, dan di antara protein ini, 21 protein unik adalah signifikan dalam tisu tumor. Tambahan pula, 81 fosfoprotein unik telah dikenalpasti, dan di antara protein ini, 12 fosfoprotein unik telah dikesan signifikan dalam tisu tumor. Selain itu, sejumlah 73 glikoprotein unik telah dikenalpasti, dan di antara protein ini, 10 glikoprotein unik telah dikesan signifikan dalam tisu tumor. Secara amnya, protein-protein ini didapati terlibat dalam proses-proses fisiologi seperti proliferasi, kemandirian, pergerakan, invasi, angiogenesis, metastasis dan pembaharuan diri sel stem dan laluan isyarat struktur sel. Seterusnya, analisis jaringan terhadap data interkasi protein menunjukkan laluan biologi yang berkemungkinan terkesan oleh perubahan profil protein pada tumor

payudara. Data menunjukkan laluan yang terubah akibat perubahan porfil protein ialah laluan isyarat *Notch*, *Hippo*, *Met*, *Hedgehog*, mekanisma pemproses protein rangka sel, pelengkap, biosintesis asid amino, dan pelekatan fokal. Hasil analisis bioinformatik terhadap profil protein terekspres secara berbeza pada tahap jaringan mendedahkan perkaitan yang banyak di antara laluan yang terkesan dengan perkembangan dan metastasis kanser payudara. Ciri-ciri yang menarik dalam kajian semasa ialah ia juga telah menemui beberapa biomarker baru kanser payudara yang berpotensi, iaitu, Peptidyl-Prolyl Cis-Trans Isomerase FKBP10, TAR DNA-mengikat protein 42, topoisomerase DNA 1, jari-jari Zinc CCCH domain yang mengandungi protein 18, 45 kDa protein pengikat kalsium, immunoglobulin kappa variable 3-11 dan protein luminal epididimis 189. Kajian ini melaporkan buat kali pertama mengenai protein yang terekspres tinggi dalam sampel tumor payudara. Secara amnya, kajian ini telah menyediakan pemahaman asas untuk memahami laluan biologi yang telibat dalam karsinogenesis tisu payudara dan seterusnya manjadi garis dasar untuk penanda-bio spesifik-tisu untuk tujuan diagnosis dan rawatan kanser payudara yang efektif. Dari penemuan ini, kesimpulan yang boleh dibuat ialah perbezaan profil proteomik yang direkodkan dan dijelaskan dalam kajian ini boleh digunakan sebagai penanda-bio berpotensi untuk membangunkan tatasusunan protein terhadap tumor payudara yang baru dan boleh dipercayai. Tambahan pula, biomarker dan laluan yang berkaitan mempunyai potensi dieksploitasi sebagai sasaran untuk diagnosis atau terapi kanser payudara.

COMPARATIVE PROTEOMIC, GLYCOPROTEOMIC AND PHOSPHOPROTEOMIC ANALYSES OF BREAST CANCER TISSUES

ABSTRACT

Breast cancer is the second-most leading cause of death among women worldwide. This is true in the Malaysian population as well; breast cancer is one of the most common types of cancer among women. This study was aimed to identify potential tumour proteins that are characteristic of breast carcinogenesis. In order to detect the tumour-specific proteins, differential proteomic profiling was conducted using advanced proteomics techniques in healthy and tumour breast tissue samples. In this study, protein profiles of breast tumour and their adjacent normal tissues from Malaysian women were obtained and examined using GELFREE separation, LTQ-Orbitrap LC-MS/MS identification, and bioinformatic analyses. Findings of the current proteomic profile showed that among the proteins that were detected, several proteins were found with post-translation modifications (PTMs) such as phosphorylation and glycosylation. The analysis of proteins resulted in the identification of 137 proteins, out of that 21 proteins were significantly abundant in tumour tissues. In addition, 81 phosphoproteins were identified, out of that 12 proteins were found significantly higher in abundance in tumour tissues than compared to that their normal counterparts. Furthermore, the analysis identified a total of 73 glycoproteins, out of these 10 glycoproteins were recorded to be significantly abundant in tumour tissues. Generally, most of the identified proteins were found to be involved in physiological processes such as proliferation, cell survival, motility, invasion, angiogenesis, metastasis and stem cell self-renewal and cell structure signalling pathways. Further, the network analysis of protein interaction data revealed about the

biological pathways that could probably have affected due to the changes in the protein profile in breast tumour. The result demonstrated that the pathways associated with the altered protein profile were Notch, Hippo, Met, and Hedgehog signalling pathways, as well as cytoskeletal protein processing mechanisms, complement systems, amino acid biosynthesis and the focal adhesion. Results of the integrated bioinformatics analysis on the differentially expressed proteins profile at network levels revealed a considerable association of the affected pathways with the development and metastasis of breast cancer. The striking feature of the current study is that it has also discovered several novel potential biomarkers of breast cancer, namely, Peptidyl-Prolyl Cis-Trans Isomerase FKBP10, TAR DNA-binding protein 42, DNA topoisomerase 1, Zinc finger CCCH domain-containing protein 18, 45 kDa Calcium-binding protein, immunoglobulin kappa variable 3-11 and epididymis luminal protein 189. The present study reported for the first time about these proteins that were overexpressed in breast tumour samples. Overall, this study has provided a foundational basis to understand the biological pathways that are involved in carcinogenesis of breast tissue and eventually to establish a baseline for tissue-specific biomarkers. From the present findings, it can be concluded that the differential proteomic profile recorded and elucidated in the study can be used as potential biomarkers to develop novel and reliable breast tumour-specific protein-arrays. The biomarkers and the related pathways can be potentially exploited as targets for breast cancer diagnosis or therapy.

CHAPTER 1

INTRODUCTION

1.1 Background

Cancer is a leading cause of death worldwide and one of the most serious public health problems globally. The disease entails the abnormal and irrepressible growth of body cells that subsequently spread out of control to other organs, resulting in the interference with body functions that are necessary for healthy living. In the United State, cancer accounted for 25 % of all mortalities and indicated an upward trend in the rate of its incidence, which is projected to reach 26.4 million with 17 million cancer-related deaths by 2030 (Boyle et al., 2008; Siegel et al., 2019).

Breast cancer is the second leading cause of cancer-related deaths and is the fifth-most prominent cause of death worldwide (Lee et al., 2014; Lee & Oh, 2014). It is the most widespread malignancy among women globally (Molina-Montes et al., 2014), with an annual estimate of 2,088,849 (11.6 %) new cases and about 626,679 (6.6 %) deaths worldwide (Bray et al., 2018). The high morbidity of breast cancer is related to late diagnosis in which cancer has reached aggressive stages (Jassem et al., 2013). It is a common cause of cancer death among women (Harhra & Basaleem, 2012; Chahil et al., 2015); Malaysia's cancer profile resembles those of most Asian countries. According to Malaysia's National Cancer Registry (MNCR), there were 18,206 (32.1 %) breast cancer cases from 2007 to 2011. This profile makes it be the most frequently diagnosed type of cancer among Malaysian women with an age-standardised incidence rate (ASR) of 31.1 per 100,000 women (Azizah et al., 2016). Malaysia is expected to witness an increase in cancer cases because of changes in lifestyles and an improved standard of living that lead to increased life expectancy. Therefore, it is imperative to note that higher proportion of breast cancer cases (about

43 %) was diagnosed at an advanced stage (III and IV) of the disease (Azizah et al., 2016). Although the cause of breast cancer is largely unknown, and the precise prevention approaches have yet to be developed. However, the primary prevention strategy for many years has relied on early detection and early intervention to increase survival rates (Abdullah et al., 2013).

Sharma et al. (2005) asserted that treatment options are improving for current breast cancer subjects, which improves prospects for long-term survival. Breast cancer treatment thrives on continuous research and development of new interdisciplinary treatment modalities. Modern advances and exploits in proteomics can support this process via the identification of biomarkers. Proteomics research represents a novel molecular approach to the study of cellular or tissue protein identification and expression profiling (Ramm et al., 2015). Wilkins et al. (1996) defined a proteome "as the protein complement expressed by a genome." An organism's proteome content is the complete protein expressed in every cell and tissue. Human bodies are estimated to contain over two million proteins, normally coded by only 20,000–25,000 genes.

Biomarkers are substances present in body fluids or tissues that can be employed to determine an individual's risk for developing cancer or an indicator that is used to assess the presence of disease (Henry & Hayes, 2012; Kamel & Al-Amodi, 2016). In this regard, cancer can be detected based on how cancer tissue profile differs from normal tissue; these differing characteristics are biomarkers of tumourigenesis (Victor & Levenson, 2007). Cancer biomarkers are generated by either a tumour or the body as a response mechanism to tumour growth. Pritzker (2004) reported that studies on proteomics had identified many cancer biomarker candidates that go beyond proteins. They can be genes, gene products, specific cells, molecules, enzymes, or hormones which can be detected in blood, urine, tissues, or other body fluid (Rhea &

Molinaro, 2011). Other biological molecules potentially used as biomarkers include ribonucleic acids (RNA), deoxyribonucleic acids (DNA), Messenger RNA (mRNA) and micro (mi) RNA, lipids, carbohydrates, and polyamines, respectively (Kurebayashi et al., 2006; Hamam et al., 2017).

The most common biomarkers used in the prognosis and forecast of cancer are glycosylated proteins, such as HER2/NEU in breast cancer (Cabioglu et al., 2005) and carcinoembryonic antigen (CEA) in colon, lung, breast, and pancreatic cancers (Adamczyk et al., 2012). Proteomic and genomic studies have also identified some biomarker options for breast cancer prognosis, including steroid receptors, growth factor receptors, p21, p53, Ki-67, cyclins, BRCA1, BRCA2, urokinase plasminogen activators, and pro- and anti-apoptotic factors (Mandal et al., 2009). It is relevant to mention that despite the surge in biomarker candidates within the past decade, the United States Food and Drug Administration (FDA) has acknowledged only a limited number of these new biomarkers as correlated with research and diagnosis; CA15.3, CA27-29, and specifically for stage IV breast cancer, human epidermal growth factor receptor 2 (HER2/NEU) (Kamel & Al-Amodi, 2016).

Generally, biomarkers help to differentiate between individuals who are affected by cancer to those who are cancer-free. These differences in biomarkers usually occur as a result of germline or somatic mutations, transcriptional changes, and post-translational modifications (PTMs) (Henry & Hayes, 2012). For instance, patients with a positive family history for ovarian cancer can submit to a genetic examination to determine whether they carry a germline mutation, such as BRCA1 (Breast cancer 1), that increases the risk of developing breast and/or ovarian cancer (Easton et al., 1995). Biomarkers also help in determining prognosis, or the probability of disease

relapse independent of treatment. The use of biomarkers also potentially strengthens the effectiveness of treatment and detection.

Glycosylation and phosphorylation together are the most common types of PTMs in eukaryotic cells (Gavrilov et al., 2015). Glycosylated proteins form when an enzyme introduces carbohydrate side chains onto existing protein molecules to produce complex oligosaccharide sequences and associated structural variability (Byrne et al., 2007). Go et al. (2018) reported that over half of all known protein sequences have the potential to be glycosylated. Glycosylation involves in the regulation of numerous biological events and processes. For example, glycosylation inhibits E-cadherin, which ordinarily promotes cell-cell adhesion and limits metastatic potential. Pinho et al. (2013) attributed the loss of E-cadherin function in patients with gastric carcinoma to glycosylation; specifically, the addition of β1,6GlcNAc-branched N-glycans. In addition, glycosylation induces angiogenesis via both the vascular endothelial growth factor (VEGF, through O-GlcNAcylation) and the receptor tyrosine kinases (VEGFRs, through galectins) (Munkley & Elliott, 2016). The wide range of glycan functions include the control of vascular permeability (Croci, et al., 2014), regulation of Notch signalling (Pakkiriswami et al., 2016), maintenance of endothelial cell survival (Kitazume et al., 2010), and connection of blood and lymphatic vessels (Cheng et al., 2018).

The phosphorylation of certain amino acids residues, including tyrosine, serine, and threonine, occurs post-translation and is considered a reversible and dynamic alteration that is central to the regulation of cellular signalling pathways. Phosphorylation controls numerous biological processes, including (but not limited to) cell growth, differentiation, and apoptosis. Thus, the outcomes of diseases (such as cancer) can be linked with abnormal protein phosphorylation (Wang et al., 2017;

Ardito et al., 2017). Phosphorylation of eIF4E, the mRNA 5′ cap-binding protein, promotes tumourigenesis and is implicated in the development and progression of cancer (D'Abronzo & Ghosh, 2018).

Even though early diagnosis improves a patient's probability of undergoing successful breast cancer treatment, the United States FDA has not recently reported any approved tissue tests for early breast cancer detection. There is a need to find dependable breast cancer biomarkers in patient tissues (Herzig & Tsikitis, 2015). But the American Society of Clinical Oncology (ASCO) has updated its recommendations for the use of promising biomarkers include Carcinoembryonic antigen (CEA), Estrogen receptor (ER), CA 15-3, CA 27.29, Progesterone receptor (PR), Human epidermal growth factor receptor 2 (HER2), Urokinase plasminogen activator (uPA), Plasminogen activator inhibitor 1 (PAI-1) and multianalyte testing of tissue for individuals diagnosed with breast cancer and body fluid analysis for both women at risk and to monitor individuals after treatment (Kabel, 2017; Sauter, 2017).

In the systems biology era, information must be integrated from sets of multiple molecules that participate in a process. Therefore, candidate genes, transcripts, and proteins derived from –omic experiments can be analysed using curated databases or repositories. A relevant approach uses pathway analysis tools to predict which biological substrates are affected by the candidate molecules. Examples of these databases and repositories of curated annotations of molecules include gene ontology terms (The Gene Ontology Consortium, 2018), protein-protein interactions (Kerrien et al., 2012), biochemical pathways (Kanehisa et al., 2015), and gene expression (Clough & Barrett, 2016). The standard protocol for pathway analysis involves: (i) uploading a list of molecules obtained from genomic and proteomic experiments, (ii) retrieving annotations of these molecules using one or several of the commented databases and

repositories, (iii) obtain the over-represented pathways compared to background (normally, the full genome), and (iv) generate representative cellular networks that contain the identified input molecules. Database for Annotation, Visualization and Integrated Discovery (DAVID) is a classic tool that allows the user to compare multiple molecules (from genes to proteins) with many repositories with heterogeneous information (Jiao et al., 2012). Also, STRING (Szklarczyk et al., 2015) uses multiple repositories to seek significant relations between pairs of molecules. Finally, CYTOSCAPE (Fabregat et al., 2017) is a powerful tool to represent biological networks. With the addition of REACTOME FI plugin, the user can generate a full network that encompasses the input molecules as well as linker nodes (Crowgey et al., 2017). This approach generates functional networks in which nodes are molecules and edges are functional relations, allowing the initial list of molecules to be interpreted in terms of a molecular pathway.

New research advances into biomarkers continue to present more effective methods for breast cancer diagnosis. For instance, the method of mass spectrometry (MS) and Orbitrap have advanced our understanding of glycosylation and phosphorylation beyond the uses of glycan structures. Recently, developers of proteomics technologies presented the prospect of identifying new breast cancer biomarkers. Advanced MS technologies are already being applied to characterise PTMs, particularly glycosylation and phosphorylation. Here, the glycoproteomic profiles of breast tissues are comparatively analysed against non-tumour tissues to identify modulated proteins, which could characterise potential markers related to clinical features of breast cancer. This research represents a critical step toward fulfilling the exigent need to differentiate and detect more sensitive markers for the early detection and diagnosis of breast cancer (Dotz et al., 2015).

1.2 Problem statement

Breast cancer is the most common cancer among Malaysian women, and it constitutes approximately 31.2% of all cancers as reported by the Malaysian National Cancer Registry in 2007 (Zainal Ariffin & Nor Saleha, 2011). However, the majority of breast cancer cases in Malaysia are diagnosed at an advanced phase of the disease (Lim et al., 2011). Increased rates of mortality due to breast cancer also recorded in many parts of Malaysia (Youlden et al., 2014). The higher morbidity is linked to a late diagnosis of breast cancer where cancer has reached to the aggressive stages (Jassem et al., 2013). Breast cancer is dominant among Malaysian women aged between 40 and 49 (Chahil et al., 2015). The exact cause of breast cancer is yet to be determined, so precise prevention and early detection are of the utmost concern (Abdullah et al., 2013). Moreover, regulating bodies such as the US-FDA and Malaysian National Pharmaceutical Regulatory Agency (NPRA) have not recently approved any tissue tests for early breast cancer detection. But few tumour markers that showed evidence of clinical utility were recommended by FDA that include CA 15-3, CA 27.29, HER-2/neu, and circulating tumour cells analysis of EpCAM, CD45, CK8, 18, respectively (Sauter, 2017).

Since breast cancer patients have a promising prognosis only if their disease is diagnosed as early as possible, and before advancing of cancer to the extent that even modern medicine systems cannot address the issue (Herzig & Tsikitis, 2015). Therefore, there is an urgent need to find reliable tissue-specific targets for breast cancer. The identification of potential proteins among breast cancer patients could aid in early detection, better prognosis, effective treatment, thorough monitoring and comprehensive understanding of the process of carcinogenesis and its different stages.

1.3 Research questions

This research considers the discussion in the preceding sections and addresses the following research questions:

1. What are the differences between the normal breast tissue and tumour breast tissue with respect to the tissue-specific proteomic profile?
2. How is the tumour-specific protein profile implicated in the breast carcinogenesis?
3. Can the differential protein profiles be used as the reliable and valid biomarkers for early detection of breast cancer?

1.4 Significance of the study

This study may detect a new protein(s) that can support the accurate and early diagnosis of breast cancer at different stages of growth. The use of proteins as biomarkers of breast cancer is critical for determining the risk factors for breast cancer growth and optimising courses of therapy in the patients with different stages of cancer (Duffy et al., 2017). Protein biomarkers will help to differentiate accurately between affected persons and cancer-free individuals. Therefore, researchers have asserted that there is a crucial need to distinguish and detect more sensitive markers for early breast cancer detection and diagnosis (Dotz et al. 2015).

1.5 Research objectives

This study was designed to investigate and determine protein expressions in human normal and tumour breast tissue samples using advanced proteomic approaches. In addition, the study was planned to investigate the physiological

pathways affected due to the changes in protein expression. Finally, the objective of the study was to examine the role and implication of the protein profile in tumourigenesis.

1.5.1 Study aims

The aim of the present study was to investigate tissue biomarkers using Orbitrap mass analyser (LTQ Orbitrap MS/MS) to distinguish patients with breast cancer from the healthy population. In addition, the study was intended to obtain a comprehensive understanding of the altered protein profile in breast cancer at the pathway and network levels, therefore, an integrated bioinformatics analysis of differentially expressed proteins in breast tumour was conducted. Assessment of protein expression profiles in tumour tissues using proteomic technologies can be helpful for the discovery of novel and reliable biomarkers for the detection of breast cancer; therefore the main aim of this study was to identify a biomarker signature for early screening detection, prognosis, and treatment of breast cancer.

1.5.2 Specific objectives

1. To identify the proteins, phosphoproteins, and glycoproteins in tumour and normal breast tissues.
2. To study the differential expression of the proteins, phosphoproteins, and glycoproteins in tumour and normal breast tissues.
3. To study the correlation of the selected proteins, phosphoproteins and glycoproteins with known breast cancer pathways.

CHAPTER 2

LITERATURE REVIEW

2.1 Breast Cancer

2.1.1 Definition of breast cancer

Breast cancer is a malignant tumour that originates in breast tissue and is characterised by uncontrolled propagation and accretion of abnormal cells. This disease is more prevalent in women, but men aged between 60 and 70 years also can develop breast cancer, with a ratio of 1 male affected per 125 female diagnoses (Ferlay et al., 2010; Giordano, 2018). Approximately 85% of breast cancer cases occur in the mammary ducts, the channels that drain milk from the lobules (milk-producing glands) to the nipple, while the remaining 15% originate in the lobules themselves (Stolier and Wang, 2008; Tchouala, 2014). In rare cases, breast cancer originates in the stromal tissues, an umbrella term for the breast's fatty and fibrous connective tissues (Kopans, 2007; Chattoraj & Vishwakarma 2018). The different kinds of breast cancer are categorised based on their origin (Sharma et al., 2010). Invasive/infiltrating ductal carcinoma (IDC) is the most pervasive; it originates and develops in the lactiferous duct, where it disrupts the duct tube and then invades or infiltrates the adjacent tissues of the breast (Hamrita et al., 2008; Sharma et al., 2010). This most common type is also one of the most severe kinds of breast cancer because it can metastasise by proliferating into the lymph or blood systems (Barrière et al., 2012).

2.1.2 Incidence of breast cancer

Among women, breast cancer is the most common invasive cancer (Siegel et al., 2016) worldwide. On a global scale, it is a significant cause of ill health and death in women aged over 45 years (Youlden et al., 2012). In 2013, approximately 23% (1.38 million) of new cancer cases and 14% (458,400) of cancer-related mortalities were attributable to breast cancer. Also, approximately 50% of all breast cancer cases and 60% of breast cancer-related deaths occur in countries that are less technologically advanced (Kooshyar et al., 2013).

Malaysia is a middle-income county with a population of around 30 million that falls into the sub-Saharan and Asia-Pacific regions. The NCR of Malaysia reported that breast cancer is the most common cancer in Malaysia, with an age-standardised incidence rate (ASR) of 47.4 per 100,000 adult females from 2003 to 2005, which decreased to 39.3 per 100,000 in 2006 (Norsa'adah et al., 2012; Sulaiman et al., 2014). Though breast cancer is the most common cancer among all cultures of Malaysian women (Al-Dubai et al., 2012), the incidences do vary between the three major ethnic groups (Malay, Chinese, and Indian) in Peninsular Malaysia. The incidence rate found to be highest in Chinese women at 59.9 per 100,000, compared to 54.2 per 100,000 Indian women and 39.4 per 100,000 Malay women (Al-Dubai et al., 2011 and Abdullah et al., 2013). The significant disparity across Malaysia's three major ethnic groups may be due to dissimilarities in lifestyle, nutrition, and reproductive behaviours (including pregnancy and breastfeeding practices) (Tan et al., 2018)

The ASR of breast cancer in Malaysian women is still significantly higher than in other Asian countries (for comparison: Beijing, China: 24.6 per 100,000; Hiroshima, Japan: 36.6 per 100,000; Chennai, India: 23.9 per 100,000; Seoul, Korea: 20.8 per 100,000) (Dahlui, et al., 2011; Sajahan & Omar 2018). Breast cancer accounted for

31.3% of new cancer cases in women in Malaysia (Dahlui et al., 2011). Despite the downward trend in incidence in 2006, breast cancer remains the most common cancer among Malaysian women as well as in Peninsular Malaysia, irrespective of gender. For example, Breast cancer is the most common cancer among Malaysian women and accounted for 31% of total female cancers, and the most common cancers affecting females in Malaysia are breast (Lim et al. 2008; Zainal Ariffin & Nor Saleha, 2011; Tan et al., 2018).

Tan et al. (2018) reported that breast cancer is the most common form of cancer among Malaysian women. One in nine women is expected to have breast cancer. It was also observed that the lifetime risk of healthy women was highest among Chinese by 1 in 22, followed by Indians by 1 in 24, and lowest among Malays and 1 in 35. The majority of cancer patients in Malaysia were diagnosed at the later stages of the disease: 15.45% at stage I, 46.9% at stage II, 22.2% at stage III, and 15.5% at stage IV (Dahlui et al., 2013). The five-year survival rate following diagnosis was found to be 59.1% in Kuala Lumpur, the capital city of Malaysia, and 25.8% in the less developed east coast state of Kelantan (Moore, 2013; Zaridah, 2014).

Figure 2.1 shows the frequencies of the ten most common cancers among Malaysian women, with the highest rates recorded for breast cancer (17.7), and the lowest rates recorded for stomach cancer (3.4). The new cancer cases were distributed among the Malaysian states as follows: 18.8% in Penang, 18.4% in Johor, 11.3% in Selangor, 8.7% in Perak and Sarawak, 8.2% in Sabah, 7.1% in the Federal Territory of Kuala Lumpur, 5.5% in Kedah, 4.9% in Pahang, 4.7% in Kelantan, 3.9% in Malacca, 3.5% in Terengganu, 3.4% in Negeri Sembilan, 0.6% in Perlis, and 0.01% in the Federal Territory of Labuan (Omar et al., 2011). In 2008, the Third NCR Report of Malaysia stated that the incidence rate for breast cancer was highest among the persons aged 50

to 60, with the exception of Indians, for whom breast cancer peaked after 60 years of age. Malay women are diagnosed less frequently than Chinese and Indian women (Norsa'adah et al., 2012), with only 5,410 diagnoses of primary breast cancer documented among Malaysian women in 2012 (Yip et al., 2014). There have also been significant increases in breast cancer mortality rates in many regions, particularly Malaysia (Youlden et al., 2014). Higher morbidity is associated with late diagnosis of breast cancer, where cancer has reached advanced stages (Jassem et al., 2013). Breast cancer is prevalent among Malaysian women between the ages of 40 and 49 (Chahil et al., 2015). Therefore, it is necessary to improve the strategies for early detection and intervention.

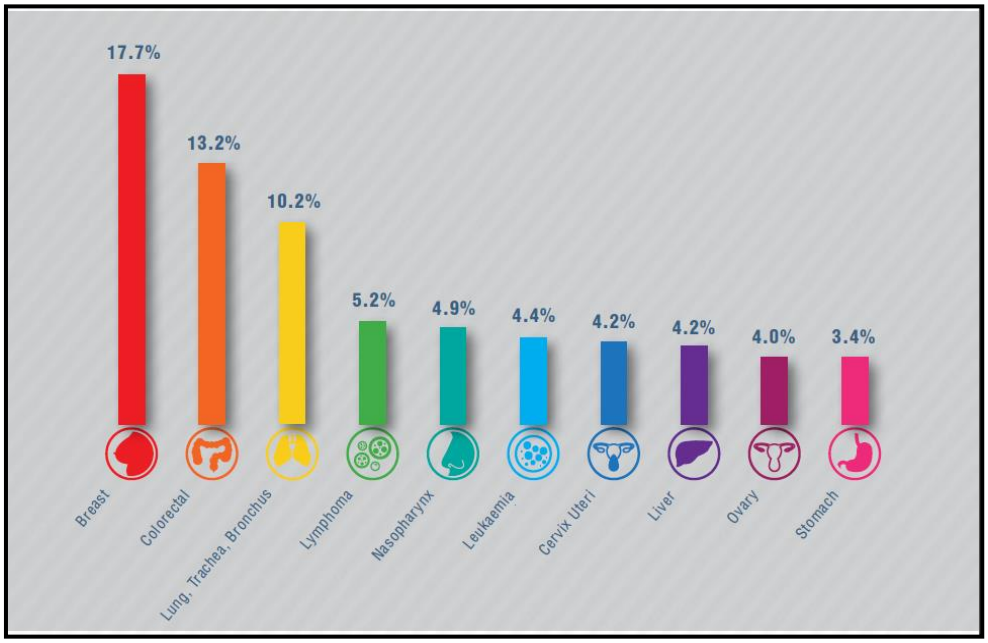

Figure 2.1 Ten most common cancer in Malaysia, 2007-2011

Adapted from the summary of the Malaysian National Cancer Registry Report 2007 – 2011.

2.1.3 Types of breast cancer

The majority of breast cancers are ductal carcinomas, but other less common forms exist and affect various regions of the breast. These include inflammatory breast cancer (IBC), medullary carcinoma, Paget's disease of the breast, tubular carcinoma, phyllodes tumour, metaplastic carcinoma, sarcoma, micropapillary carcinoma, and adenoid cystic carcinoma. Categorical distinctions between types of breast cancers include epithelial or non-epithelial, invasive or non-invasive, and multifocal or multicentric (Saslow et al., 2007). It is imperative to mention that most tumours are initiated by benign changes within the breast and usually do not cause cancer (American Cancer Society, 2010).

After ductal carcinoma, the next most common form of breast cancer is found in the lobules and hence is called lobular carcinoma (Azim et al., 2011). There are two subtypes of non-invasive breast cancer: ductal carcinoma *in situ* (DCIS) is the most common, while lobular carcinoma *in situ* (LCIS) is less common. DCIS is denoted as stage 0, indicating its non-invasive potential, because its proliferation does not extend beyond the milk duct into any adjacent normal breast tissues. Furthermore, DCIS is not fatal, but it does confer an increased risk of invasive breast cancer as the individual grows older (Li et al., 2006). Conversely, LCIS consists of a region of anomalous cell growth and is not considered breast cancer, though it does indicate the patient’s risk of developing invasive breast cancer in the future (Figure 2.2).

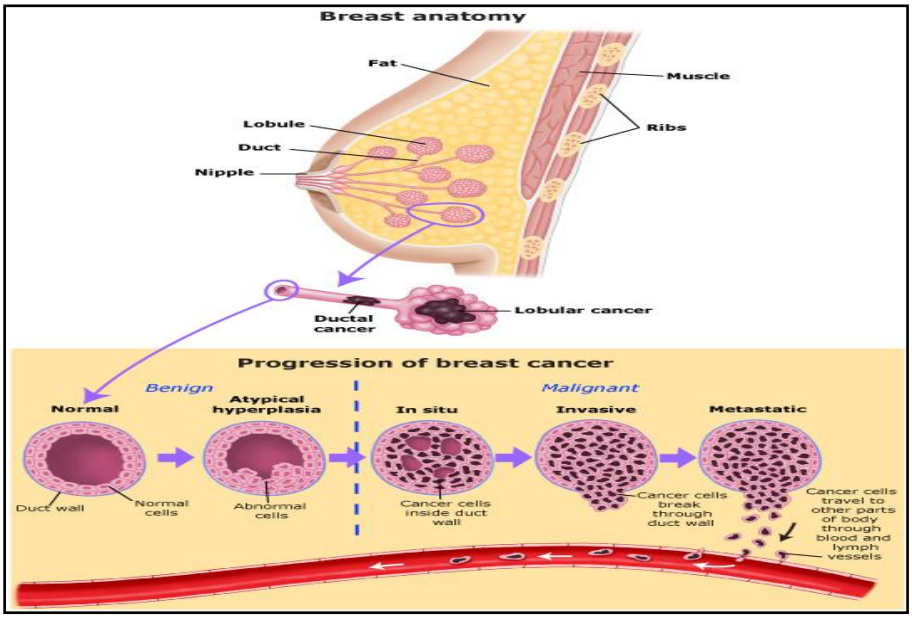

Figure 2.2 Breast anatomy and progression of cancer

(Adapted from http://www.brestcancer.org)

The remaining stages of breast cancer are all considered invasive because the cancer cells spread into or invade the normal breast tissue adjacent to the ducts or lobules (Ma & Jemal, 2013). Invasive breast cancer may be one of three subtypes: invasive (or infiltrating) ductal carcinoma (IDC) is the most common, followed by invasive lobular carcinoma (ILC) and IBC. IDC comprises about 80% of all breast cancers (Lakhani et al., 2006). IBC is aggressive and uncommon, constituting less than 5% of all breast cancer cases in the United States. In this type, the earliest manifestations of cancer are reddening and swelling of the breast tissue, rather than a distinct tumour. These skin changes are prone to be misdiagnosed as an infection rather than a malignancy. IBC tends to propagate and proliferate rapidly, causing symptoms to deteriorate within days or even hours. Therefore, the identification of early symptoms and pursuit of immediate treatment can be life-prolonging. Nonetheless, it is important to mention that while IBC is a fatal diagnosis, the currently available treatments are better than previous options at controlling the disease (Levine & Veneroso, 2008).

2.1.4 Triple negative breast cancer

Triple-negative breast cancer (TNBC) is a heterogeneous disease which is associated with poor prognosis and has a lack of expression of the estrogen receptor (ER), progesterone receptor (PR) and epidermal growth factor receptor type 2 (HER2) (Chen et al., 2015). TNBC is a heterogeneous disease which is associated with poor prognosis; and breast cancer type that lacks the expression of the estrogen receptor (ER), progesterone receptor (PR) and epidermal growth factor receptor type 2 (HER2) (Chen et al., 2015).

TNBC is characterised as aggressive tumour subtype represents around 15% of breast cancer diagnoses (Mundt et al., 2018). Based on gene expression analysis, it was established that the tumour subtype breast cancer represents a heterogeneous disease of 6 identified distinct TNBC subtypes, to which each subtype displays unique biology (Elsawaf et al., 2013; Hubalek et al., 2017). Exploring novel approaches for the treatment of these subtypes is critical, especially because the median survival for women with metastatic TNBC is less than 12 months, and virtually all women with metastatic TNBC ultimately will die of their disease despite systemic therapy. To date, not a single targeted therapy has been approved for the treatment of TNBC, and cytotoxic chemotherapy remains the standard treatment (Abramson et al., 2015)

TNBC is a particularly aggressive type of breast cancer lacking clear diagnostic approach and targeted therapies. Consequently, more than 50% of the TNBC patients die of the metastatic BC within the first 6 months of the diagnosis (Golubnitschaja et al., 2018). Besides the unclear diagnosis of this cancer type as described before, this cancer (TNBC) is considered a significant unsolved clinical case, that has a propensity to be more aggressive, with a worse prognosis than receptor-positive subtypes (Hudis & Gianni, 2011). The problem of this type of cancer is generally among young women

of various races; and the various ages of American women of African origin (Howlader et al., 2014). Identifying subtypes within the TNBC type, and proteins within those subtypes that can serve as therapeutic targets will be extremely valuable (Lawrence et al., 2015). The identification of several specific subtypes characterised by different biologic pathways and various sensitivities to chemotherapy is instrumental in delivering a more personalised therapy for TNBC.

2.1.5 Stages of breast cancer

The stage of breast cancer is determined by the size and the extent of involvement of the lymph nodes and/or other organs (National Cancer Institute, 2019). Importantly, the stage determines prognosis and treatment selections. According to the American Joint Committee on Cancer (AJCC), the TNM classification system is the most common method to categorise stage; it describes different characters of cancer-based on specific standard criteria (Greene et al., 2002). The TNM acronym stands for the three criteria incorporated by this classification: (a) the extent of the tumour (T), (b) the extent of lymph node (N) involvement, and (c) the presence of metastasis (M) (National Cancer Institute, 2019). Breast cancer stage is determined by either the clinical stage (through a physical exam, biopsy, and imaging) or the pathologic stage (which also incorporates the results of surgery) (Jian et al., 2017). After T, N, and M are identified, they are combined, and an overall breast cancer stage is allocated. Table 2.1 designates the definition of each stage of breast cancer according to the TNM system.

Table 2.1 Stages of breast cancer

Stage	Definition
Stage 0	Tumour cells confined to the breast duct without extension into surrounding tissue
Stage I	Tumour diameter < 2cm without the involvement of any lymph nodes
Stage IIA	Tumour diameter < 2cm with the involvement of axillary lymph node(s), OR tumour diameter 2–5cm without the involvement of any lymph nodes
Stage IIB	Tumour diameter 2–5cm with the involvement of axillary lymph node(s), OR tumour diameter > 5cm without the involvement of any lymph nodes or extension into the chest wall
Stage IIIA	No tumour in the breast itself but lymph nodes (potentially including those near the breastbone) contain cancer and are annealing to themselves or other structures
Stage IIIB	Tumour extends into the chest wall and/or skin of the breast; lymph nodes (potentially including those near the breastbone) may contain cancer and are annealing to themselves or other structures. The diagnosis of IBC starts at this stage.
Stage IIIC	Involvement of lymph nodes either above or below the collarbone AND lymph nodes in the axilla or near the breastbone; may or may not be signs of cancer in the breast itself
Stage IV	Involvement of other organs (metastasis)

(Adapted from Kumar et al., 2005; Jung, 2009; Singletary et al., 2002)

The characterisation of clinicopathological properties that indicates the patient's prognosis comprises of lymphovascular invasion diagnosis. It also includes the tumour size, lymph node metastasis, histological subtype, and grade diagnosis. These diagnosis characterisations (Clark, 2008), are based on the histological analysis of primary breast cancer samples called TNM (Tumour size, Nodes, and Metastasis) system. Moreover, the scale of categorisation for TNM stage of breast cancer was accomplished/done, following the procedures as explained by Sobin et al. (2009). Tables 2.2 - 2.4 show the

physical examination and imaging (i.e., mammography) procedure for assessing T, N, and M categories. Table 2.5 shows the TNM system that has been integrated into tumour stages with given prognostic scale values.

Table 2.2 T – Primary tumour for breast cancer

TX	Primary tumour cannot be assessed
T0	No evidence of primary tumour
Tis	Carcinoma *in situ*
Tis (DCIS)	Ductal carcinoma *in situ*
Tis (LCIS)	Lobular carcinoma *in situ*
Tis (Paget)	Paget disease of the nipple not associated with invasive carcinoma and/or carcinoma *in situ* (DCIS and/or LCIS) in the underlying breast parenchyma.
T1	Tumour 2 cm or less in greatest dimension T1mi Microinvasion 0.1 cm or less in greatest dimension*
T1a	More than 0.1 cm but not more than 0.5 cm in greatest dimension
T1b	More than 0.5 cm but not more than 1 cm in greatest dimension
T1c	More than 1 cm but not more than 2 cm in greatest dimension
T2	Tumour more than 2 cm but not more than 5 cm in greatest dimension
T3	Tumour more than 5 cm in greatest dimension
T4	Tumour of any size with direct extension to the chest wall and/or to the skin (ulceration or skin nodules)
T4a	Extension to the chest wall (does not include pectoralis muscle invasion only)
T4b	Ulceration, ipsilateral satellite skin nodules, or skin oedema (including peau d'orange)
T4c	Both 4a and 4b, above
T4d	Inflammatory carcinoma

Table 2.3 N – Regional lymph nodes for breast cancer

NX	Regional lymph nodes cannot be assessed (e.g., previously removed)
N0	No regional lymph node metastasis
N1	Metastasis in movable ipsilateral Level I, II axillary lymph node(s)
N2	Metastasis in ipsilateral Level I, II axillary lymph node(s) that are clinically fixed or matted; or in clinically detected* ipsilateral internal mammary lymph node(s) in the absence of clinically evident axillary lymph node metastasis
N2a	Metastasis in axillary lymph node(s) fixed to one another (matted) or to other structures
N2b	Metastasis only in clinically detected* internal mammary lymph node(s) and in the absence of clinically detected axillary lymph node metastasis
N3	Metastasis in ipsilateral infraclavicular (Level III axillary) lymph node(s) with or without Level I, II axillary lymph node involvement; or in clinically detected* ipsilateral internal mammary lymph node(s) with clinically evident Level I, II axillary lymph node metastasis; or metastasis in ipsilateral supraclavicular lymph node(s) with or without axillary or internal mammary lymph node involvement
N3a	Metastasis in infraclavicular lymph node(s)
N3b	Metastasis in internal mammary and axillary lymph nodes
N3c	Metastasis in supraclavicular lymph node(s)

Table 2.4 M – Distant metastasis for breast cancer

M0	No distant metastasis
M1	Distant metastasis

Table 2.5 Stage grouping for breast cancer

Stage 0	Tis	N0	M0
Stage IA	T1*	N0	M0
Stage IB	T0, T1*	N1mi	M0
Stage IIA	T0, T1	N1	M0
	T2	N0	M0
Stage IIB	T2	N1	M0
	T3	N0	M0
Stage IIIA	T0, T1*, T2	N2	M0
	T3	N1, N2	M0
Stage IIIB	T4	N0, N1, N2	M0
Stage IIIC	Any T	N3	M0
Stage IV	Any T	Any N	M1

Note: *T1 includes T1mi

2.1.6 Diagnosis of breast cancer

The majority of women can detect breast masses on their own through self-examination, but most of the identified masses are non-tumour. Self-examination alone has been proven insufficient to reduce breast cancer mortality, so it is advisable to use more than one method for detecting tumours (Hackshaw & Paul, 2003). Multiple methods combined can allude to the presence and stage of breast cancer; common techniques include self-examination, mammography, ultrasound, biopsy, and imaging like magnetic resonance imaging (MRI) or positron emission tomography (PET) scans.

Mammography is a non-invasive radiographic technique (X-ray) that examines the breast tissue. It is considered the benchmark method for the screening and diagnosis of breast cancer (Nass et al., 2000). Ultrasound uses high-frequency sound waves that create an image of the breast tissue that complements the information provided by

additional tests, such as mammography. Despite these advantages, the ultrasound is inadequate because it cannot examine the whole breast at one time and cannot detect tumours in the earliest stages. Hence, it is useful for small areas only. Moreover, it is not as comprehensive as a mammogram. In fact, 1–2% of breast cancers are identified with some method other than ultrasound or mammography (Berg et al., 2008).

PET scans identify cancer by injecting radioactive material into the tissue of interest and scanning the tissue for areas of rapid proliferation (i.e., the cells that uptake the most radioactive material). However, PET scans are limited in their capacity to identify small tumours (van der Hoeven et al., 2002), and the test is not readily available in most medical facilities; the scans tend to be expensive and require exceptional knowledge. MRI uses magnetic fields to image the body and has known efficacy for the detection of breast cancer (Smith & Andreopoulou, 2004). The biopsy method removes cells for histological examination under a microscope to evaluate for the presence of malignant cells and assess the extent of tumour tissue. The specific biopsy technique depends on the features of the breast abnormality (Srinivas et al., 2002). At this time, no single method is sufficiently reliable, cost-efficient, or accessible enough for widespread use in breast cancer screening. A biomarker with a high specificity could supplement the diagnostic process tremendously, bolstering the chances of diagnosing breast cancer with precision and detecting the disease in its earliest stages.

2.2 Stages of cancer development

Carcinogenesis is a complex process with distinct molecular and cellular modifications that occur in discrete but closely related stages—initiation, promotion, and progression—that ultimately lead to the development of a malignant tumour (Siddiqui et al., 2015; Moga et al., 2016). The tumour progresses through successive

stages that are shown in Figure 2.3. The first step involves physical, biological, or chemical alterations, or the spontaneous conversion of normal cells due to exposure to carcinogenic agents that mutate the cellular genome and create the potential for neoplastic development (Devi, 2004). The promotion step is a reversible process in which stimuli cause initiated cells to undergo selective clonal expansion, creating a pre-malignant tumour cell population that is dividing and propagating actively. The last stage is tumour progression, which is an irreversible process where the additional genetic mutations result in the formation of new tumour cell clones that possess heightened invasive cellular phenotypes and metastatic potential (Moga et al., 2016; Maru et al., 2016; Kotecha et al., 2016).

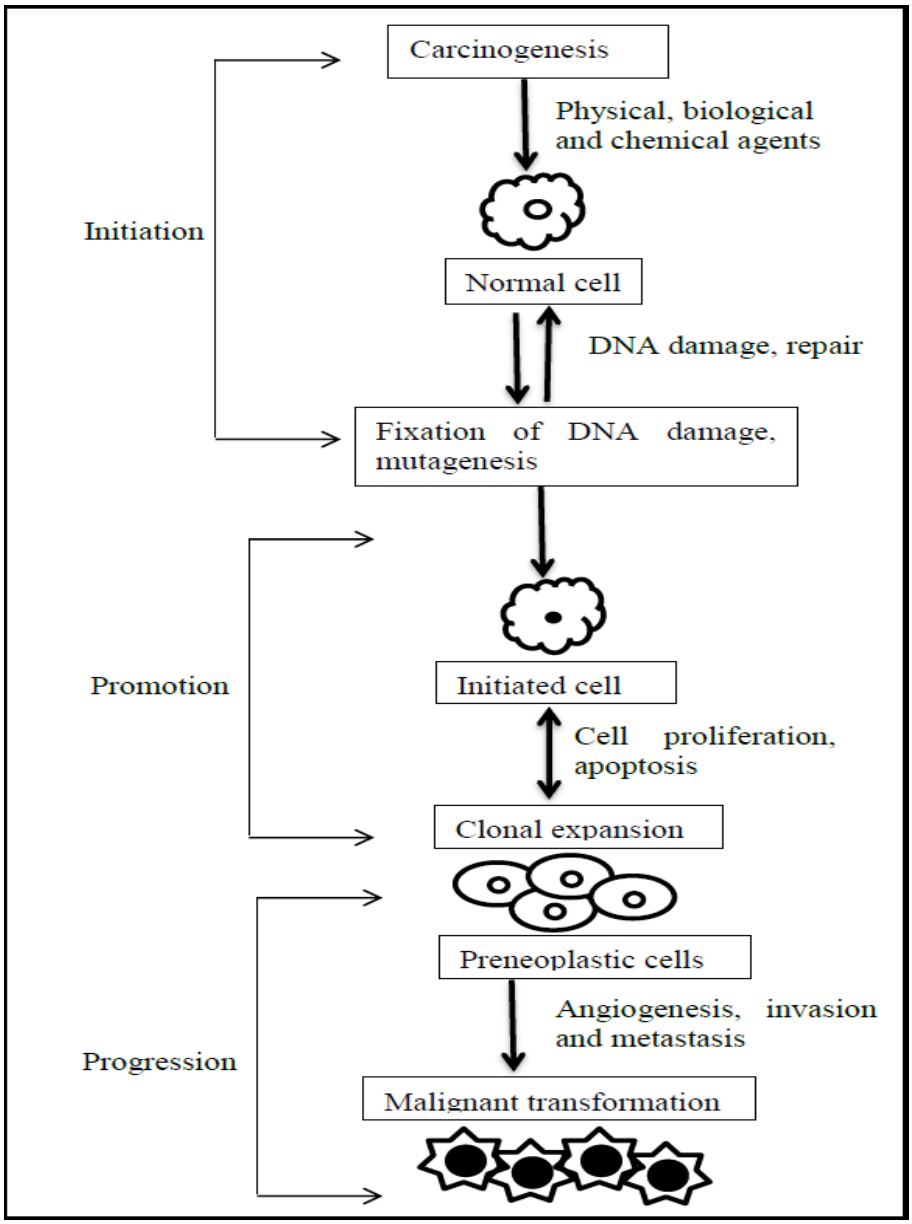

Figure 2.3 Stages of cancer development

(Adapted from Siddiqui et al., 2015)

2.3 Breast cancer biomarkers

A biomarker is a gene, protein, peptide, or metabolite that exists in tissue, cells, and/or fluids in the body. It is used to indicate pathological and physiological changes during the incidence of a disease. Biomarkers must have characteristic features objectively measured and evaluated as indicators of normal biological processes,

pathogenic processes, or pharmacologic responses to therapeutic interventions so that they can be applied as indicators of disease traits and aid in diagnosis and prognosis (Henry & Hayes, 2012; Kuo et al., 2018).

Biomarkers hold promise to aid breast cancer assessment, they have become crucial for both healthy patients (in determining a patient's risk of developing breast cancer) and diagnosed patients (in selecting the appropriate therapy and managing cancer diagnosis) (Duffy et al., 2017). Breast cancer cells that are positive for the estrogen receptor (ER) provide an index for sensitivity to endocrine treatment. The expression of this biomarker (ER) increases as the breast cancer patient ages (Duffy et al., 2017). The expression of ER-α in breast cancer was found to be approximately 70%, thus predicting a potentially positive response to endocrine (hormonal) therapy (Lauring & Wolff, 2016). Testing for progesterone receptor (PR) expression as a potential biomarker in breast cancer was done to determine the likelihood of successful cancer treatment with endocrine (hormonal) therapy, in which hormones are interrupted from aiding cancer growth. The progesterone receptor (PR) becomes activated when it interacts with the hormone progesterone, and 65 % of breast cancers that are ER-positive are also PR-positive. Therefore, PR expression likely depends upon ER expression and are usually measured collectively via immunohistochemical (IHC) assay although PR can be independently prognostic in breast cancer (Salmen et al., 2014; Van Belle et al., 2010; Sauter, 2017).

Similarly, human epidermal growth factor receptor-2 (HER2) is a transmembrane tyrosine kinase receptor that is part of a larger family of epidermal growth factor receptors (Banin et al., 2014). This protein (HER2) is found in all breast cells at different intensities due to HER2 gene amplification (Rakha & Green, 2017). It accelerates cell growth at a high level, prompting an increase in size by approximately