UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Proteomic characterisation of primary breast tumour extracellular matrix

by

Chanelle Monique Pillay

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

Doctor of Philosophy

in

Pharmacology

in the

Faculty of Health Sciences

at the

University of Pretoria

Supervisor

Prof. A.D. Cromarty

Co-supervisors

Dr S.H. Stoychev

Dr B.A. Stander

2021

Declaration

University of Pretoria

Faculty of Health Sciences

Department of Pharmacology

I, Chanelle Monique Pillay,

Student number: 27127053

Subject of the work: Proteomic characterisation of primary breast tumour extracellular matrix

Declaration

1. I understand what plagiarism entails and am aware of the University's policy in this regard.
2. I declare that this thesis is my own, original work. Where someone else's work was used (whether from a printed source, the internet or any other source) due acknowledgement was given and reference was made according to departmental requirements.
3. I did not make use of another student's previous work and submit it as my own.
4. I did not allow and will not allow anyone to copy my work with the intention of presenting it as his or her own work.

Signature:

Acknowledgements

Firstly, I would like to thank my Heavenly Father for His eternal love and grace. I would not have been able to complete this degree without my God, who deserves all the glory, honor and praise. Philippians 4:13 – I can do all things through Christ who strengthens me.

I would like to acknowledge and give sincere thanks to the following people for their assistance and support throughout the duration of this study:

- My supervisor Prof. Duncan Cromarty, thank you for your constant support and encouragement throughout my postgraduate studies. It has been a long journey and your continuous motivation, guidance, dedication and belief in me has meant so much over the years. You have been an amazing supervisor and mentor.
- My co-supervisor Dr Stoyan Stoychev, thank you for your encouragement, guidance and constant willingness to help. Your technical advice was highly valued and appreciated.
- My co-supervisor, Dr Andre Stander, thank you for your assistance with this research project.
- Dr Sindisiwe Buthelezi, thank you for your assistance, support and patience while training me in the laboratory. Thank you for always encouraging and motivating me especially when it was hard to see the end. You are a great mentor and friend.
- The Department of Pharmacology at the University of Pretoria, staff members and especially my dear friends Hafiza, Shamiso, Machel and Bongai.
- The Department of Surgery at the Steve Biko Academic Hospital, medical personnel and surgeons especially, Prof. Meshack Ntlhe, Dr Fathima Docrat and Dr Athena Maliakel. Your willingness to assist with sample collection was greatly appreciated.
- My colleague and friend, Precious Setlai, thank you for your assistance with sample collection and your continuous encouragement during this research project.
- The Department of Anatomical Pathology at the University of Pretoria, especially Prof. Meshack Bida and staff members for allowing me to make use of departmental resources and facilities. Your assistance was greatly appreciated.
- A heartfelt thank you to Susra van Biljon and Litshani Nemakhavhani from the NHLS for their assistance on the cryotome. Thank you for always being so willing to help whenever I needed assistance.

- Dr Marissa Muller, thank you for your assistance and expert advice with the histological analysis. I really appreciate your support and guidance.
- The Council of Scientific and Industrial Research (CSIR) and staff members especially, Dr Ireshyn Govender, Dr Previn Naicker and Mr Sipho Mamputha. Thank you for your technical advice and support during experiments at the CSIR.
- Dr Kim Sheva, we started out on this journey together and what a journey it has been my friend! I truly cannot thank you enough for your constant support and motivation over the years.
- Amrita, thank you for always encouraging me and giving the best advice.
- Niki, thank you for your continuous support and prayers throughout my studies.
- My family, especially my sisters Casandra and Annuscha and my brother, Emlyn. Thank you for your constant love, support and motivation.
- My parents, Siva and Valerie, to whom I dedicate this thesis to. Thank you for your unconditional love and unwavering support. Thank you for always believing in me and encouraging me to achieve my goals. I love you very much.

Abstract

Breast cancer is the most commonly diagnosed cancer in women and is the leading cause of female cancer mortality worldwide. High cancer mortality rates, mostly due to late-stage diagnosis and the lack of appropriate personalised therapy, highlights treatment failure that prompts the need for continued research to identify new and improved breast cancer detection methods and treatment. The purpose of this study was to use advanced mass spectrometry-based proteomics to characterise and compare the proteome but especially extracellular matrix (ECM) protein components from solid invasive ductal carcinoma tumours to matched non-tumorous breast tissue with the aim of identifying potential prognostic markers or new therapeutic drug targets for breast cancer treatment.

Breast tumours are dense, complex tissue masses made up from a number of different proliferating cell types that are embedded in an intricate tumour microenvironment. Several studies have highlighted the role of the tumour microenvironment, more specifically the ECM, in tumour development and progression from localised invasion to advanced metastasis. The ECM consists of numerous protein components that provide a scaffold for both cell and growth factor binding, where ECM changes have been associated with tumour advancement. By implication, characterisation of tumour ECM components can potentially be used as prognostic or staging markers for breast cancer or to identify new targets for anticancer therapies.

In this research study, cryotome cut slices of snap-frozen tumour biopsies resected from patients diagnosed with invasive ductal carcinoma (IDC) were used to characterise the primary breast tumour proteome especially for the ECM. Haematoxylin and eosin staining, the gold standard for routine histopathological diagnosis of cancer, was used to visualise tissue morphology and to confirm the clinical IDC diagnosis.

An optimised protein extraction method involving high pressure cycling technology was used for tissue homogenisation and protein solubilisation of tumour biopsies. Proteomics analysis involving innovative semi-automated and cutting-edge sample preparation and liquid chromatography tandem mass spectrometry-based methods that are at the fore-front of drug target validation, drug discovery and prognostic marker identification, were used to acquire proteomic data from the tissue isolated from both tumour biopsies and equivalent non-

tumorous breast tissue. A semi-automated sample preparation method using hydrophilic affinity-based protein capture, clean-up and off-bead trypsin digestion was used to produce peptides, followed by analysis using a Dionex Ultimate 3000 RSLC system coupled to an SCIEX 6600 TripleTOF mass spectrometer. Data independent acquisition using sequential window acquisition of all theoretical mass spectra (SWATH) data was collected and bioinformatic protein identification and relative quantification was performed. SWATH data provided reliable proteomic assessment, could identify low abundance proteins as well as provide relative quantitation of differentially expressed proteins in tumour samples. Tumour associated ECM changes were classified through STRING pathway analysis comparing the relative protein abundance between non-tumorous and tumour masses.

Pathway analysis revealed that ribosomal, spliceosome and endoplasmic reticulum protein processing pathways with associated protein components were significantly upregulated in breast tumour samples. Proteomic data confirmed that protein homeostasis, associated with protein synthesis, protein folding and alternative splicing, is severely affected in solid tumours in order to meet the demands of uncontrolled tumour growth and promotion of tumour metastasis. SWATH-based quantification and pathway enrichment analysis did identify several ECM protein networks containing a number of differentially expressed ECM proteins in breast tumour samples. These ECM proteins within the tumour microenvironment are involved in several cancer related biological processes that include structural integrity, cancer cell proliferation, tumour growth, tumour tissue invasion, and metastasis. These differentially expressed ECM proteins could potentially be used as putative biological prognostic signatures for breast cancer or be used as new drug targets to slow or completely inhibit breast cancer advancement and progression.

This exploratory study provides valuable proteomic data for breast cancer research associated with the tumour microenvironment and has laid the foundation for prognostic and pharmacological based studies for cancer therapeutics by identifying putative ECM protein candidates that can be further assessed in independent verification and validation breast cancer studies.

Keywords: Breast cancer, Invasive ductal carcinoma, tumour microenvironment, extracellular matrix proteome, proteomics, LC-MS/MS, SWATH-based quantification.

Table of Contents

Glossary of Abbreviations

°C	Degrees Celsius
µg/µL	Micrograms per microlitre
µL	Microlitre
µm	Micrometre
%	Percent
2-D	Two dimensional
3-D	Three dimensional
AUC	Area under curve
ATF6	Activating transcription factor-6
AI	Aromatase inhibitors
BLBC	Basal-like breast cancer
BMI	Body mass index
BSA	Bovine serum albumin
CAF	Cancer associated fibroblast
CDK	Cyclin dependent kinase
CV	Coefficient of variation
DCIS	Ductal carcinoma *in situ*
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
DLU	Ductal lobular unit
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ECM	Extracellular matrix
EFGR	Epidermal growth factor receptor
ER	Oestrogen receptor
ERAD	Endoplasmic reticulum associated degradation
ESI	Electrospray ionisation
FDR	False discovery rate
FFPE	Formalin-fixed paraffin-embedded

g	Grams
g	Gravity
GAG	Glycosaminoglycans
H&E	Haematoxylin and eosin
HCl	Hydrochloric acid
HER-2	Human epidermal growth factor 2
HILIC	Hydrophilic interaction liquid chromatography
hnRNPs	Heterogenous nuclear ribonucleoproteins
IAA	Iodoacetamide
IDC	Invasive ductal carcinoma
ILC	Invasive lobular carcinoma
IRE1α	Inositol-requiring enzyme 1α
iTRAQ	Isobaric tags for relative and absolute quantification
Kpsi	Kilo-pound per square inch
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LOX	Lysyl oxidase inhibitor
M	Molar (moles per litre)
m	Metres
m/z	Mass-to-charge ratio
mg	Milligrams
mg/mL	Milligrams per millilitre
min	Minutes
mL	Millilitre
mM	Millimolar
MMPs	Matrix metalloproteinases
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
ms	Milliseconds
nL	Nanolitre
nm	Nanometre
NOS	Not otherwise specified
OCT	Optimal cutting temperature liquid

PARP	Poly ADP-ribose polymerase
PCT	Pressure cycling technology
PDI	Protein disulphide isomerases
PD-1	Programmed death receptor 1
PD-L1	Programmed death receptor ligand-1
PERK	Protein kinase RNA-like endoplasmic reticulum kinase
pH	Negative logarithm of the hydrogen ion concentration
PI3K	Phosphatidylinositol-3-kinase
PR	Progesterone receptor
PSM	Peptide-spectrum match
RP	Reverse phase
RPs	Ribosomal proteins
rRNA	Ribosomal ribonucleic acid
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
sec	second
SILAC	Stable isotope labelling of amino acids in cell culture
SLRP	Small-leucine-rich-proteoglycan
snRNPs	Small nuclear ribonucleoproteins
SPARC	Secreted protein acidic and rich in cysteine
SWATH	Sequential Window Acquisition of all Theoretical Mass Spectra
TGF_{β}	Transforming growth factor beta
TMT	Tandem mass tagging
TNBC	Triple negative breast cancer
TNC	Tenascin C
TNM	Tumour node metastasis
UPR	Unfolded protein response
UV	Ultraviolet
V	Volts
VEGF	Vascular endothelial growth factor
VEGFR-1	Vascular endothelial growth factor 1
VEGFR-2	Vascular endothelial growth factor 2
vs	Versus
XIC	Extracted ion chromatogram

Chapter 1

1.1 Literature review

1.1.1 Introduction

Cancer is currently the second leading cause of death in both developed and developing countries. This form of non-communicable disease has a significant impact on declining life expectancy and is expected to rank as the number one cause of deaths worldwide in the late 21st century. The International Agency for Research on Cancer (IARC) provided the latest global cancer data for incidence and mortality rates for thirty-six types of cancers in a total of one hundred and eighty-five countries as shown in Figure 1.1 (Bray *et al.*, 2018). The incidence of cancer has reached an alarmingly high number, with an estimated 18.1 million new cases and 9.6 million deaths recorded in 2018. It has been reported that approximately 43.8 million people are currently living within the five-year period after initial diagnosis of some form of cancer. Overall, breast cancer is ranked as one of the three most prevalent types of cancers, with the highest incidence rate and is classified within the top five cancer types having a high global mortality rate. Furthermore, breast cancer is the most frequently diagnosed type of malignancy in women and is ranked as the leading cause of female cancer mortality worldwide (Sung *et al.*, 2021).

Despite the lower incidence of cancer in sub-Saharan Africa, the mortality rate among cancer patients on the African continent is significantly higher when compared to cancer patients from other regions of the world (Bray *et al.*, 2018). Although this is partly due to the fact that a higher occurrence of cancers associated with poorer survival are prevalent in Africa, the main cause of high mortality rates can be attributed to the lack of social and economic development. Socioeconomic factors include the high cost of cancer treatment, unaffordable or limited access to suitable screening methods resulting in late-stage diagnosis, the complexity and duration of cancer treatment, chemotherapy resistance and the lack of personalised therapy in conjunction with the limited number of available skilled and highly trained medical specialists within this speciality (Morhason-Bello *et al.*, 2013; Edge *et al.*, 2014).

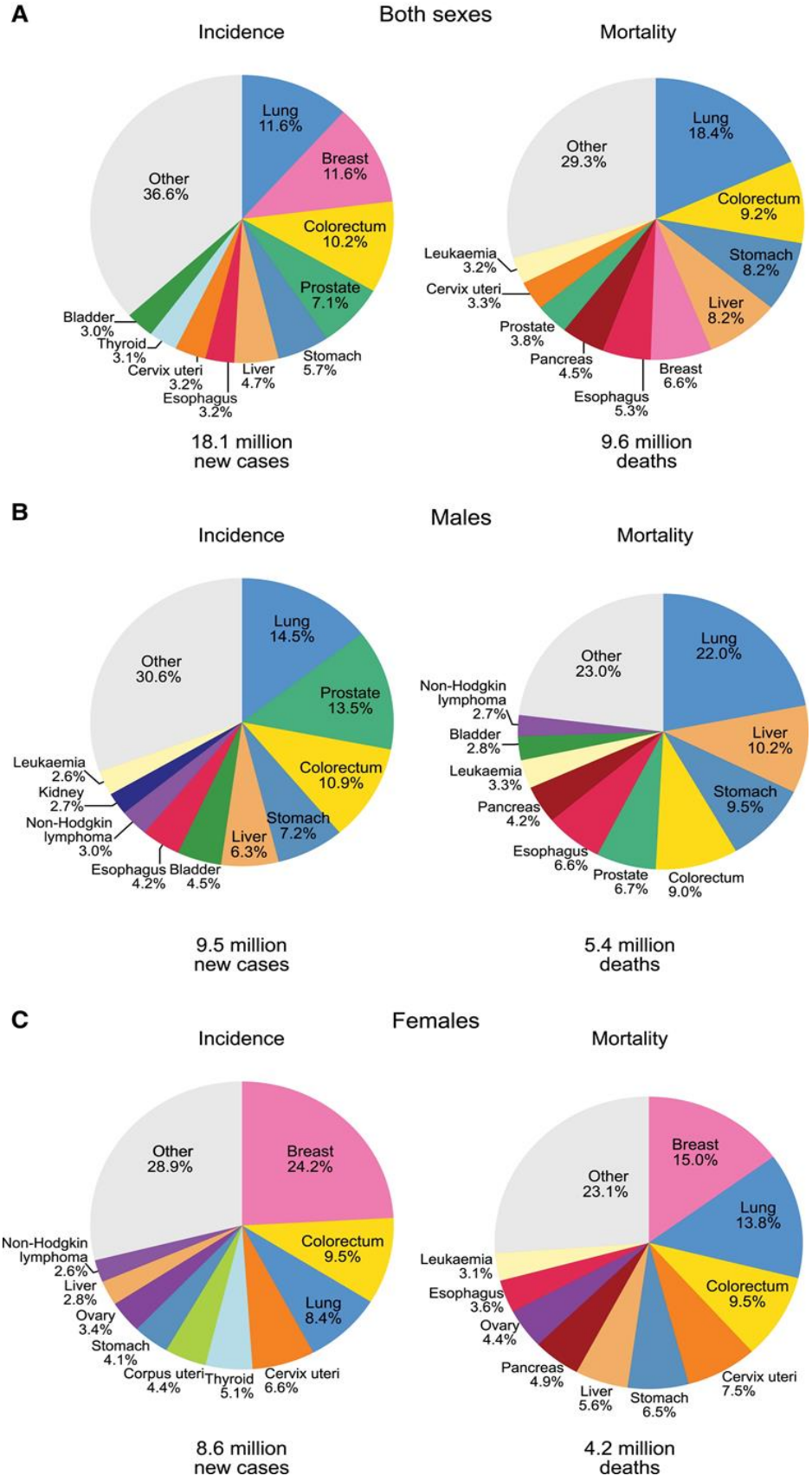

Figure 1.1: The distribution of cancer cases and deaths by region in 2018 for (A) Both sexes, (B) Males, and (C) Females. For each sex, the area of the pie chart reflects the proportion of the total number of cases or deaths (Bray *et al.*, 2018) (with permission).

Breast tumours are complex tissue masses made up of a number of distinct cell types surrounded by an intricate tumour microenvironment. Several studies have highlighted the role of the tumour microenvironment, more specifically the extracellular matrix (ECM), in tumour development and progression (Egeblad *et al.*, 2010a; Lu *et al.*, 2012; Oskarsson, 2013; Giussani *et al.*, 2015; Wishart *et al.*, 2020). The ECM consists of numerous components, including proteins that provide a scaffold for growth and which exhibit tumour promoting properties. Based on these characteristics, ECM components can potentially be targeted by new or existing anticancer therapies and can be classified as prognostic or staging markers for breast cancer (Insua-Rodríguez and Oskarsson, 2016). Furthermore, due to the tumour microenvironment's growth-promoting properties and the hypothesised role in some chemotherapy treatment failures, tumours should not be studied and characterised according to cell susceptibility only, but should also include the contribution of the surrounding microenvironment in order to develop new effective chemotherapy while also counteracting drug resistance (Albini and Sporn, 2007; Tsai *et al.*, 2014).

Proteomic studies have in recent years become prominent in cancer research, specifically with regards to identifying potential biomarkers indicating poor prognosis and for discovery of potential drug targets (Geiger *et al.*, 2012). Liquid chromatography tandem mass spectrometry (LC-MS/MS) based proteomics is the technique of choice for both targeted and non-targeted proteome analysis as LC-MS/MS provides high sensitivity, reproducibility and mass accuracy. Tryptic digestion of proteins followed by LC-MS/MS, also known as bottom-up proteome profiling, can be used to generate amino acid sequence data for peptides from essentially all proteins found within the ECM. Amino acid sequence data can subsequently be used to infer identity and characterise relevant proteins within a complex protein mixture with the potential to identify key prognostic markers and drug targets within small samples of human breast cancer. Additionally, useful proteomic data can be used to lay the foundation for the development of new anticancer treatments or guide the improvement in the efficacy of existing therapies by facilitating personalised treatment for individual patients (Angel *et al.*, 2012).

1.2 Embryological development and anatomical structure of the breast

Basic knowledge of the anatomical structure and maturation of the breast is essential in order to understand the development and progression of breast cancer. The embryological development of the human breast, as shown in Figure 1.2, commences as early as the fourth week of gestation and continues to develop, both physically and functionally, until adulthood. The organogenesis of the mammary gland is induced and controlled by both genetic and hormonal factors (Pandya *et al.*, 2011). The ectodermal layer, the outer germ layer surrounding the developing embryo, forms ectodermal thickenings known as mammary ridges which differentiate into mammary buds during the fifth week of gestation. Formation of secondary mammary buds and further differentiation into mammary lobules can be observed by the twelfth week of foetal development. Subsequently, secondary mammary buds differentiate into long epithelial branches to form lactiferous ducts that join the developing mammary lobules to the nipple of the developing foetus. Adipose cells, ligaments, vascular and lymphatic structures, which make up the stromal compartment of the breast, continue to develop throughout foetal development (Watson and Khaled, 2008; Bae *et al.*, 2013).

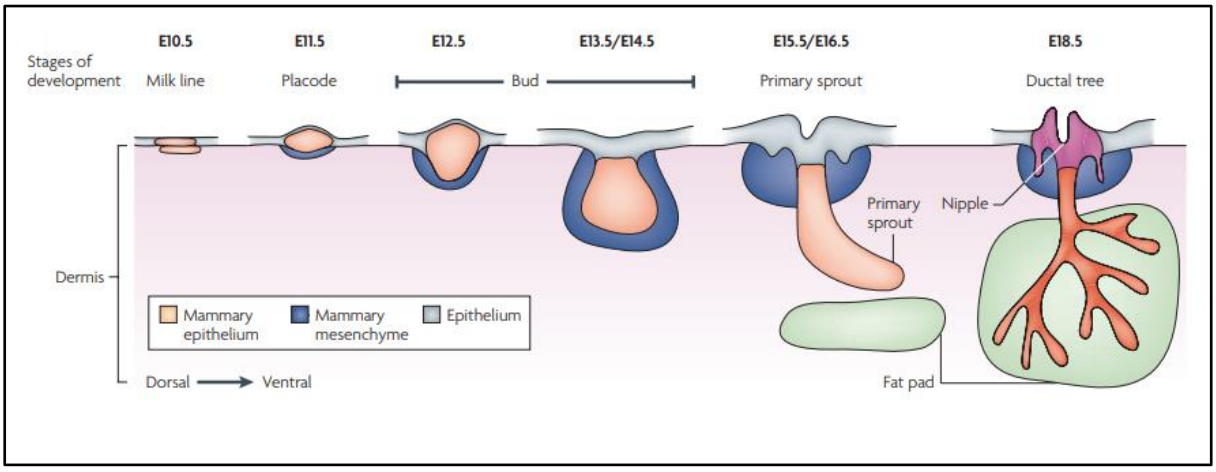

Figure 1.2: Embryological development of the human mammary gland (Robinson, 2007) (with permission).

Due to the absence of maternal hormonal influence, the development of the breast ceases after birth and proliferation of the ductal and lobular system, as well as the surrounding breast stroma is only stimulated again by increased oestrogenic effects during the onset of puberty (Ellis and Mahadevan, 2013). The mature female breast, located anteriorly to the

pectoralis major muscle of the thoracic cavity and extending superiorly from the second rib to the sixth rib inferiorly, consists of a well-developed but complex stromal compartment which surrounds approximately 15 - 25 lobes that each contain individual tubule-acinar glands that are connected by interlobular milk ducts as shown in Figure 1.3 (Macéa and Fregnani, 2006).

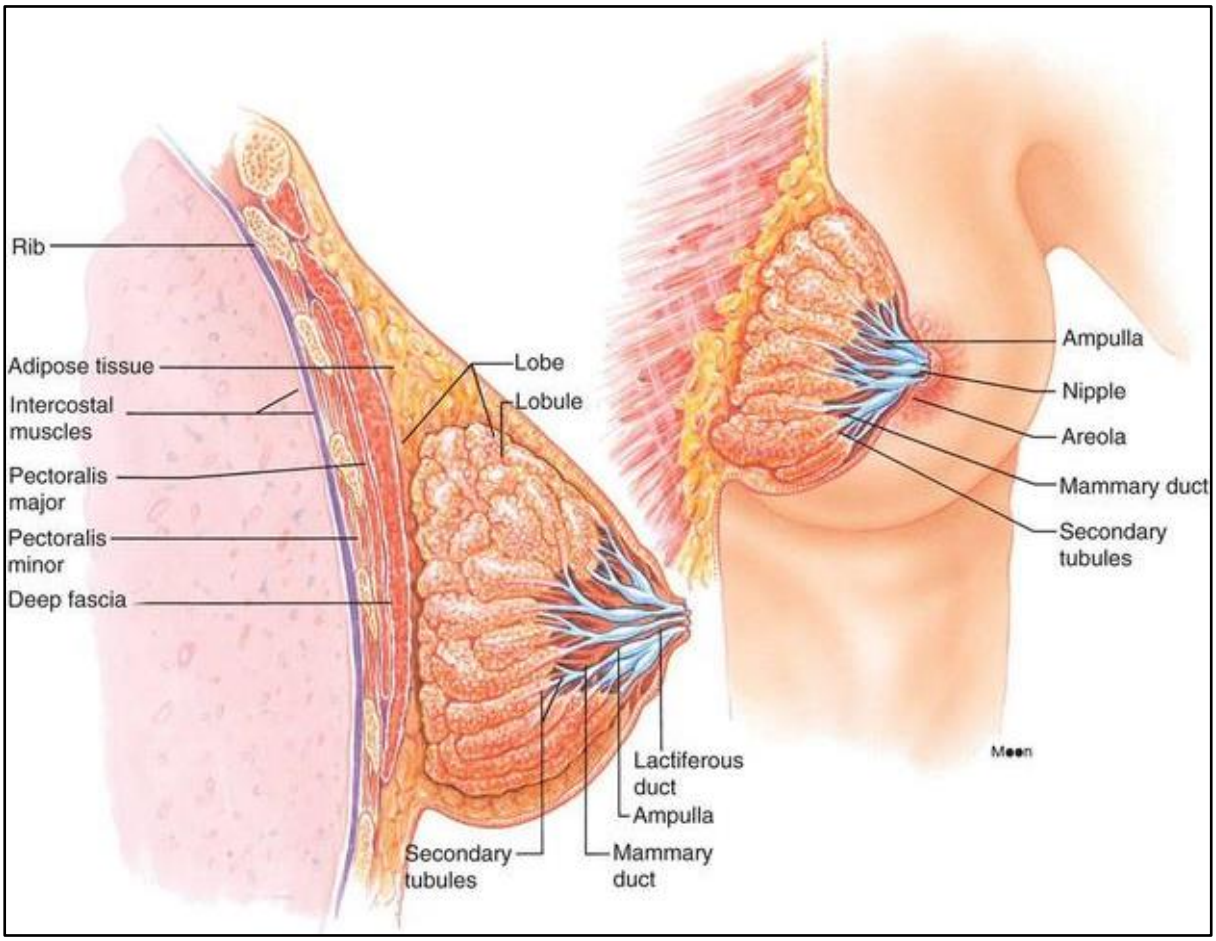

Figure 1.3: Anatomical structure of the female breast. (Image from https://medika.life/the-mammary-glands/) (with permission)

1.3 Classification of breast cancer

Breast ducts and lobules are lined by two types of cell layers, a monolayer of luminal epithelial cells and a basal layer of myoepithelial cells, which are separated from the surrounding stroma by a basement membrane. The majority of breast cancers arise from neoplasms, resulting from molecular alterations that lead to uncontrolled cellular growth and immortality

of epithelial cells, which line the ductal and lobular structures of the breast (Wellings and Jensen, 1973; Wellings, 1980). Breast cancers display complex heterogeneity which plays a prominent role in the risk of disease progression and response to treatment. Histopathological type, receptor status, molecular characteristics, tumour grading and staging are commonly used to characterise the diverse types of breast tumours according to their general morphology and structural organisation (Bertos and Park, 2011; Polyak, 2011).

1.3.1 Histopathological typing

Histopathology performed on tumour biopsies is used to classify the tumours and their growth pattern by assessing morphological and cytological patterns that are associated with specific types of breast cancers and to identify the cellular origin. Histopathology is therefore a fundamental tool used by pathologists for diagnosis and phenotypic classification of breast cancers (Li *et al.*, 2005). Carcinomas *in situ* refers to the neoplastic proliferation of epithelial cells that are confined within the cellular epithelial lining, whereas a breach of the basement membrane and subsequent invasion of cancer cells into the surrounding breast tissue is collectively referred to as invasive carcinomas (Bertos and Park, 2011). The most common histological subtype, accounting for approximately 75% of all breast tumours, is the invasive ductal carcinoma (IDC) not otherwise specified (NOS) implying no further subtyping (Sinn and Kreipe, 2013). Ductal carcinoma *in situ* (DCIS), which affects the same anatomical regions as IDC, but which is retained within an intact basement membrane and layer of myoepithelial cells, is hypothesised to be a precursor of IDC and often presents in conjunction with IDC (Cowell *et al.*, 2013). Invasive lobular carcinoma (ILC) is the second most commonly diagnosed type of breast cancer and makes up 10% of reported cases (Li *et al.*, 2005). Other histological subtypes of invasive breast cancers, which include micropapillary, medullary, mucinous (A and B), adenoid cystic, neuroendocrine, tubular, apocrine, and metaplastic tumours, make up the minority of diagnosed breast cancer cases, with each type having distinct histological features. Histological images of the common and rarer subtypes of invasive breast cancers are shown in Figure 1.4 (Polyak, 2011; Mayrhofer *et al.*, 2013).

Invasive ductal carcinomas exhibit diffuse clusters of cells arranged in cords or trabeculae. Variability in size and shape of neoplastic cells can be observed, with distinct nuclei and a number of mitoses with a stromal compartment that varies in abundance (Figure 1.4A).

Invasive micropapillary carcinomas are rare forms of IDC where tumours consist of small epithelial cells located within distinct clear stromal spaces. Reverse polarity, also referred to as inside-out structure, is a characteristic feature of invasive micropapillary carcinomas and alludes to the fact that the apical pole of the tumour cells faces the stromal compartment instead of the luminal ductal space (Figure 1.4B). Lobular carcinomas are tumours that consist of small round neoplastic cells that are distributed in a single file, commonly referred to as an Indian file pattern, which infiltrates the surrounding stroma (Figure 1.4C). Medullary carcinomas present as a sheet of poorly differentiated cells that show variations in size, shape and the number of nuclei. A syncytial growth pattern consisting of multinucleated cells can be observed (Figure 1.4D). As indicated by its name, tubular carcinomas appear microscopically as elongated or angular tubules consisting of a single layer of epithelial cells that lie openly in the lumina and can infiltrate the desmoplastic stroma (Figure 1.4E). Mucinous carcinomas are groups of uniform epithelial tumour cells that cluster together in an extracellular compartment with abundant mucin (Figure 1.4F). Inflammatory breast cancer is characterised by dermal lymphatic invasion and surrounding cellular inflammatory infiltrate (Figure 1.4G). Observed histological abnormalities associated with each type of breast cancer can be compared and contrasted microscopically to healthy breast tissue which comprises of ductal and lobular units that are lined by inner epithelial and outer myoepithelial cell layers which are surrounded by an extracellular stromal compartment (Figure 1.4H) (Mayrhofer *et al.*, 2013).

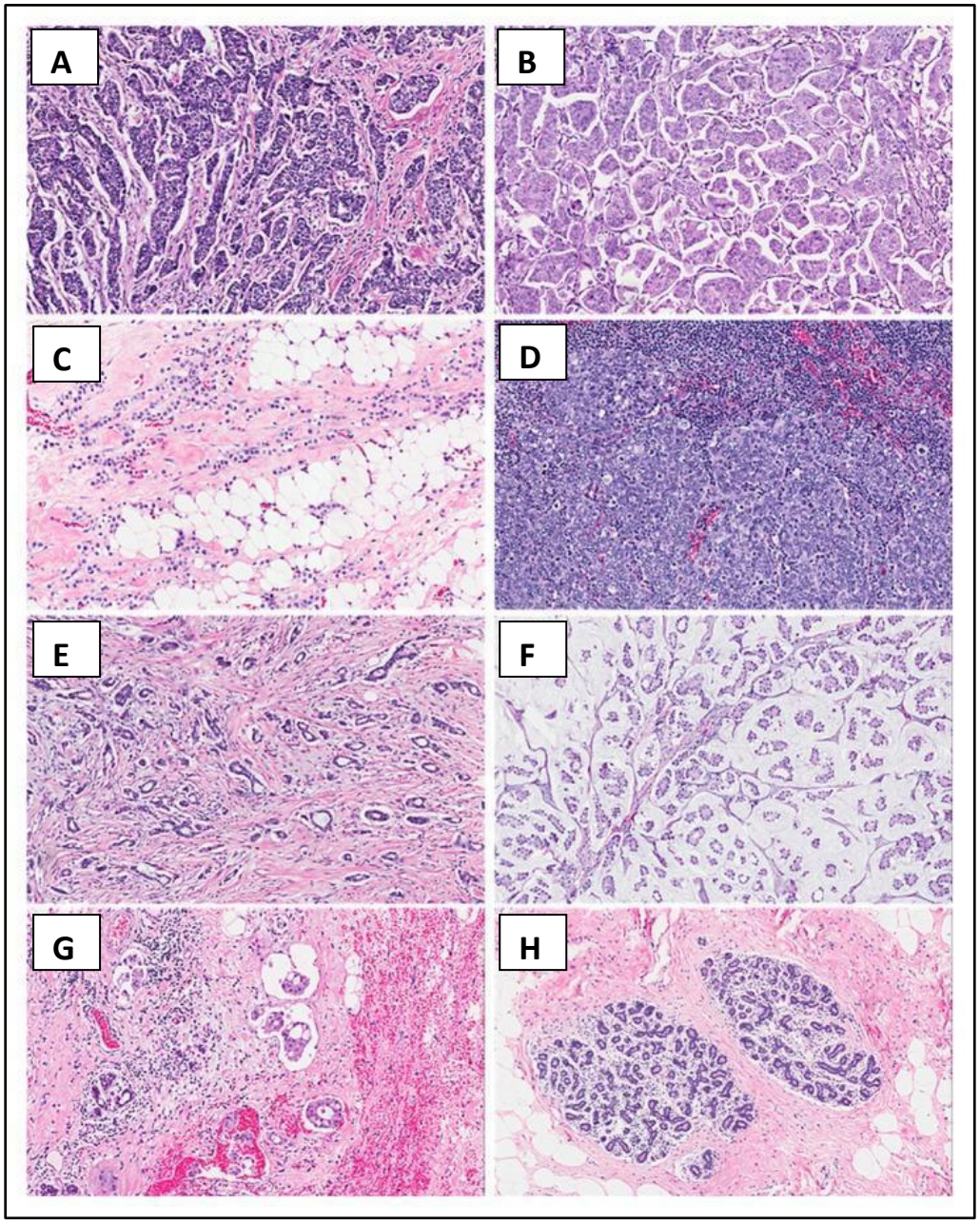

Figure 1.4. Haematoxylin and eosin (H&E) histological photomicrographs of different breast cancer types at 40× magnification (A) Invasive ductal carcinoma**.** (B) Invasive micropapillary carcinoma (C) Lobular carcinoma (D) Medullary carcinoma (E) Tubular carcinoma (F) Mucinous carcinoma. (G) Inflammatory breast cancer (H) healthy breast tissue (Mayrhofer *et al.*, 2013) (with permission).

1.3.2 Receptor status and molecular subtype

Gene expression profiling methods of cells from tumours have led to molecular subtype characterisation of breast tumours into several unique clusters, that have different overall biological and clinical characteristics. Accurate molecular subtype characterisation combined with clinical data can be used to predict patient prognosis (Eliyatkın *et al.*, 2015). Furthermore,

immunohistochemistry is utilised to stain cancer cells for the presence of specific receptor status that is used to predict tumour susceptibility to targeted chemotherapy (Tang and Tse, 2016). Classification of breast cancer tumour subtypes based on differences in gene expression patterns was first proposed by Perou *et al.* and Sorlie *et al.* after conducting studies to categorise breast carcinomas and correlating molecular characteristics to clinical outcomes. Five distinct molecular subtypes were identified; Luminal A, Luminal B, basal, human epidermal growth factor 2 (HER-2) positive (ErbB2/HER-2 amplification) and normal breast-like, (Perou *et al.*, 2000; Sørlie *et al.*, 2001; Sorlie *et al.*, 2003). Luminal A subtype is associated with the positive expression of oestrogen (ER) and progesterone (PR) receptors with minimal expression of HER-2, whereas luminal B subtype is linked to the positive expression of ER, PR and HER-2 receptors. HER-2 subtype is associated with the amplification and overexpression of HER-2 or (ErbB2) cell surface receptors that have growth factor ligands. Basal subtypes are often referred to as triple negative breast cancers (TNBC) as intracellular ER, PR as well as HER-2 cell surface receptors are essentially absent from these cancer cells. Normal breast-like carcinomas are described as breast cancers that have a similar gene expression profile to normal breast cells (Eliyatkın *et al.*, 2015; Ahn *et al.*, 2016). Characterising breast tumours according to molecular subtypes has been successfully used to determine best treatment options and as a prognostic tool, with ER positive tumours showing the most successful prognosis. Patients, testing positive for HER-2 tumours, show relatively good prognosis if treated with HER-2 targeting therapeutic agents (Smith *et al.*, 2007; Prat *et al.*, 2015). However, TNBC breast cancer has poor clinical outcome when compared to the other breast cancer subtypes (Nishimura and Arima, 2008). A summary of the major breast cancer subtypes and associated prognosis and treatment susceptibility is summarised in Table 1.1

Table 1.1: A summary of major breast cancer subtypes with associated prognosis and treatment susceptibility (adapted from (Eliyatkın *et al.*, 2015)).

Molecular subtype				
	Luminal A	**Luminal B**	**HER2/neu**	**Basal like[a]**
Gene expression pattern	Expression of luminal (low molecular weight) cytokeratin, high expression of hormone receptors and related genes	Expression of luminal (low molecular weight) cytokeratin, moderate-low expression of hormone receptors and related genes	High expression of HER2/neu, low expression of ER and related genes	High expression of basal epithelial genes and basal cytokeratin, low expression of ER and related genes, low expression of HER2/neu
Clinical and biologic properties	50% of invasive breast cancer, ER/PR positive, HER2/neu negative	20% of invasive breast cancer, ER/PR positive, HER2/neu expression variable, higher proliferation than Luminal A, higher histologic grade than Luminal A	15% of invasive breast cancer, ER/PR negative, HER2/neu positive, high proliferation diffuse TP53 mutation, high histologic grade and nodal positivity	15% of invasive breast cancer, most ER/PR/HER2/neu negative (triple negative), high proliferation, diffuse TP53 mutation, BRCA1 dysfunction (germline, sporadic)
Histologic correlation	Tubular carcinoma, cribriform carcinoma, low grade invasive ductal carcinoma, NOS, classic lobular carcinoma[b]	Invasive ductal carcinoma, NOS, micropapillary carcinoma	High grade invasive ductal carcinoma, NOS	High grade invasive ductal carcinoma, NOS, metaplastic carcinoma, medullary carcinoma
Response to treatment and prognosis	Response to endocrine therapy, variable response to chemotherapy, good prognosis	Response to endocrine therapy (tamoxifen and aromatase inhibitors) not as good as Luminal A variable response to chemotherapy (better than Luminal A), prognosis not as good as Luminal A	Response to trastuzumab (Herceptin), response to chemotherapy with anthracycline, usually unfavourable prognosis	No response to endocrine therapy or trastuzumab sensitive to platinum group chemotherapy and PARP inhibitors, not all, but usually worse prognosis.

PARP poly-adenosine diphosphate ribose polymerase
[a]Basal like tumour group includes a low-grade group with low proliferation but expression of basal type (high molecular weight) cytokeratin and triple negative phenotype (like adenoid cystic carcinoma, secretory carcinoma).
[b]Classical lobular carcinoma generally exhibits Luminal A properties, while pleomorphic lobular carcinoma usually shows features of other molecular subtypes.

1.3.3 Histological grading and staging of breast cancer

Grading and staging of breast carcinomas by histological investigation is essential for pathology reports of breast cancer patients, which allows clinicians to confidently diagnose or monitor tumour progression in patients and to predict the overall prognosis (Henson *et al.*, 1991; Rakha *et al.*, 2010). Histological grading describes the morphological differences between normal breast cells and breast cancer and includes key features, such as cancer cell surface markers, tubule formation, nuclear pleomorphism and mitotic count, which are indicators of tumour differentiation and excessive cellular proliferation, are assessed by scoring according to the widely used Nottingham Modification of the Bloom-Richardson system which has been shown to be a strong predictor of patient survival or relapse (Galea *et al.*, 1992; Rakha *et al.*, 2008). This histological grading is based on the cumulative scores of three tumour features, and this grading fairly accurately identifies patients with low-grade, intermediate-grade or higher-grade tumours that present with well, moderately or poorly differentiated cells, respectively (Rakha *et al.*, 2010). Higher graded tumours indicate patients who have an inferior prognosis compared to patients with low and intermediate graded tumours. High grade tumours are associated with poor cellular differentiation showing less uniform cellular nuclei and disorganised cell grouping from uncontrolled cellular division.

Staging of breast cancer gives an indication of the invasive extent and whether the tumour has spread and if so, does this spread involve only local tissue, draining lymphatics or distant organs. This involves measuring the size of the initial tumour, determining the extent of draining lymph node involvement and whether the cancer has established metastasises in other regions or organs of the body. The tumour, node, metastasis (TNM) system, first suggested by Pierre Denoix *et al.* in 1952, is most commonly used to stage breast cancer. Grading and staging coupled to gene expression combinations are used to predict patient treatment options, individualised chemotherapy treatment regimens, and survival (Denoix *et al.*, 1952; Veronesi *et al.*, 2006).

1.4 Risk factors of breast cancer

The incidence of breast cancer is multifactorial and disease aetiology has been extensively studied whereby a number of causative factors relating to genetic influences, hormonal

exposure, age, obesity and lifestyle choices, have been identified (Mcpherson *et al.*, 2000; Ataollahi *et al.*, 2015; Barnard *et al.*, 2015). It is well established that germline mutations of the two breast cancer associated genes, BRCA1 and BRCA2, are associated with genetic susceptibility to developing breast cancer and are at high risk (Easton *et al.*, 1993b; Wooster *et al.*, 1995; Frank *et al.*, 2002). The first breast cancer related gene to be identified was the BRCA1 gene, a penetrant autosomal dominant gene that often shows mutations. Females who are carriers of the BRCA1 gene mutation have been reported to have a 60-80% risk of developing familial breast cancer (Easton *et al.*, 1993a; Miki *et al.*, 1994). Clinical characteristics that distinguish BRCA1 gene related breast cancers from other gene mutation induced breast cancer types include early age of onset, increased risk of developing bilateral breast malignancy and a higher incidence of associated malignancies such as ovarian cancer (Martin and Weber, 2000).

Expansive research has confirmed the link between an increased risk of breast cancer and irregularly high levels of oestrogen or biological processes, such as early onset of menstruation, nulliparity, not breast feeding, late onset of menopause, which result in prolonged exposure to hormones (Pike *et al.*, 1979; Paffenbarger Jr *et al.*, 1980; Clemons and Goss, 2001). Other risk factors include: age, race, poor diet, high alcohol intake, obesity, exposure to ionising radiation, sedentary life style, increased breast tissue density, certain oral contraceptives and hormone replacement therapy (Sun *et al.*, 2017).

1.5 Diagnosis and treatment of breast cancer

Screening methods and early detection are essential for the overall effectiveness of a treatment plan for breast cancer patients. Several diagnostic techniques, including clinical examination, mammography, breast ultrasound and breast magnetic resonance imaging (MRI), are used to screen for and visualise breast carcinomas. Following a preliminary diagnosis of breast cancer using one or more of the above-mentioned techniques, ultrasound guided core needle biopsies followed by histological analysis of tumour biopsies are carried out and is used to confirm patient diagnosis and to detect the presence or absence of malignancy (Singh *et al.*, 2008; Becker, 2015).

Surgical procedures such as radical mastectomy, first introduced by Halsted in 1894 and the primary treatment option for breast cancer in the early 20th century (Halsted, 1894), that involve the complete resection of the breast together with draining lymphatic structures and the pectoralis muscle, have evolved substantially over the years. The initial approach that metastatic breast cancer can be cured solely with aggressive surgery has evolved and a less aggressive approach is now preferred when treating breast cancer patients (Halsted, 1894; Cotlar *et al.*, 2003). Supporting evidence from clinical data obtained from numerous clinical studies has indicated a reduced risk of relapse and improved overall patient survival when incorporating radiation, neoadjuvant or adjuvant chemotherapy in combination with surgical intervention as the standard of care for metastatic breast cancer (Anampa *et al.*, 2015).

The National Surgical Adjuvant Breast and Bowel Project (NSABP), that was the first randomised clinical trial that assessed the role of adjuvant chemotherapy in breast cancer, reported a significant decrease in local recurrence and secondary metastatic tumours when administering an alkylating agent, such as thiotepa, to pre-menopausal women who had undergone a radical mastectomy with positive lymph node involvement (Fisher *et al.*, 1968) The findings from the NSABP was further supported when similar results were observed in a second trial involving another alkylating agent, L-phenylalanine mustard (Fisher *et al.*, 1975). The implementation of polychemotherapy was first supported by a study that showed significant improvement in patient survival when alkylating agents such as cyclophosphamide and antimetabolite agents such as methotrexate and 5-fluorouracil were administered in combination to breast cancer patients (Bonadonna *et al.*, 1976). Several other studies, involving pre- or post-menopausal women who were either axillary lymph node positive or negative, provided further evidence of the significant benefits of combinational chemotherapy and clinical data from these individual studies collectively resulted in the recommendation by the US National Institute of Health Consensus Panel in 2001 that adjuvant polychemotherapy should be administered to breast cancer patients diagnosed with localised breast carcinomas, independent of lymph node involvement, menopausal status or tumour subtype (Mansour *et al.*, 1989; Fisher *et al.*, 1997; Mansour *et al.*, 1998; Abrams, 2001; Albain *et al.*, 2009).

Chemotherapeutic drug classes summarised in Figure 1.5, such as antimetabolites or enzyme inhibitors, DNA alkylators, anti-mitotic agents, immunologic agents, hormonal therapy and

ion modulators, are currently used as adjuvant therapy for breast cancer treatment (Abotaleb *et al.*, 2018).

Chemotherapeutic agents, both cell cycle and non-cell cycle specific drugs, are used to reduce localised and distant recurrence of the disease by targeting one of the four cell cycle phases, ultimately interfering with the mitotic mechanisms involved in irregular cell division (Malhotra and Perry, 2003; Cattley and Radinsky, 2004). A summary of the various chemotherapeutic agents including endocrine and hormonal targeted therapies is listed in Table 1.2. Breast cancer is one of the few cancer types where classification of heterogenous tumours with subsequent individualised treatment regimen design has resulted in a significant improvement in treatment response and patient survival (Higgins and Baselga, 2011; Perez *et al.*, 2011; Dai *et al.*, 2015).

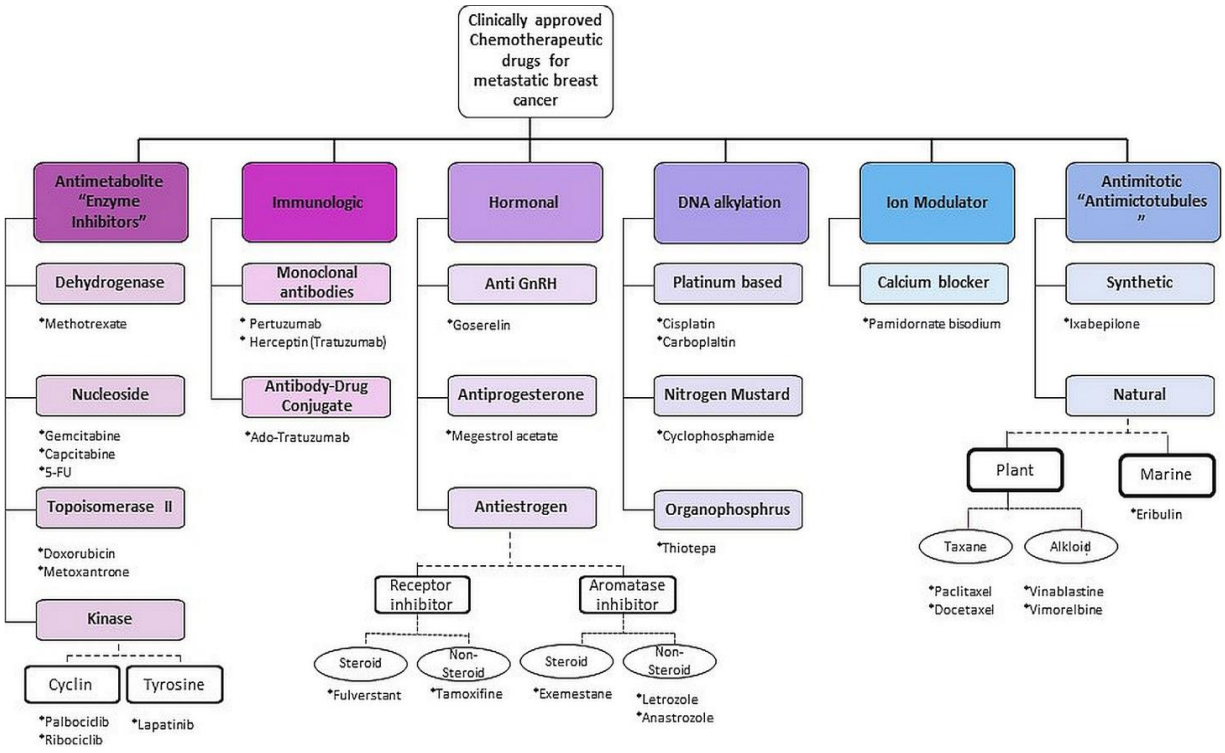

Figure 1.5. Summary of chemotherapeutic agents used for the treatment of metastatic breast cancer (Abotaleb *et al.*, 2018) (with permission).

Hormone receptor-expressing disease, which includes oestrogen positive tumours, respond to endocrine based therapy such as anti-oestrogen agents, ER antagonists and aromatase inhibitors (Jordan and Brodie, 2007). Clinical trials have provided evidence of a significant

decline in tumour recurrence and increased patient survival when tamoxifen, an ER antagonist, is administered to patients for a minimum of 5 years after initial treatment success, independent of involvement of lymph nodes, with significant improvement observed in only patients with ER subtype tumours (Baum, 1998; Levine *et al.*, 1998; Anampa *et al.*, 2015). Aromatase inhibitors (AIs), a class of agents that prevents the synthesis of oestrogen, have also been found to be more effective when given in combination with tamoxifen for extended adjuvant therapy in the treatment of postmenopausal women diagnosed with early stage breast cancer (Goss *et al.*, 2003; Perez, 2007; Hamadeh *et al.*, 2018).

Overexpression or amplification of the HER-2 oncogene has been observed in 25% of breast cancers thus making the HER family of receptors an appropriate drug target for breast tumours (Arteaga, 2002; Wolff *et al.*, 2013). Several studies have shown that the concurrent or sequential addition of monoclonal antibody treatment, such as trastuzumab that induces antibody-dependant cytotoxicity, to adjuvant chemotherapy is beneficial in lowering the risk of local or distant disease recurrence when administered to patients with tumours expressing elevated levels of HER-2 receptors (Piccart-Gebhart *et al.*, 2005; Romond *et al.*, 2005; Slamon *et al.*, 2011).

Despite the effectiveness of targeted therapies, a major challenge to the current treatment of breast cancer is primary and acquired drug resistance (Higgins and Baselga, 2011). Primary resistance can occur through the activation of compensatory pathways that continue to drive cancer cell proliferation and progression in the presence of an inhibitory cancer drug directed at a specific target or pathway (Serra *et al.*, 2011). Alternatively, underlying mechanisms for acquired drug resistance may include loss of target expression or mutations due to prolonged treatment courses (Berns *et al.*, 2007; Mittendorf *et al.*, 2009). Although combination chemotherapy improves overall patient treatment response, it is extremely important to investigate new ways to counteract the underlying mechanisms of drug resistance that threaten the efficacy of the current chemotherapeutic agents.

Moreover, although several chemotherapeutic agents have been clinically approved for breast cancer treatment, treatment failure is high and therefore continued research to identify novel drug targets and develop new cancer therapies is important (Higgins and Baselga, 2011; Rivenbark *et al.*, 2013; Thomas *et al.*, 2018; Schickli *et al.*, 2019).

Table 1.2. Summary of chemotherapeutic agents used in the treatment of cancer.

Drug class	Mechanism of action	Drug examples	References
Antimetabolites "Enzyme inhibitors"			
Anthracyclines	• Topoisomerase II inhibitor • DNA base pair intercalation	Doxorubicin; Epirubicin	(Sinha and Politi, 1990), (Arcamone, 2012)
Antimetabolite agents	• Inhibits DNA replication resulting in cell death	Capecitabine; Methotrexate	(O'shaughnessy *et al.*, 2005), (Walko and Lindley, 2005)
Poly ADP-ribose polymerase (PARP) inhibitors	• Inhibits PARP enzyme and DNA repair pathways	Talazoparib; Olaparib	(Yap *et al.*, 2011)
Cyclin dependent kinases	• Inhibits cell cycle CDK 4 and 6	Palbociclib; Abemaciclib; Ribociclib	(Kwapisz, 2017), (Poratti and Marzaro, 2019)
Tyrosine kinase inhibitors	• Compete for ATP-binding domain of protein kinases • Inhibits EGFR and HER-2	Lapatinib; Neratinib; Gefitinib	(Burstein *et al.*, 2010), (Segovia-Mendoza *et al.*, 2015)
Phosphatidylinositol -3- kinase (PI3K) inhibitors	• Inhibits PI3K and decreases cellular proliferation • Sensitises cancer cells anticancer therapies	Alpelisib	(Massacesi *et al.*, 2016), (Chen *et al.*, 2017)
Immunologic			
Monoclonal antibodies	• Binds to extracellular domain of HER- 2 receptor • Inhibits downstream tyrosine kinase signalling pathway	Trastuzumab; Pertuzumab	(Slamon *et al.*, 2011), (Hudis, 2007)
Programmed death receptor-1/ligand-1 (PD-1/PD-L1) inhibitor	• Blocks PD-1 receptors on surface of T-lymphocytes and tumour cell ligand PD-L1	Atezolizumab	(Gong *et al.*, 2018)

Hormonal			
Oestrogen-receptor antagonist	• Selective oestrogen receptor modulators	Tamoxifen; Raloxifene; Roremifene; Fulverstant	(Goss *et al.*, 2003), (Howell *et al.*, 2004), (Begam *et al.*, 2017)
Aromatase inhibitors	• Inhibits aromatase enzymatic step in oestrogen biosynthesis • Prevents oestrogen induced cellular proliferation	Letrozole; Anastrozole; Exemestone	(Miller *et al.*, 2008) (Rimawi *et al.*, 2018)
DNA Alkylation			
Alkylating agents	• DNA alkylation and cross-linking • DNA synthesis inhibition and cell death	Cyclophosphamide; Carboplatin	(Fleming, 1997; Huitema *et al.*, 2002; Lehmann-Che *et al.*, 2010)
Ion Modulator			
Calcium blocker	• Binds to hydroxyapatite crystals of bone which leads to reduction of bone resorption in breast cancer patients with bone metastasis	Pamidronate bisodium	(Glover *et al.*, 1994; Abotaleb *et al.*, 2018)
Antimitotic "Anti-microtubules"			
Vinca alkaloids	• Inhibits tubulin polymerisation • Disrupts microtubule function	Vinorelbine; Vinblastine	(Zelek *et al.*, 2001; Moudi *et al.*, 2013)
Antimicrotubular agents	• Disrupts and destabilises mitotic spindles and microtubules. • Apoptotic cell death due to G2/M cell cycle blockade	Docetaxel; Paclitaxel; Eribulin	(Jordan and Wilson, 1998; Mcgrogan *et al.*, 2008)

1.6 Tumour microenvironment

Tumours are considered to be separate "organs" that make use of normal biological processes associated with regular tissue development, to sustain their growth and progression. However, in contrast to normal organs, tumours are abnormal in structure as well as functionality. Tumours influence the body's biological processes and interact with surrounding organs, whereby the resulting systemic effects prove to be detrimental to the survival of an affected individual (Egeblad *et al.*, 2010a). Tumours are not only made up of irregular cell types but also a complex and dynamic tumour microenvironment. The 'seed and soil' theory, first proposed by Steven Paget in 1889, highlighted the pivotal role of the tumour microenvironment by stating that tumour growth and subsequent metastasis is a direct result of intricate interactions between tumour cells ("seeds") and an essential fertile niche such as the surrounding tumour microenvironment ("soil") (Paget, 1889; Fidler, 2003). Additionally, it has been suggested that the tumour microenvironment plays a significant role in drug resistance by impacting on chemotherapeutic drug delivery and distribution, usually resulting in a sub-optimal therapeutic response (Minchinton and Tannock, 2006; Trédan *et al.*, 2007).

1.7 The extracellular matrix (ECM)

The tumour microenvironment consists of a stromal compartment held in place by a supportive framework known as the ECM. The stroma comprises of glycosaminoglycans, growth factors, cytokines and a diverse number of cells which include mesenchymal supporting cells such as fibroblasts and adipocytes as well as vascular and immune cells (Egeblad *et al.*, 2010a). Stromal components are essential for normal mammary gland development but have been implicated in tumorigenesis (Wiseman and Werb, 2002).

The ECM, the non-cellular component of the tumour microenvironment as illustrated in Figure 1.6, forms an intricate network around cells and plays a major role in regulating and facilitating normal biological processes. In addition to providing a supporting scaffold for tissues, the ECM is an essential part of cellular biology by providing biochemical and biomechanical signals that facilitate cellular growth, survival, migration, differentiation, controls vascular development as well as influences immune function (Hynes, 2009; Lu *et al.*, 2012). The ECM in different tissue and cellular environments provides a specialised

pericellular compartment boundary known as the basement membrane, consisting of mainly type IV collagen, fibronectin and laminins.

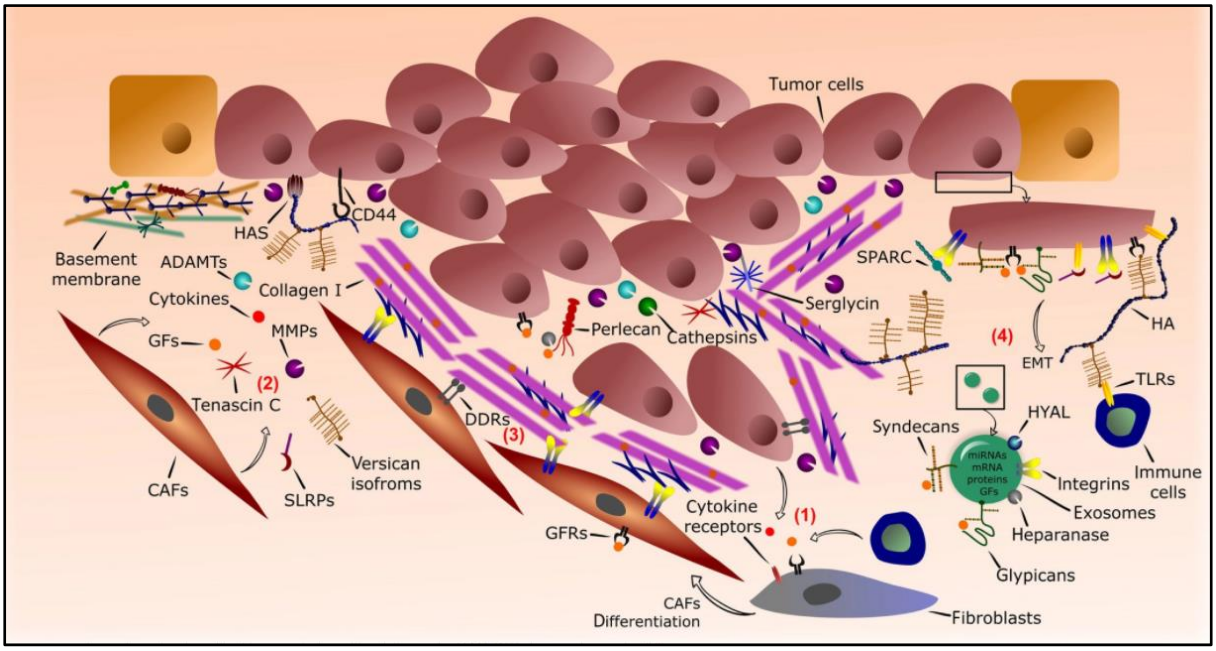

Figure 1.6. A diagram showing a typical tumour microenvironment consisting of ECM components. CAFs= cancer associated fibroblasts; DDRs= discoidin domain receptors; ECM= extracellular matrix; EGF= epidermal growth factor; GFs= growth factors; GFRs= growth factor receptors; HA= Hyaluronan; MMPs= matrix metalloproteinases; SLRPs= small leucine-rich proteoglycans; SPARC = secreted protein acidic and rich in cysteine; TLRs= toll-like receptors (Theocharis *et al.*, 2019) (with permission).

The basement membrane is specialised and different for specific tissue types fulfilling important biological roles as it serves as the boundary between epithelial, endothelial, fat, muscle and nerve cells in addition to providing structural and mechanical support. The interstitial matrix compartment of the ECM comprises of mostly fibrillar collagens, glycoproteins, proteoglycans, matricellular proteins, polysaccharides and specific ECM remodelling enzymes (Egeblad *et al.*, 2010b). Mechanisms of ECM function, as depicted in Figure 1.7, include blocking or promoting cellular migration whereby cellular adhesion and migration is facilitated through ECM associated receptors such as integrins and syndecans (Figure 1.7, stage 1-3). The ECM directs physiological influences and facilitates cell to cell communication by providing growth factor binding sites and promoting cell surface receptor interactions through signal transduction cascades that influence the regulation of gene expression and subsequent changes in cell behaviour (Figure 1.7, stage 4-6). Enzymatic remodelling of the ECM by matrix metalloproteinases (MMPs), results in modified protein

fragments having cell signalling and functional effects, as well as maintaining ECM homeostasis (Figure 1.7, Stage 7). Moreover, biomechanical characteristics of the ECM such as elasticity and stiffness influence cellular migration as well as altering cellular behaviour and function, in response to changes in biomechanical force of the ECM (Figure 1.7, stage 8).

Evidently the ECM constituents play a functional role in the normal development and homeostasis of breast tissue, however, studies have found that irregular levels of certain ECM components such as fibrillar collagens, remodelling enzymes and matricellular proteins contribute to the initial development of breast carcinomas, tumour progression and are implicated in drug therapy resistance (Hu *et al.*, 2017). Furthermore, alterations to the ECM which include disruption of the basement membrane and an increase in tissue stiffness have been implicated in cancer cell survival, invasion and proliferation as depicted in Figure 1.8. Although it is well known that genetic mutations in addition to environmental factors contribute to the occurrence of malignancy, it has been shown that the disruption of the ECM is extensively associated with the well-defined hallmarks of cancer (Pickup *et al.*, 2014). Consequently, the ECM has become a growing focus area for cancer research and a suitable target for anticancer therapeutics (Giussani *et al.*, 2015).

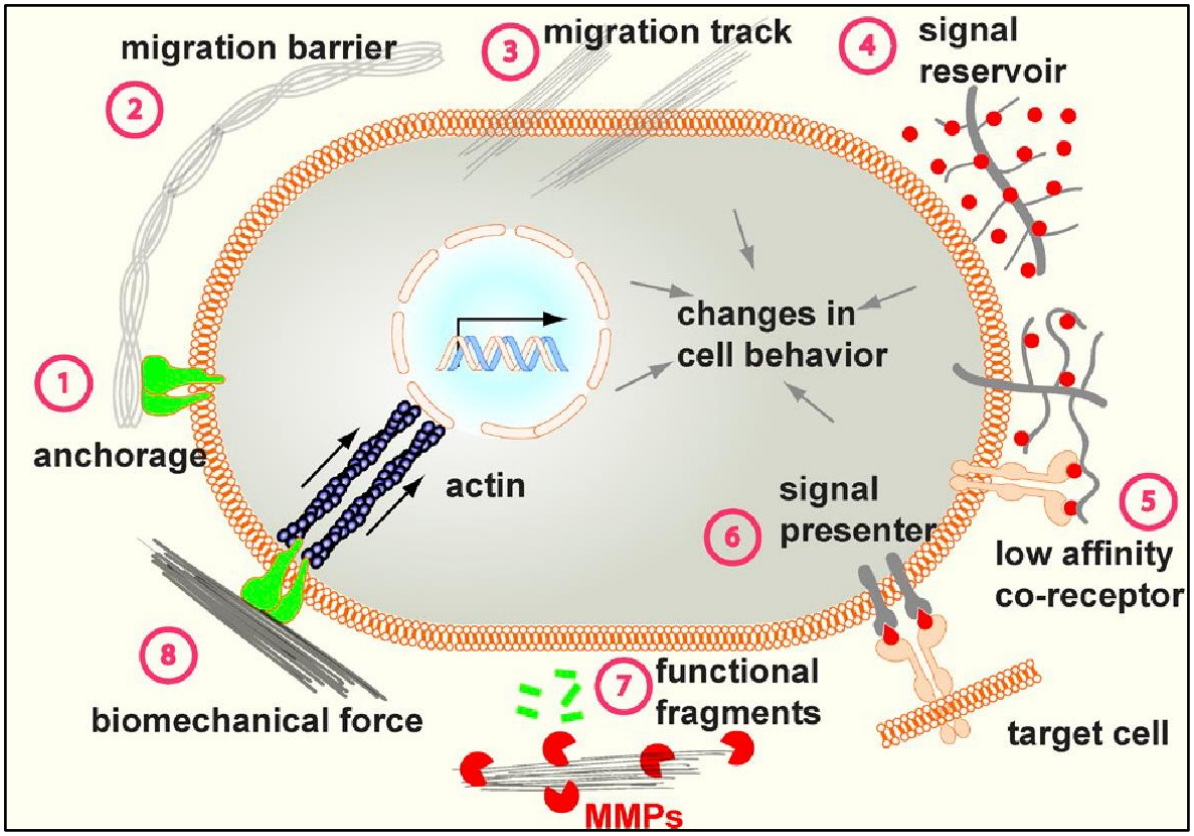

Figure 1.7. Mechanisms of ECM function. (Stage 1-3) block or facilitate cell migration. (Stage 4-6) Signal transduction involved in regulating gene transcription and cell behaviour. (Stage 7) ECM remodelling results in functional fragments involved in cell signalling and ECM homeostasis. (Stage 8) Changes in cell behaviour in response to biomechanical properties of the ECM (Lu *et al.*, 2012) (with permission).

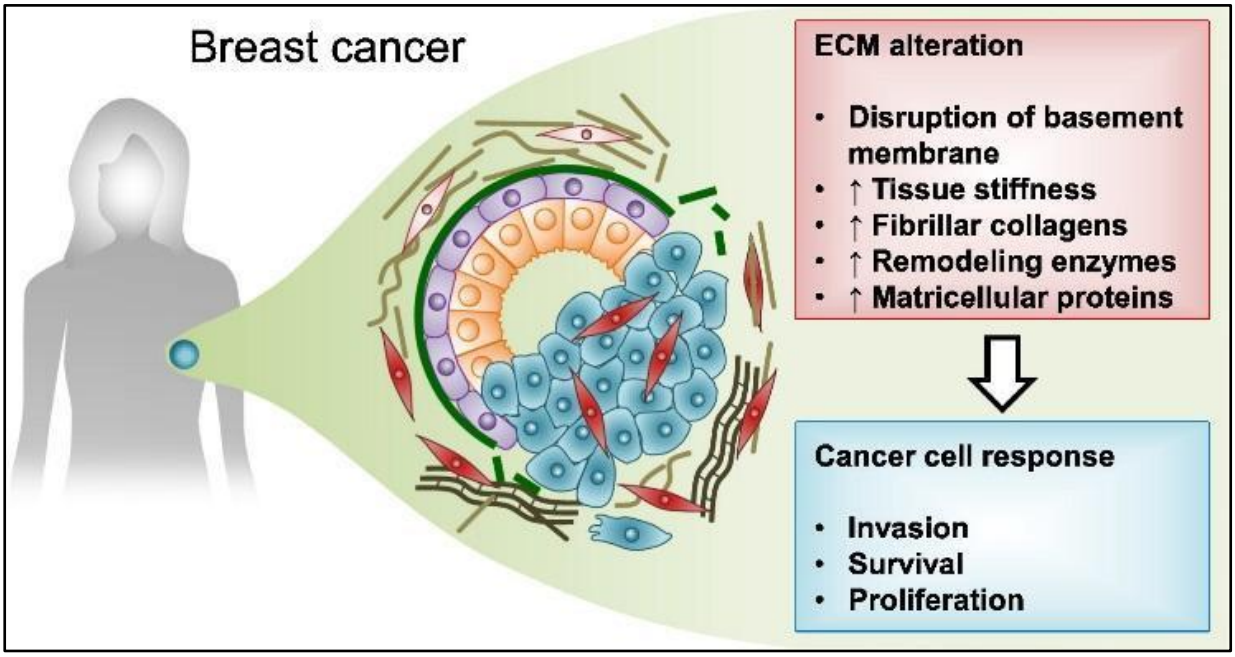

Figure 1.8. Alteration of the extracellular matrix (ECM) leads to cancer cell invasion, survival and proliferation (Insua-Rodríguez and Oskarsson, 2016) (with permission).

1.8 ECM components associated with breast cancer development and progression

1.8.1 Collagens

Collagens, which are synthesised by fibroblasts, are the most abundant fibrous protein constituents of the ECM and make up 30% of total human protein mass (Kielty and Grant, 2002). Collagens consist of three polypeptide alpha chains, with an amino acid sequence motif rich in proline and 4-hydroxyproline repeats and stabilising glycine residues, that fold into a triple stranded helix. Polypeptide chains form the building blocks of collagen fibrils, that combine to form fibres that again combine to form bundles that play a prominent role in maintaining the biological and structural components of the ECM (Ricard-Blum *et al.*, 2005; Kadler *et al.*, 2007). The superfamily of collagens, encoded by forty-two different genes, consists of twenty-eight (I–XXVIII) different collagens types which are further subdivided into four main subfamilies based on their supramolecular assembly of collagen fibrils and filaments. The different classes of collagens include fibril-forming collagens (I, II, III, V, XI, XXIV, XXVII), fibril-associated collagens (IX, XII, XIV, XVI, XIX, XX, XXI, XXII), network-forming collagens (IV, VIII, X) and membrane collagens (XIII, XVII, XXIII, XXV) (Ricard-Blum, 2011; Nielsen and Karsdal, 2016).

Various changes to the composition and structure of collagens are observed during breast cancer and studies have shown that specific fibrillar types of collagen, such as type I, III and V, are significantly increased with a possible association to tumour invasion and aggressive phenotypes (Kauppila *et al.*, 1998; Egeblad *et al.*, 2010b). Alternatively, decreased levels of non-fibrillar type IV collagen, which is an important structural component of the basement membrane, have been linked to an increased risk of metastasis. This is due to degradation of type IV collagen enabling the invasion of cancerous cells through the basement membrane to other sites or organs in the body (Duffy *et al.*, 2000). Collagens provide a scaffold for the cancer cells along which they can move while also significantly contributing to tissue density through increased collagen deposition and enzymatic crosslinking, a prominent clinical feature of advanced breast cancer. Increased tissue density and stiffness is associated with poor response to chemotherapy and lower survival rates (Provenzano *et al.*, 2009; Bonnans *et al.*, 2014). Furthermore, increased tissue density has been linked to a higher risk of developing breast cancer, more specifically the DCIS type, and promoting cancer progression.

Although there is an association between increased tissue density and breast carcinomas, the direct correlation between irregular collagen levels in particular and breast cancer progression or its potential as an indicator of poor prognosis must still be investigated (Gill *et al.*, 2006; Provenzano *et al.*, 2008; Barcus *et al.*, 2013).

1.8.2 Fibronectin

Fibronectin is a fibrous glycoprotein that is extensively involved in structural organisation, such as the formation and organisation of collagen fibrils of the ECM. Fibronectin is also involved in cell-matrix and cell-cell signalling that is specifically related to growth factor and cellular adhesion interactions in normal and tumorous breast tissue (Pankov and Yamada, 2002). Cancer-associated fibroblasts (CAF) show upregulated expression of fibronectin which is induced by surrounding tumour cells and certain primary tumour cytokines. Alterations in fibronectin levels in the ECM leads to enhanced tumorigenic properties such as increased tissue stiffness, viscosity and angiogenesis which is ultimately associated with the pre-metastatic tumour microenvironment (Kaplan *et al.*, 2005; Wang and Hielscher, 2017). Studies have identified the presence of increased levels of fibronectin in the ECM of breast tumours whereby irregular levels of fibronectin are associated with poor treatment outcomes and increased mortality risk for breast cancer patients, thus identifying fibronectin as a possible prognostic marker for breast cancer (Ioachim *et al.*, 2002; Bae *et al.*, 2013; Fernandez-Garcia *et al.*, 2014).

1.8.3 Laminins

Integrated networks consisting of collagens and laminins, which are interconnected by crosslinkers such as nidogen, are mostly found in the lamina dense layer of the basement membrane. Laminins, first discovered in mouse Engelbreth-Holm-Swarm sarcoma cells, are a family of glycoproteins that act as major constituents in the assembly of the basement membrane (Timpl *et al.*, 1979). Approximately sixteen known laminins, consisting of various combinations of alpha, beta and gamma chains, interact with numerous molecules of the ECM to facilitate cellular adhesion and migration (Tzu and Marinkovich, 2008). Irregular levels of laminins have been observed in lung, colon and squamous cell cancers thus characterising

laminins as potential diagnostic or clinical signatures for cancer classification (Määttä *et al.*, 1999; Katayama and Sekiguchi, 2004). A number of specific laminins such as laminin-332 and laminin-511 have been implicated in breast cancer development and progression (Zahir *et al.*, 2003; Kwon *et al.*, 2012) where increased laminin-322 expression has been found in aggressive types of breast cancer involving cell migration and invasion through its interaction with alpha-3 integrin (Carpenter *et al.*, 2009). Furthermore, laminin-322 is associated with cancer cell invasion by promoting the transition of epithelial cells exhibiting increased motility into mesenchymal cells (Kim *et al.*, 2011). Studies have also shown that neoplastic cellular adhesion, migration and invasion are all promoted through the interaction of laminin-511 with cell surface integrin receptors (Chia *et al.*, 2007).

1.8.4 Glycosaminoglycans and proteoglycans

Glycosaminoglycans (GAGs) are carbohydrate-based polymers produced by fibroblasts and deposited into the ECM where they may be combined with core proteins to form proteoglycans, which are covalently protein-bound GAGs. Both GAGs and proteoglycans are extensively involved in providing mechanical support and water retention for tissue. Hyaluronan is a GAG primarily involved in tissue repair and inflammation. Hyaluronan is considered a potential biological marker for breast cancer as studies have shown that serum hyaluronan levels are significantly increased in patients diagnosed with metastatic breast cancer (Delpech *et al.*, 1990; Karousou *et al.*, 2014). Increased concentrations of a number of proteoglycans, such as versican, decorin, lumican and syndecan have also been implicated in breast cancer development and progression (Eshchenko *et al.*, 2007; Kischel *et al.*, 2010). One study reported a correlation between irregular levels of versican and an increased relapse rate in patients diagnosed with node-negative primary breast cancer. It was suggested that fibroblast deposition of versican into the peritumoral stroma and subsequent ECM remodelling promoted both local cancer cell invasion and metastasis (Ricciardelli *et al.*, 2002).

1.8.5 Matricellular proteins

Cell motility and invasive features of cancer cells are promoted by the expression of matricellular proteins. High levels of these proteins have been found in breast tumours and

have been linked to inducing tumorigenic signalling associated with metastasis (Chong *et al.*, 2012). Increased tenascin C (TNC) expression levels is associated with breast cancers with some studies suggesting that this matricellular protein can potentially be used as a prognostic biomarker for breast cancer (Ishihara *et al.*, 1995; Ioachim *et al.*, 2002). Increased levels of periostin, another matricellular protein, and the associated overexpression of vascular endothelial growth factor receptor 2 (VEGFR-2) promoted primary tumour growth in a xenograft tumour study (Shao *et al.*, 2004). Moreover, expression of periostin was found to be significantly higher in breast tumours than in normal human breast tissue, with some studies proposing that overexpression of periostin may promote metastasis (Zhang *et al.*, 2010; Xu *et al.*, 2012; Wang *et al.*, 2013).

1.8.6 Matrix metalloproteinases

Matrix metalloproteinases (MMPs) are enzymes involved in remodelling ECM where the resulting changes in ECM may allow or enhance cancer cell invasion and migration. This association implicates MMPs in breast cancer progression and metastasis. There are a number of MMPs each with a different protease selectivity, but overexpression of MMP-2, 3, 9 and 14 have been reported as the most prevalent in breast cancers (Schedin, 2006; Oskarsson, 2013). MMP-2 and MMP-9 are involved in degradation of collagen IV facilitating invasion of cancer cells through the basement membrane. Studies have also shown that high plasma levels and expression of MMP-2 in breast tumours are indicative of poor clinical prognosis in patients suffering from node positive breast carcinoma (Duffy *et al.*, 2000; Leppä *et al.*, 2004). Figure 1.9 summarises some changes in ECM components that are involved in both breast cancer development and metastasis.

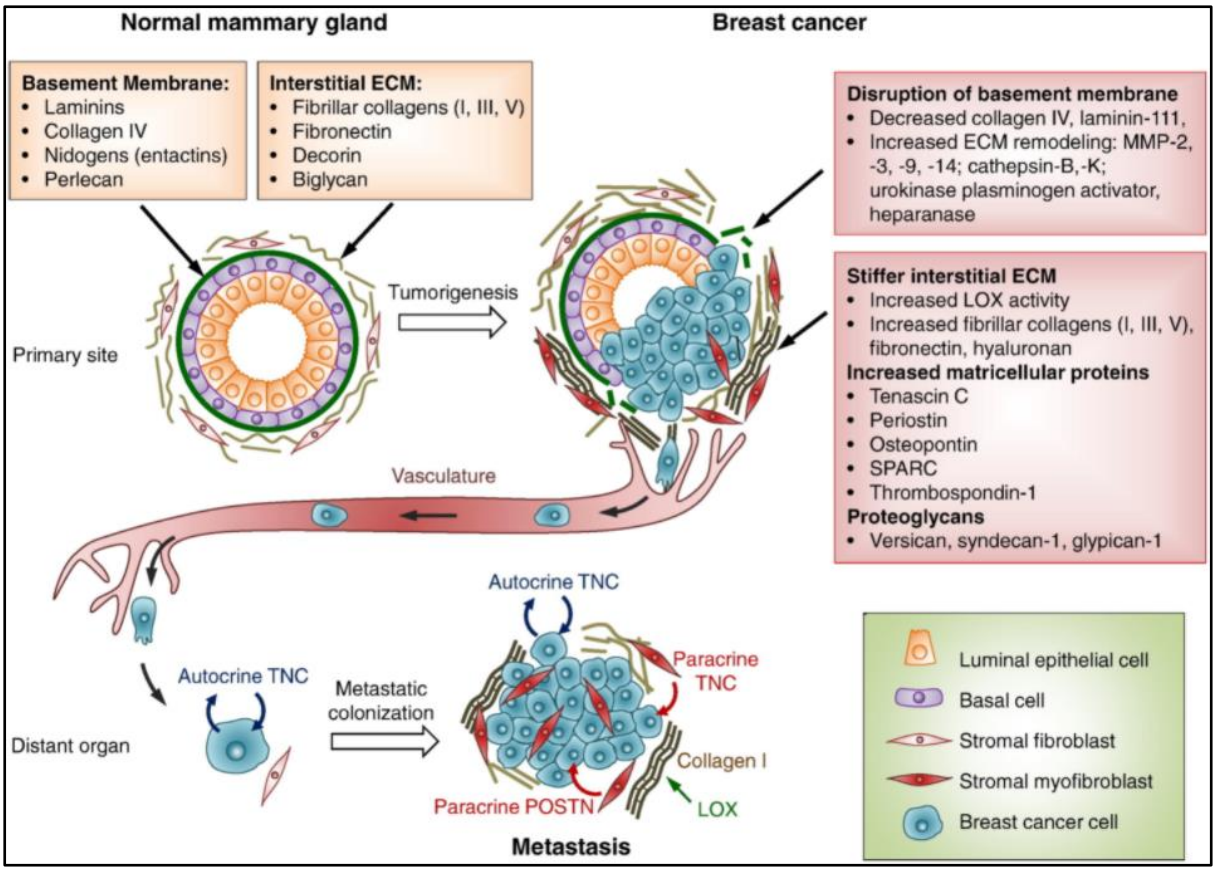

Figure 1.9. A summary of the main components of the ECM in normal mammary glandular tissue and components associated with the dysregulation of the extracellular matrix (ECM). Disruption of the basement membrane is caused by decreased levels of collagen IV and an increase in certain matrix metalloproteinases (MMPs) such as MMP 2, 3, 9 and 14. Stiffness of the interstitial ECM, which leads to tumour aggressiveness, occurs as a result of collagen and fibronectin deposition as well as increased collagen crosslinking by lysyl oxidase (LOX). Increased expression of matricellular proteins and proteoglycans also contributes to breast cancer development and metastasis (Insua-Rodríguez and Oskarsson, 2016) (with permission).

1.9 Extracellular matrix targeted research

Dysregulation of the ECM is a major contributor to the development and progression of malignancy (Pupa *et al.*, 2002). Consequently, monitoring of ECM components in cancer patients can prove useful in facilitating personalised therapy as the presence of various ECM components can be used to identify individuals who appear to be at risk of developing an aggressive form of cancer (Bergamaschi *et al.*, 2008). Moreover, tumours that show early signs of highly invasive behaviour can be identified through ECM biomarkers, and treated accordingly, ultimately leading to improved prognosis (Giussani *et al.*, 2015). However, in order to achieve this, a thorough understanding of the ECM and the cancer associated changes of the components are essential. Achieving this understanding requires that effective

methods to characterise the ECM from human breast tumours must be developed and optimised then used to identify and quantitate the various constituents found within the ECM.

Due to the distinct role of the ECM in breast cancer, the ECM has also been identified as a viable therapeutic target for both chemical and biological drug compounds (Insua-Rodríguez and Oskarsson, 2016). The use of drug compounds to inhibit over-production of relevant ECM components has been extensively investigated in pre-clinical studies and results suggest that this can potentially be achieved in clinical practice (Yasuo Kunugiza *et al.*, 2011; Urakawa *et al.*, 2012). Another avenue of anticancer drug research linked to the ECM involves the use of ECM components to concentrate tumour suppressive drugs or to absorb radioactivity within specific regions of tumours, thus limiting the exposure of normal tissues. However, the success of this would be dependent on identifying ECM components that are expressed at levels significantly higher in tumours than in normal tissues (Insua-Rodríguez and Oskarsson, 2016). Matricellular proteins such as TNC, which is found at relatively higher levels in tumours than in healthy tissues, would thus be a feasible target type for this type of approach. Studies have demonstrated an improved selectivity of radiation in xenograft tumours by making use of radio-labelled anti-TNC antibodies in colon cancer and glioblastoma models (Petronzelli *et al.*, 2005; De Santis *et al.*, 2006; Lingasamy *et al.*, 2019). Communication between cell membrane associated integrins and ECM components facilitates a number of signalling processes that involve cancer cell proliferation and migration, suggesting that the disruption of these ECM-receptor interactions could be a useful therapeutic strategy. Studies have shown that when integrin β1 was targeted with an inhibitory antibody in a human cancer cell line, it resulted in decreased cell proliferation and initiated apoptosis (Varner and Cheresh, 1996; Desgrosellier and Cheresh, 2010). It was previously discussed that mammary tissue density is indicative of poor prognosis in breast cancer patients and that ECM remodelling along with collagen crosslinking contribute to the ECM changes that induce breast cancer progression and metastasis. Significant advances have been made in the development of lysyl oxidase (LOX) inhibitors, an enzyme involved in collagen crosslinking, whereby studies have shown a decrease in primary tumour growth and metastasis with the use of LOX inhibitor compounds (Levental *et al.*, 2009). Several potential therapeutic points in ECM synthesis or

modification as well as ECM receptor targets are summarised in Figure 1.10. Drugs effectively targeting these sites could prove to be beneficial in new and effective anticancer therapies.

Besides a prognostic and therapeutic role, the characterisation of the ECM can play a significant role in cancer research, as with the ECM derived from decellularized tissue commonly used for tissue engineering or three-dimensional ECM scaffolds for *in vitro* culturing (Crapo *et al.*, 2011). The current drug discovery and development pipeline for cancer drugs remains a wasteful process involving exorbitant amount of funding that may or may not result in the investigational compound reaching the market (Moreno and Pearson, 2013). Certain pre-clinical cancer models are poor representations of the actual *in vivo* pathological processes associated with cancers and or predictors of drug efficacy and safety which is one of the main obstacles during cancer drug development (Johnson *et al.*, 2001). Pre-clinical models tend to ignore or exclude the tumour microenvironment and therefore poorly represents the true nature of tumour biology. This impacts negatively on maintaining pre-clinical drug efficacy during clinical studies (Moreno and Pearson, 2013). Two-dimensional (2D) or monolayer cell culturing methods using either primary cell cultures or established cell lines are commonly used for pre-clinical drug screening and mechanistic studies. However, 2D models are limited by the fact that generally only one cell type associated with a tumour type is represented and where the microenvironment has no ECM, and very limited cell-contact signalling, therefore failing to exhibit typical *in vivo* characteristics of a solid tumour (Unger *et al.*, 2015).

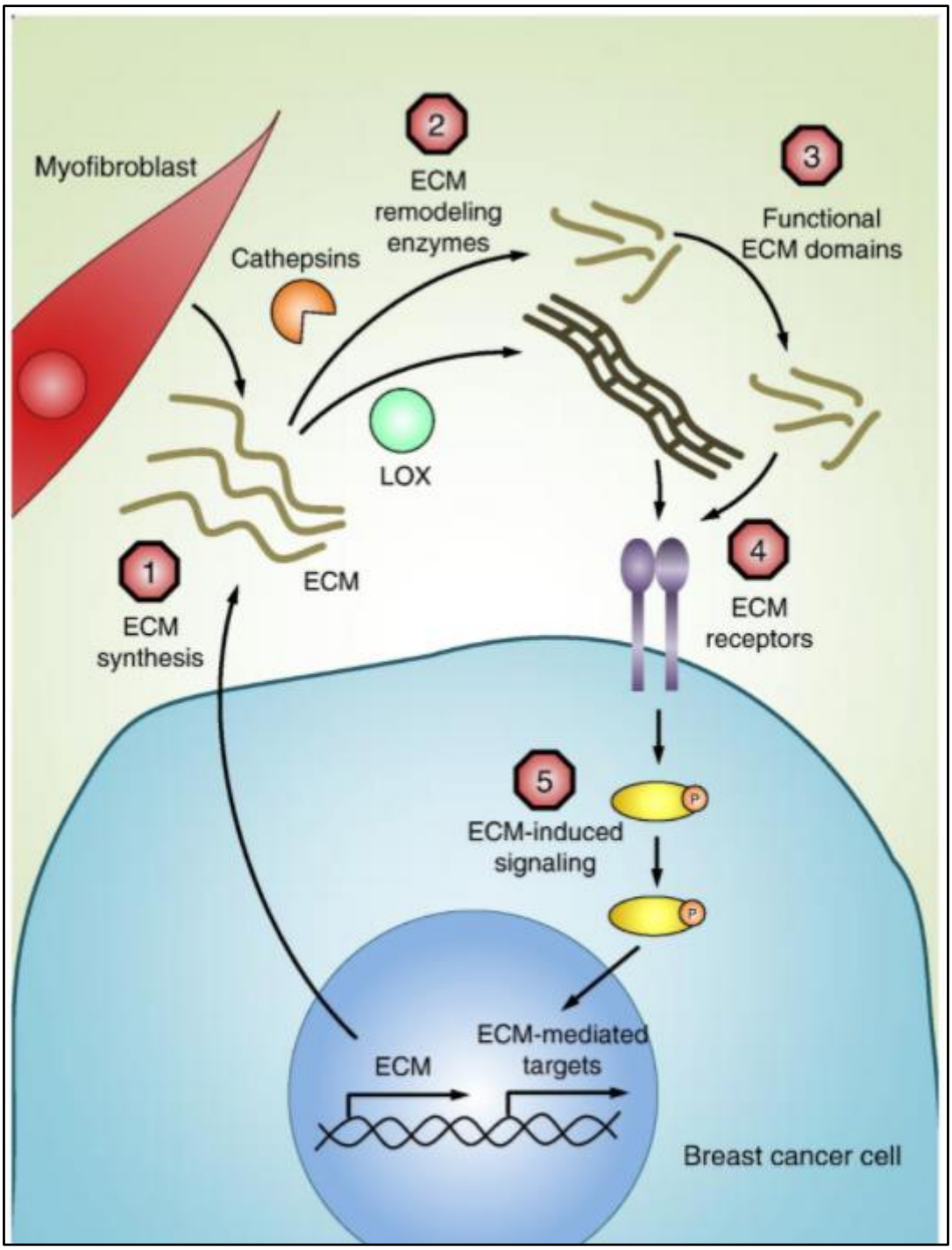

Figure 1.10. A diagram showing potential therapeutic drug targets associated with cancer cell associated extracellular matrix (ECM). 1) Inhibiting the synthesis of ECM components by cancer associated fibroblasts and breast cancer cells. 2) Small molecules can be used to inhibit ECM remodelling enzymes such as matrix metalloproteases and lysyl oxidase (LOX). 3) Use of labelling techniques of cancer specific ECM components to selectively accumulate radiation and to improve drug delivery to tumours. 4) Use of antibodies or peptides to block ECM-receptor interactions. 5) Kinase or phosphatase inhibitors can be utilised to target intracellular signalling pathways that are prompted by ECM components (Insua-Rodríguez and Oskarsson, 2016) (with permission).

Multicellular tumour spheroids are grown from supporting stromal cells and tumour cell lines to resemble the tumour and its microenvironment as closely as possible (Thoma *et al.*, 2014). Although spheroids are currently being used for drug screening and efficacy studies, it is yet to be established whether data and downstream clinical predictions obtained from 3D models are more accurate than the standard 2D culture system. Moreover, tumour spheroids fail to represent the complex histological features of tumours and are unable to capture the intra- and inter- tumour heterogeneity (Unger *et al.*, 2015). In an attempt to address the limitations of *in vitro* and *ex vivo* cancer models, Majumder *et al.* carried out a study that involved the development of an *ex vivo* tumour ecosystem for head and neck squamous carcinomas (Majumder *et al.*, 2015). Tumour sections obtained from patients with head and neck carcinomas were cultured in culture plate wells that were coated with tumour matched extracellular matrices. Drug efficacy assays were then carried out directly on these cultured tumours. Processed data was used to successfully predict drug efficacy in patients. The developed model was found to be robust and reliable thus indicating its potential to predict possible clinical outcomes to new anticancer therapy and facilitate the personalisation of cancer treatment. A critical factor claimed to be behind the success of the study was the characterisation of the ECM of resected tumours, followed by the use of ECM protein ratios that matched that of the tumour which ensured optimal growth of the resected tumours (Majumder *et al.*, 2015). The study illustrated an additional advantage of isolating and characterising the ECM of tumours for drug efficacy studies.

1.10 Mass spectrometry

1.10.1 Mass spectrometry-based proteomics

The field of proteomics can be broadly defined as the use of biotechnology and various scientific disciplines to study the complete protein complement expressed by a cell, tissue or organism also known as the proteome. The term ‘proteomics’ was initially formalised in 1995 by Prof. Marc Wilkins and since then has become a distinctive and resourceful area of research (Wasinger *et al.*, 1995; Wilkins *et al.*, 1996; Graves and Haystead, 2002; Aslam *et al.*, 2017). Proteomics involves various assay types including immunology-based assays, crystallographic assays and mass spectrometer-based assays in applications that includes

qualitative and quantitative assays, structural and functional proteomics, spatial distribution and dynamics, profiling of protein expression under different conditions, protein-protein interaction studies, post-translational modification studies and many more. A particular avenue of research where the role of proteomics has been significantly established is that of drug discovery and drug target identification and validation. Exploitation of the structural, chemical and physical properties of proteins can be utilised in experimental techniques that are used for protein separation prior to mass spectrometry based proteomics (Graves and Haystead, 2002).

An example of protein separation techniques includes sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), an essential technique of protein molecular size analysis for many years and still widely used for successfully resolving complex protein mixtures prior to protein identification and characterisation (Laemmli, 1970; Dunn, 1993). Gel electrophoresis was a critical step in the progression of protein studies, however major limitations include the lack of sensitivity, poor dynamic range of concentrations that can be visualised, poor resolution of proteins of similar molecular mass and the inability to directly identify proteins represented as single bands on gels (Dunn, 1993; Garfin, 2009).

Due to a number of analytical constraints within the proteomics field, there was an increased demand for more sensitive and selective analytical tools for protein studies and this essentially resulted in significant developments and improvements in mass spectrometry-based proteomics. The mass spectrometer (MS), which was invented in the early 20th century and was initially restricted to small molecules, has been extensively developed since then and is currently the leading analytical instrumentation for targeted protein identification and characterisation. MS provides the required sensitivity, reproducibility and mass accuracy in addition to the high through-put required for current proteomic applications (Angel *et al.*, 2012; Aslam *et al.*, 2017). Proteomic workflows that include gel electrophoresis followed by mass spectrometry is still a very useful tool in the field of proteomics (Paulo, 2016; Kim and Cho, 2019).

Shotgun proteomics, also referred to as bottom-up proteomics, is currently the preferred method for large scale proteomic studies and involves the characterisation of proteins that have been hydrolysed into peptides through enzymatic proteolysis (Yates, 1998) followed by chromatographic separation with final amino acid sequencing of these individually separated

peptides by tandem mass spectrometric fragmentation. An overview of a typical shotgun MS-based proteomic workflow is illustrated in Figure 1.12 and briefly summaries the steps involved in a bottom-up proteomic experiment whereby proteins, which are extracted from a biological sample of interest, are proteolytically cleaved into peptides using enzymes such as trypsin. The resulting peptide mixture is then separated using reverse phase (RP) chromatography and analysed using LC-MS/MS. This involves peptide separation on a non-polar (hydrophobic) stationary phase (column) of hydrocarbon chain (e.g. C18) bonded silica particles and a mobile phase made up of a mixture of water and acetonitrile. Peptides with increasing hydrophobicity are eluted from the column due to the stationary phase chemistry and the mobile phase gradient. An adaption of the use of reverse phase chromatography that increases the number of identified proteins is the application of two-dimensional (2-D) column chromatography, as used in this study. This method involves two completely separate chromatographic separations, an initial enrichment and pre-fractionation of sample peptides on a C18 column at a high-pH (alkaline) for the first dimension of separation, followed by a second separation at a low-pH (acidic) (Mcqueen and Krokhin, 2012). Two-dimensional RP chromatography is widely used in proteomic experiments due to its improved resolution of peptides before introduction into the MS, ultimately producing high quality and extensive analytical data. Protein identification is subsequently determined by comparing the acquired peptide mass spectra to theoretical peptide mass spectra generated from a protein sequence database (Dupree *et al.*, 2020). A significant advance in MS technology involved the development of soft ionisation techniques, such as electrospray ionisation (ESI), which has allowed for the analysis of large biomolecules such as peptides (Yamashita and Fenn, 1984). Following elution of separated peptides from a reverse phase column, peptides are subjected to a high voltage that is used to convert peptides in solution to ions in a gaseous phase. Peptide ions are mass separated in the first, usually a quadrupole, mass analyser by scanning through mass-to-charge (m/z) values and intensities of peptides are captured as the MS1 spectra. Selected peptide ions, also referred to as precursor ions, are then fragmented in the collision cell of the mass spectrometer, and the resulting combination of peptide fragment ions are scanned and referred to as the MS2 or MS/MS spectra. The compilation of the m/z values and intensities of all fragment ions from a single precursor ion, is used to identify the resulting amino acid sequence from the selected precursor peptide (Ho *et al.*, 2003). The amino acid sequence is determined from the combination of the masses of the “y” and “b”

ions that form from the peptides during random collision induced fragmentation in the collision cell or through high-energy collisional dissociation on Orbitrap type instruments with this feature. The experimental fragment ion spectra are interpreted to give the amino acid sequence that is compared to a database of protein sequences from which protein identities can be inferred using bioinformatic software. Protein inference is determined by software algorithms that provide a score which ultimately provides an indication of the confidence of the inferred protein identity. A core feature of the analytical software used for peptide identification involves calculating a value that measures the quality between an acquired MS/MS spectrum and a theoretical spectrum also referred to as a peptide-spectrum match (PSM). Every MS/MS spectrum in a particular data set is matched to the chosen protein database and the highest scoring PSM is determined (Nesvizhskii, 2010; Hubler *et al.*, 2019). However, false positive identities of proteins from peptide PSMs may occur when carrying out protein database searches (Patterson, 2003; Aggarwal and Yadav, 2016; Bogdanow *et al.*, 2016). The false discovery rate (FDR) is defined as the ratio between the number of false or incorrect PSMs and the total number of PSMs (Aggarwal and Yadav, 2016). The FDR approach, first proposed by Benjamini and Hochberg, is widely used for multiple hypothesis testing to correct for multiple comparisons in proteomic experiments (Benjamini and Hochberg, 1995). Target-decoy search strategy is commonly used to assess the FDR in MS-based proteomic experiments by providing a simple yet effective way to provide an estimation of false positive discovery rate (Elias and Gygi, 2007). Briefly, decoy sequences are constructed either by reversing, shuffling or randomising the protein search database and then linked to the original unaltered protein database consisting of target sequences. The combined target-decoy sequences are then submitted to search engines and the number of false positives may be determined based on the number of matches to the decoy sequences. A score threshold based on a predetermined FDR can then be used to reduce the number of false identifications in the dataset. Therefore, the target-decoy search strategy not only allows for the estimation of false positives but also provides useful information when considering the filtering criteria used to accurately distinguish between true and false PSMs (Elias and Gygi, 2010; Levitsky *et al.*, 2017).

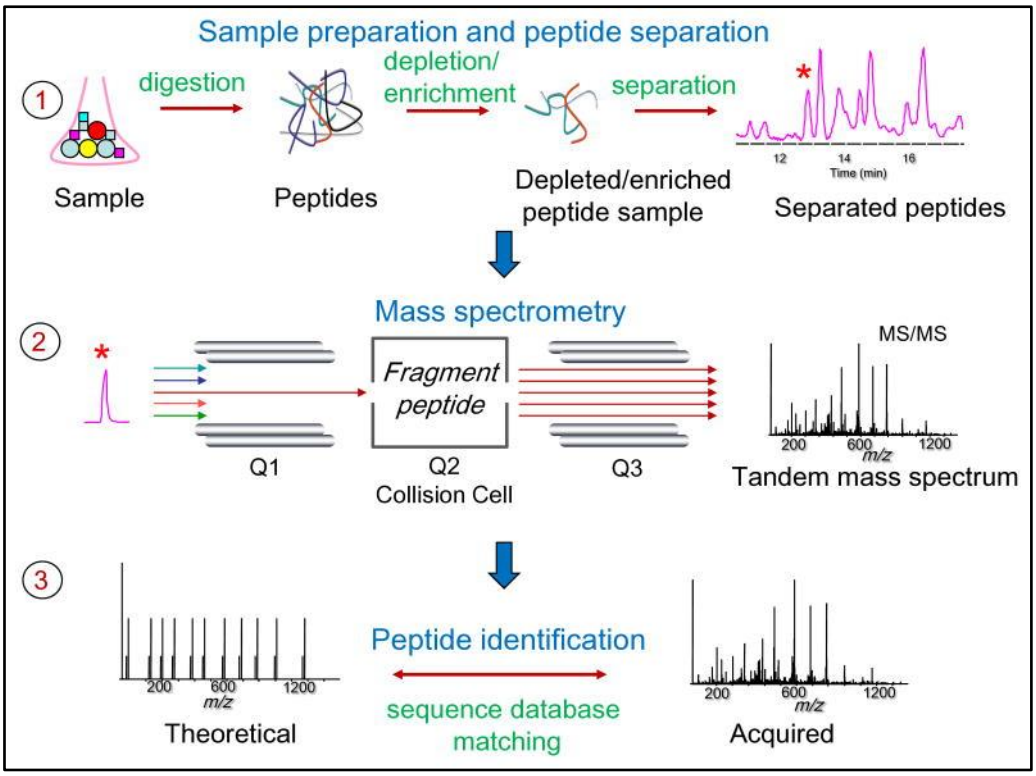

Figure 1.12. An overview of a shotgun proteomic mass spectrometry workflow involving (1) sample preparation and peptide separation, (2) mass spectrometry and (3) peptide identification (Nesvizhskii, 2010) (with permission).

1.10.2 Mass spectrometry data acquisition methods

The type of MS data acquisition method is determined by the manner in which precursor (MS1) and product ions (MS2) are scanned by the mass spectrometer as shown in Figure 1.13 (Hu *et al.*, 2016; Sinha and Mann, 2020). Multiple reaction monitoring (MRM), also referred to as selected reaction monitoring (SRM), and parallel reaction monitoring (PRM) are common strategies used in targeted proteomic experiments to analyse specified proteins of interest. For both MRM and PRM, a single peptide precursor ion is selected in an MS1 scan and fragmented, however during MS2 scans only prominent resulting fragment ions are selected and detected in MRM acquisition mode whereas all fragment ions are measured for PRM (Hu *et al.*, 2016). Precursors and fragment ions are selected based on m/z and elution times or signal intensities that have been predefined by the user and correspond to most intense or dominant peptides eluting at any specific retention time. Specific selection of a peptide precursor ion and corresponding post-fragmentation ion, referred to as a transition, allows for extremely high sensitivity and specificity as well as reproducible and accurate protein quantification when using this type of targeted acquisition methods such as MRM and

PRM. However potential drawbacks of targeted acquisition strategies may include the need for prior knowledge of a sample and time-consuming method optimisation for each transition specific to a protein of interest (Borràs and Sabidó, 2017).

Data-dependent acquisition (DDA), commonly used for discovery-based proteomics, is characterised by the MS instrument sequentially selecting and scanning precursor ions based on their abundance within predefined mass and charge state ranges (Goldfarb *et al.*, 2016). This approach does not require any prior knowledge of the sample source nor input by the operator, however the automatic selection of precursors in DDA mode may lead to irreproducibility and signal variability. This is due to insufficient scanning time of precursor ions across replicates and the under-sampling of low abundant peptide species in the presence of highly abundant where more intense peptide species may generate missing values for low abundance peptides that may be of particular interest. However, despite these disadvantages, DDA is still the most commonly used data acquisition method for bottom-up proteomics (Pino *et al.*, 2020).

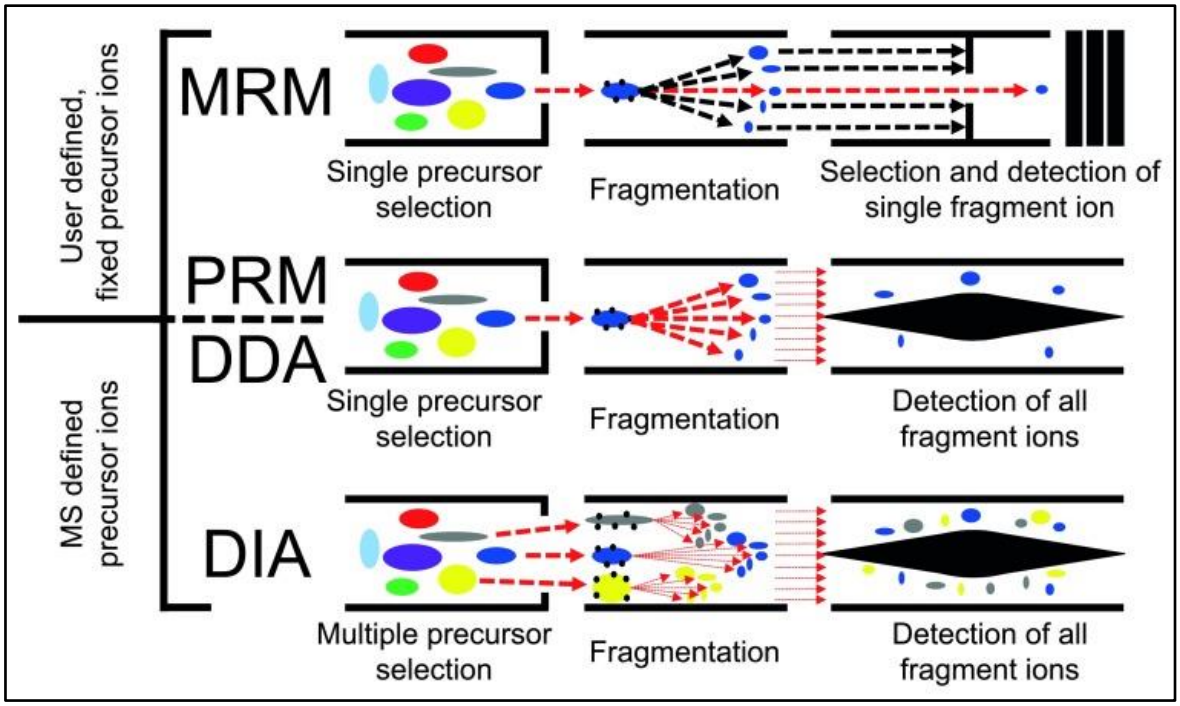

Figure 1.13. A schematic representation of the different LC-MS/MS acquisition modes at ion fragment level, multiple reaction monitoring (MRM), parallel reaction monitoring (PRM), data-dependent acquisition (DDA) or data-independent acquisition (DIA), that are used in mass spectrometry-based proteomics (Hu *et al.*, 2016) (with permission).

Data-independent acquisition (DIA) and Sequential Windowed Acquisition of all Theoretical Fragment Ion Mass Spectra (SWATH) methodologies combines the extent of protein identification that is attainable with DDA and the reputed accuracy and reproducibility that is characteristic of MRM and PRM, thus making it a suitable option for large scale, discovery based bottom-up proteomic experiments (Hu *et al.*, 2016). DIA involves the division of the entire m/z range into sequentially small precursor isolation windows, whereby the MS simultaneously fragments multiple precursor ions within the specified mass range as shown in Figure 1.13. This results in a more complex but complete fragment ion spectrum, resulting in fewer missing values compared to DDA where only the most intense precursors selected for MS/MS (Ludwig *et al.*, 2018). The fragment ions are then detected as complex MS coupled to MS/MS spectra whereby specific DIA/SWATH software packages are needed to deconvolute and assign the multiplex fragment ion spectra to peptides and search the SWATH data against public or project specific spectral libraries for protein identification. A spectral library, generated with one or two dimensional (2-D) RP chromatography and initial DDA runs, is the most common method for targeted SWATH data extraction. A spectral library can be described as a spectral map consisting of precursors, product ions with associated ion intensities and retention times against which the SWATH data is searched (Gillet *et al.*, 2012; Ludwig *et al.*, 2018). Instrument parameters such as the precursor isolation window width, fragment and precursor ion accumulation time and the chromatographic cycle time are all interrelated and affect both the selectivity and sensitivity of a SWATH-MS base method (Ludwig *et al.*, 2018). Therefore, it is imperative to determine optimum settings and find a balance between each acquisition parameter to ensure robust and reliable SWATH data (Schilling *et al.*, 2017). The precursor isolation window width refers to the mass range in which comprehensive ion fragment data is acquired during the MS/MS scan, therefore influencing selectivity and dynamic range of the method. Narrow isolation windows, when compared to wider isolation windows, may be beneficial in terms of improving ion efficiency by limiting the number of peptide precursor masses that are selected, however, it may also lead to extended cycle time resulting in fewer data points per peak which is then detrimental to the accuracy of quantitation (Amodei *et al.*, 2019). Cycle time is determined by the total sum of accumulation times for all the MS and MS/MS scans and refers to the number of data points captured for a chromatographic peak. More than 10 data points are required for chromatographic peak reconstruction and to obtain reliable SWATH-based protein

quantification (Schilling *et al.*, 2017). Although DIA/SWATH MS data analysis is highly complex, this systematic MS acquisition method has become a widely used approach for extensive quantitative proteomics in cancer research and biomarker studies due to its ability to provide, comprehensive, high quality and reproducible data (Jylhä *et al.*, 2018; Koopmans *et al.*, 2018; Pino *et al.*, 2020).

1.10.3 Protein quantification techniques

Protein quantification strategies may be categorised into label-based or label-free approaches (Sinha and Mann, 2020). Label-based techniques utilise metabolic or chemical labelling of protein or peptides. Stable isotope labelling of amino acids in cell culture (SILAC) is an example of metabolic labelling whereby isotopically labelled amino acids (light or heavy arginine and/or lysine) are incorporated into cell culture medium and are subsequently taken up by growing cells (Chen *et al.*, 2015). Cell lysates can be pooled and proteins from each lysate are distinguished by LC-MS/MS which is able to determine the mass difference as a result of differential isotope labelling. Furthermore, relative protein abundance between samples can be determined based on the intensities of the heavy and light labelled peptides (Wang *et al.*, 2018). Technical variation is significantly minimised in SILAC experiments due to the fact that labelled samples can be pooled together at the beginning of the experimental workflow and can be carried through each step of the sample preparation process at the same time ultimately resulting in increased robustness and quantitative accuracy (Zhang and Neubert, 2009; Deng *et al.*, 2019). Isobaric tags for relative and absolute quantification (iTRAQ) and tandem mass tags (TMT) are widely used for relative MS protein quantitation and involve chemical labelling at the peptide level, allowing for multiple samples to be pooled together which are then subsequently analysed using LC-MS/MS (Thompson *et al.*, 2003; Wiese *et al.*, 2007). Briefly, tagged peptides which are identical but originate from different samples elute from the chromatographic column at the same time and are analysed by the MS. The tags are released during fragmentation and their respective signal intensities are measured and correlated back to the individual samples that were initially pooled together thus allowing for relative protein quantification between the original sample sources. Advantages of chemical labelling techniques such as iTRAQ and TMT also include limited

technical variation as well as the ability to analyse multiple biological samples simultaneously (Jylhä *et al.*, 2018; Zecha *et al.*, 2019).

Spectral counting and area under the curve (AUC) are two categories for MS label-free quantification (Nahnsen *et al.*, 2013). Spectral counting determines relative protein quantification by counting the number of identified MS/MS spectra of each of the peptides from the same protein found in the different samples being compared. Spectral counting is based on the principle that an increase in protein abundance leads to a greater number of proteolytic peptides which ultimately results in an increase in the number of fragment ion spectra detected (spectral count) (Lindemann *et al.*, 2017). For AUC, relative protein abundance is measured by comparing chromatographic peak areas of the same peptide identified in each sample. The AUC can be determined from extracted ion chromatograms (XIC), consisting of ion peak intensities and retention times, which correlate linearly to peptide concentration and protein abundance (Megger *et al.*, 2013). Label-free quantification methods are commonly used for large scale quantitative proteomic-based clinical studies due to its simple and economical application, however high quality and reproducible LC-MS/MS data is critical for accurate protein quantification (Pham *et al.*, 2012; Zhao *et al.*, 2020).

1.11 Scope of study

1.11.1 Study aims and purpose

The primary aim of this research project was to use advanced mass spectrometry-based proteomic profiling to characterise and compare protein components, including ECM from invasive ductal carcinomas and matching non-tumorous breast tissue to potentially identify prognostic markers or new therapeutic drug targets.

1.11.2 Study objectives

- To collect resected tumorous and matching non-tumorous tissue as frozen sections from chemotherapy naïve patients diagnosed with invasive ductal carcinoma.
- To use microscopic techniques to confirm patient diagnosis and differentiate between tumour types and to confirm matched non-tumorous tissue sections by histological means.
- To extract the protein complement of frozen histological sections of resected breast tissue using an optimised barocycler extraction method.
- To perform automated HILIC-bead assisted in-solution tryptic digestion on extracted proteins and to analyse these digests using LC-MS/MS.
- To compare the relative abundance of different proteomic components in matched non-tumorous breast tissue and in solid invasive ductal tumours using peptide sequencing to identify and quantitate relevant proteomic components including the recalcitrant ECM protein complement.

Chapter 2

2 Materials and Methods

2.1 Introduction

Non-tumorous and tumorous breast tissue samples were resected from breast cancer patients diagnosed with grade I-III invasive ductal carcinoma at the Steve Biko Academic Hospital, Pretoria, South Africa. Histopathological screening of the resected samples was essential for the study and was performed through screening haematoxylin and eosin (H&E) stained histological slides prepared from the resected tissue and were used to confirm the patient diagnosis and to distinguish between non-tumorous and tumorous tissue prior to further experiments being carried out. Histological analysis of tissue consisted of cryo-sectioning of tissue samples, H&E staining followed by classification of stained histological samples by a qualified pathologist from the Department of Anatomical Pathology at the University of Pretoria and the National Health Laboratory Service (NHLS). Tissue samples were also cryo-sectioned (60 µm) for downstream proteomic analysis that involved protein extraction using either detergents or repeated cycles of extreme hydrostatic pressure using a Barocycler® 2320EXT instrument (Pressure Biosciences Inc., Massachusetts, USA). Protein extraction was carried out on adjacent histological sections from the same non-tumorous and tumorous tissue samples. A semi-automated hydrophilic affinity-based protein capture, sample clean-up and off-bead trypsin digestion sample preparation sequence was used and tryptic digests were analysed using a Dionex Ultimate 3000 RSLC system (Thermo Fisher Scientific, Waltham, Ma, USA) coupled to an SCIEX 6600 TripleTOF mass spectrometer (MS) (SCIEX, Massachusetts, USA). Spectral data was analysed using SCIEX Protein Pilot. Advanced sequential window acquisition of all theoretical mass spectra (SWATH) technology was used for data independent acquisition and analysis of peptide amino acid sequences. Amino acid sequence data was used to identify proteins of interest as well as post translationally modified proteins and to provide relative quantitation of ECM proteins present in different samples. Tumour associated proteome changes were identified through comparison of the relative protein levels between non-tumorous and tumour masses.

2.2 Ethical approval

Ethical approval for the study was obtained from the Faculty of Health Sciences Research Ethics Committee at the University of Pretoria (Ref: 86/2017) (Appendix I) and a permission letter to collect samples from clinical patients at the Steve Biko Academic Hospital was obtained (Appendix II). Permission for the analysis of human tissue samples was also granted by the CSIR Research Ethics Committee (Ref: 213/2017) (Appendix III).

2.3 Sample collection

A cohort of patients diagnosed with invasive ductal type breast cancer who had not undergone prior neoadjuvant chemotherapy and were to be surgically treated at the Steve Biko Academic Hospital, Pretoria, South Africa were recruited for this study. An information leaflet was provided and discussed with each patient prior to the patient signing an informed consent form (Appendix IV) which consented to accessing medical records and to use the resected tissue for research purposes. Patient anonymity and confidentiality was ensured by allocating each patient with a study code number. There was no incentive for participation in the study and the patient's treatment regimen was not altered in any way due to participation in the study. A patient profile, consisting of age, tumour stage, immunophenotype and lymph node status was recorded (Appendix V). Surgeons from the Department of Surgery at the Steve Biko Academic Hospital were responsible for resecting matching non-tumorous breast tissue and tumour samples after performing modified radical mastectomies. Directly after resection, tissue samples were snap frozen by immersion in liquid nitrogen to preserve the biological structure and functional components of tissue samples. Samples were snap frozen within a maximum of 2 hours of starting surgery. Immediate snap freezing prevented any further morphological changes and enzyme induced lysis. Samples were then stored at -80°C until further processing by cryotome sectioning that was performed within one week of sample collection. The frozen 60 µm sections cut using the cryotome were then individually stored in Eppendorf Protein LoBind tubes (Hamburg, Germany) tubes at -80°C until batch processed for protein extraction and automated digestion with trypsin.

2.4 Histological analysis

Frozen breast tissue (non-tumorous and tumour) samples from individual patients were separately embedded and secured on the chuck of a cryotome using Tissue-Tek OCT (optimal cutting temperature) compound (Sakura, Tokyo, Japan). Tissue-Tek OCT is a commercially available polymer that is used in routine histological sample preparation to ensure samples remain in a fixed position on the cryotome chuck to allow for effective and optimal sectioning of frozen tissue. For proteomic analysis, breast tissue was sectioned at 60 µm thickness using a Cryotome E cryostat (Thermo Fisher Scientific, Waltham, Ma, USA) at the Department of Anatomical Pathology of the National Health laboratory Service (NHLS) at the University of Pretoria. Individual cryotome cut breast tissue sections were transferred into Eppendorf Protein LoBind tubes (Hamburg, Germany) as shown in Figure 2.1 and stored at -80°C for later downstream proteomic analysis.

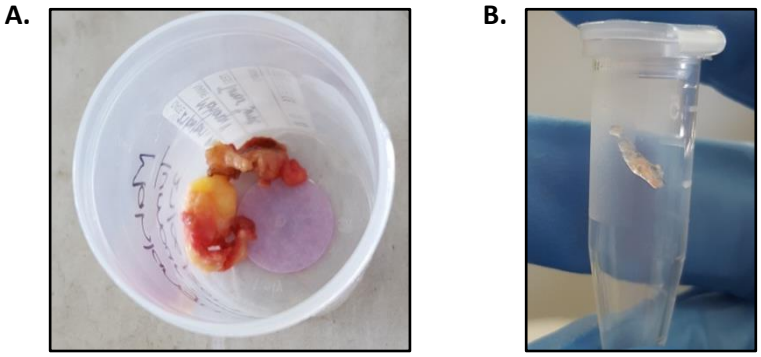

Figure 2.1: Breast tissue mass collected after a radical mastectomy surgery and a tissue section. (A) Breast cancer tumour and tissue samples were collected from patients diagnosed with invasive ductal carcinoma. (B) Tissue samples were sectioned at 60 µm thickness for downstream proteomic analysis.

After cryo-sectioning tissue sections for proteomic analysis, the remaining resected breast tissue was placed face down into a plastic cassette in preparation for standard histological processing. Tissue samples were orientated so that when the wax block was sectioned, the H&E-stained tissue was a visual representation of the sample that was used to characterise the proteome of the breast tissue sample. Conventional tissue processing involves placing samples in a buffered aqueous formalin fixative to preserve tissue integrity and structure, to prevent autolysis and putrefaction as well as impart mechanical rigidity for subsequent processing. Formaldehyde is the most commonly used fixative in pathology laboratories and is used as a 10% neutral buffered formalin solution. Formalin is able to diffuse into the various

layers of tissue allowing for the preservation of cellular composition and biomolecules such as proteins through chemical crosslinking (Kiernan, 2000; Thavarajah *et al.*, 2012). Before embedding in paraffin wax, the tissue is subjected to gradual dehydration to remove the aqueous fixative and any remaining water. Following dehydration, the dehydrating agent is cleared or removed with a substance that is completely miscible with both dehydrating and embedding agents. The final step involves impregnation of the tissue with a molten embedding medium which offers sufficient support and rigidity needed for thin sections of tissue to be cut, and at the same time making the processed tissue soft enough to allow the microtome blade to cut through the tissue without causing damage to tissue morphology or structure (Baskin, 2014). Breast tissue samples were placed in formalin (10% buffered solution) overnight then processed using an automated tissue fixation and wax embedding system (Tissue Processor VIP-6, The Scientific Group, Johannesburg, SA) for histological analysis using an overnight protocol routinely used by the Department of Anatomical Pathology, University of Pretoria. Solutions and bottle sequence for the protocol are summarised in Table 2.1. Breast tissue samples fixed in 10% buffered solution were dehydrated in increasing concentrations of ethanol (70% and 96% for 40 minutes each, 96% for 45 minutes and twice in 100% for 60 minutes). Samples were then placed in a 1:1 xylene: ethanol solution for 60 minutes, followed by 100% xylene for 2 hours in total. The samples were placed in molten paraffin wax for 3.5 hours at 60°C and finally embedded in wax blocks which were allowed to cool to 4°C prior to sectioning.

Formalin-fixed paraffin-embedded (FFPE) tissue samples were sectioned at 3-5 µm thickness using a Leica RM2255 microtome (Wetzler, Germany) and mounted onto standard microscope slides in preparation for H&E staining at the Department of Anatomical Pathology, NHLS, University of Pretoria. H&E is the most commonly used histological stain to differentiate between cellular components, whereby haematoxylin stains cellular nuclei a blue colour and eosin stains the cytoplasm and ECM a pink colour (Fischer *et al.*, 2008). This contrast in colour allows for the visualisation of general tissue structure and was used in this study to assess non-tumorous and tumour tissue slices obtained from breast cancer patient samples. FFPE samples were de-waxed in 100% xylene for 5 minutes and then in a 1:1 xylene: ethanol solution for 1 minute.

Table 2.1: A summary of the sequence for tissue fixation using the standard automated protocol consisting of solutions and bottle sequence for formalin-fixed paraffin-embedded tissue using the Tissue Processor VIP-6.

Process	Solution and bottle sequence	Duration
1	Formalin (10% buffered solution)	1 hour
2	Formalin (10% buffered solution)	1 hour
3	70% ethanol	40 minutes
4	96% ethanol	40 minutes
5	96% ethanol	45 minutes
6	100% ethanol	1 hour
7	100% ethanol	1 hour
8	50/50 xylene/ethanol	1 hour
9	Xylene	1 hour
10	Xylene	1 hour
11	Paraffin wax	45 minutes
12	Paraffin wax	45 minutes
13	Paraffin wax	1 hour
14	Paraffin wax	1 hour

Samples were then rehydrated in decreasing concentrations of ethanol (100 % for 1 minute, 96% ethanol for 1 minute and distilled water for 1 minute). Tissue sections were stained with Mayer's Haematoxylin stain for 5 minutes and rinsed with distilled water for 5 minutes. Samples were then dipped in eosin stain for 1 minute and washed in distilled water for 1 minute. Dehydration of samples was carried out by dipping each microscope slide in increasing concentrations of ethanol (96% ethanol and 100% ethanol for 1 minute each). Samples were then placed in a 1:1 ratio xylene: ethanol solution for 1 minute and 100% xylene for 1 minute before mounting using a xylene-based mounting media and a glass coverslip.

After samples were successfully stained, they were viewed under a microscope to confirm the presence of tumorous tissue and the matching of the tissue type. The cancer type was confirmed by a specialist pathologist from the NHLS who characterised and confirmed tumorous breast and matching of the non-tumorous tissue to support the inclusion of each patient's sample into the proteomic research study.

2.5 Proteomic analysis

2.5.1 Off-bead protein digestion

2.5.1.1 Protein extraction using pressure cycling technology

Reagents

i. Extraction buffer

The protein extraction buffer consisted of 4% SDS (Sigma-Aldrich, St Louis, USA), 100 mM Tris (Sigma-Aldrich, St Louis, USA), 20 mM Dithiothreitol (DTT) (Sigma-Aldrich, St Louis, USA) and a protease inhibitor cocktail (cOmplete™ Protease Inhibitor Cocktail Tablets, Roche, Basel, Switzerland). DTT and the protease inhibitor was prepared fresh and added to the lysis buffer. The pH was adjusted to 8.0 with a 1 M HCl solution.

Method

Pressure cycling technology (PCT) utilising the Barocycler 2320EXT instrument (Pressure Biosciences Inc., Massachusetts, USA) which cycles between high pressure and ambient pressure approximately every minute, was used for tissue homogenisation and protein extraction. The barocycler is a benchtop instrument that is able to process up to 16 samples simultaneously using an array of barocycler microtubes with micropestles. Tissue samples in the microtubes are trapped between the bottom of the tube and micropestle tip which thoroughly homogenises the tissue sample on every cycle where the combined effect of the mechanical action and the extraction ability of the detergent-based buffer under high pressure results in effective extraction of proteins from tissue samples (Lucas *et al.*, 2018). Frozen tissue sections (60 µm) were carefully transferred individually to Barocycler microtubes. A volume of 30 µL of extraction buffer (4% SDS, Tris, DTT protease inhibitor) was added to each tissue section and briefly centrifuged to drive the tissue section and extraction buffer to the bottom of the narrow Barocycler tube. The Barocycler® 2320EXT instrument was preheated to the desired temperature before the multi-sample holder was loaded and one of two extraction methods run. Appropriate pestle and caps were inserted into the tubes and loaded into the multi-sample holder. Two Barocycler extraction methods were compared where the first used the vendor recommended parameters while the second method used an

in-house optimised method for histological slices that included twice the number of extraction cycles and a final long high-pressure cycle that was shown to increase both the extracted protein concentration and number of proteins identified. A limited number of samples were tested to confirm the improved extraction using both the methods. Parameters for in-solution extraction using the barocycler one-step and two-step methods are summarised in Table 2.2. After completion of the extraction cycles, samples were carefully removed from the barocycler multi-tube holder and the lids were removed. Barocycler tubes were then placed upside down in 0.5 mL Eppendorf Protein LoBind tubes and briefly centrifuged to transfer the extracts. Barocycler microtubes were removed and the LoBind tubes centrifuged at 16 000 *g* for 10 minutes and the supernatants were transferred to fresh tubes. An aliquot of 5 µL from each sample extract was transferred to a separate Eppendorf Protein LoBind tube for protein quantification and the remaining samples stored at -80°C for further analysis using the automated HILIC bead based tryptic digestion and LC-MS/MS as described in Sections 2.5.1.3 and 2.52 below.

Table 2.2 Parameters for the one-step and two step barocycler method for tissue protein extraction.

Method	Temperature (°C)	Pressure (Kpsi)	No. of cycles	Duration of 1 cycle
Barocycler one-step method	95	45.0	35	High pressure for 30 seconds with stepwise low pressure for 15 seconds
Barocycler two-step method	95	45.0	70	High pressure for 30 seconds with stepwise low pressure for 15 seconds
	95	45.0	1	Continuous high pressure for 60 minutes

2.5.1.2 2-D Quant Protein Assay

The 2-D Quant Protein Assay Kit (GE Healthcare, Amersham, UK) was used to accurately determine protein concentrations for each sample extract. All reagents were provided premade in the 2-D Quant Assay Kit and used per manufacturer instructions. All reagents were stored at 4°C until used. This assay gives a negative slope calibration curve with a maximum quantifiable amount of 50 µg protein.

Reagents

i. Precipitant

The precipitant solution renders proteins insoluble.

ii. Co-precipitant

The co-precipitant solution contains reagents that co-precipitate with proteins and enhances their removal from solution.

iii. Copper solution

Precipitated protein is resuspended using the copper solution.

iv. Colour reagent A and B

Colour reagent A is mixed with colour reagent B to prepare the colour reagent used to measure unbound copper ion.

v. Bovine serum albumin standard solution

Bovine serum albumin (BSA) standard solution is used to prepare a standard curve.

Method

The 2-D Quant Protein Assay Kit works by quantitatively precipitating proteins out of solution, using precipitant and co-precipitant solutions, while leaving behind interfering contaminants. The 2-D Quant assay is based on the specific binding of cupric ions to the polypeptide backbone of proteins and a colorimetric agent that reacts with unbound cupric ions whereby the colour density is inversely proportional to the protein concentration which can be extrapolated from a standard curve. A standard curve series was prepared by aliquoting volumes of 0 µL, 5 µL, 10 µL, 15 µL, 20 µL and 25 µL (volumes containing 0 µg, 10 µg, 20 µg, 30 µg, 40 µg and 50 µg quantities of BSA respectively) into separate 2 mL Eppendorf LoBind Protein tubes. A volume of 500 µL of the precipitant solution was added to each BSA standard solution and to 5 µL of sample including a blank. Samples were vortex mixed and incubated at ambient temperature for 3 minutes. A volume of 500 µL of the co-precipitant solution was then added and samples were briefly vortex mixed and centrifuged for 5 minutes at 10 000 *g*. The supernatants were removed and samples were briefly centrifuged and any remaining visible liquid removed. A volume of 100 µL of copper solution and 400 µL of deionised water was added to each tube and samples were vortex mixed until the precipitated protein pellet

was dislodged from the bottom of the tubes. Samples were vortex mixed again so that the protein pellet was completely re-suspended. A volume of 1 mL of working colour reagent was rapidly added to each tube to ensure instantaneous mixing and this was followed by an incubation period of 15-20 minutes at ambient temperature. A volume of 300 µL of each standard and sample was then plated into a 96-well plate and the absorbance for the standards and samples was read at 480 nm using an ELx800 UV universal microplate reader (Bio-Tek Instruments Inc. Vermont, USA). A standard curve was generated by plotting the absorbance of the standards (A_{480}) against the concentration of BSA protein standards. Protein concentration of each sample was then extrapolated from the BSA standard curve.

2.5.1.3 Hydrophilic interaction liquid chromatography (HILIC)

Reagents

i. NH_4HCO_3 solutions

A 50 mM solution of NH_4HCO_3 (Sigma-Aldrich, St Louis, USA) was prepared by dissolving 200 mg of NH_4HCO_3 in 50 mL of deionised water. A 25 mM solution of NH_4HCO_3 was made by adding 25 mL of the 50 mM NH_4HCO_3 solution to 25 mL of deionised water. The solutions were stored at 4°C and were used within two weeks.

ii. Iodoacetamide solution

A 500 mM stock solution of Iodoacetamide (IAA) (Sigma-Aldrich, St Louis, USA) was prepared by dissolving 92.5 mg in 1 mL of 25 mM NH_4HCO_3. The IAA solution was prepared fresh just before use and stored in a dark container.

iii. DTT solution

A 1 M stock solution of DTT (Sigma-Aldrich, St Louis, USA) was prepared by dissolving 154.3 mg in 1 mL of 25 mM NH_4HCO_3. The solution was prepared fresh just before use.

iv. High affinity equilibration buffer I or "Binding buffer"

A final volume of 30 mL of high affinity (HA) equilibration buffer I consisted of 200 mM ammonium acetate (NH_4AC) (Sigma-Aldrich, St Louis, USA) in a 30% MS-grade acetonitrile in MS-grade water solution.

v. High affinity Equilibration Buffer II

HA equilibration buffer II consisted of 100 mM (NH_4AC) and 15% acetonitrile. The buffer was prepared by diluting the HILIC equilibration buffer I with MS-grade water at a 1:1 dilution ratio.

vi. Acetonitrile solution

A volume of 500 µL of MS-grade water was mixed with 9.50 mL of 100% acetonitrile to produce a 95% acetonitrile solution.

vii. MagReSyn HILIC beads

A 20 mg/mL MagReSyn HILIC bead suspension was provided by ReSyn Biosciences (Pty) Ltd (CSIR, Pretoria, South Africa) and stored at 4°C until used.

viii. Trypsin

Sequence grade modified trypsin (Promega, Madison, USA) was added to protein samples at a 1:10 ratio. A mass of 200 µg of trypsin was diluted with 200 µL of provided diluent to make up a final concentration of 1 µg/µL. A final volume of 2 µL of the trypsin solution was then added to 20 µg of protein samples in 50 mM NH_4HCO_3 solution.

Method

Sample clean-up and digestion was done using the automated KingFisher™ Duo magnetic liquid handling station (Thermo Fisher Scientific, Waltham, Ma, USA) and MagReSyn HILIC microparticles (ReSyn Biosciences, Pretoria, South Africa). MagReSyn HILIC is a proprietary polymeric magnetic microparticle support designed for the solid phase extraction and recovery of biomolecules from common sample contaminants that may interfere with downstream analytical procedures, primarily aimed at sample preparation for mass spectrometry. MagReSyn HILIC involves the use of an organic mobile phase to drive amphiphilic biomolecules into a water-rich layer formed on the surface of a stationary phase allowing biomolecules to be separated from less polar compounds. Retention of biomolecules at the aqueous layer is a result of hydrogen and weak electrostatic interactions. The MagReSyn beads selectively bind proteins and facilitate the removal of interfering substances

or contaminants that may affect protein solubility or denaturing that could ultimately hinder trypsin digestion thereby affecting downstream MS analysis. Ammonium acetate and ammonium formate are volatile ionic additives used to control the pH and ionic strength of the sample. Before MagReSyn bead sample clean-up was carried out, soluble protein samples were reduced with DTT, a dithiol-containing reagent used to reduce protein disulphide bonds, followed by alkylation of free thiols using a final concentration of 60 mM IAA and incubated in the dark for 30 minutes. Excess IAA was quenched with a further addition of DTT to a final concentration of 20 mM. A final amount of 20 µg of protein from each sample was used for the HILIC sample clean-up procedure. A 96 deep-well HILIC plate was then set up by adding reagents and samples from row A to G as shown in Figure 2.2. All steps prior to trypsinolysis were carried out at room temperature (23°C – 25°C). The automated system has a 12-pin robotic magnetic head with a plastic comb, initially placed in Row H, that allows for the mixing and the transferring of reagents and samples between wells during the different steps carried out by the magnetic handling station. MagReSyn HILIC beads were washed in Row G, and transferred to Row F for equilibration (one minute) followed by protein binding in Row E (30 minutes). Two successive washes in Rows D and C (one minute each) were carried out to remove potential non-bound contaminants. Proteins were then digested with trypsin and the reaction was allowed to proceed for 4 hours at 37°C in Row A which was temperature controlled using a Kingfisher™ Duo Peltier heating block. After protein digestion with trypsin, the HILIC beads were removed from protein digests and transferred to the original storage position in Row G. Peptide solutions were harvested from Row A and transferred to 0.5 mL Eppendorf Protein LoBind tubes and vacuum dried at ambient temperature using a CentriVap (Labconco, Missouri, USA) before being stored at -80°C for later LC-MS/MS analysis.

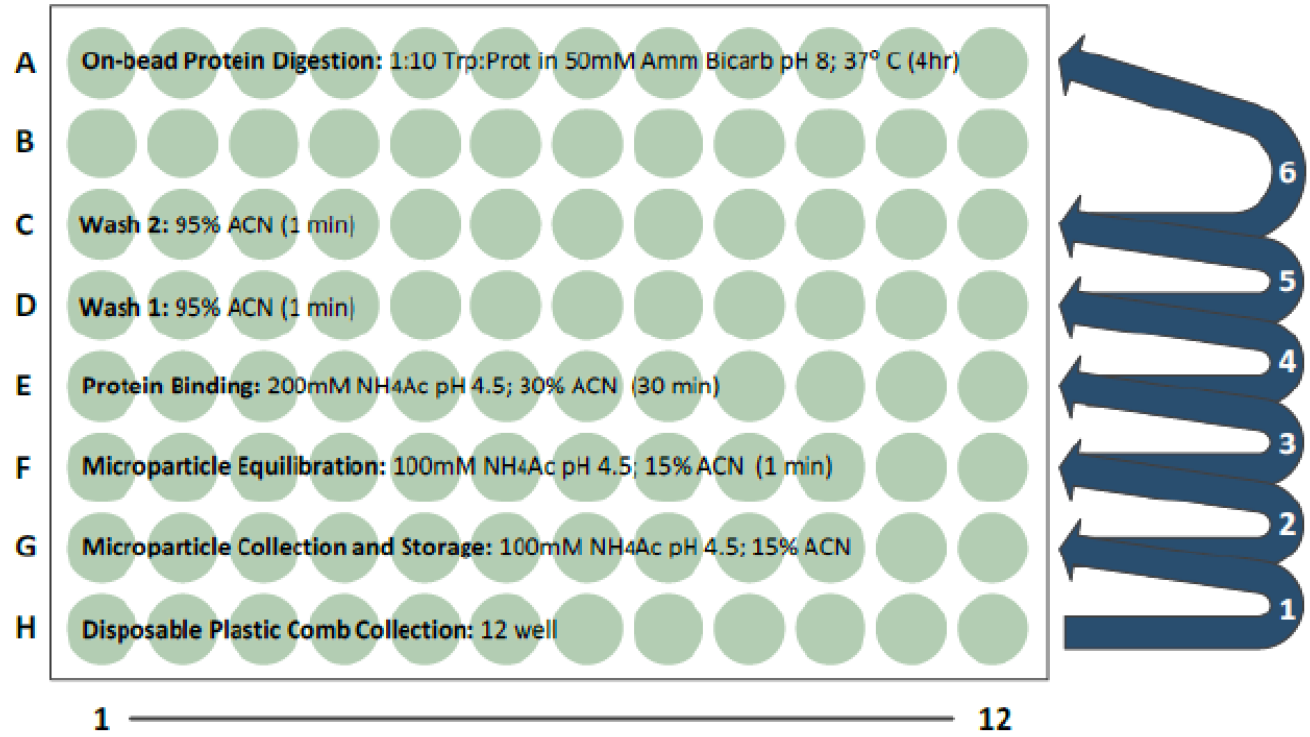

Figure 2.2: Plate setup for the KingFisher™ Duo magnetic handling station and method summary for on-bead digestion protocol. Protein clean-up and trypsin digestion was performed using MagReSyn HILIC beads. Digestion was performed in a 96 deep-well plate. A 12-tip disposable magnetic comb was placed in Row H. The 12-tip magnetic head picked up the comb (Row H) and collected the magnetic HILIC beads (Row G) for equilibration in Row F. Protein binding was carried out in Row E. Protein bound beads were washed in Rows D and C. On-bead protein digestion, using trypsin (1:10; enzyme: protein), was performed for 4 hours at 37 °C (Row A) (Stoychev *et al.*, 2012) (with permission).

2.5.1.4 Quantitative Colorimetric Peptide Assay

The Pierce Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific, Waltham, Ma, USA) was used to accurately determine peptide concentrations for each trypsin digested sample. All reagents were provided premade in the assay kit and used per manufacturer instructions. All reagents were stored at 4°C until used.

Reagents

i. Colorimetric Digest Assay Standard (1mg/mL)

A dilution series of the peptide digest assay standard was prepared as shown in Table 2.3 and used to generate an eight- point standard curve, with a concentration range between 0- 1000 µg/mL.

Table 2.3 Dilution series of the peptide digest assay standard for the standard curve used to determine the unknown peptide concentration for individual trypsin digestion samples

Tubes	Volume of water (µL)	Volume of Peptide Digest Assay Standard (µL)	Final concentration of Digest Assay Standard (µg/mL)
A	0	150 of Stock	1000
B	75	75 of tube A	500
C	75	75 of tube B dilution	250
D	75	75 of tube C dilution	125
E	75	75 of tube D dilution	62.3
F	75	75 of tube E dilution	31.3
G	75	75 of tube F dilution	15.6
Blank		Use water for blank	0

ii. Colorimetric Peptide Assay working reagents

Colorimetric peptide assay working reagents A, B and C were prepared by combining the reagents in 50:48:2 ratio respectively. The final working reagent was made up fresh and used within 30 minutes.

Method

The Pierce Quantitative Colorimetric Peptide Assay is based on copper reduction (Cu^{2+} to Cu^{+1}) by the amide backbone of peptides under alkaline conditions followed by the coupling of the reduced copper ions to a supplied chelator which essentially forms a bright red complex. The colorimetric signal can be read at 480 nm wavelength and peptide concentrations can be extrapolated from a standard curve. A volume of 40 µL of each prepared standard concentration (A-G), individual samples and a blank were pipetted into Eppendorf LoBind Protein tubes. A volume of 360 µL of working reagent was then added to each tube and samples were briefly vortex mixed. Samples were heated at 37°C for 15 minutes and then plated into a 96-well plate. The absorbance for each standard and sample was read at 480 nm using an ELx800 UV universal microplate reader (Bio-Tek Instruments Inc. Vermont, USA). A

standard curve was generated by plotting the absorbance values (A_{480}) of the standards against the concentration of the peptide digest standards (µg/mL). The peptide concentration of each sample was then extrapolated from the peptide digest standard curve.

2.5.2 First dimension UPLC fractionation

Reagents

i. Acetonitrile, ammonium hydroxide and mass spectrometer grade water were all of MS-grade and were purchased from Fluka Analytical (Basel, Switzerland).

Method

In the first dimension, 20 µg of tryptic digest peptides collected from the wells of the Kingfisher system and diluted to a total volume of 20 µL was injected and fractionated using high-pH reverse phase chromatography. This was performed using a Dionex UltiMate 3000 RSLC system (Thermo Fisher Scientific, Waltham, Ma, USA) with an Acclaim PA II column (1.0 mm x 15 cm, C18, 3 µm, 120 Å) (Thermo Fisher Scientific, Waltham, Ma, USA). Mobile phase A contained 20 mM ammonium hydroxide in water (pH 9.6) and Mobile phase B contained 20 mM ammonium hydroxide (pH 9.6) in 80% acetonitrile in water. Peptide separation was carried out at 50 µL/min at 21°C using a linear gradient (4% B for 5 min, from 4% B to 60% B in 35 min, to 90% B in 5 min, and thereafter held at 4% B for 15 min to re-equilibrate. Fractions were collected every 30 sec from 12.5 to 27.5 min and a total of 30 fractions were collected. The collected fractions were pooled into 0.5 mL Protein LoBind tubes (Eppendorf, Germany) to give 10 new combined fractions according to the following pooling scheme: (F1= [1, 11, 21]; F2= [2, 12, 22]; F3= [3, 13, 23]; F4= [4, 14, 24]; F5= [5, 15, 25]; F6= [6, 16, 26]; F7= [7, 17, 27]; F8= [8, 18, 28]; F9= [9, 19, 29], F10= [10, 20, 30]). Pooled fractions were vacuum dried using a CentriVap (Labconco, Missouri, USA) and stored at -80°C until further analysis. These samples were used to generate the DDA analysis based project specific protein library for later analysis by SWATH.

2.5.3 Nano LC-MS/MS

Reagents

Acetonitrile/formic acid solutions

ii. Acetonitrile, formic acid and water were all of MS-grade and were purchased from Fluka Analytical (Basel, Switzerland).

Methods

The pooled peptide fractions and individual samples from non-tumorous and tumour samples were prepared for low pH reverse phase chromatography by re-suspending the dried samples in 25 µl of 2% acetonitrile/0.2% formic acid and spiked with iRT peptide standards (Biognosys, Zurich, Switzerland). Peptide analysis was performed on a Dionex Ultimate 3000 RSLC system (Thermo Fisher Scientific, Massachusetts, USA) coupled to a SCIEX 6600 TripleTOF mass spectrometer (SCIEX, Massachusetts, USA). Injected peptides (0.5 µg) were first de-salted online using an Acclaim PepMap C18 trap column (75 µm × 2 cm) for 5.5 minutes at 5 µL/minute using 2% acetonitrile/0.2% formic acid. Trapped peptides were then transferred and separated on an Acclaim PepMap C18 nanoRSLC column (75 µm × 15 cm, 2 µm particle size). Peptide elution was achieved by using a flow-rate of 0.5 µL/minute using the following gradient: 2.5% - 45% B over 45 min (A: 0.1% formic acid; B: 80% acetonitrile/0.1% formic acid).

An electrospray voltage of 2.5 kV was applied to the fused silica emitter (New Objective: 20 µm ID x 5 cm, 10 µm tip). Initially the Sciex 6600 TripleTOF mass spectrometer was operated in data dependent acquisition (DDA) mode to accumulate a project specific library of the expected peptides and proteins from the pooled high pH reverse phase separation of control sample digests (Nweke *et al.*, 2020). For DDA runs, precursor (MS) scans were acquired from m/z 400-1500 using an accumulation time of 250 milliseconds (ms) followed by 80 fragment ion (MS/MS) scans, acquired from m/z 100-1800 with 25 ms accumulation time each for a total scan time of 2.3 sec. Multiply charged ions (2+ - 5+, 400 -1500 m/z) were automatically fragmented in Q2 collision cells using nitrogen as the collision gas. Collision energies were adjusted automatically as a function of m/z and charge. The mass spectrometer was auto-recalibrated after every injection using a [Glu1]-Fibrinopeptide B standard (Sigma-Aldrich, Missouri, USA).

Peptide samples from individual non-tumorous tissue and tumour were analysed using data-independent acquisition (DIA) by Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH) (Röst *et al.*, 2014; Nweke *et al.*, 2020). The same LC gradient as above used for DDA was used, whilst the SWATH method consisted of a series of 70 fragment ion (MS/MS) scans of overlapping sequential precursor isolation windows (variable m/z isolation width, 1 m/z overlap, high sensitivity mode) covering the 400 to 900 m/z mass range, with a precursor MS scan for each cycle. The accumulation time was 250 ms per MS scan (from 400 to 900 m/z) and 25 ms for each product ion scan (100 to 1800 m/z), thus making a 2.05 sec total cycle time. The raw *.wiff files generated by the mass spectrometer were then processed using bioinformatics software as described in Section 2.5.4.

2.5.4 Data processing

2.5.4.1 Data-dependant database search

Initial data processing was performed using Protein Pilot (v 5.0.1) (Seymour, 2017) where raw data (*.wiff) files were searched against the human UNIPROT database of human protein sequences concatenated with a list of common contaminating proteins as well as the sequences of the Biognosys iRT peptide retention time standards. In the search settings, trypsin was selected as the proteolytic enzyme, iodoacetamide based alkylation and thorough search effort with biological modifications allowed also selected. A false discovery rate (FDR) analysis was set at 0.1% local FDR cut-off applied at PSM, peptide and protein levels.

2.5.4.2 SWATH library generation and data analysis using Spectronaut™ 11

Raw SWATH™ data files (.wiff) were converted into Spectronaut™ 11 (Biognosys Inc, Beverly, Massachusetts, USA) compatible files (*.HTRMS) and subsequent comparison analysis was carried out using default settings in Spectronaut™ 11 (Khurana and George, 2008) as summarised in Table 2.4. Briefly, dynamic iRT retention time prediction was used with the correction factor for window set at 1. The decoy method was set as scrambled database and the FDR was set at 0.01 (1%) at the peptide level. The ID picker algorithm, using default Spectronaut™ settings, was used for protein inference. Interference correction at MS1 and MS2 levels was enabled to consistently exclude ions with interferences across all runs. Quantification was done at MS2 level and quantities were calculated using the area under the curve between the XIC peak boundaries for each targeted ion. Data filtering was enabled and

the Q-value threshold was set at ≤ 0.01 and used to select peptides for protein group quantitation. Cross run global normalisation was enabled and individual runs were normalised using the median quantities of all peptides selected for normalisation. The in-house generated spectral library was used as a reference for data extraction. A volcano plot (log2 fold change vs log_{10}p-value) was generated for comparison analysis using a t-test that was performed on the log_2 ratio of peptide intensities attained for individual proteins. The p-values were corrected using the multiple hypothesis testing correction approach to regulate the false discovery rate (FDR) (Storey, 2002). A candidate protein was only selected if a log_2 fold change greater than or equal to 1 (2-fold difference), Q-value less than or equal to 0.01% (1% FDR) and greater than or equal to 2 unique peptides were reported for a specific protein.

Table 2.4: Summary of Spectronaut™ experimental settings for quantitation.

Interference Correction	Yes
Exclude All Multi-Channel Interferences	Yes
MS1 Min	2
MS2 Min	3
Proteotypicity Filter	None
Major (Protein) Grouping	By Protein Group ID
Minor (Peptide) Grouping	By Stripped sequence
Major Group Quantity	Mean peptide quantity
Major Group Top N	Yes
Max	3
Min	1
Minor Group Quantity	Mean precursor quantity
Major Group Top N	Yes
Max	3
Min	1
Quantity MS-Level	MS2
Quantity Type	Area
Data Filtering	Q-value
Cross Run Normalisation	Yes
Normalisation Strategy	Global Normalisation
Normalise on	Median
Row selection	Automatic

2.5.4.3 Sample comparison using Perseus v1.6.1.1

The filtered list of protein candidates from Spectronaut™ was exported directly into Perseus v1.6.1.1 (Max Planck Institute of Biochemistry), to generate principal component (PCA) plots for comparisons between individual patient samples. For analysis, all missing values were assigned from a normal distribution and a two-sample test was conducted using an FDR of 0.05 with 250 randomisations and no weighting of the mean (s = 0). PCA was done using a Benjamini-Hochberg cut-off method with an FDR of 0.05 and 5 enrichment components.

2.5.4.4 Annotation of biological pathways using Cytoscape v3.8.0

A list of candidate proteins with associated average $\log_2$ ratios representing differential abundance was exported directly into Cytoscape v3.8.0 to generate a STRING enrichment analysis using the STRING App v1.5.1. A protein query was generated using imported enrichment data from the STRING database using a confidence cut-off score of 0.7 for the interactions retrieved. A Cytoscape STRING enrichment table, which included enriched terms with corresponding FDR values and gene components, was then generated.

Chapter 3

3 Results and discussion

3.1 Histological analysis

Histological analysis of breast tissue samples was an important component of this study as it was used to assess the inclusion of patients based on the visual confirmation of an invasive ductal carcinoma (IDC) diagnosis. Furthermore, histological analysis was utilised to determine whether the samples collected from the same study participant were in fact representative of a tumour and an equivalent matching non-tumorous tissue. Haematoxylin and eosin (H&E) staining, the gold standard for routine histopathological diagnosis of cancer, is used to visualise tissue morphology and to confirm the presence of cancer cells (Fischer *et al.*, 2008). The H&E-staining was an effective means to screen patient samples and differentiate between non-tumorous and tumour tissue, whereby 27 of the 58 patient samples collected during this study were disregarded after histological analysis clearly showed cancer cells in the surgically excised breast tissue supposedly from non-tumorous tissue regions. A series of H&E micrographs of histological slices of tumour and similar but adjacent non-tumorous tissue obtained from individual patients are illustrated in Figures 3.1 – 3.8 where the cancer cells and general morphology are indicated. Additionally, histological analysis was a complementary technique to downstream proteomic analysis, as micrographs of sequentially sectioned H&E-stained sections provided a visual representation of the tissue from the sequential histological sections from both non-tumorous and tumour tissue to show the extent of the cancerous changes in each sample processed for LC-MS/MS proteomic analysis.

Terminal ductal-lobular units (DLU), shown in Figure 3.4C, are characteristic features of normal breast tissue. Breast ducts and lobules have an inner luminal layer of epithelial cells and an outer basal layer of myoepithelial cells, where cellular nuclei stain dark blue with haematoxylin. Ductal-lobular units are surrounded by a stromal compartment (S) of reticular arranged collagen fibres, which are not clearly seen at the low magnification of these micrographs, but is represented by the pink eosin-stained areas in both the non-tumorous and tumour sample histology images. The ductal and lobular structures are surrounded by both a relatively dense stroma and adipocyte tissue (A) as shown in several of the non-tumorous micrographs (e.g. Figure 3.3A, Figure 3.5A and Figure 3.5C).

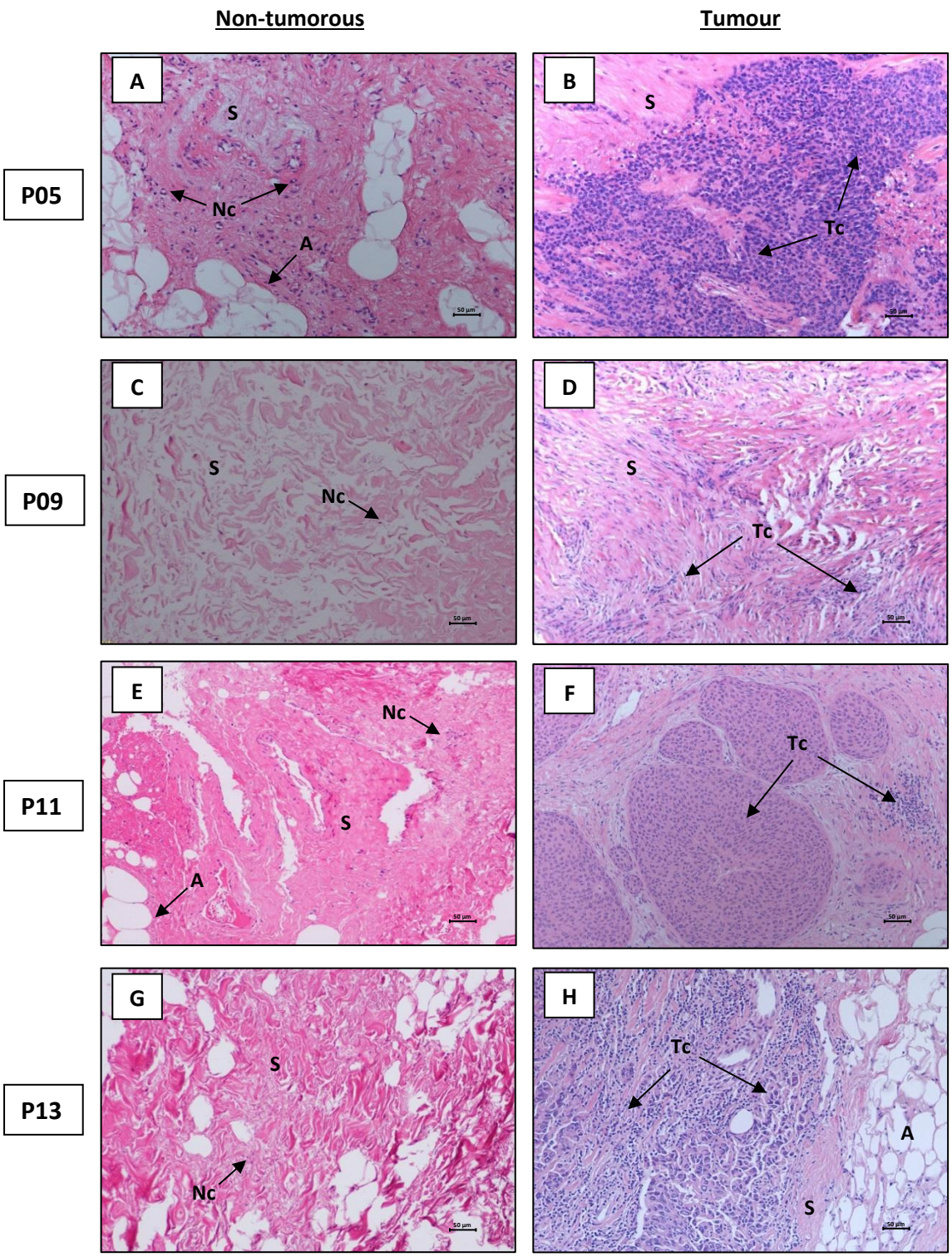

Figure 3.1: Histological analysis of patient samples P05 (A and B), P09 (C and D), P11 (E and f) and P13 (G and H) representing non-tumorous and tumour tissue stained with H&E and visualised at 10x magnification. Micrographs show non-tumorous breast tissue sections consisting of adipocytes **(A)** and normal epithelial cells **(Nc)** that are surrounded by an organised stromal compartment **(S)** consisting of collagen fibres and other extracellular matrix components. Extensive pockets of invasion of malignant epithelial cells **(Tc)** into normal breast stroma are visible in stained invasive ductal carcinoma tumour sections.

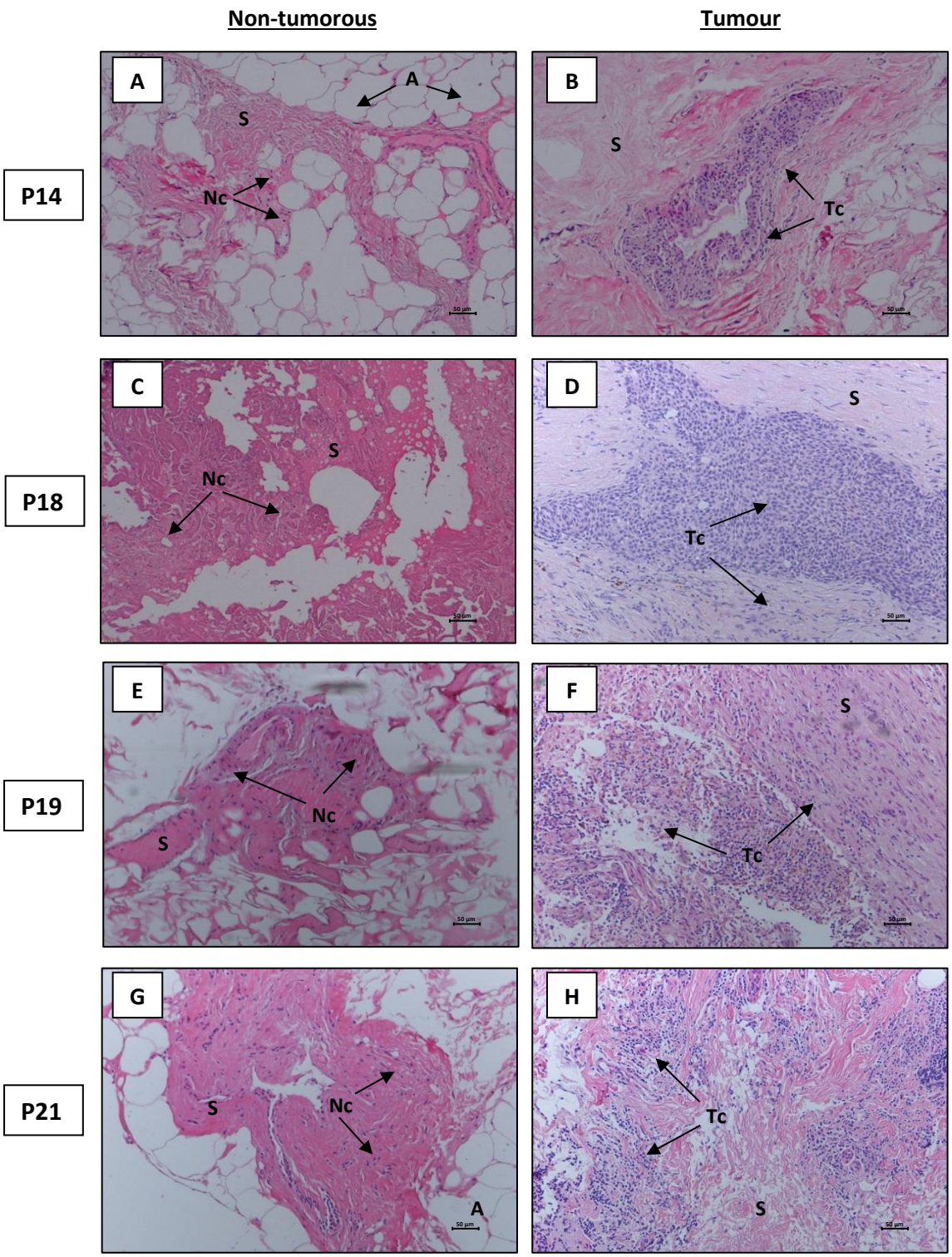

Figure 3.2: Histological analysis of patient samples P14 (A and B), P18 (C and D), P19 (E and F) and P21 (G and H) representing non-tumorous and tumour tissue stained with H&E and visualised at 10x magnification. Micrographs show non-tumorous breast tissue sections consisting of adipocytes **(A)** and normal epithelial cells **(Nc)** that are surrounded by an organised stromal compartment **(S)** consisting of collagen fibres and other extracellular matrix components. Extensive pockets of invasion of malignant epithelial cells **(Tc)** into normal breast stroma are visible in stained invasive ductal carcinoma tumour sections.

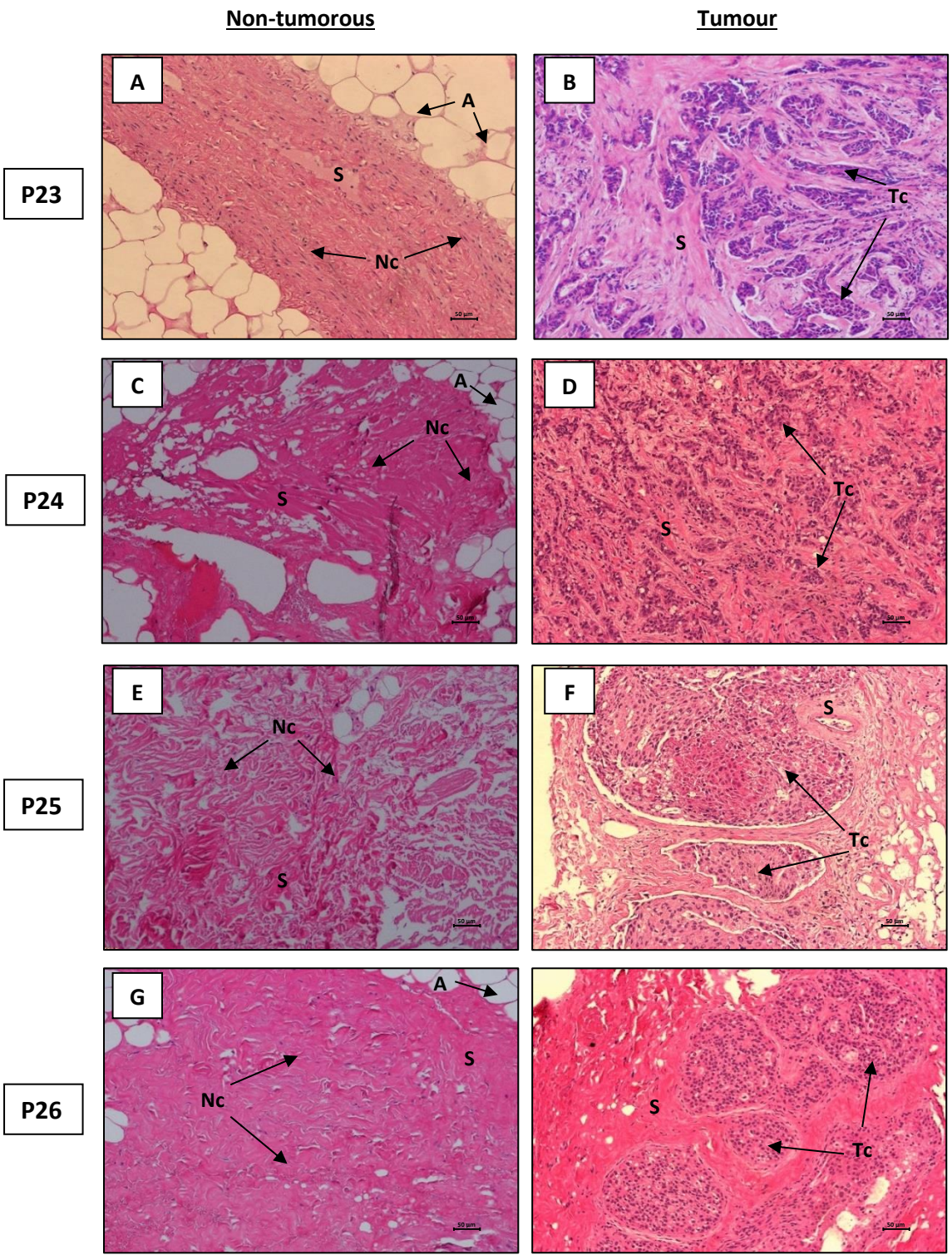

Figure 3.3: Histological analysis of patient samples P23 (A and B), P24 (C and D), P25 (E and F) and P26 (G and H) representing non-tumorous and tumour tissue stained with H&E and visualised at 10x magnification. Micrographs show non-tumorous breast tissue sections consisting of adipocytes **(A)** and normal epithelial cells **(Nc)** that are surrounded by an organised stromal compartment **(S)** consisting of collagen fibres and other extracellular matrix components. Extensive pockets of invasion of malignant epithelial cells **(Tc)** into normal breast stroma are visible in stained invasive ductal carcinoma tumour sections.

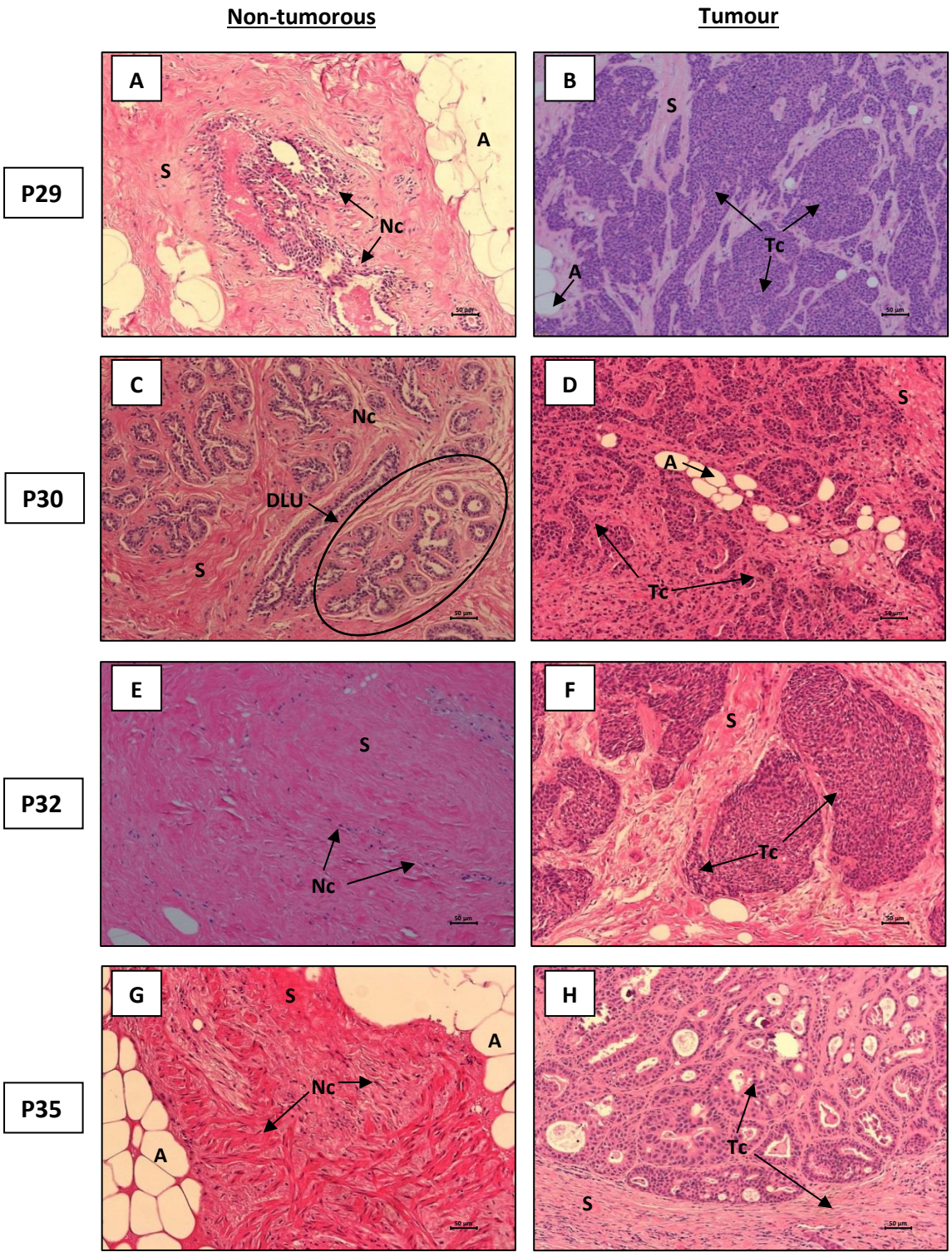

Figure 3.4: Histological analysis of patient samples P29 (A and B), P30 (C and D), P32 (E and F) and P35 (G and H) representing adjacent non-tumorous and tumour tissue which was stained with H&E and visualised at 10x magnification. Micrographs show non-tumorous breast tissue sections consisting of adipocytes **(A)** and normal epithelial cells **(Nc)** that are surrounded by an organised stromal compartment **(S)** consisting of collagen fibres and other extracellular matrix components. Micrograph (c) shows an example of ductal lobular units **(DLU)** that are a characteristic feature of normal breast tissue. Extensive pockets of invasion of malignant epithelial cells **(Tc)** into normal breast stroma are visible in stained invasive ductal carcinoma tumour sections.

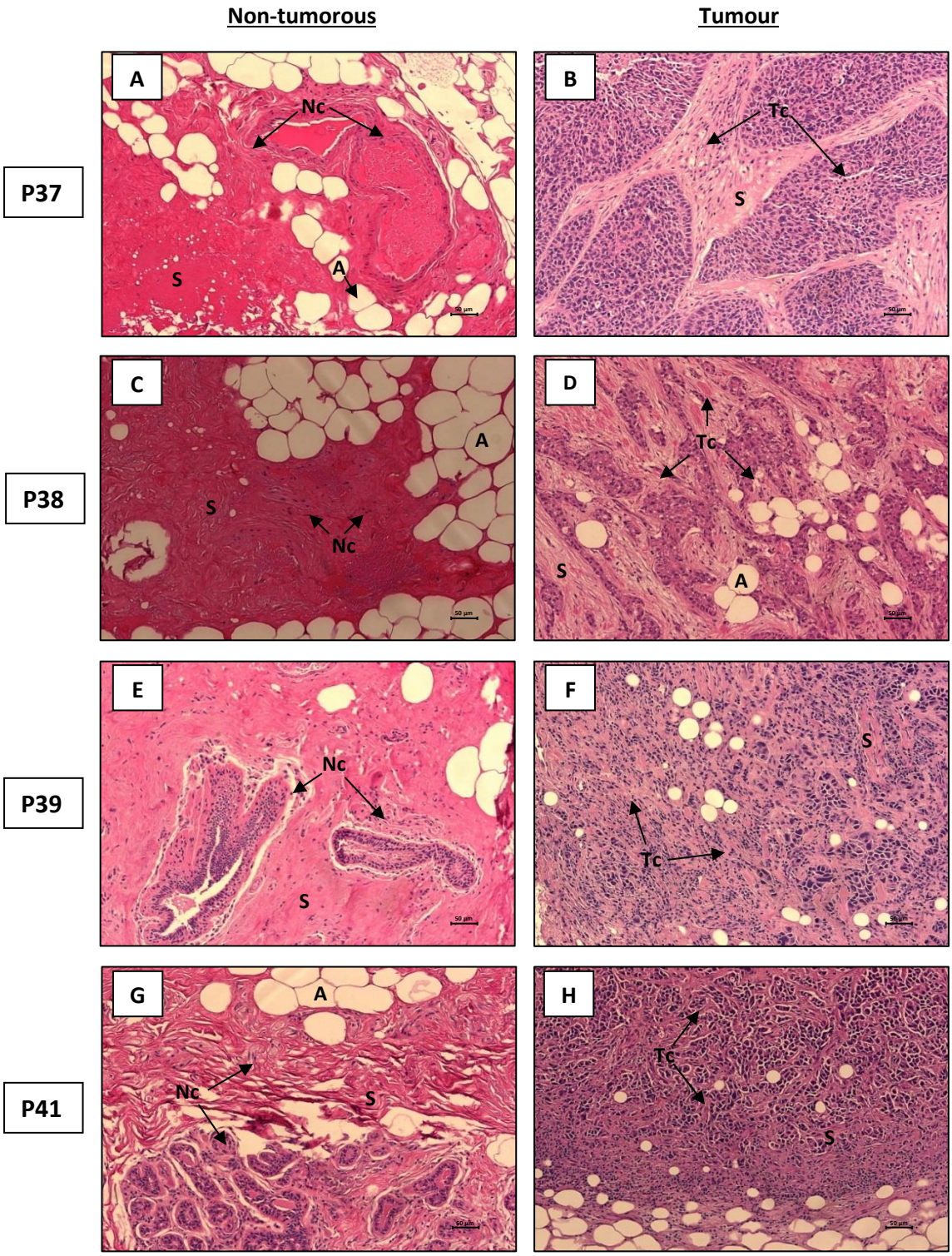

Figure 3.5: Histological analysis of patient samples P37 (A and B), P38 (C and D), P39 (E and F) and P41 (G and H) representing non-tumorous and tumour tissue stained with H&E and visualised at 10x magnification. Micrographs show non-tumorous breast tissue sections consisting of adipocytes **(A)** and normal epithelial cells **(Nc)** that are surrounded by an organised stromal compartment **(S)** consisting of collagen fibres and other extracellular matrix components. Extensive pockets of invasion of malignant epithelial cells **(Tc)** into normal breast stroma are visible in stained invasive ductal carcinoma tumour sections.

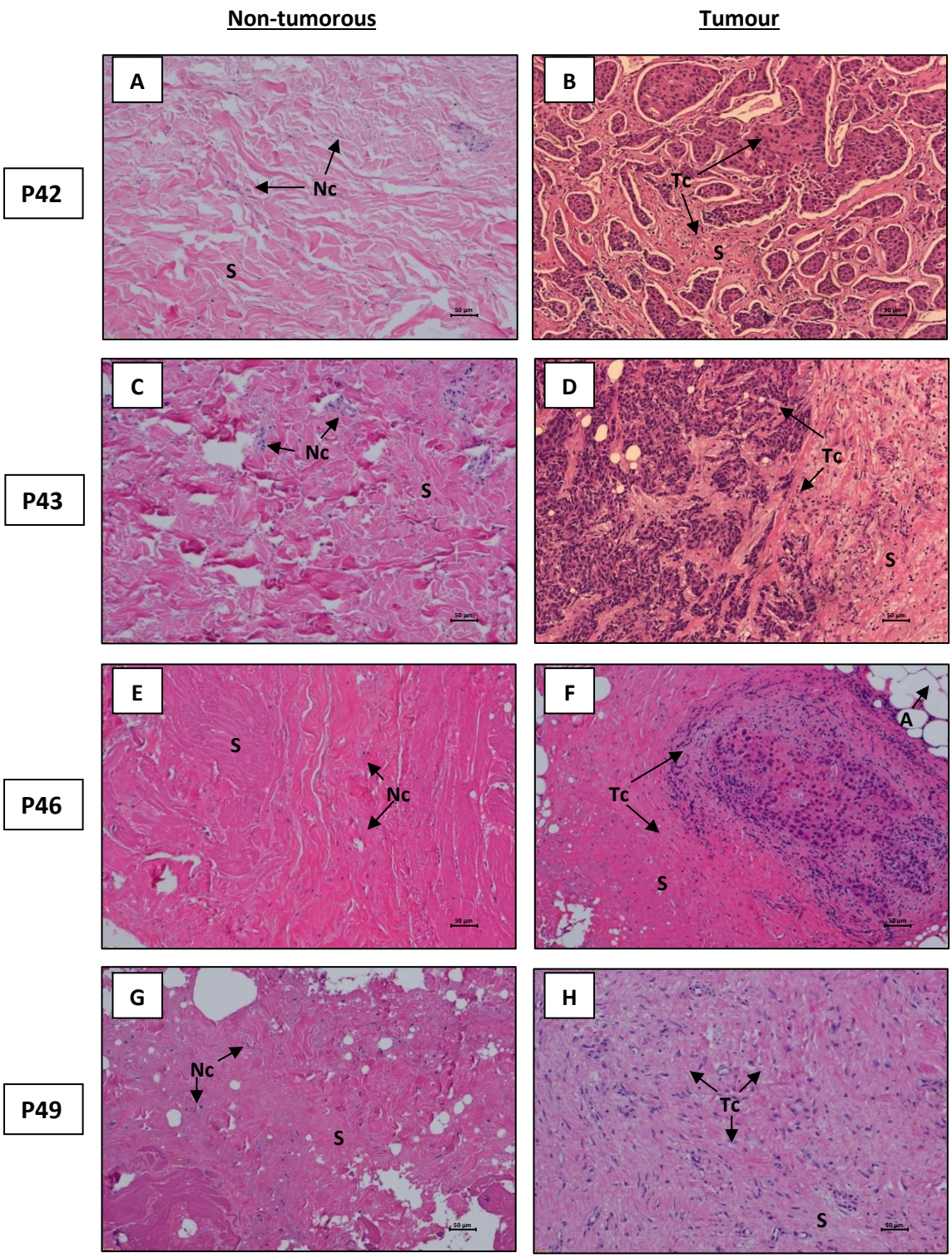

Figure 3.6: Histological analysis of patient P42 (A and B), P43 (C and D), P46 (E and F) and P49 (G and H) representing non-tumorous and tumour tissue stained with H&E and visualised at 10x magnification. Micrographs show non-tumorous breast tissue sections consisting of normal epithelial cells **(Nc)** that are surrounded by an organised stromal compartment **(S)** consisting of collagen fibres and other extracellular matrix components. Extensive pockets of invasion of malignant epithelial cells **(Tc)** into normal breast stroma with very few adipocytes **(A)** are visible in stained invasive ductal carcinoma tumour sections.

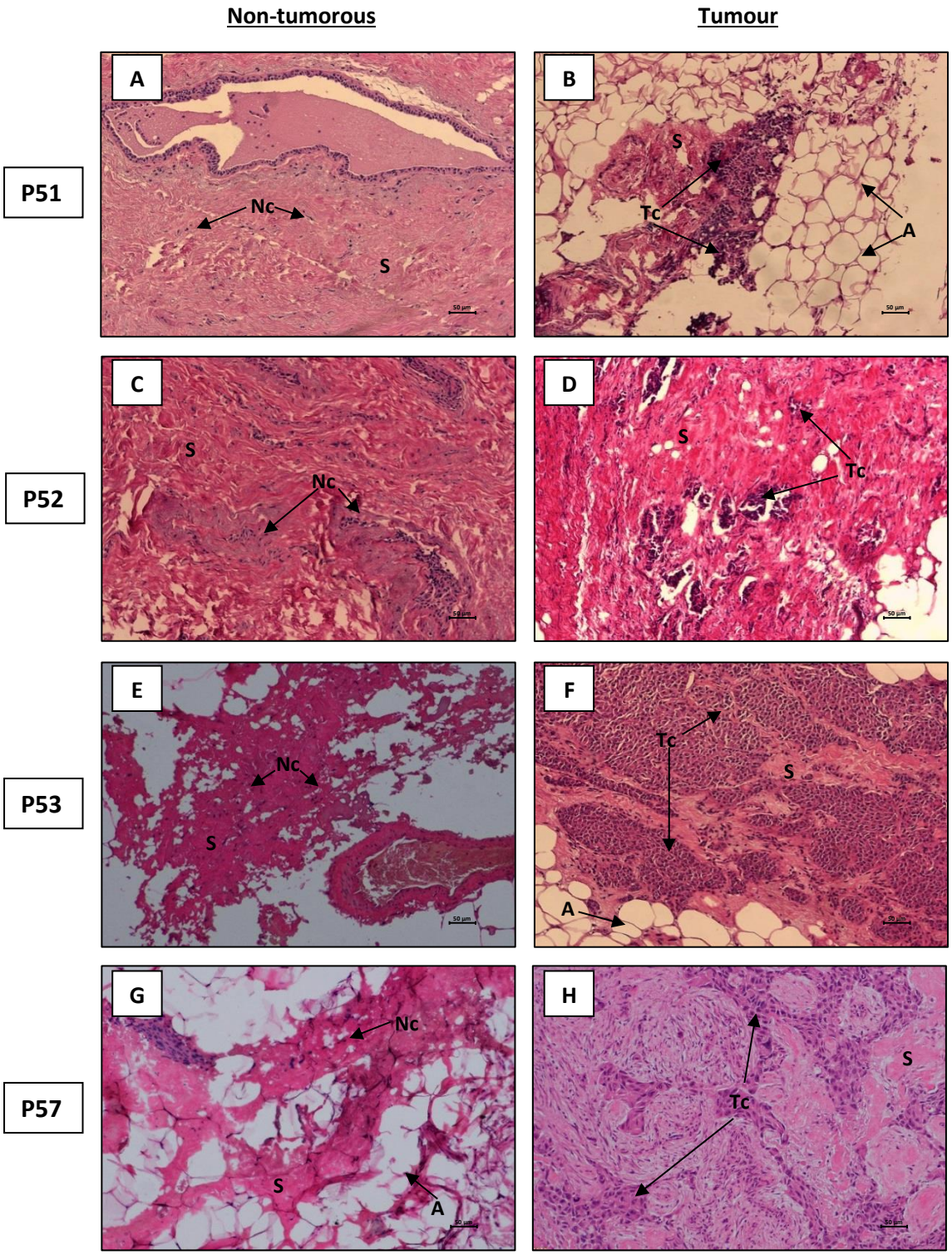

Figure 3.7: Histological analysis of patient samples P51 (A and B), P52 (C and D), P53 (E and F) and P57 (G and H) representing non-tumorous and tumour tissue stained with H&E and visualised at 10x magnification. Micrographs show non-tumorous breast tissue sections consisting of adipocytes **(A)** and normal epithelial cells **(Nc)** that are surrounded by an organised stromal compartment **(S)** consisting of collagen fibres and other extracellular matrix components. Extensive pockets of invasion of malignant epithelial cells **(Tc)** into normal breast stroma are visible in stained invasive ductal carcinoma tumour sections.

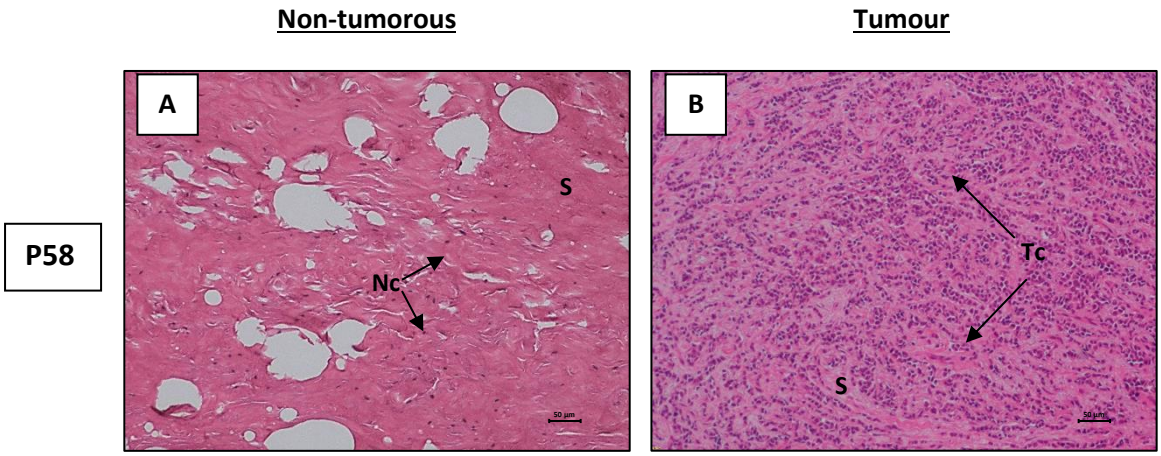

Figure 3.8: Histological analysis of patient sample P58 (A and B) representing adjacent non-tumorous and tumour tissue stained with H&E and visualised at 10x magnification. (A) showing a non-tumorous breast tissue section consisting of normal epithelial cells **(Nc)** surrounded by an organised stromal compartment **(S)** consisting of collagen fibres and other extracellular matrix components. (B) Extensive cell invasion of normal breast stroma by malignant epithelial cells are indicated **(Tc)** in a stained tumour section obtained from an invasive ductal carcinoma tumour.

Breast tissue composition is complex and heterogenous in nature with high inter-individual variation associated with the relative areas of epithelium, stroma and adipose tissue as illustrated in the above histological micrographs. Furthermore, studies have shown that breast composition is dynamic with changes occurring continuously with age and hormonal changes linked to the regression of epithelial and stromal compartments with an increase in proportions of adipose cells (Sun *et al.*, 2014; Taroni *et al.*, 2015). H&E-stained sections obtained from tumour samples clearly illustrate the dysregulated tissue structure in IDC, where numerous closely packed cellular nuclei, typical of uncontrolled cancer cell proliferation, form distinct dense cell clusters that vary in size and shape. Furthermore, extensive invasion into the surrounding stroma by clusters of malignant epithelial cells results in varied cell abundance within tumour stromal compartments that contrasts to the uniform appearance of equivalent non-tumorous stroma, as illustrated in the micrographs of non-tumorous breast tissue in Figures 3.1 -3.8.

Clinical patient information regarding tumour grading, hormone receptor expression (oestrogen (ER), progesterone (PR) and human epidermal growth factor 2 (HER-2)), Ki67 cellular proliferation index and lymph node status was obtained from patient medical reports. A summary of patient information is presented in Appendix V. Histological tumour sections were graded by an experienced pathologist using the Nottingham modification of the Bloom-

Richardson Index. Tumour grade assessment classified three (10%) of the 29 tumour samples as grade 1, nine (31%) as grade 2, eight (28%) as grade 3 and nine (31%) tumour samples were unspecified as shown in Figure 3.9A. Tumour grading is a crucial component of histopathological reports and is routinely utilised to predict patient prognosis and to determine treatment plans, with higher-grade tumours, commonly associated with poor long-term patient survival, characterised by poor cellular differentiation and uncontrolled cellular growth (Galea *et al.*, 1992; Rakha *et al.*, 2008).

Gene expression profiling studies have led to the molecular classification of breast cancer into six distinct subtypes namely Luminal A, Luminal B, basal-like (BLBC) or triple negative (TNBC), HER-2 overexpressing and normal-like breast tumours (Perou *et al.*, 2000). Molecular classification of breast cancer is used to determine patient prognosis and to provide vital information when determining personalised treatment regimens for breast cancer patients (Sørlie *et al.*, 2001). However, due to the high cost, the lack of general availability and technical complexities, gene expression profiling is not commonly implemented in many public clinical settings, as was the case in this study. In the absence of gene expression profiling data, alternative assessment involving immunophenotyping of ER, PR and HER-2 receptor expression and the Ki67 index of cellular proliferation was used to classify breast cancer subtypes. This immunohistochemical classification for breast cancer subtyping and treatment planning is endorsed by the St Gallen International Expert Consensus 2011 (Goldhirsch *et al.*, 2011; Goldhirsch *et al.*, 2013). Furthermore, several studies have provided support for the use of the Ki67 cellular proliferation index plus immunohistochemical characterisation of ER, PR and HER-2 expression as a surrogate for gene expression profiling for breast tumour subtyping (Cheang *et al.*, 2009; Mccafferty *et al.*, 2009; Al-Thoubaity, 2020). The immunophenotype criteria applied in this study for molecular classification is summarised in Table 3.1 and the various percentages of the four main breast cancer immunophenotypes, Luminal A, Luminal B, TNBC and HER-2 overexpressing that were identified in this study are summarised in Figure 3.9B.

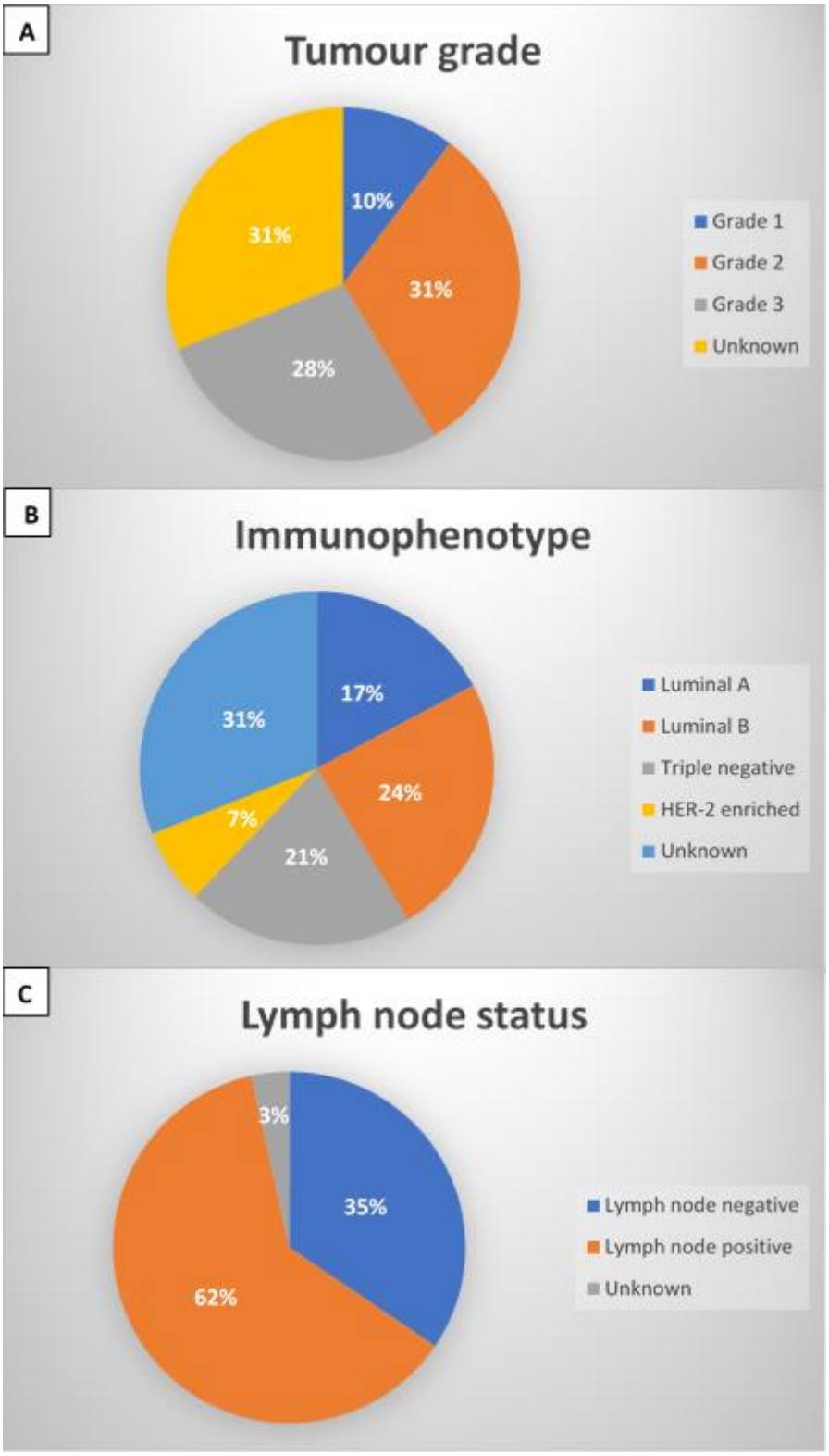

Figure 3.9: Pie graph summaries of (A) Tumour grade, (B) Immunophenotype and (C) Lymph node status expressed in terms of percentage patients diagnosed with breast carcinomas. Samples taken from patients who had missing information from their clinical files were grouped together into an unknown category. Tumour grades were determined using the Nottingham Modification of the Bloom-Richardson Index. Tumour subtyping was characterised using receptor expression profiles and the Ki67 proliferation index. Lymph node positive refers to the presence of cancer cells in lymph nodes and is associated with the risk of tumour metastasis.

Immunophenotyping classified five (17%) of the tumour samples as Luminal A, seven (24%) as luminal B, six (21%) as TNBC, two (7%) as HER-2 overexpressed and nine (31%) were unclassified due to insufficient data in clinical patient reports. Luminal A and B cancers are characterised by the positive expression of ER and PR hormonal receptors and negative HER-2 expression with a small proportion of luminal B cancers testing positive for HER-2. Ki67 is a nuclear protein that is upregulated during the proliferation phase of the cell cycle and commonly quantitated by the monoclonal antibody, MIB-1, to assess cellular proliferation rates in breast cancer (Urruticoechea *et al.*, 2005). Lower expression of Ki67 nuclear protein has been observed in Luminal A breast cancers compared to luminal B and is therefore used to discriminate between these two luminal type breast cancers (Cheang *et al.*, 2009; Mccafferty *et al.*, 2009). Although the St Gallen consensus endorses a Ki67 index of 20% as a cut-off, and was used in this study for distinguishing between Luminal A and B cancers, index variation between different laboratories does occur, with some studies using a Ki67 index of 14% as the distinguishing cut-off (Al-Thoubaity, 2020; Tsang and Tse, 2020).

Table 3.1: A summary of the immunohistochemical classification of samples into four breast cancer molecular subtypes (Goldhirsch *et al.*, 2011; Tsang and Tse, 2020)

Molecular subtype	Immunophenotype
Luminal A	ER and/or PR positive; HER-2 negative, Ki67 < 20% (low)
Luminal B (HER-2 negative) Luminal B (HER-2 positive)	ER and/or PR positive; HER-2 negative, Ki67 ≥ 20% (high) ER and/or PR positive; HER-2 positive, Ki67 > 2%
Triple negative	ER and PR negative; HER-2 negative
HER-2 overexpression	ER and PR negative; HER-2 positive
ER = oestrogen receptor; PR= progesterone receptor; HER-2= human epidermal growth factor receptor 2.	

The triple-negative breast cancer phenotype, characterised by the absence of both ER and PR hormonal receptors, as well as low HER-2 expression, is associated with poor prognosis and higher grading, which was evident in this study where TNBC tumours were either classified as grade 2 or 3 (See Appendix V). HER-2 overexpressing tumours, also identified by negative hormonal receptor expression but with higher HER-2 status, are also commonly of higher grades and this was observed in both HER-2 overexpressing tumours that were classified as grade 3 in this study. It must be stated that due to the heterogeneous nature of breast cancer, immunohistochemical-based molecular classification is considered to be an over-simplified approach with limitations for tumour subtyping and is therefore only used to predict tumour behaviour, treatment modalities or patient prognosis in a clinical setting (Inic *et al.*, 2014; Zhao *et al.*, 2015; Soliman and Yussif, 2016). Furthermore, additional immunohistochemical identification of basal cytokeratin markers such as CK5 and CK6, which are located in the basal region of the myo-epithelial cell layer in breast tissue, is advised to provide a more comprehensive assessment of tumour subtypes, specifically for basal-like or triple negative breast cancers (Choccalingam *et al.*, 2012; Al-Thoubaity, 2020). However, basal cytokeratin markers are not yet routinely assessed during pathology analysis of tumour biopsies from cancer patients by the NHLS and hence the characterisation of these additional tumour marker was considered to be a limitation for distinguishing between breast cancer molecular subtypes in this study.

Lymph node involvement was observed in 18 (62%) patients, with no lymph node involvement seen in 10 (35%) patients and nodal status was unknown for 1 (3%) patient as shown in Figure 3.9C. Lymph node involvement refers to the presence of cancer cell migration from the primary breast tumour into the draining lymph nodes located in the axillary region of the body. Cancer cell infiltration of axillary lymph nodes indicates metastasis having occurred, with nodal status used to determine prognosis and the requirement for systemic adjuvant therapy for cancer patients (Colleoni *et al.*, 2005). Several factors, such as age, tumour size and tumour grade have been reported to be associated with an increased risk of positive lymph node disease in invasive breast cancers (Abner *et al.*, 1998; Gajdos *et al.*, 1999; Aquino *et al.*, 2017). Several other cohort studies have reported a high percentage of axillary metastasis in patients diagnosed with IDC similar to the observations in this study, (Borst and Ingold, 1993; Keihanian *et al.*, 2019; Ahadi *et al.*, 2020).

3.2 Proteomic analysis

3.2.1 Sample preparation and proteome coverage optimisation

Due to the cellular and tissue complexity of the breast and especially the breast tumour microenvironment, as described in Chapter 1, many technical challenges may arise when attempting to perform proteomic characterisation studies of protein mixtures that include highly soluble and less soluble proteins like those of the ECM. Mass spectrometry (MS) technology has evolved and improved significantly over the years, thus making liquid chromatography tandem mass spectrometry (LC-MS/MS) an indispensable analytical tool to characterise the proteome of complex biological samples. However, the key to any successful MS based proteomic experiment is robust and reproducible sample preparation which involves protein extraction or solubilisation, digestion and sample clean-up (Klont *et al.*, 2018). Extraction and solubilisation of proteins are key steps in ensuring successful downstream analysis, as only soluble proteins can be accurately quantified. In this study, a sample preparation protocol that effectively extracted and solubilised breast tissue proteins, using a high concentration of sodium dodecyl sulphate (SDS) in conjunction with pressure cycling technology (PCT) extraction method developed specifically for histological sections, followed by an automated trypsin digestion protocol using a liquid handling system and standardised LC-MS/MS analysis were performed to assess the most efficient method in terms of the depth of proteome coverage for sample analysis. Four replicates of non-tumorous breast tissue were initially used for each sample preparation method and the efficiency of protein extraction was gauged by monitoring the number of identified peptides and proteins when using the same protein concentrations after the barocycling extraction. A summary of protocol parameters and the results, comparing the efficiency of the methods are summarised in Table 3.2. The protein concentration of each sample extract following the barocycler extraction for all the samples analysed are given in Table A1 in Appendix VI. The two methods are referred to as “one-step barocycler method” and “two-step barocycler method”.

The one-step barocycler method was found to be the less efficient method with 2958 peptides and 469 proteins identified by LC-MS/MS in the frozen 60 µm thick histology sections as summarised in Table 3.2. There was a significant increase in the number of LC-MS/MS

identified peptides and proteins, 7716 and 1663 respectively, when using the two-step barocycler method. The volume of extract required for samples extracted using the two-step barocycler method was almost half of that required when using the one-step barocycler method, due to a higher total protein extraction for the two-step barocycler method. The extracts were subjected to exactly the same clean-up and digestion steps using an automated HILIC-based bead coupled protocol to perform tryptic digestion. The released peptides were analysed by a standardised DIA based LC-MS/MS method. The barocycler temperature (95°C) as well as pressure (45 Kpsi) were the same for both extraction methods. However, the number of pressure cycles differed between the two sample preparation methods, whereby the one-step barocycler method consisted of only 35 PCT cycles while the two-step barocycler method consisted of 70 PCT cycles with an additional final high-pressure cycle of 60 minutes.

Table 3.2: Sample analysis method efficiency showing protocol parameters with the total number of peptides and proteins identified for each method.

Sample analysis	**Protocol parameters**				**No. peptides or proteins identified**	
Method	**Temperature (°C)**	**Pressure (Kpsi)**	**No. of cycles**	**Duration of 1 cycle**	**No. Peptides**	**No. Proteins**
One-step barocycler method + LC-MS/MS	95	45	35	High pressure for 30 seconds and low pressure for 15 seconds	2958	469
Two-step barocycler method + LC-MS/MS	95	45	70	High pressure for 30 seconds and low pressure for 15 seconds	7716	1163
	95	45	1	High pressure for 60 minutes		

The increase in the number of proteins identified with the two-step barocycler method suggests that an increase in the number of high-pressure cycles and increased cycle time improved total protein extraction and solubilisation, leading to a significant improvement in protein coverage. Making use of a lysis buffer containing a high concentration of SDS and

subjecting protein samples to PCT, with repeated cycles of ambient and high hydrostatic pressure at high temperatures resulted in the rapid destabilisation of intermolecular interactions between cellular components that aided in protein solubility (Gross *et al.*, 2008). As mentioned previously, sample preparation is a crucial step in any proteomic experiment but is also a labour-intensive multistep process that is associated with the highest contribution to experimental variation. Studies have reported that the programmable multi-sample processing barocycler instrument can minimise possible sample degradation and potential technical variation, which is commonly seen in biological sample preparation methods used prior to LC-MS/MS analysis, by enabling automated batchwise tissue homogenisation and protein solubilisation in a temperature-controlled vessel exposed uniformly to repeated low- and high-pressure cycles (Shao *et al.*, 2016; Lucas *et al.*, 2018).

3.2.2 Block design for sample analysis

Sample analysis "blocking", illustrated in Figure 3.10, was done according to tumour receptor expression profiles to minimise potential experimental bias. Matched samples were allocated into four blocks and samples within individual blocks were processed and analysed groupwise.

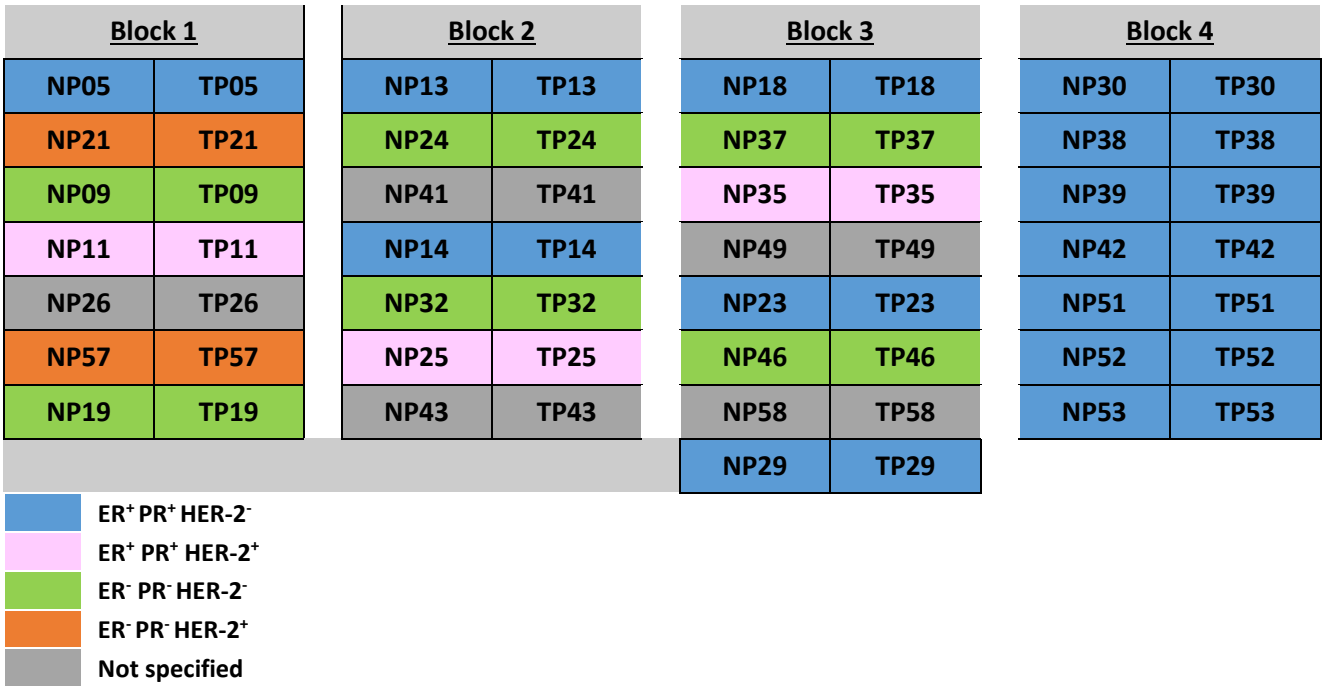

Figure 3.10: Sample block design for analysis of patient samples according to tumour receptor expression profiles. The matched patient samples were allocated into four blocks with pairing of non-tumorous (NP) and tumour (TP) tissue samples obtained from each individual patient.

3.2.3 SWATH-based quantification

Peptide samples from individual non-tumorous and tumour tissue were analysed using data-independent acquisition (DIA) or Sequential Window Acquisition of all Theoretical Mass Spectra™ (SWATH). SWATH Analysis consisted of acquisitions of 70 fragment ion (MS/MS) scans of overlapping sequential precursor isolation windows with variable m/z isolation width, 1 m/z overlap and high sensitivity mode selected, covering the 400 to 900 m/z mass range, with a precursor MS scan for each cycle. The accumulation time was 250 ms for the MS scan between the range of 400 to 900 m/z and 25 ms for each product ion scan with a range of 100 to 1800m/z, thus making a 2.05 sec total cycle time. Data processing, involved the conversion of raw SWATH (*.wiff format) data files into Spectronaut™ compatible files and subsequent analysis carried out using default settings in Spectronaut™ 11 software.

A comprehensive project specific spectral library was initially generated using two dimensional (2-D) RP-RP LC-MS/MS data dependent analysis of pooled samples and used as a reference for data extraction and protein identification. Two-dimensional RP chromatography is widely used in proteomic experiments due to its ability to improve the chromatographic resolution and enhance the depth of analytical data obtained from complex biological systems (Yang *et al.*, 2012). Moreover, the orthogonal 2-D separation of peptides based on their hydrophobicity at different pH is often used to simplify the complexity of protein samples consisting of a wide dynamic range of analytes (Yang *et al.*, 2012). Although 2-D proteomic analysis is quite labour intensive, this approach can be extremely beneficial in ECM proteomic studies, as it is well known that in-depth analysis of the ECM is challenging due to its complexity, low solubility of several components, with low abundant ECM constituents spanning a wide concentration range further complicated by the presence of high abundance cellular components within the same protein sample (Byron *et al.*, 2013; Ma *et al.*, 2019). Furthermore, generation of a project specific protein library reduced the number of possible false and unlikely protein hits that lead to a more confident proteome coverage, as several studies have highlighted the added benefit of using project specific libraries generated for SWATH-MS experiments (Zi *et al.*, 2014; Griss, 2016; Govaert *et al.*, 2017).

The spectral library consisted of a total of 17939 peptides associated with 3066 proteins that were validated at a 1.0% False Discovery Rate (FDR) estimated using the decoy hit distribution. A post-analysis overview of protein identifications for each sample can be found in Appendix

VII and a summary of peptide and protein recovery for individual samples are summarised in the bar graphs in Figure 3.11. The average number of peptides and proteins identified for non-tumorous tissue was 6449 and 1475 respectively. There was almost 40% more peptides and proteins matched to the spectral library for tumour samples with an average of 9836 peptides and 2126 proteins.

Figure 3.11: Bar graphs illustrating the number of (A) peptides and (B) proteins that were matched to the project specific spectral library for non-tumorous (blue) and tumour breast tissue samples (orange).

3.2.4 Post analysis overview and data quality control

3.2.4.1 Technical variation

Relative quantitation of protein levels in biological samples forms a central part of proteomic studies, however the accurate measurement of protein abundance may be affected by many factors including overall technical variation and experimental bias that may occur at any point during the experimental workflow (Molloy *et al.*, 2003). Therefore, the level of technical variation associated with each step in the experimental design needs to be established and appropriate measures need to be taken to minimise the effects of this variation prior to any analysis of the clinical samples. Ultimately, the technical variance in any proteomic experiment needs to be smaller than the expected biological variation to ensure that accurate and reliable conclusions can be drawn from the analytical data. Studies have shown that increasing the number of biological replicates, obviously taking into consideration sample availability and associated costs, may significantly reduce the overall experimental variation leading to more robust and reliable analytical data (Oberg and Vitek, 2009).

In this study, the level of technical variation was determined by assessing four sequential histological slices of non-tumorous breast tissue, also referred to as technical replicates, that were obtained from the same patient. Technical replicates were processed independently, but in parallel and taken through the complete experimental workflow from sample preparation to MS analysis. The complete optimised experimental workflow displayed a low technical variation with a median coefficient of variance (CV) for precursors (9.8%), peptides (10.2%) and proteins (9.1%) as shown in the graphs in Figure 3.12 (A-C). The percentage of precursors (71.0%), peptides (68.6%) and proteins (75.4%) with an acceptable CV of less than 20% was also calculated. The optimised workflow from sample processing to final analysis was deemed to have acceptable technical variation and the same complete workflow was used for analysis of the clinical samples without further optimisation, as the acceptable cut-off for technical CV for clinical mass spectrometry workflows is ≤20% (Lynch, 2016; Fu *et al.*, 2018).

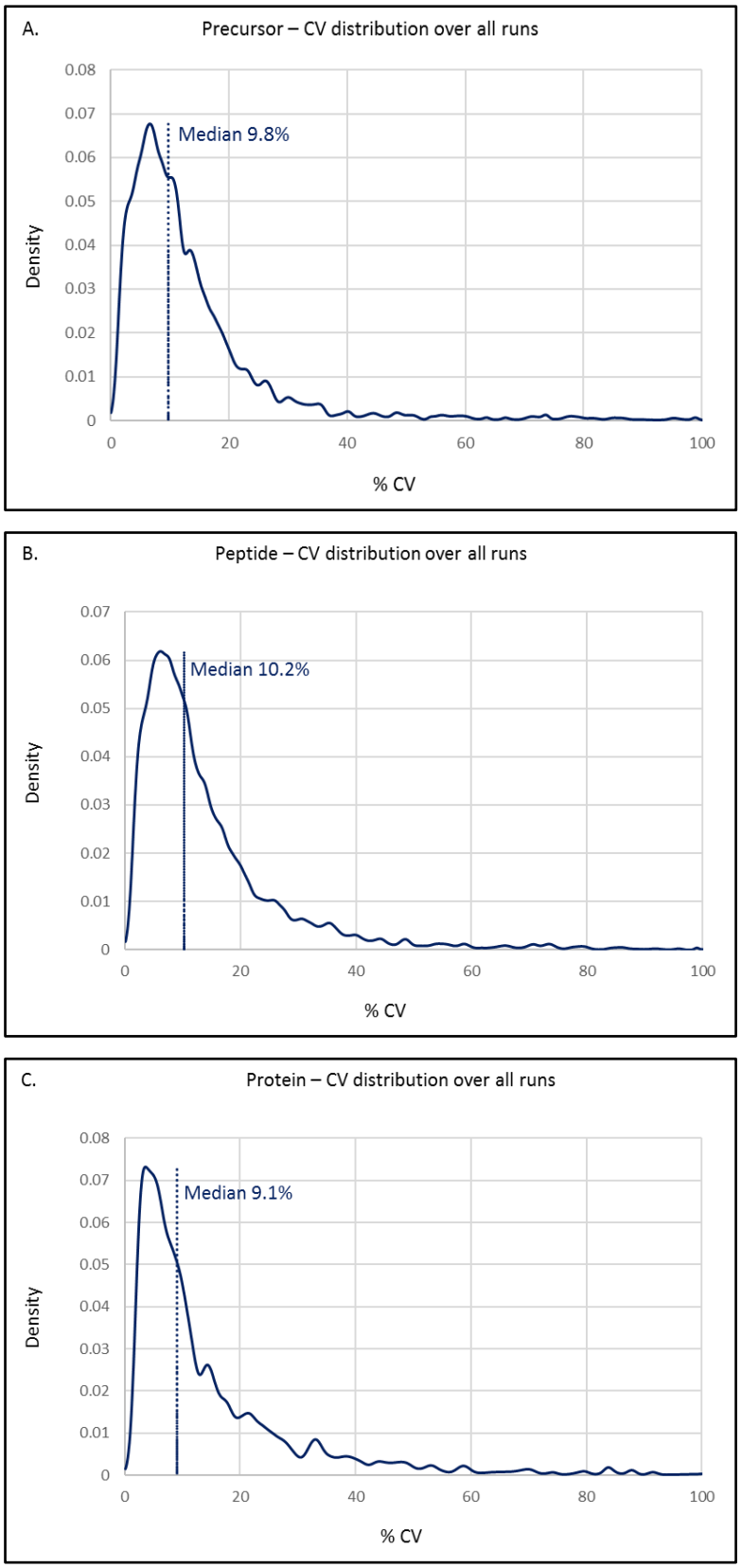

Figure 3.12: Coefficient of variation (CV) distribution for (A) Precursors, (B) Peptides and (C) Proteins for non-tumorous histological slices that were used to assess technical variation. The CV median value is indicated for precursors (9.8%), peptides (10.2%) and proteins (9.1%) for all technical runs.

3.2.4.2 Post-analysis normalisation

Systemic variation is inherent to any proteomic workflow, therefore post-analysis normalisation is an essential strategy that is routinely used in proteomic analysis to ensure that analytical data obtained from individual patient samples can be comparable to each other. (Välikangas *et al.*, 2018). Data normalisation, in addition to using a sample analysis blocking design, ensured that observed differences in protein abundance were most likely to result from true biological changes and not due to method-introduced bias or technical variation. Post-acquisition data normalisation using the median of the total precursor ion intensity, as shown in Figure 3.13, is commonly used in label-free quantification and was thus implemented in this study to ensure accurate protein quantitation (O'rourke *et al.*, 2019).

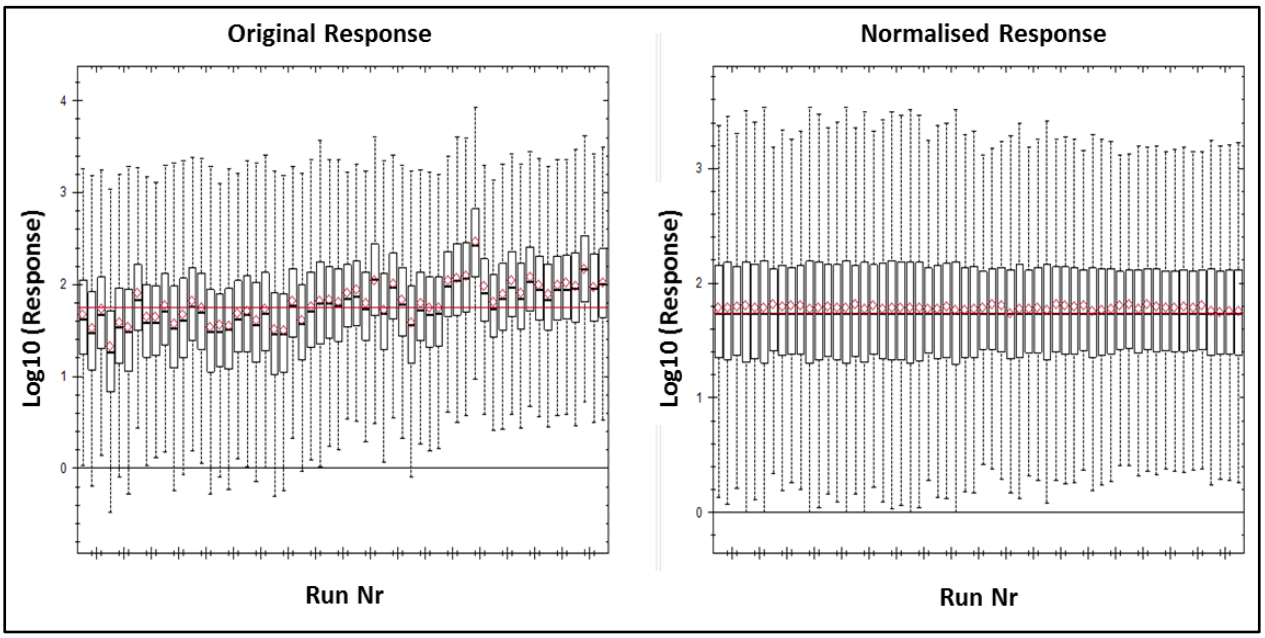

Figure 3.13 Post-acquisition data normalisation. Data normalised using the median value of the total ion intensity with individual sample whisker plots showing the raw (left) and normalised (right) responses.

3.2.4.3 Biological variation and sample correlation

Biological variation was also assessed in this study and a 42% CV median of distribution was reported for non-tumorous samples and a 51% CV for tumour samples as shown in Figure 3.14. Biological variance is not only influenced by genetic factors, but qualitative and quantitative changes in breast tissue composition are also linked to age, body mass index (BMI), menopausal status, use of oral contraceptives, previous pregnancies or lactation status

(Sun *et al.*, 2014; Sandhu *et al.*, 2016). Therefore, inherent heterogeneity of biological samples, especially in human tissue samples extracted from numerous patients of diverse ethnic backgrounds, was expected to be relatively high in this clinically based proteomic study. Furthermore, tumour heterogeneity was evident from the differences in closely surrounding tissue type, tumour size, tumour grade, immunophenotypes and nodal status of the tumours obtained from patients, although all were diagnosed with IDC to be included in this study. This heterogeneity was expected to contribute significantly to the collective biological variance observed in the tumour sample group (Turashvili and Brogi, 2017).

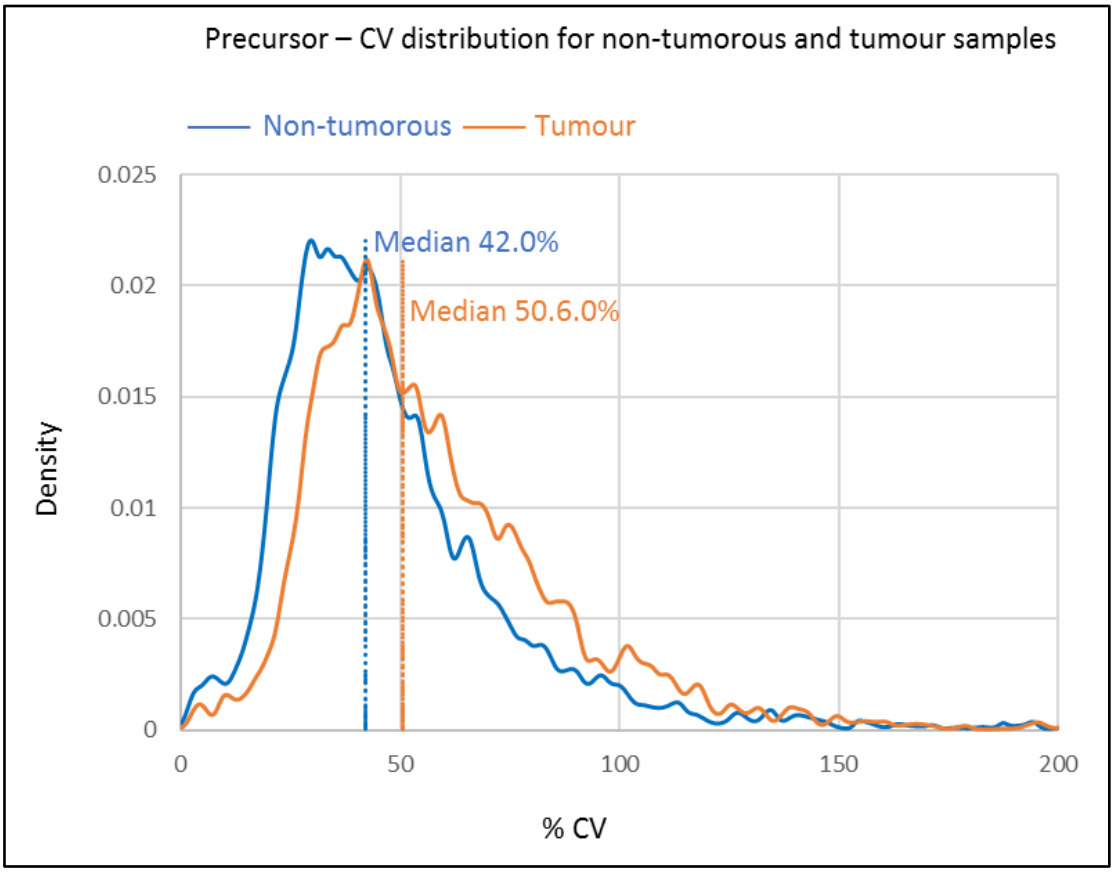

Figure 3:14. Coefficient of variation (CV) distribution for precursor ion intensities for non-tumorous and tumour samples. The CV median value for non-tumorous samples was 42% and 50.6% for tumour samples.

A correlation heatmap (Figure 3.15) showing the relationship between sample datasets analysed in this study, where the red colour indicates a stronger correlation and blue colour represents a weaker correlation. A stronger correlation was observed for non-tumorous samples, which could be expected due to the more homogenous nature of non-tumorous

tissue compared to that of the different types and stages of the tumour tissue. Significant heterogeneity in tumour morphology was observed in this study and this would explain the weaker correlation between different tumour samples as shown in Figure 3.15.

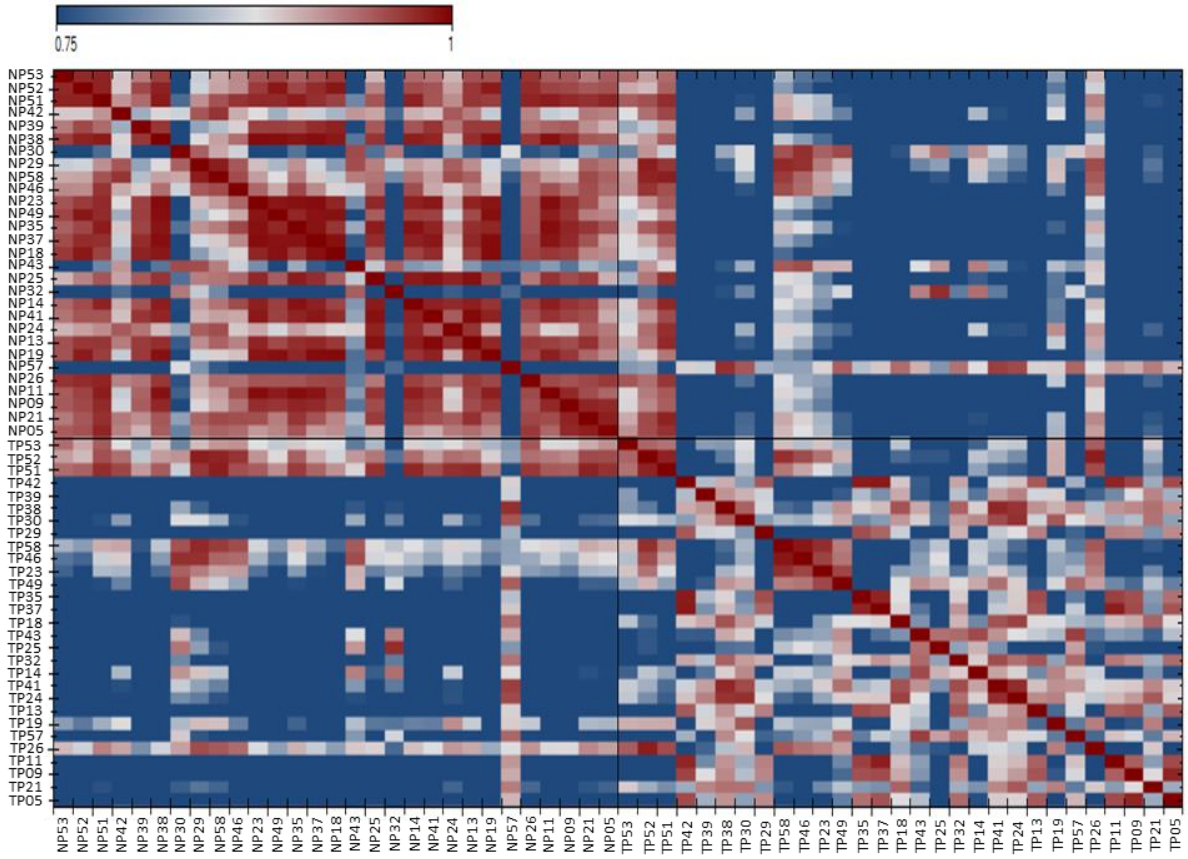

Figure 3:15. Sample correlation heatmap illustrating the relationship between an individual sample dataset compared to every other sample dataset in this study. Red indicates a strong correlation while blue shows a weak or distant correlation between datasets. With a few obvious individual exceptions, a stronger correlation is observed between the non-tumorous sample group, whereas the tumour sample group showed a weaker correlation.

3.2.4.4 Statistical analysis

A statistical power analysis was conducted using MSstats software version 3.13.7 to estimate the number of patient samples required, at a power of 0.8 and 0.01% FDR, for a two-fold change in abundance for differentially expressed proteins. Based on the computed results, 58 samples consisting of 29 non-tumorous and 29 matched tumour samples obtained from patients who were diagnosed with IDC were included in this study. Taking into consideration the measured sample variability and cohort size, a minimum two-fold change was considered

to be significant when assessing differentially expressed proteins. A volcano plot (log_2 fold change vs log_{10} p-value) was generated for comparison analysis using a t-test performed on the log_2 ratio of protein intensities measured for individual proteins and the data was plotted as shown in Figure 3.16. Multiple-hypothesis correction was applied by correcting p-values using the q-value approach to control FDR (Storey, 2002). Candidate proteins were selected only if a log_2 fold abundance change of greater than or equal to 1 (2-fold difference), a q-value of less than or equal to 0.01% (1% FDR) was reported and where 2 or more unique peptides were reported with 100% confidence for that specific protein. A final list of all candidate proteins meeting the above-mentioned criteria is included in Appendix VIII. All proteins with a q-value below 0.01 % were exported directly into Perseus v1.6.1.1, to generate hierarchical clustering and principal component analysis (PCA) plots. Only differentially expressed proteins were functionally annotated and subject to further pathway enrichment analysis using Cytoscape_v3.8.0.

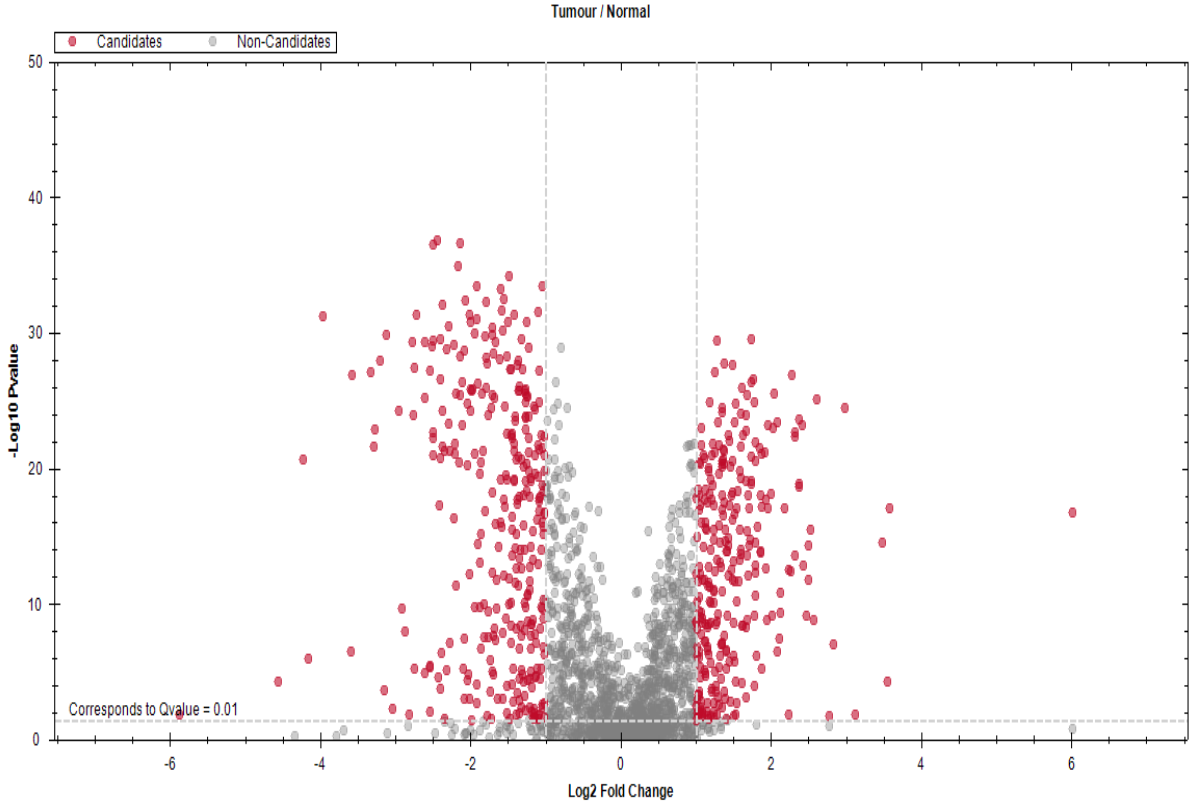

Figure 3.16: Volcano plot of proteins of interest after comparing all tumour samples to all non-tumorous samples. Non-tumorous samples were indicated as the reference or control. A single red point represents an identified protein of interest that corresponds to a q-value ≤ 0.01, Log_2 fold-change ≥ 1 (2-fold difference) and having at least 2 unique peptides.

PCA plots are extremely useful for the analysis of large datasets and are often used to visualise the grouping of samples as well as identify potential outliers or data variation, thus serving as a tool for quality control (Ringnér, 2008). A PCA plot for non-tumorous and tumour samples is illustrated in Figure 3.17 and clearly shows two distinct groupings consisting of non-tumorous tissue and tumour samples which were compared in this study. However, patient samples NP30, NP32, NP42, NP43, NP57, TP46, TP51, TP52 and TP58 failed to group within their respective sample groups as shown in Figure 3.17. Patient tumour samples TP46, TP51, TP52 and TP58 grouped closely with non-tumorous samples and these tumour samples also had a higher correlation with non-tumorous samples as shown in the sample correlation heatmap in Figure 3.15. This may be due to several reasons including the relative proportion of non-tumorous to tumour tissue in the histological slice or biological contamination from adjacent non-tumorous tissue may have influenced the proteomic characterisation of these tumour sections. Histological analysis of TP51 (Figure 3.7B) clearly shows a small cluster of tumour cells surrounded by a relatively larger proportion of adipose tissue, TP46 (Figure 3.6F) and TP52 (Figure 3.7D) shows small clusters of tumour cells that are surrounded by relatively larger areas of epithelial tissue and stroma thus offering a possible reason for the non-grouping of tumour sample with its respective tumour tissue group. Phenotypic characteristics may also have played a role as patient pathology reports confirmed that TP51, TP52 and TP58 samples were Grade 1 tumours with TP46 unspecified. Tumour grade is used to describe tumours or cancer cells in relation to the appearance of normal cells and how likely the tumour is to proliferate and metastasise (Rakha *et al.*, 2008). Grade 1 tumours are phenotypically more similar to normal cells than higher grade tumours and this could offer a possible explanation as to why TP51, TP52 and TP58 clustered with the non-tumorous group of samples in this study. Biological contamination may also be a possible explanation as to why patient samples NP30, NP32, NP43, and NP57 failed to cluster with the non-tumorous samples. Tumour metastasis was reported in all patient samples except for patient no. 57 (not specified) with tumours either assigned a Grade 2 or Grade 3 except for patient no. 32 (not specified), thus indicating rapid and aggressive tumour growth. It may be suggested, although with caution, that the non-tumorous section resected from each of these patients may have been "contaminated" with tumour cell exudates that may have possibly influenced the phenotypical characteristics of these non-tumorous sample, thus influencing the grouping of samples. The sample correlation heatmap in Figure 3.15 also showed that non-tumorous

samples NP30, NP32 and NP57 correlated better with tumour samples than the non-tumorous group of samples. Technical errors regarding lower recovery were also considered, whereby a reduced number of peptides and proteins matched to the spectral library was observed for NP42 (Figure 3.11) and this could be a possible reason for the non-grouping of NP42 with non-tumorous samples as shown in the PCA plot in Figure 3.17

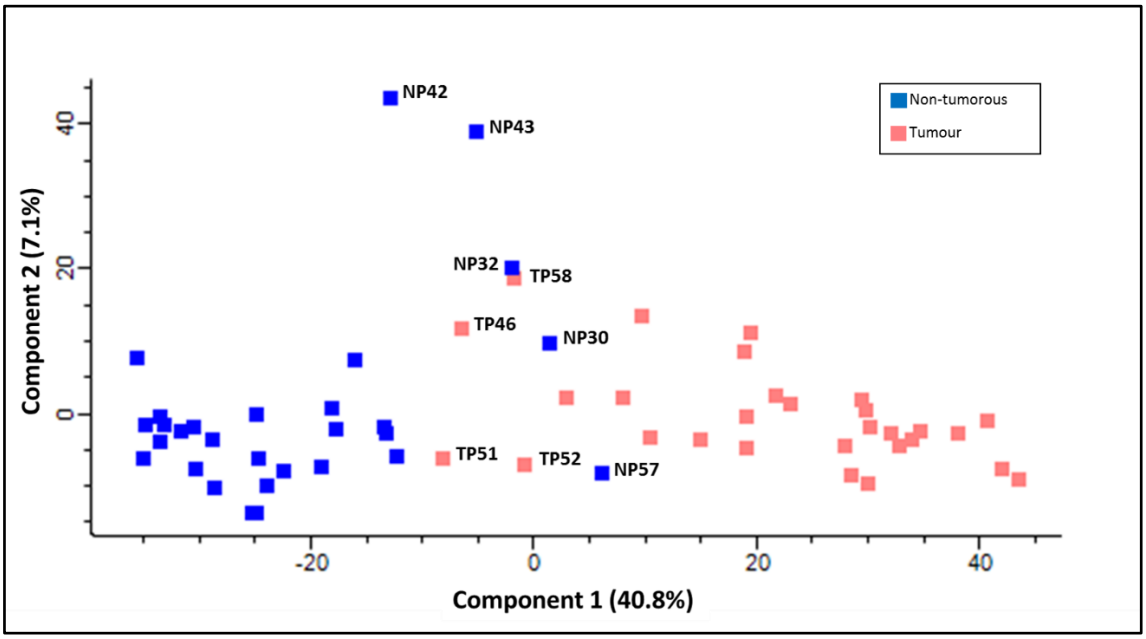

Figure 3.17: Principal component analysis of non-tumorous indicated in blue and tumour samples indicated in pink. Samples that failed to group with their respective sample groups are annotated.

3.2.5 Annotation of biological pathways involved in tumorigenesis

Proteomic experiments are known to generate large datasets which need to be analysed in several ways including within a biological context to ensure relevant interpretation. Cytoscape software provides several analytical tools, such as STRING, that can group and visualise proteins within large and complex networks according to their function or biological pathways in which the proteins are known to be involved. The STRING tool was used to functionally annotate the proteomic data generated in this study based on known protein-protein interactions and experimental data obtained from various resources. Integration of STRING networks with experimental data using Cytoscape software is an effective way to perform pathway analysis that allows for visualisation of up- or downregulated biological pathways, protein-protein interactions and highlighting biologically relevant abundance

changes for individual proteins in specific biological pathways (Doncheva *et al.*, 2018). A list of 613 candidate proteins, meeting the previously described identification criteria together with their associated average $\log_2$ ratios, representing differential abundance between sample groups, was exported directly into Cytoscape to generate a STRING enrichment analysis of significantly altered protein expression that could be associated with different biological pathways as shown in Figure 3.18. Individual proteins of biological pathways are represented by nodes and the various protein-protein associations between specific nodes are represented by black lines which are referred to as edges. The change in relative abundance of each protein present in tumour samples compared to non-tumorous samples is represented by the blue nodes indicating down-regulation or red nodes, representing up-regulation of protein expression. Clustered nodes represent many proteins with strong associations.

Although there were many proteins that showed large fold changes, either as a multifold increase or decrease in abundance between the non-tumorous tissue and tumours, many of these were "isolated" proteins that did not link to any other proteins within specific metabolic pathways and were considered as orphan or random increases that could not be explained through the upregulation of specific processes or biological effects.

Another aspect that became evident was that the very low abundant proteins appeared to show very high fold changes in abundance during the comparison of the non-tumorous tissue verses tumours. This appeared to be due to the low abundance or even where the protein was not detected in some of the samples but was detected in some of each group. This caused distorted fold change calculations which could be eliminated by manual inspection. A third aspect that showed very high fold changes were from samples that had obvious contamination by extraneous fluid such as blood where the haemoglobin sub-units were shown to have fold change ratios in the thousands although these are obviously from blood contamination in a limited number of samples.

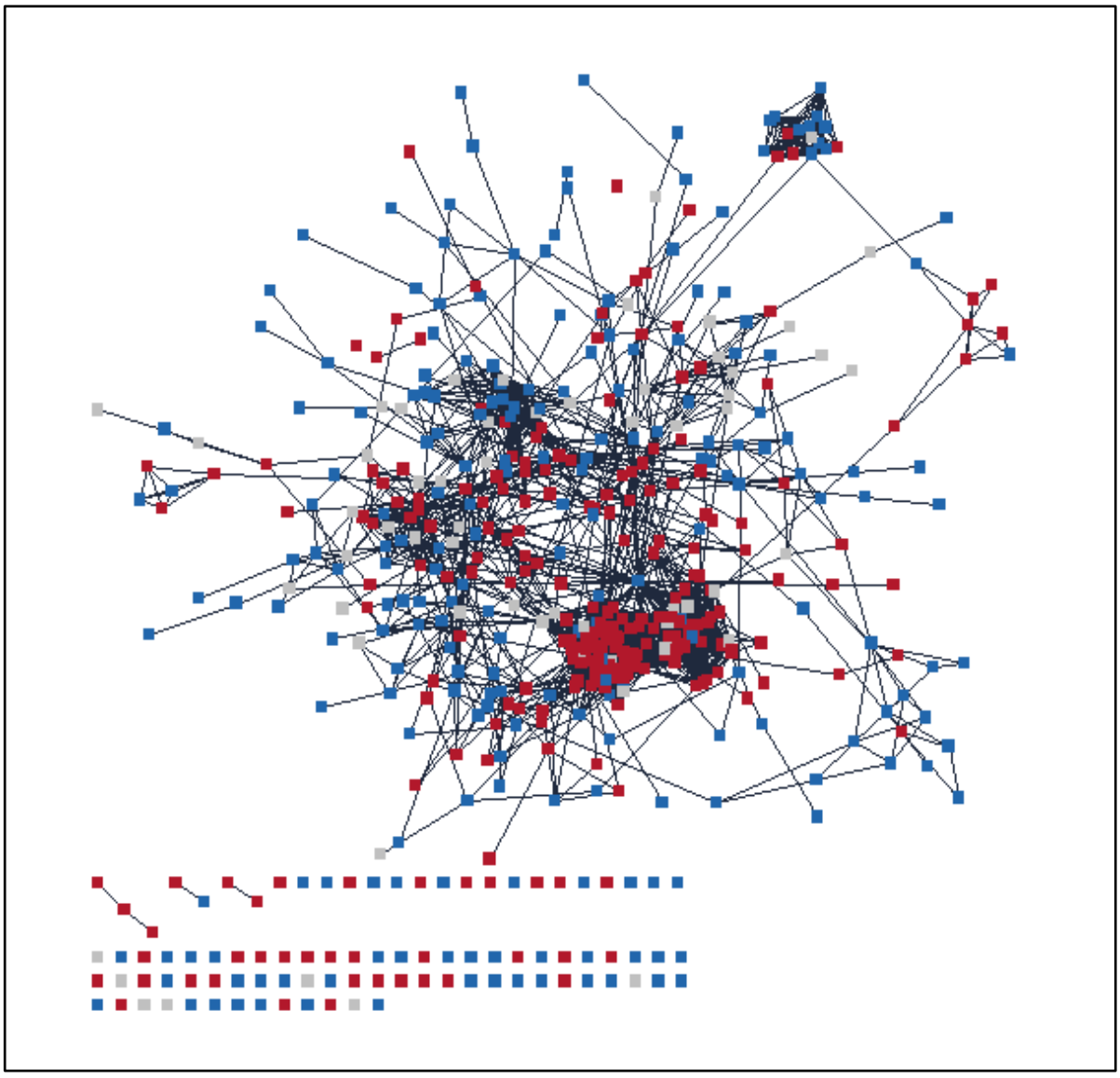

Figure 3.18: STRING network diagrams illustrating the complete functional enrichment analysis for significantly expressed protein candidates that were identified in this study. Individual proteins in biological pathways are represented by nodes and the various protein-protein associations between specific nodes are represented by black lines which are referred to as edges. The relative abundance changes of each protein present in tumour samples compared to non-tumorous samples is represented by blue indicating down-regulation or red, representing up-regulation of protein expression.

Enrichment analysis of KEGG pathways was carried out in order to identify key biological pathways that are up- or down-regulated in breast tumour samples versus non-tumorous breast tissue. The enrichment analysis identified a number of pathways which included ECM receptor interactions and cellular processes associated with protein translation, splicing machinery as well as protein homeostasis and quality control. Protein translation and splicing machinery consisting of RNA and splicing proteins were significantly enriched and are

represented by the two large red clusters that consist of upregulated proteins illustrated in the main STRING network in Figure 3.18. These cancer related pathways have also been identified in another proteomic based study that involved the proteomic profiling of luminal breast cancer progression (Pozniak *et al.*, 2016). The identified molecular networks are listed with their respective FDR-values, number of genes, nodes and edges and each pathway is briefly discussed below. The ECM receptor pathway will be discussed in conjunction with the other ECM related sub-networks identified as this is the main focus of this study.

Table 3.3: List of prominent cancer related pathways. Identified molecular pathways are listed with their respective FDR-values, number of genes, nodes and edges.

Pathway description	FDR value	Number of genes	Number of nodes	Number of edges
Ribosome pathway	1.66E-34	52	52	1242
Extracellular matrix receptor interaction	2.40E-11	22	22	103
Spliceosome pathway	6.99E-9	23	23	226
Endoplasmic reticulum protein processing	3.6E-9	21	21	41

3.2.5.1 Ribosome pathway

The eukaryotic ribosome, referred to as the 80S ribosome, is made up of a 40S small subunit and a 60S large subunit with both subunits playing a unique role in protein synthesis, by either decoding messenger RNA or facilitating peptide bond formation (Ben-Shem *et al.*, 2011)**.** The smaller 40S subunit consists of an 18S ribosomal RNA (rRNA) component with 33 ribosomal proteins (RPs) and the larger 60S subunit is made up of the 5S, 5.8S and 28S rRNAs together with 47 RPs (Ramakrishnan, 2011). Synthesis of ribosomes is a complex process that involves RNA polymerases (polymerase I, II and III) and consists of coordinated steps involving the 80 ribosomal proteins that are structural components of ribosomes as well as acting as RNA chaperones to facilitate the correct folding of rRNAs needed for the assembly of mature ribosomal subunits. Furthermore, RPs have a fundamental role in ribosome biogenesis but also have extra-ribosomal functions associated with regulation of cellular proliferation and apoptosis (Xu *et al.*, 2016).

Amplified ribosomal biogenesis is critical to meet the increased protein synthesis demands of malignant cells to sustain and drive uncontrolled cell proliferation and tumour progression (Sulima *et al.*, 2017; Penzo *et al.*, 2019)**.** Dysregulation or hyperactivation of the ribosomal biogenesis pathway, as a result of oncogenic activity or the loss of tumour suppressor genes, such as retinoblastoma protein or p53, is associated with uncontrolled cellular growth, neoplastic transformation as well as cancer cell migration and tissue invasion (Xu *et al.*, 2016). Oncogenic activity of c-Myc, an oncoprotein that induces protein synthesis, is overexpressed in several cancers. c-Myc upregulates protein synthesis by promoting RNA polymerases I, II and III activity which consequently results in increased rRNA synthesis and mRNA translation (White, 2005; Xu *et al.*, 2016). Furthermore, defects in ribosomal biogenesis and function can lead to congenital disorders collectively known as ribosomopathies that are linked to an increased risk of developing cancer in adulthood (Narla and Ebert, 2010; Sulima *et al.*, 2017). In this study, an upregulation of numerous proteins associated with ribosomal biosynthesis was observed in tumour samples when compared to non-tumorous breast tissue as shown in Figure 3.19. A total of 52 RPs and ribonuclear proteins showing at least two-fold changes in abundance were identified and a list of these proteins is summarised in Table 3.4. It is of interest that only one protein in the whole group was down regulated and that this protein is associated with mitochondrial ribosomal protein 1 despite the other mitochondrial ribosomal proteins being up-regulated.

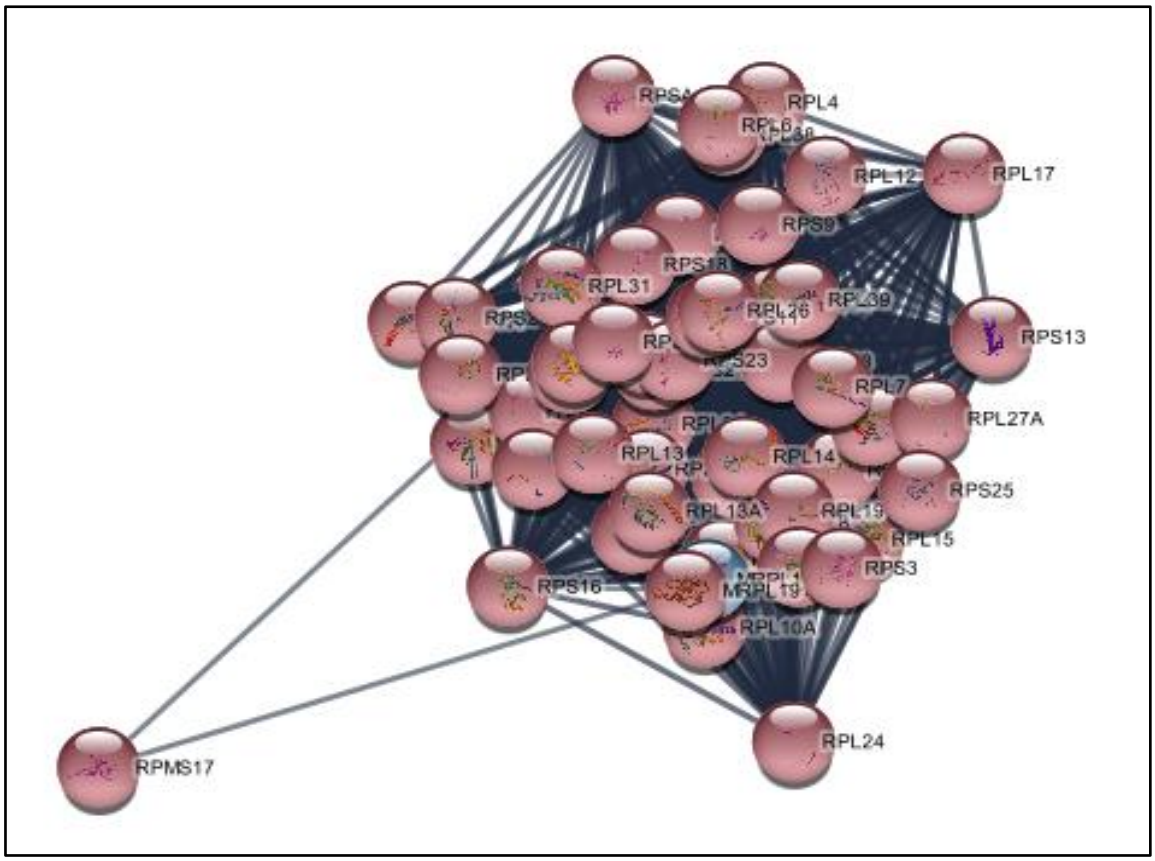

Figure 3.19: STRING network showing significant differentially expressed ribosomal proteins indicating their relative abundance in tumour samples compared to non-tumorous samples. Each protein is represented by a node, which are labelled with their corresponding gene names, and connected by black lines that represent relevant protein-protein interactions. Nodes are either blue representing down-regulation or red indicating up-regulation of the specific protein expression.

Table 3.4: A list of differentially expressed ribosomal proteins in tumour samples compared to non-tumorous breast tissue. Uniprot accession numbers, gene names corresponding to the above STRING network (Figure 3.22), average $\log_2$ ratios and q-values are listed in the table below. Negative values indicate down-regulation and positive values indicate up-regulation of protein expression.

Uniprot Accession number	Gene name	Protein description	Average Log_2 Ratio	q-value
P62861	FAU	40S ribosomal protein S30	2.38	8.28E-24
Q9BYD6	MRPL1	39S ribosomal protein L1 (mitochondrial)	-2.53	2.15E-06
Q9Y3B7	MRPL11	39S ribosomal protein L11 (mitochondrial)	1.38	1.10E-04
P49406	MRPL19	39S ribosomal protein L19 (mitochondrial)	1.40	3.07E-04
P62906	RPL10A	60S ribosomal protein L10a	1.53	6.24E-25
P30050	RPL12	60S ribosomal protein L12	1.10	2.59E-21
P26373	RPL13	60S ribosomal protein L13	1.85	7.94E-22
P40429	RPL13A	60S ribosomal protein L13a	1.73	2.17E-15
P50914	RPL14	60S ribosomal protein L14	1.07	3.43E-23
P61313	RPL15	60S ribosomal protein L15	1.36	9.08E-21

P18621	RPL17	60S ribosomal protein L17	1.72	1.10E-17
Q07020	RPL18	60S ribosomal protein L18	1.69	1.83E-25
P84098	RPL19	60S ribosomal protein L19	1.41	1.18E-07
P46778	RPL21	60S ribosomal protein L21	1.67	5.39E-23
P62750	RPL23A	60S ribosomal protein L23a	1.29	4.20E-10
P83731	RPL24	60S ribosomal protein L24	1.68	4.18E-24
P61254	RPL26	60S ribosomal protein L26	1.35	6.00E-21
P61353	RPL27	60S ribosomal protein L27	1.63	9.00E-23
P46776	RPL27A	60S ribosomal protein L27a	1.24	2.17E-15
P46779	RPL28	60S ribosomal protein L28	1.35	4.55E-08
P47914	RPL29	60S ribosomal protein L29	1.57	1.80E-12
P62899	RPL31	60S ribosomal protein L31	1.74	1.55E-19
P49207	RPL34	60S ribosomal protein L34	1.02	4.36E-04
P42766	RPL35	60S ribosomal protein L35	2.41	1.90E-23
Q9Y3U8	RPL36	60S ribosomal protein L36	1.67	1.48E-19
P61513	RPL37A	60S ribosomal protein L37a	1.24	6.38E-11
P63173	RPL38	60S ribosomal protein L38	1.96	1.98E-23
Q59GN2	RPL39	60S ribosomal protein L39	2.98	1.36E-24
P36578	RPL4	60S ribosomal protein L4	2.28	8.81E-27
Q02878	RPL6	60S ribosomal protein L6	1.92	1.54E-21
P18124	RPL7	60S ribosomal protein L7	1.60	3.51E-24
P62917	RPL8	60S ribosomal protein L8	1.56	7.64E-19
Q9Y2R5	RPMS17	28S ribosomal protein S17	1.03	8.50E-08
P46783	RPS10	40S ribosomal protein S10	1.25	1.60E-21
P62280	RPS11	40S ribosomal protein S11	1.48	3.08E-13
P62277	RPS13	40S ribosomal protein S13	2.38	2.23E-19
P62263	RPS14	40S ribosomal protein S14	1.79	3.98E-13
P62841	RPS15	40S ribosomal protein S15	1.34	5.09E-08
P62249	RPS16	40S ribosomal protein S16	1.03	5.73E-19
P62269	RPS18	40S ribosomal protein S18	1.74	2.73E-26
P39019	RPS19	40S ribosomal protein S19	1.38	9.37E-21
P15880	RPS2	40S ribosomal protein S2	1.33	3.16E-15
P63220	RPS21	40S ribosomal protein S21	1.23	3.75E-22
P62266	RPS23	40S ribosomal protein S23	1.26	5.14E-27
P62847	RPS24	40S ribosomal protein S24	1.47	1.33E-20
P62851	RPS25	40S ribosomal protein S25	1.80	5.56E-21
P62854	RPS26	40S ribosomal protein S26	1.45	2.40E-22
P23396	RPS3	40S ribosomal protein S3	1.18	5.08E-25
P62701	RPS4X	40S ribosomal protein S4	1.29	1.38E-23
P62081	RPS7	40S ribosomal protein S7	1.37	9.90E-22
P46781	RPS9	40S ribosomal protein S9	1.62	5.62E-26
P08865	RPSA	40S ribosomal protein SA	1.08	4.89E-22

These results are in agreement with studies that have reported a significant enhancement of various RPs in prostate cancer whereby small ribosomal proteins, RPS19, RPS21, RPS24 and RPS2 were detected at significantly higher levels in cancerous human prostate tissue compared to non-malignant tissue (Wang *et al.*, 2009; Arthurs *et al.*, 2017). Moreover, RPL15 which is a structural protein of the large ribosomal subunit was also found to be overexpressed in breast tumour tissue when compared to non-tumorous breast tissue. Overexpression of RPL15 has also been previously reported in a breast cancer *in vivo* genomic study where it was shown that the increase in RPL15 gene expression was associated with breast cancer metastasis in multiple organs and was directly linked to enhanced synthesis of other ribosomal proteins. The same study also showed that activation of RNA polymerase I and ribosome biogenesis was increased in aggressive human breast cancer cell types (Belin *et al.*, 2009). Expression of the RPS19 protein was observed to be increased by 2.6 fold in the breast tumour samples, supporting a previous finding where an increased RPS19 was seen in pre-clinical studies involving breast and ovarian cancer cells (Markiewski *et al.*, 2017). Furthermore, upregulation of RPS19 has been reported in colon cancer and gene mutations of RPS19 have been linked to an increased risk of developing cervical cancer (Safaeian *et al.*, 2012; Kimura *et al.*, 2020). It has been suggested that RPS19 may potentially be considered a therapeutic drug target due to its immunosuppressive properties involving interaction with complement C5a receptor 1 in the tumour microenvironment and by also promoting immunosuppressive cytokines such as transforming growth factor beta (TGFβ) (Markiewski *et al.*, 2017).

The above-mentioned proteins are just a few examples of RPs that were found to be upregulated in breast tumour samples and the role they may play in cancer development was briefly highlighted as it is beyond the scope of this current study. The overall upregulation of the ribosome pathway and its protein components observed in tumorous breast tissue is in agreement with similar findings reported in literature and provides further evidence that components from the ribosomal biosynthesis pathway should be explored further in breast cancer research (Xu *et al.*, 2016; Penzo *et al.*, 2019).

3.2.5.2 Spliceosome pathway

Gene transcription involves the synthesis of pre-mature messenger RNA molecules that are converted to mature messenger RNA (mRNA) through a process called RNA splicing. RNA

splicing is an essential and intricate post-transcriptional process catalysed by a ribonuclear protein complex known as the spliceosome, whereby non-coding base sequences (introns) are excised and coding sequences (exons) are merged before mRNA molecules are translated into functional proteins (Wahl *et al.*, 2009; Wang and Aifantis, 2020). Several proteins representing the spliceosome were identified in this study as shown in the STRING network in Figure 3.20 and listed with their log2 ratio changes in Table 3.5.

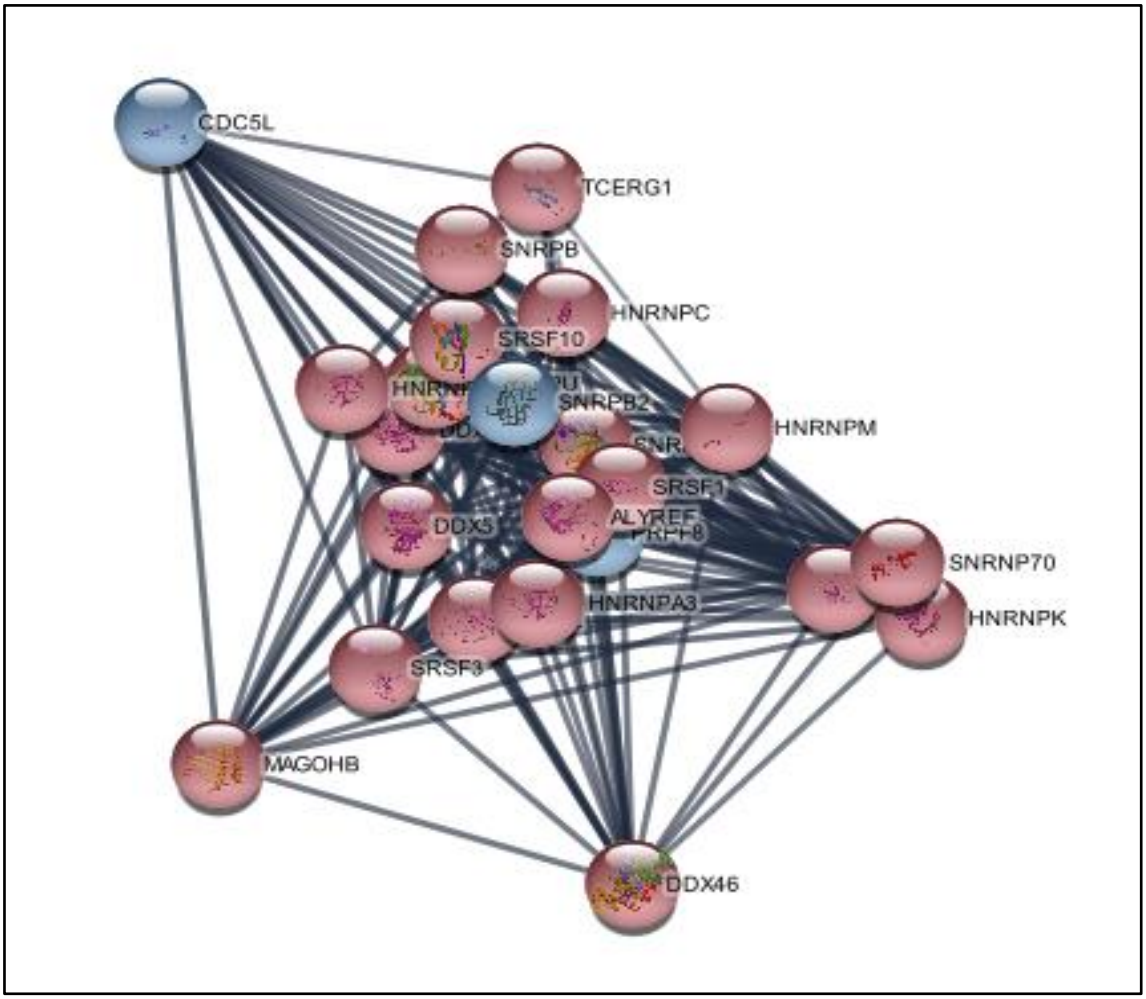

Figure 3.20: STRING network showing significant differentially expressed spliceosome proteins indicating their relative abundance in tumour samples compared to non-tumorous samples. Each protein candidate is represented by nodes, which are labelled with their corresponding gene names, and are connected by black lines that represent relevant protein-protein interactions. Nodes are either coloured in blue to illustrate down-regulation or red to indicate up-regulation in protein expression.

The spliceosome is important in protein synthesis as variable forms of mRNA are generated through alternative RNA splicing and leads to numerous proteins with unique functions being

synthesised from a single gene (Kelemen *et al.*, 2013). The spliceosome is a large multiprotein complex that consists of small nuclear ribonucleoproteins (snRNPs) as well as small nuclear RNA proteins (snRNA) that interact with various splicing proteins such as heterogenous nuclear ribonucleoproteins (hnRNPs) and serine/arginine-rich proteins where each class of proteins playing a unique role in the mechanisms associated with RNA splicing (Will and Lührmann, 2011).

Table 3.5: A list of differentially expressed spliceosome proteins in tumour samples compared to non-tumorous breast tissue. Uniprot accession numbers, gene names corresponding to the above STRING network (Figure 3.23), average log_2 ratios and q-values are listed in the table below. Negative values indicate down-regulation and positive values indicate up-regulation of protein expression.

Uniprot Accession number	Gene name	Protein description	Average Log2 Ratio	q-value
Q86V81	ALYREF	THO complex subunit 4	2.11	1.87E-08
Q99459	CDC5L	Cell division cycle 5-like protein	-1.29	1.51E-20
Q13838	DDX39B	Spliceosome RNA helicase DDX39B	1.38	1.43E-27
Q7L014	DDX46	Probable ATP-dependent RNA helicase DDX46	1.31	9.44E-04
P17844	DDX5	Probable ATP-dependent RNA helicase DDX5	2.05	1.40E-25
P09651	HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1	1.22	3.07E-19
P51991	HNRNPA3	Heterogeneous nuclear ribonucleoprotein A3	1.88	9.73E-18
P07910	HNRNPC	Heterogeneous nuclear ribonucleoproteins C1/C2	1.86	1.47E-18
P61978	HNRNPK	Heterogeneous nuclear ribonucleoprotein K	1.34	1.55E-13
P52272	HNRNPM	Heterogeneous nuclear ribonucleoprotein M	2.13	1.19E-11
Q00839	HNRNPU	Heterogeneous nuclear ribonucleoprotein U	1.60	9.46E-20
Q96A72	MAGOHB	Protein mago nashi homolog 2	1.29	4.19E-03
Q6P2Q9	PRPF8	Pre-mRNA-processing-splicing factor 8	-1.10	4.93E-03
P08621	SNRNP70	U1 small nuclear ribonucleoprotein 70 kDa	1.38	1.39E-08
P14678	SNRPB	Small nuclear ribonucleoprotein-associated proteins B, B’ and N	1.16	4.50E-18
P08579	SNRPB2	U2 small nuclear ribonucleoprotein B''	-2.31	3.67E-06
P62318	SNRPD3	Small nuclear ribonucleoprotein Sm D3	1.04	1.61E-12
Q07955	SRSF1	Serine/arginine-rich splicing factor 1	1.20	1.14E-06
O75494	SRSF10	Serine/arginine-rich splicing factor 10	1.53	3.86E-03
Q01130	SRSF2	Serine/arginine-rich splicing factor 2	1.16	3.43E-12
P84103	SRSF3	Serine/arginine-rich splicing factor 3	1.29	1.17E-09
Q16629	SRSF7	Serine/arginine-rich splicing factor 7	1.70	6.49E-13
O14776	TCERG1	Transcription elongation regulator 1	2.24	4.12E-03

Pathway enrichment analysis showed an upregulation of the spliceosome pathway and most of its associated protein components thus suggesting an increase in splicing activity in tumour tissue compared to non-tumorous breast tissue. Although RNA splicing is a vital process needed for enhancing protein complexity and functionality needed for cellular survival and differentiation, it is well known that irregularities associated with this dynamic and complex process may lead to various human diseases including cancer (Srebrow and Kornblihtt, 2006; Liu and Cheng, 2013; Wang and Aifantis, 2020). Tumour cells are able to alter and exploit the RNA splicing process resulting in the synthesis of cancer-specific protein isoforms that promote cancer cell survival and tumour growth (Escobar-Hoyos *et al.*, 2019; Wang and Aifantis, 2020). For example, upregulation of proto-oncogenes such as c-Myc is directly linked to an increase in the transcription of snRNP genes and other splicing factors involved in RNA splicing thus impacting splicing regulation and protein control (Koh *et al.*, 2015).

Several RNA splicing proteins belonging to the hnRNP family that have been linked to cancer development (Geuens *et al.*, 2016) were found to be upregulated in the breast tumour tissue compared to non-tumorous tissue in this study, as shown in Figure 3.20. Heterogenous nuclear ribonucleoprotein C (hnRNPC) is a highly abundant nuclear protein consisting of two isoforms, hnRNPC1 and hnRNPC2, that play a fundamental role in pre-RNA processing and RNA splicing (Geuens *et al.*, 2016). Several studies have reported abnormally high levels of hnRNPC and highlighted the role that hnRNPC may play in breast cancer, lung adenocarcinoma, ovarian cancer, hepatocellular carcinoma and glioblastoma (Sun *et al.*, 2007; Park *et al.*, 2012; Kleemann *et al.*, 2018; Sarbanes *et al.*, 2018; Guo *et al.*, 2021). Heterogenous nuclear ribonucleoprotein K (hnRNPK) is a multi-functional protein associated with tumour formation and metastasis, with elevated protein levels being reported in breast cancer, colon cancer, oral squamous cell carcinoma and prostate cancer (Lu and Gao, 2016). An average 4-fold increase in protein expression of heterogenous nuclear ribonucleoprotein M (hnRNPM) was found in breast tumour tissue when compared to non-tumorous breast tissue as shown in Table 3.5. Increased protein expression of hnRNPM has been detected in breast carcinomas and is associated with enhanced cellular proliferation, tumour growth and increased oncogenic activity linked to breast cancer progression via the axin/β-catenin signalling pathway (Yang *et al.*, 2018). Furthermore, increased expression of hnRNPM has been reported in aggressive forms of breast cancer and is associated with breast cancer

metastasis via hnRNPM-mediated alternative splicing mechanisms involved in promoting epithelial- mesenchymal transition (Xu *et al.*, 2014b). Increased expression of hnRNPM has also been found in colorectal carcinomas and identified as a potential biomarker for colorectal cancer (Chen *et al.*, 2014).

Furthermore, alternative splicing also results in the generation of specific splice variants of ECM proteins, such as fibronectin, tenascin, versican and osteopontin, that all play a key role in ECM remodelling, and are associated with cancer development and progression (Weidle *et al.*, 2011; Robertson, 2016; Efthymiou *et al.*, 2020). Spliceosome-targeted breast cancer therapies are being explored as the spliceosome and its splicing proteins are extensively involved in several hallmarks of cancer including angiogenesis, uncontrolled proliferation, invasion and metastasis, evading the immune system and growth suppressor signals, genomic instability as well as unchecked cellular replication (Coltri *et al.*, 2019; Koedoot *et al.*, 2019).

3.2.5.3 Endoplasmic reticulum protein processing pathway

The intracellular endoplasmic reticulum plays an essential role in many cellular processes that include lipid biosynthesis and calcium homeostasis (Oakes and Papa, 2015). In addition to these biological processes, the endoplasmic reticulum is considered to be a specialised cellular organelle involved in the synthesis, folding, post-translational modification and quality control of proteins that is required before secretory and membrane proteins are exported to their final destination located either in the cellular or extracellular compartment (Wang and Kaufman, 2014; Oakes and Papa, 2015). Accumulation of misfolded proteins occurs when proteins fail to pass the endoplasmic reticulum quality control process, which activates the endoplasmic reticulum associated degradation (ERAD) pathway. This results in proteasomal degradation of misfolded proteins in order to maintain endoplasmic reticulum and protein homeostasis (Moon *et al.*, 2018). However, endoplasmic reticulum stress may occur if the capacity to carry out protein folding, due to the increasing protein synthesis, is exceeded or when the ERAD system is overwhelmed due to the accumulation of misfolded proteins (Avril *et al.*, 2017). Endoplasmic reticulum stress is a common feature in cancer as protein synthesis is significantly increased in tumour cells in order to meet the demands of uncontrolled cellular proliferation and tumorigenic processes that are promoted by

oncogenic activation (Oakes, 2020). Upregulation of the endoplasmic reticulum protein processing pathway associated proteins was observed in the breast tumour tissue analysed in this study and is shown as a STRING network in Figure 3.21 and listed with log2 ratios in Table 3.6.

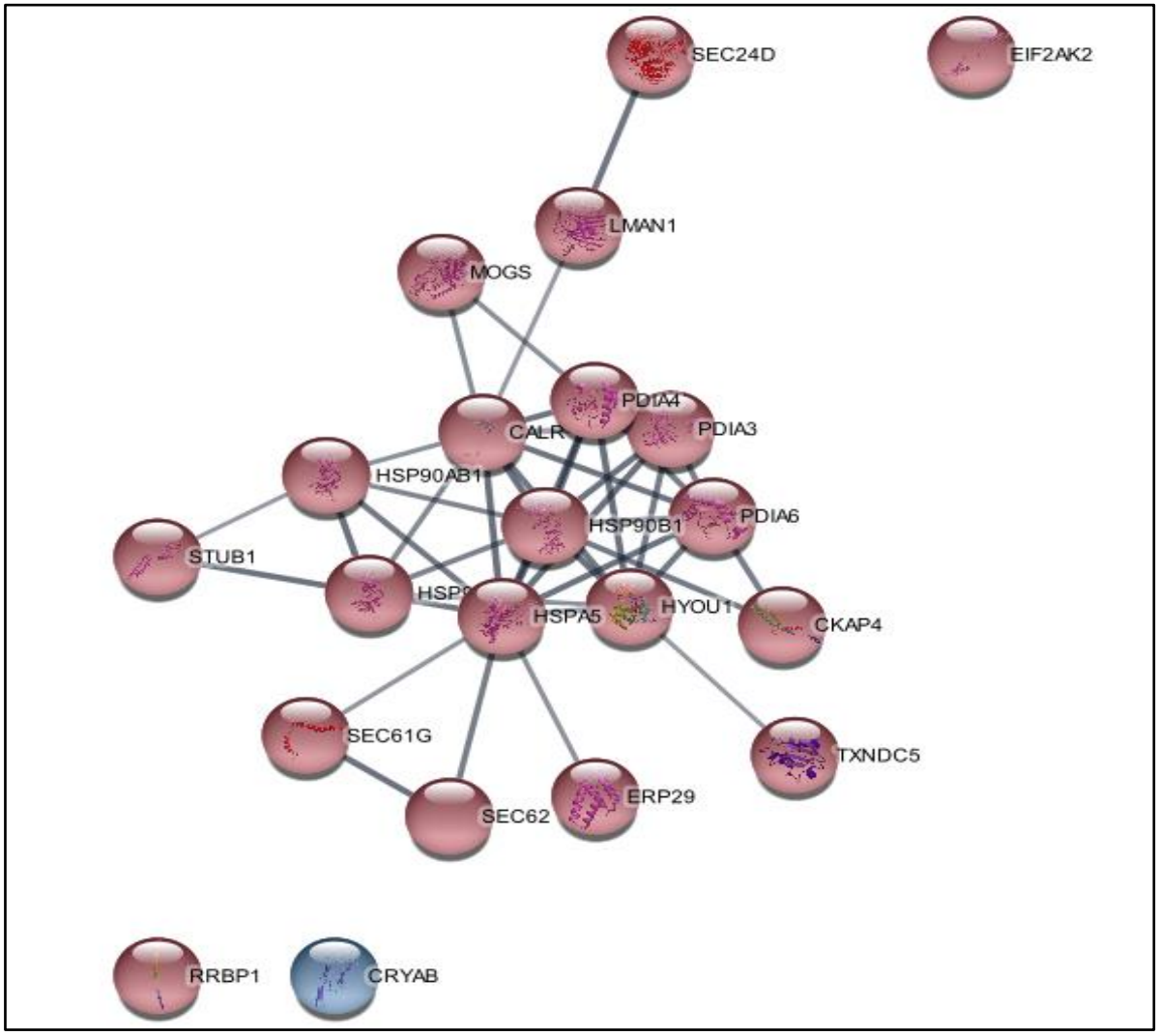

Figure 3.21: STRING network showing significant differentially expressed proteins associated with endoplasmic reticulum protein processing and quality control indicating their relative abundance in tumour samples versus non-tumorous samples. Each protein candidate is represented by nodes, which are labelled with their corresponding gene names, and connected by black lines that represent relevant protein-protein interactions. Nodes are all red indicating up-regulation in expression of all these endoplasmic proteins.

The unfolded protein response (UPR) is an adaptive response that is activated to alleviate endoplasmic reticulum stress and to maintain homoeostasis by reducing the accumulation of improperly folded proteins by temporarily limiting protein translation, increasing the transcription of endoplasmic reticulum chaperone proteins involved in protein folding, to

amplify the ERAD processes and if endoplasmic reticulum stress remains uncontrolled, triggering apoptosis (Hetz *et al.*, 2015; Oakes, 2017).

Table 3.6: A list of differentially expressed endoplasmic reticulum proteins involved in protein processing and homeostasis in the analysed tumour samples compared to non-tumorous breast tissue. Uniprot accession numbers, gene names corresponding to the above STRING network (Figure3.24), average log_2 ratios and q-values are listed in the table below. Negative values indicate down-regulation and positive values indicate up-regulation of protein expression.

Uniprot Accession Number	Gene name	Protein description	Average Log2 Ratio	q-value
P27797	CALR	Calreticulin	1.15	3.00E-09
Q07065	CKAP4	Cytoskeleton-associated protein 4	2.08	1.25E-23
P19525	EIF2AK2	Interferon-induced, double-stranded RNA-activated protein kinase	1.87	1.24E-14
P02511	CRYAB	Alpha-crystallin B chain	-2.13	1.87E-25
P30040	ERP29	Endoplasmic reticulum resident protein 29	1.35	2.85E-24
P07900	HSP90AA1	Heat shock protein HSP 90-alpha	1.12	3.16E-18
P08238	HSP90AB1	Heat shock protein HSP 90-beta	1.41	1.27E-14
P14625	HSP90B1	Endoplasmin	1.33	9.23E-14
P11021	HSPA5	Endoplasmic reticulum chaperone BiP	1.20	1.02E-19
Q9Y4L1	HYOU1	Hypoxia up-regulated protein 1	1.61	1.11E-14
P49257	LMAN1	Protein ERGIC-53	1.18	2.21E-12
Q13724	MOGS	Mannosyl-oligosaccharide glucosidase	1.14	2.66E-03
P30101	PDIA3	Protein disulfide-isomerase A3	1.15	3.73E-16
P13667	PDIA4	Protein disulfide-isomerase A4	1.50	1.51E-09
Q15084	PDIA6	Protein disulfide-isomerase A6	1.47	1.37E-18
Q9P2E9	RRBP1	Ribosome-binding protein 1	1.82	2.17E-16
O94855	SEC24D	Protein transport protein Sec24D	1.05	3.00E-03
P60059	SEC61G	Protein transport protein Sec61 subunit gamma	1.63	2.63E-09
Q99442	SEC62	Translocation protein SEC62	1.32	1.72E-16
Q9UNE7	STUB1	E3 ubiquitin-protein ligase CHIP	1.00	1.30E-09
Q8NBS9	TXNDC5	Thioredoxin domain-containing protein 5	1.02	2.94E-07

The UPR and associated endoplasmic reticulum protein components have been reported to be up-regulated in response to cancer related endoplasmic reticulum stress in a number of cancers including breast cancer (Fernandez *et al.*, 2000; Andruska *et al.*, 2015; Avril *et al.*, 2017; Mcgrath *et al.*, 2018). Tumour cells are known to activate pro-survival pathways, such as the UPR, in response to harsh tumour microenvironment conditions characterised by hypoxia, glucose deprivation as well as oxidative stress, all of which are known to cause

endoplasmic reticulum stress and which affect protein processing. Tumour cells are able to survive under chronic environmental stress and avoid apoptosis or autophagy by activating and altering the UPR (Corazzari *et al.*, 2017; Oakes, 2020).

Endoplasmic reticulum chaperone BiP (HSPA5), also known as heat shock protein A5, was upregulated in breast tumour samples when compared to non-tumorous breast tissue as shown in Figure 3.21. Endoplasmic reticulum chaperone BiP is an endoplasmic reticulum stress sensor whereby misfolded proteins bind competitively to endoplasmic reticulum chaperone BiP leading to its dissociation from three endoplasmic reticulum transmembrane proteins; protein kinase RNA-like endoplasmic reticulum kinase (PERK), inositol-requiring enzyme 1α (IRE1α) and activating transcription factor 6 (ATF6), ultimately leading to a cascade of events that activate the UPR (Pobre *et al.*, 2019). Several studies have reported that elevated endoplasmic reticulum chaperone BiP abundance is associated with resistance to chemotherapy, specifically to topoisomerase inhibitors, and therefore can potentially be used as an indicator to predict the success of chemotherapy and prognosis of breast cancer patients (Reddy *et al.*, 2003; Lee *et al.*, 2006; Roller and Maddalo, 2013; Gifford and Hill, 2018).

A specific family of proteins known as protein disulphide isomerases (PDIs) are a group of endoplasmic enzymes involved in the UPR by playing a role in the formation of disulphide bonds, acting as protein chaperones for protein folding and mediating post-translational modification of synthesised proteins (Depuydt *et al.*, 2011; Lee, 2017). Five members of the PDI family; protein disulphide isomerase A3, protein disulphide isomerase A4, protein disulphide isomerase A6, endoplasmic reticulum resident protein 29 and thioredoxin domain-containing protein 5 were found to be upregulated in breast tumour tissue compared to non-tumorous breast tissue in this present study. PDIs have been reported to be up-regulated in several cancers and is associated with tumour invasion and metastasis (Zong *et al.*, 2012; Lee, 2017). Increase in expression of PDIs, such as PDAI3 which was also observed in this study, has previously been reported in breast cancer and may potentially be considered a staging marker for lymph node metastasis in breast cancer (Lee *et al.*, 2012; Ramos *et al.*, 2015). Furthermore, PDIs are associated with several ECM components and have been implicated in metastasis due to its role in protein folding and activation of matrix metalloproteases that are

involved in ECM remodelling. PDIs also prepare cell surface integrins that are associated with cellular adhesion or migration (Khan *et al.*, 2012; Xu *et al.*, 2014a). PDIs regulate the expression of proteases, Cathepsin B and D, that are also involved in ECM remodelling and cancer cell migration (Dumartin *et al.*, 2011). Moreover, TXNDC5 which was also found to be upregulated in breast tumour tissue in this study, is involved in upregulated angiogenesis and is associated with remodelling ECM components through MMP-9 and Cathepsin B (De Lucca Camargo *et al.*, 2013; Lee, 2017). Endoplasmic reticulum chaperone BiP and PDIs have been explored as prognostic markers for cancer as well as implicated in chemoresistance (Xu *et al.*, 2014a; Lee, 2017). The upregulation of these endoplasmic reticulum proteins in this current study provides additional evidence that these proteins should be explored further in breast cancer research as possible new drug targets.

3.2.6 Characterisation of breast tumour extracellular matrix

The complete enrichment STRING network shown in Figure 3.18 was filtered for ECM only related pathways, whereby sub-networks were extracted from the main protein network in order to reduce the complexity of the large protein network and to focus on only ECM proteins that were the main proteins of interest in this study. A list of filtered ECM enrichment terms is shown in Table 3.7 and contains the description of enriched terms, term categories with corresponding FDR values, the number of genes, nodes and edges associated with each ECM protein sub-network.

A final protein network containing ECM proteins and their respective protein-protein associations and abundance was generated by merging all ECM sub-networks. The merging of all ECM networks eliminated recurring proteins from individual sub-networks and allowed visualisation of ECM proteins and associated protein-protein interactions on a single enrichment network analysis consisting of 59 ECM proteins as shown in Figure 3.22. Table 3.8 summarises all differentially expressed ECM protein components (nodes) that are labelled with their respective gene names, average $\log_2$ ratios and q-values. The ECM proteins that were identified in this study have multiple biological roles, however they will be discussed within the context of their potential role in breast cancer for the purpose of this study.

Table 3.7: A STRING enrichment analysis filtered for extracellular matrix functional enrichment terms that were extracted from the complete STRING network generated from the final candidate protein list containing all proteins that meet the specified selection criteria. The description of enriched terms, term category with corresponding FDR values in addition to the number of genes, nodes and edges associated with the selected ECM protein sub-network are listed.

Pathway description	Category	FDR value	Number of genes	Number of nodes	Number of edges
Extracellular matrix	UniProt Keywords	1.26E-12	38	38	146
Extracellular matrix	GO Component	3.25E-10	35	35	124
Extracellular matrix binding	GO Function	4.72E-02	6	6	5
Extracellular matrix component	GO Component	1.29E-10	18	18	58
Extracellular matrix disassembly	GO Process	2.41E-02	7	7	3
Extracellular matrix organization	Reactome Pathways	6.34E-18	48	48	262
Extracellular matrix organization	GO Process	1.67E-17	48	48	251
Extracellular matrix proteoglycans	Reactome Pathways	1.29E-12	22	22	105
Extracellular matrix receptor interaction	KEGG Pathways	2.40E-11	22	22	103
Extracellular matrix structural constituent	GO Function	2.70E-04	12	12	21

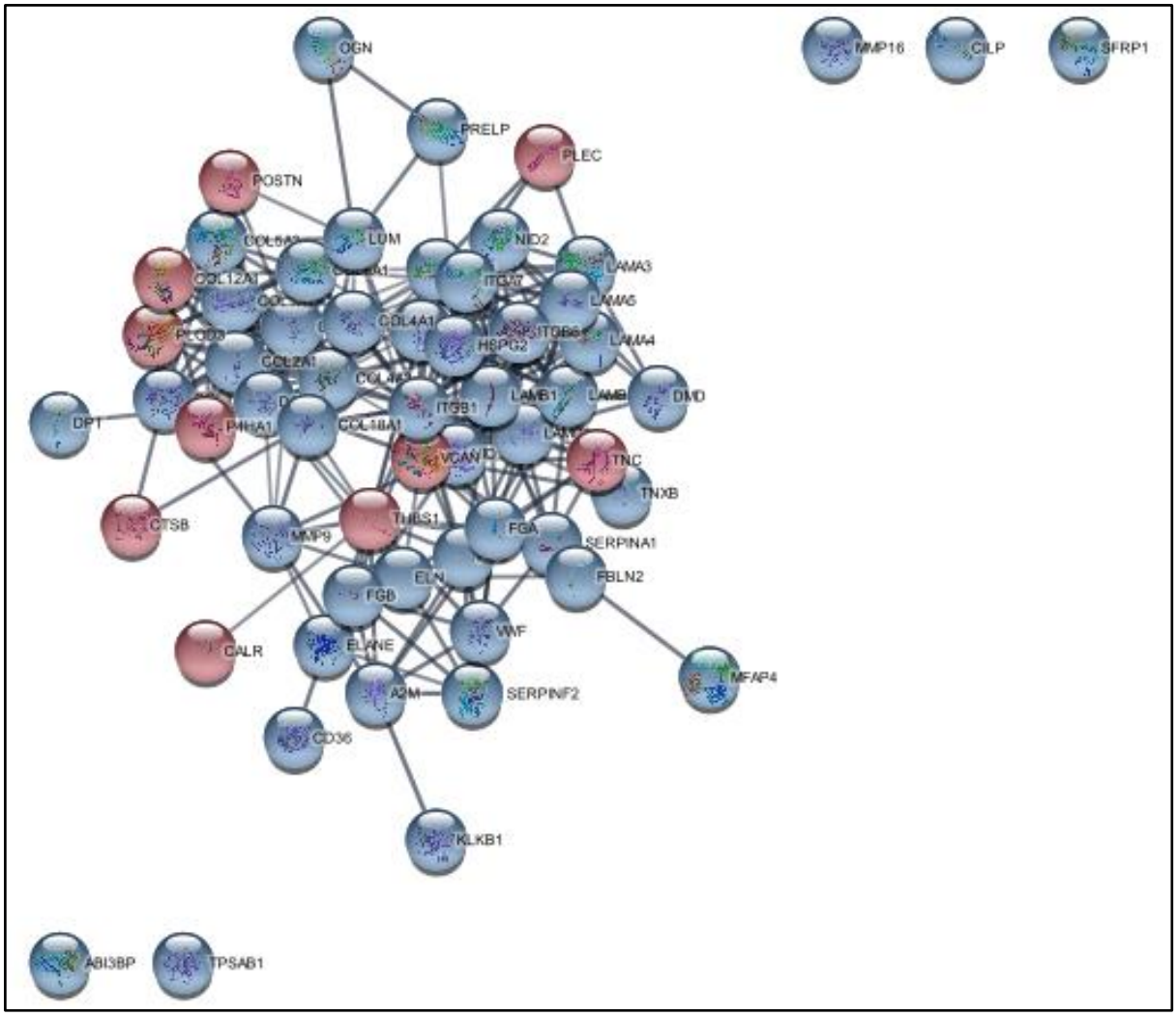

Figure 3.22: STRING network of merged ECM protein sub-networks showing significant differentially expressed ECM proteins indicating their relative abundance in tumour samples compared to non-tumorous samples. Each protein candidate is represented by nodes, which are labelled with their corresponding gene names, and are connected by black lines that represent relevant protein-protein interactions. Nodes are either coloured in blue to illustrate down-regulation or red to indicate up-regulation in protein expression.

Table 3.8: A list of differentially expressed ECM proteins in tumour samples compared to non-tumorous breast tissue. Uniprot accession numbers, gene names corresponding to the above ECM STRING network (Figure 3.25), corresponding average log_2 ratios and Q-values are listed in the table below. Negative values indicate down-regulation and positive values indicate up-regulation of protein expression.

Uniprot Accession number	Gene name	Protein description	Average Log_2 Ratio	Q-value
P01023	A2M	Alpha-2-macroglobulin	-1.81	2.64E-29
Q7Z7G0	ABI3BP	Target of Nesh-SH3	-1.48	1.18E-12
P27797	CALR	Calreticulin	1.15	3.00E-09
P16671	CD36	Platelet glycoprotein 4	-2.50	9.66E-29
O75339	CILP	Cartilage intermediate layer protein 1	-1.27	2.02E-04
Q99715	COL12A1	Collagen alpha-1(XII) chain	2.13	2.91E-10
P39059	COL15A1	Collagen alpha-1(XV) chain	-2.04	7.38E-25
P39060	COL18A1	Collagen alpha-1(XVIII) chain	-1.36	8.95E-28
P02452	COL1A1	Collagen alpha-1(I) chain	-1.51	4.23E-05
P02458	COL2A1	Collagen alpha-1(II) chain	-2.43	1.16E-05
P02461	COL3A1	Collagen alpha-1(III) chain	-1.35	1.07E-07
P02462	COL4A1	Collagen alpha-1(IV) chain	-2.19	1.42E-25
P08572	COL4A2	Collagen alpha-2(IV) chain	-2.08	2.06E-28
P25940	COL5A3	Collagen alpha-3(V) chain	-1.07	1.20E-05
P12109	COL6A1	Collagen alpha-1(VI) chain	-1.03	7.23E-16
P07858	CTSB	Cathepsin B	2.02	4.45E-10
P07585	DCN	Decorin	-1.85	8.55E-16
P11532	DMD	Dystrophin	-1.16	1.58E-03
Q07507	DPT	Dermatopontin	-1.36	3.27E-12
P08246	ELANE	Neutrophil elastase	-2.42	6.58E-18
P15502	ELN	Elastin	-1.34	4.77E-04
P98095	FBLN2	Fibulin-2	-2.09	2.21E-08
P02671	FGA	Fibrinogen alpha chain	-1.26	1.28E-10
P02675	FGB	Fibrinogen beta chain	-1.87	9.04E-14
P02679	FGG	Fibrinogen gamma chain	-1.45	3.93E-16
P98160	HSPG2	Basement membrane-specific heparan sulfate proteoglycan core protein	-1.92	2.27E-30
P23229	ITGA6	Integrin alpha-6	-1.04	1.11E-16
Q13683	ITGA7	Integrin alpha-7	-1.56	2.92E-18
P05556	ITGB1	Integrin beta-1	-1.34	8.66E-26
P16144	ITGB4	Integrin beta-4	-1.14	7.11E-03
P18564	ITGB6	Integrin beta-6	-1.41	1.04E-30
P03952	KLKB1	Plasma kallikrein	-2.08	3.81E-04
Q16787	LAMA3	Laminin subunit alpha-3	-1.03	7.36E-08
Q16363	LAMA4	Laminin subunit alpha-4	-1.27	5.08E-25
O15230	LAMA5	Laminin subunit alpha-5	-1.38	4.97E-21
P07942	LAMB1	Laminin subunit beta-1	-1.27	5.28E-24

P55268	LAMB2	Laminin subunit beta-2	-2.14	4.33E-35
P11047	LAMC1	Laminin subunit gamma-1	-2.16	1.08E-33
P51884	LUM	Lumican	-1.71	4.57E-13
P55083	MFAP4	Microfibril-associated glycoprotein 4	-1.62	6.66E-15
P51512	MMP16	Matrix metalloproteinase-16	-1.02	2.16E-16
P14780	MMP9	Matrix metalloproteinase-9	-3.58	1.72E-07
P14543	NID1	Nidogen-1	-1.99	3.07E-30
Q14112	NID2	Nidogen-2	-1.51	5.32E-28
P20774	OGN	Mimecan	-1.91	4.54E-15
P13674	P4HA1	Prolyl 4-hydroxylase subunit alpha-1	1.50	5.23E-03
Q15149	PLEC	Plectin	1.22	2.51E-21
O60568	PLOD3	Multifunctional procollagen lysine hydroxylase and glycosyltransferase LH3	1.94	1.98E-13
Q15063	POSTN	Periostin	2.57	1.06E-09
P51888	PRELP	Prolargin	-1.81	1.97E-17
P01009	SERPINA1	Alpha-1-antitrypsin	-1.23	1.41E-22
P08697	SERPINF2	Alpha-2-antiplasmin	-1.37	1.22E-04
Q8N474	SFRP1	Secreted frizzled-related protein 1	-1.02	6.33E-03
P07996	THBS1	Thrombospondin-1	1.41	1.56E-07
P24821	TNC	Tenascin	1.62	2.49E-09
P22105	TNXB	Tenascin-X	-1.94	1.19E-10
P20231	TPSB2 TPSAB1	Tryptase beta-2 Tryptase alpha/beta-1	-1.17	4.95E-16
P13611	VCAN	Versican core protein	1.01	2.42E-18
P04275	VWF	von Willebrand factor	-1.25	2.71E-30

The integrity of ductal basement membrane, a compact structural sheet of ECM protein constituents, plays a key role in the prevention of breast cancer cells invading the surrounding stroma and then spreading to distant organs (Chang and Chaudhuri, 2019). The abundance of several major basement membrane proteins, namely collagen IV (alpha-1 and alpha-2 chains), laminins (laminin subunit alpha-3, laminin subunit alpha-4, laminin subunit alpha-5, laminin subunit beta-1, laminin subunit beta-2, laminin subunit gamma-1), basement membrane-specific heparan sulphate proteoglycan core protein (perlecan), nidogen-1 and nidogen-2 were significantly reduced relative to levels observed in adjacent non-tumorous breast tissue (See Table 3.8). Lower levels or complete absence of collagen IV, the structural backbone of the basement membrane, and laminins have been reported in IDC thus supporting the findings in this study (Hewitt *et al.*, 1997; Diaz *et al.*, 2005; Wu *et al.*, 2016). Nidogen-1 and nidogen-2 are glycoproteins that non-covalently crosslink collagen IV to laminin molecules

thus stabilising the laminin-collagen network within the basement membrane. The downregulated expression of nidogens along with the proteoglycan core protein, perlecan, further alludes to the loss of structural stability of the basement membrane in tumour samples (Yurchenco, 2011). Perlecan, a key basement membrane component, binds to several structural ECM basement proteins including basement membrane anchoring molecule, prolargin (PRELP), which was also detected at lower levels in tumour tissue as shown in Figure 3.22 (Iozzo, 2005). These results are in agreement with other studies directed towards the basement membrane in cancer research (Yurchenco *et al.*, 2004; Wu *et al.*, 2016) and provide further evidence of the complete aberration, or loss, of basement membrane ECM components that leads to significant structural weakness associated with tumour invasion into surrounding tissue.

Non-fibrillar basement membrane associated collagens, collagen alpha-1 (XV) chain and collagen alpha-1 (XVIII) chain, were both found to be downregulated in tumour tissue. Collagen XV is a proteoglycan that has been implicated in metastasis where studies have reported that invasion past the basement membrane by ductal cancer cells most likely occurs after the degradation of collagen XV (Amenta *et al.*, 2003; Mutolo *et al.*, 2012). Endostatin, a product of collagen XVIII cleavage, is a potent anti-angiogenic mediator (Folkman, 2006). Endostatin directly binds to vascular endothelial growth factor receptors-1 and -2 (VEGFR-1 and VEGFR-2) thereby inhibiting downstream angiogenic signalling pathways (Kim *et al.*, 2002). Furthermore, endostatin partially exerts its anti-angiogenic activity by inhibiting metalloproteinase activity involved in the degradation of ECM components to allow blood vessel formation for tumour growth (Kim *et al.*, 2000). The down-regulation of collagen XVIII expression in tumour samples suggests decreased levels of endostatin that may play a role in promoting tumour growth and cancer cell survival through increased angiogenesis.

Studies have shown that breast tumour stroma undergoes a range of structural and cellular changes in response to altered molecular signalling that occur during malignancy (Walker, 2001; Conklin and Keely, 2012; Declerck, 2012). These stromal changes can paradoxically vary from reduced collagen formation and low expression of ECM proteins with an increase in number of stromal cells to the development of a dense collagenous stroma with increased levels of ECM proteins and fewer but larger stromal cells, the latter response referred to as the desmoplastic response or stromal reaction. Although assessing the correlation between

stromal changes and breast tumour grading has been attempted, further studies are needed to better understand the complex molecular mechanisms involved in breast cancer desmoplasia and to determine whether stromal changes are unique to different types or stages of breast carcinomas. These changes may include the contrasting decrease or increase in collagen levels, changes in overall ECM protein composition, tumour grade or molecular subtypes in breast cancer (Walker, 2001; Iacobuzio-Donahue *et al.*, 2002a; Khamis *et al.*, 2012; Catteau *et al.*, 2019). Desmoplasia is a prominent feature of cancers such as pancreatic ductal adenocarcinoma (PDAC), where studies have reported a significant increase in ECM composition and upregulation of ECM-related pathways (Aghamaliyev *et al.*, 2016; Cannon *et al.*, 2018; Nweke *et al.*, 2020). However, differences in gene expression profiles of breast and pancreatic tumours related to stromal characteristics have been reported, suggesting that desmoplasia and associated stromal changes may vary for different cancer types (Iacobuzio-Donahue *et al.*, 2002a; Iacobuzio-Donahue *et al.*, 2002b)

A desmoplastic response not only involves an increase in collagen by tumour-associated cells, but also the upregulation of collagen-modifying enzymes such as lysine hydroxylases (PLODs) and prolyl-4-hydroxylase (PH4) that may be indicative of increased collagen deposition in the tumour microenvironment (Gilkes *et al.*, 2013; Xiong *et al.*, 2014). The increased expression of prolyl 4-hydroxylase subunit alpha-1 (P4HA1), multifunctional procollagen lysine hydroxylase, glycosyltransferase LH3 (PLOD3) and collagen alpha-1(XII) chain (COL12A1) in tumour samples, as reported in Table 3.8, would suggest that desmoplasia most likely occurred in the majority of stroma within the invasive ductal carcinomas assessed in this study. However, contrary to these findings, fibrillar collagens such as collagen alpha-1(I) chain (COL1A1), collagen alpha-1(II) chain (COL2A1), collagen alpha-1(III) chain (COL3A1), collagen alpha-3(V) chain (COL5A3), and collagen alpha-1(VI) chain (COL6A1) were found to be expressed at lower levels in the tumour samples compared to matching non-tumorous tissue. Interestingly, although it has been reported that increased desmoplasia is involved in IDC progression there is also evidence that depleted levels of collagens, due to increased degradation by matrix metalloproteinases (MMPs) or reduced levels of myofibroblasts, may occur in the tumour microenvironment and in fact promotes tumour invasion into surrounding tissue (Fang *et al.*, 2014; Özdemir *et al.*, 2014; Acerbi *et al.*, 2015). Solid tumours may also develop a desmoplastic response due to an increase in several non-collagen ECM

proteins, which include fibronectin, proteoglycans and matricellular proteins such as tenascin, that leads to a "stiffer" invasive tumour stroma (Acerbi *et al.*, 2015). The desmoplastic response is in itself highly complex process involving numerous proteins and further studies are needed to investigate the exact role and cause of desmoplasia in breast cancer. Additional indications of ECM structural dysregulation in tumour samples were apparent by the downregulation of some collagen-fibre interacting proteins namely; elastin (ELN), fibulin-2 (FBLN2) and microfibril-associated glycoprotein 4 (MFAP4) expression. FBLN2 plus microfibril-associated glycoprotein MFAP4 binding to collagen provides tensile strength, elasticity and structural integrity of the interstitial matrix (Frantz *et al.*, 2010; Godwin *et al.*, 2019; Yang *et al.*, 2019).

Matrix metalloproteinase-9 (MMP-9) and matrix metalloproteinase-16 (MMP-16) downregulation was observed in breast tumour samples as reported in Table 3.8. These results are contrary to genomic studies where upregulation of mRNA levels for MMP-2, MMP-3 and MMP-11 have been reported. These MMPs are involved in collagen degradation in tumour specific stroma and are associated with the progression of DCIS to IDC (Ma *et al.*, 2019). Moreover, there is abundant literature describing the prominent role of MMP-9 in tumour invasion into surrounding tissue by degrading ECM components of the basement membrane (Duffy *et al.*, 2000; Wu *et al.*, 2008; Katari *et al.*, 2019). Several studies also reported that MMP-9 is completely absent or weakly expressed in normal breast cells with low levels detected in luminal type breast tumours, but present at relatively higher levels in HER-2 positive, triple-negative and higher-grade 3 tumours compared to normal breast tissue (Roomi *et al.*, 2009; Mehner *et al.*, 2014; Yousef *et al.*, 2014). Therefore, it must be stated that the down regulation of MMP-9 observed in the tumorous samples analysed in this study is a conflicting result between this study, where tumour invasion of surrounding tissue was observed in all histological tumour sections with an overall 62% of samples associated with lymph node involvement, and published literature. Several studies having either shown a high expression of MMP-9 in breast cancer tissue at the mRNA level using gene expression profiling methods or at the protein level using semiquantitative techniques such as ELISA and Western blot or immunohistochemistry, (Przybylowska *et al.*, 2006; Merdad *et al.*, 2014; Li *et al.*, 2017; Joseph *et al.*, 2020). The discrepancy between quantitative results may be due to a bias introduced through use of different experimental methods, with literature methods all relying

on immuno-based methods while the LC-MS/MS used in this study has several potential steps that could result in decreased quantitative results.

Cathepsin B is a lysosomal protease that is located close to the nuclear region of normal cells but often localised on the surface of tumour cells and reported to be secreted into the tumour microenvironment in breast cancers (Roshy *et al.*, 2003). Cathepsin B, is not only directly involved in the dysregulation of the surrounding tumour microenvironment by cleaving collagen IV, laminins and tenascin, but is also associated with several proteolytic cascades in ECM remodelling (Buck *et al.*, 1992; Mai *et al.*, 2002). Studies have shown that Cathepsin B stimulates selected metalloproteinases, resulting in further degradation of the ECM and facilitating tissue invasion by allowing movement of cancer cells from primary tumours to secondary sites. Further promotion of metastasis by cathepsins involves the inactivation by cleavage of E-cadherin, a cell adhesion protein, leading to decreased cell to cell adhesion, thereby promoting cancer cell spread and invasion (Cavallaro and Christofori, 2004; Gocheva *et al.*, 2006). Cathepsin B is involved in several mechanisms associated with tissue invasion by cancer cells and offers an explanation regarding the significantly increased expression of this cysteine protease in invasive breast tumours, with an average 4-fold increase in expression being observed in the tumour samples as reported in Table 3.8. The role of cathepsin B has been explored in breast cancer (Gabrijelcic *et al.*, 1992; Teo *et al.*, 2012) and the limited findings in this current study suggest that cathepsin B could potentially be evaluated further as a therapeutic target for IDC.

Two members from the tenascin protein family, tenascin-C (TNC) and tenascin-X (TNXB) were identified in breast tissue samples. Although limited protein distribution of TNC is seen in normal adult tissues, studies have revealed that tenascin gene products are overexpressed in the stroma of solid tumours and are significantly upregulated during pathological conditions involving inflammation and wound healing (Midwood *et al.*, 2009). Cancer-associated fibroblasts (CAFs) that are found within the tumour microenvironment are prominent producers of TNC, with a study identifying TNC as a potential CAF marker or prognostic indicator for breast cancer (Yang *et al.*, 2017). Additionally, the integration of TNC into the tumour microenvironment occurs via a signalling-dependent pathway involving transforming growth factor beta (TGFβ), an important cytokine that plays a role in regulating cellular differentiation and proliferation (Orend and Chiquet-Ehrismann, 2006). TNC-induced tumour

neovascularisation and the association between TNC and stem-cell signalling components, such as leucine-rich repeat–containing G protein–coupled receptor 5 (LGR5), have been linked to pulmonary metastasis in breast cancer with the presence of TNC associated with poor clinical outcomes (Oskarsson *et al.*, 2011; Oskarsson, 2013; Sun *et al.*, 2019). Moreover, TNC has been identified as a ligand for proteoglycans such as versican core protein (VCAN), which was also upregulated in tumour samples as shown in Figure 3.22. Elevated levels of TNC and VCAN have also been reported in node-negative breast cancer (Suwiwat *et al.*, 2004). The pleiotropic effects of TNC, such as the enhancement of cancer cell spread, tissue invasion and angiogenesis, strongly suggest that TNC may play a key role in cancer tumour development and metastasis (Matei *et al.*, 2011; Lowy and Oskarsson, 2015). Considering the relatively high percentage of lymph node involvement in patients in this study, the increase in TNC abundance in tumorous samples is not an unexpected finding. Few studies have reported on the other tenascin family proteins, however, sparse studies that have been conducted indicate that TNXB is significantly down-regulated in several cancer types (Liot *et al.*, 2020) , thus supporting the measured decrease in TNXB expression observed in this study. However, further studies are needed to establish the exact role that TNXB may play in breast cancer

Calreticulin, an endoplasmic reticulin protein, was found to be upregulated in tumour samples compared to levels in adjacent non-tumorous tissue. This is in agreement with several breast cancer studies that detected calreticulin at significantly higher levels in the stroma of invasive ductal breast tumours (Ezhilarasan, 2016; Zamanian *et al.*, 2016; Tariq *et al.*, 2019). Increased expression of calreticulin is associated with invasive breast cancer phenotypes by modulating the p53 and MAPK pathways in tumour formation and cancer cell migration (Zamanian *et al.*, 2016). Calreticulin is directly associated with thrombospondin-1, a matricellular protein that was observed to be overexpressed in tumorous samples as shown in Figure 3.22. Binding of calreticulin to the N-terminal domain of thrombospondin-1 mediates separation of focal adhesions, thus controlling cellular migration, but also promotes fibroblast survival, a prominent cell associated with solid tumours (Goicoechea *et al.*, 2000; Pallero *et al.*, 2008; Kalluri, 2016). Moreover, *in vitro* studies suggested that thrombospondin-1 may play a role in promoting the invasive characteristics associated with IDC (Wang *et al.*, 1996).

Lumican and decorin are both members of the small leucine-rich proteoglycan (SLRP) family involved in regulating ECM collagen assembly and considered tumour suppressors. Lumican reduces cellular migration by interacting with integrins and inhibits tumour angiogenesis by inducing endothelial cell apoptosis (Nikitovic *et al.*, 2014). Decorin suppresses tumour growth by interfering with the function of epidermal growth factor receptor as well as promoting apoptosis through activation of caspase-3 (Moscatello *et al.*, 1998; Seidler *et al.*, 2006). Down-regulation of both lumican and decorin was observed in breast tumour samples and these findings are supported by several studies, with one study in particular suggesting that lower levels of these two SLRP members are indicative of poor outcomes in node-negative invasive breast cancer (Troup *et al.*, 2003; Eshchenko *et al.*, 2007; Bae *et al.*, 2013). It must be stated that contradictory evidence of high lumican mRNA levels in breast tumour tissue have been reported insinuating increased lumican expression. (Leygue *et al.*, 1998; Leygue *et al.*, 2000). However, these conflicting results may be attributed to the documented discordance between mRNA and protein expression levels that has also been observed in breast tissue, with variation in lumican expression within different regions of the same breast tumour. Moreover, mimecan, which is another SLRP member, was also reported to be expressed at a significantly lower level in tumour samples. Mimecan exerts tumour growth suppression effects on the PI3K/AKT/mTOR pathway (Xu *et al.*, 2019). Considering the inhibitory effects these proteins have on tumour growth and cancer cell migration, the reduced protein levels of lumican, decorin and mimecan observed in tumour samples could suggest that stromal invasion and progression of the DCIS to IDC is potentiated by the lower expression of these three SLRP tumour suppressors. Substantial evidence points to the SLRP class of ECM proteins as potential agonist therapeutic targets in cancer drug discovery studies.

Periostin is a non-structural ECM protein expressed by stromal cells, cancer epithelial cells and CAFs. Dysregulation of periostin expression is associated with several malignancies (González-González and Alonso, 2018). This study identified an average 5-fold increase in periostin expression in breast tumour tissue compared to adjacent non-tumorous tissue as shown in Table 3.8. Overexpression of periostin in breast tumour tissue has been reported in previous studies, with a direct correlation between increased expression and tumour grade, with a link between high levels and poor overall survival of breast cancer patients (Ratajczak-Wielgomas *et al.*, 2016; Kim *et al.*, 2017). Periostin is involved in cancer cell proliferation and

invasion by stimulating epithelial to mesenchymal transition, enhancing the activity of ECM degrading metalloproteinases and inducing tumour angiogenesis via the upregulation of vascular endothelial growth factor receptor 2 (VEGFR-2)(Shao *et al.*, 2004; Chuanyu *et al.*, 2017; Hu *et al.*, 2017). Hence the current findings in this study further support existing evidence that periostin could be considered a biomarker or therapeutic target for breast cancer due its association with cancer cell development and progression in IDC (Ratajczak-Wielgomas *et al.*, 2017).

In summary, SWATH-based quantification and pathway enrichment analysis revealed that ribosomal, spliceosome and endoplasmic reticulum protein processing pathways with their associated protein components were significantly upregulated in breast tumour samples. These results highlight the fact that processes such as protein homeostasis, protein synthesis, protein folding and alternative splicing, are strongly affected in solid tumours in order to meet the demands of uncontrolled tumour growth and promotion of tumour metastasis. These results also serve to validate the sample preparation and automated proteomic analysis methods as data obtained through MS analysis of the breast tumour samples in this current study are similar to findings that have been reported in previous cancer-based studies.

MS-based proteomic methods have been successfully used to carry out global analysis of biological tissue specimens, however the ECM was found to be significantly under-represented in the proteomic data collected in this study. Probable reasons being due to the ECM complexity, aqueous insolubility of many ECM components, resistance of ECM proteins to tryptic digestion, and the wide dynamic concentration range of ECM proteins. Furthermore, ECM proteins undergo extensive crosslinking and a number of post-translational modifications that render the ECM proteins more resistant to solubilisation by standard detergent-based protein extraction methods, and to digestion to give appropriate length peptides for downstream LC-MS/MS analysis (Byron *et al.*, 2013; Naba *et al.*, 2016). The complexity of the ECM makes routine proteomic sample preparation techniques unsuitable when targeting the ECM and alternative preparation methods in conjunction with modified MS data interrogation techniques that selectively enrich for relevant ECM proteins should be explored when carrying out proteomic profiling of the ECM (Naba *et al.*, 2016). Alternative solubilisation and digestion methods that show increased selectivity to ECM

proteins would improve characterisation of these proteins and will be discussed in more detail in Section 4.2.

Despite these limitations and analytical constraints, several ECM networks showing differentially expressed proteins were identified in breast tumour ECM samples. The biological functions of identified ECM proteins include; involvement in several cancer related biological processes including structural integrity, cancer cell proliferation, tumour growth, tumour tissue invasion and metastasis. The involvement of the identified ECM proteins in cancer tissue can be associated with the development and progression of breast cancer, but further quantitative and validation studies of these ECM proteins will allow for their assessment as potential prognostic biomarker signatures or potential new chemotherapeutic drug targets for breast cancer.

Chapter 4

4 Final conclusions and recommendations

4.1 Final conclusions

The primary aim of this research project was originally to use advanced mass spectrometry (MS)-based proteomics to characterise and quantitate differential protein expression from non-tumorous breast tissue and tumours classified as invasive ductal carcinomas (IDC), with emphasis on extracellular matrix (ECM) proteins. This was to potentially identify prognostic markers or therapeutic drug targets for breast cancer. Breast tissue from non-tumorous regions and tumours were collected from the same patients who were diagnosed with grade I-III IDC but who had not undergone neoadjuvant chemotherapy treatment. Histological analysis of resected tumour sections of snap-frozen tissue was used to confirm the clinical diagnosis and to differentiate between non-tumorous and tumorous histological tissue sections. Protein extraction used an optimised barocycler method for histological sections carried out using high SDS detergent concentrations with repeated cycles of hydrostatic pressure using the Barocycler® 2320EXT instrument. A semi-automated hydrophilic affinity-based protein capture clean-up and off-bead tryptic digestion method was used to produce peptides from protein digests that were analysed using a Dionex Ultimate 3000 RSLC system coupled to a SCIEX 6600 TripleTOF mass spectrometer with a 45-minute increasing acetonitrile gradient on a C18 capillary column and microflow conditions. Sequential window acquisition of all theoretical mass spectra (SWATH) technology was used for data independent acquisition and analysis of amino acid sequence data of separated peptides. Amino acid sequence data was used to identify proteins of interest and SWATH-based quantification was used to compare relative protein levels between non-tumorous breast and tumour tissue samples.

Haematoxylin and eosin (H&E) staining is a routine diagnostic tool used worldwide and is considered the gold standard in cancer histopathology. H&E staining allows overall tissue structure visualisation and is able to differentiate tissue and between various cellular components (Feldman and Wolfe, 2014). Histological analysis was performed on resected tissue samples to confirm the inclusion of only patients with a positive IDC diagnosis. The

staining procedure was also successfully used to ensure the integrity of samples and to confirm that samples were in fact representative of non-tumorous and tumour tissue, as this was critical for downstream comparative proteomics. Histological micrographs showed significant inter-individual variation regarding epithelial, stromal and adipose tissue distribution for both non-tumorous and tumour tissue micrographs. Biological variation observed on sequential adjoining histological micrographs for each patient was taken into consideration when analysing final LC-MS/MS proteomic data, whereby the relative proportion of tumour volume in relation to the surrounding breast tissue may have influenced the proteomic characterisation of tissue sections. This was clearly seen for three tumour samples that clustered with the non-tumorous samples due to the relatively small volume of tumour of the histological section. Therefore, histological analysis of resected samples proved to be an essential component of this study as it allowed overall visualisation of the selected breast tissue morphology, confirmation of patient diagnosis and more importantly it provided important complementary visual information to support the downstream LC-MS/MS proteomic analysis of breast tissue sections.

In depth proteome coverage and accurate protein quantification is dependent on efficient solubilisation of proteins and extraction methods during sample preparation (Klont *et al.*, 2018). A protein extraction method consisting of a high sodium dodecyl sulphate (SDS) detergent concentration and repeated pressure cycling technology (PCT) was chosen based on the number of peptides and proteins identified. The optimised extraction method was used to effectively solubilise and extract breast tissue proteins for downstream LC-MS/MS analysis. Results from experiments using the different methods suggested that an increase in the number of high-pressure cycles as well as total cycle duration played a significant role in improving protein extraction and subsequent proteome coverage.

SWATH-MS is commonly used for relatively large-scale studies involving biomarker and drug target research and was the analytical method of choice as extensive proteome coverage and robust protein quantitation was required (Huang *et al.*, 2015; Schilling *et al.*, 2017). Technical variation for the complete experimental workflow from sample preparation through LC-

MS/MS and bioinformatic analysis was assessed and was found to be lower than the acceptable upper cut-off of 20% and therefore could be used for the sample analysis to provide acceptable accuracy and reliable protein quantification (Lynch, 2016; Fu *et al.*, 2018). Initially a comprehensive project specific spectral library of proteins was generated from 2-D RP-RP LC-MS/MS DDA runs of samples of interest and successfully used for protein identification of the main set of pooled samples. The library was derived from the analysis of samples combined from both the non-tumorous tissue and tumour samples. This process involved pooled samples of both non-tumorous tissue and tumours that were processed using the same sample extraction method and automated tryptic digestion but where the released peptides were initially separated using a reverse phase UPLC separation at high pH and fractions collected. Selected fractions were pooled into 12 pooled fractions for final DDA analysis on the LC-MS/MS system using acidic conditions and a 120 min gradient of increasing acetonitrile. The data collected was analysed using Spectronaut to develop a tissue specific spectral library for further comparison of the samples. Peptide samples from individual non-tumorous tissue and tumours following the same sample preparation methodology were then analysed using DIA by SWATH analysis following LC-MS/MS under acid conditions and a 45 min acetonitrile gradient. Interrelated instrument parameters, such as the precursor isolation window width, fragment and precursor ion accumulation time and the cycle time were adjusted in order to improve the sensitivity and selectivity of the final SWATH-MS method to produce high-quality SWATH-MS data.

Principal component analysis (PCA) of peptide/protein abundance clearly showed two distinct groups consisting of non-tumorous tissue and tumour samples. Only a few non-tumorous and tumour samples failed to group in their respective cohorts. Three of the tumour samples had relatively small volumes or zones of tumorous tissue compared to the non-tumorous tissue within the histological section and these clustered with the non-tumorous samples, most likely due to the proteins unique to the tumour being crowded out due to the small actual volume of tumour tissue. Possible explanations for the other outlier samples include; experimental error, biological contamination or differences in phenotypical characteristics associated with specific tumour grades. PCA and heatmapping were useful tools for comparative analysis of the large proteomic datasets generated in this study and could

provide visual comparisons between the two cohorts. Initial screening and data control were successfully achieved as PCA provided visualisation of grouping of samples based on protein expression profiles and also identified potential outliers within the complete proteomic dataset.

Differentially expressed proteins were functionally annotated and subject to pathway analysis. The most prominent pathways, based on the lowest FDR values, were ribosomal biosynthesis, endoplasmic reticulum protein processing pathway, RNA splicing and ECM receptor interactions. Identification of these pathways highlighted the fact that protein homeostasis and the tumour microenvironment are significantly different in solid breast cancer tumours compared to equivalent but non-tumorous tissue from the same patient.

Hyperactivation of the ribosomal biosynthesis pathway, representing increased protein synthesis, was evident from the upregulation of the numerous ribosomal and ribonuclear proteins that were identified in breast tumour samples. Overexpressed ribosomal proteins have been linked to metastasis and immunosuppression, but further studies are needed to determine whether this pathway can be a potential target for drug therapies (Pelletier *et al.*, 2018). Small ribonucleoproteins (snRNPs) and small nuclear RNA proteins (snRNA) that make up the spliceosome and splicing associated proteins such as heterogenous nuclear ribonucleoproteins (hnRNPs) and serine/arginine-rich proteins were found to be upregulated in breast tumour tissue with an average 2-fold increase identified in this study. These results suggest that RNA splicing activity is significantly enhanced in IDC. Splicing protein components are extensively involved in promoting several hallmarks of cancer, as well as ECM remodelling through alternative splicing of various ECM components (Coltri *et al.*, 2019). Endoplasmic reticulum stress, resulting from increasing protein demands due to uncontrolled cancer cell proliferation and oncogenic activation, is a common feature of cancer (Oakes, 2020). Pathway enrichment analysis showed the upregulation of the endoplasmic reticulum protein processing pathway, indicating that endoplasmic reticulum functions such as protein synthesis, folding and quality control were significantly affected in breast tumour tissue. Furthermore, proteins from the endoplasmic reticulum protein processing pathway, specifically the endoplasmic reticulum chaperone, BiP and protein disulphide isomerases (PDIs), are related to the unfolded protein response which is a pro-survival pathway that is

exploited by cancer cells to counteract endoplasmic reticulum stress caused by hypoxia, glucose deprivation and oxidative stress (Hetz *et al.*, 2015). The overall upregulation of ribosomal, splicing and endoplasmic reticulum protein processing pathways in breast tumour samples is in agreement with several other studies that have reported a close link between these pathways and the progression of a number of cancers including breast cancer (Andruska *et al.*, 2015; Oakes, 2017; Penzo *et al.*, 2019; Wang and Aifantis, 2020). Therefore, the findings in this present study provides additional evidence that protein candidates from these pathways involved in protein homeostasis should be investigated further and be considered as potential drug targets for pharmacologic agents in order to slow down the progression of tumours while others could be used as prognostic markers.

Pathway enrichment analysis also identified several ECM protein networks containing a number of differentially expressed ECM proteins in breast tumour samples that are involved in processes associated with ECM structural integrity, cancer cell proliferation, tumour growth, invasion, and metastasis.

The basement membrane is an essential barrier between epithelial cells and the interstitial matrix, effectively maintaining the organisation of cellular components. The disruption of the basement membrane plays a prominent role in cancer cells invading surrounding tissues and migrating to distant organs (Wu *et al.*, 2016). Major components of the basement membrane associated proteins, such as collagens, laminins, perlecan (basement membrane-specific heparan sulphate proteoglycan core protein) and the nidogens, were all found to be downregulated in tumour tissue compared to non-tumorous breast tissue in this study. These findings were a direct indication that substantial disruption of the ECM basement membrane had occurred in breast tumour samples and that these basement membrane ECM protein components could be considered, subject to further studies, as potential molecular markers for the progression of ductal carcinoma *in situ* (DCIS) to IDC or even the onset of tumour metastasis.

Further evidence showing that ECM structural integrity was compromised involved the downregulation of several collagen proteins as well as elastin, fibulin-2, and microfibril-associated glycoprotein 4 that are known to interact with collagen fibres to provide ECM

elasticity and tensile strength (Godwin *et al.*, 2019). Studies have reported that collagen is extensively associated with structural and functional characteristics of the ECM where both collagen degradation or deposition may play a role in promoting malignancy (Fang *et al.*, 2014; Walker *et al.*, 2018). Furthermore, upregulation of collagen-modifying enzymes that are known to promote a dense collagenous stroma conducive for cancer cell invasion was observed in breast tissue samples (Wullkopf *et al.*, 2018). Results related to collagen expression and deposition was however drawn into question and deemed to be inconclusive in this study due to several identified experimental factors which includes; low solubility, trypsin resistance and possible cross-linking that may have influenced the proteome coverage of the collagen proteins in this study. These limiting experimental factors will be discussed in Section 4.2 and possible recommendations will be suggested for further studies in order to understand the exact role and which types of collagen deposits may be associated with IDC and cancer progression.

ECM remodelling is characterised by enzymatic processes involving both qualitative and quantitative changes in the tumour microenvironment, which influence cellular proliferation, adhesion and migration. Dysregulation of the ECM, through increased ECM remodelling and degradation by proteases and matrix metalloproteinases (MMPs) is a common feature of cancer (Giussani *et al.*, 2019; Mohan *et al.*, 2020). Therefore, the decreased expression of MMPs in this current study was unexpected. Due to the differential expression of MMPs between molecular subtypes, further studies that focus on breast cancer subgroups will need to be carried out to investigate the exact role of MMPs expression in IDC breast cancer. However, ECM remodelling was evident in breast tumour tissue by the increased expression of the lysosomal protease, Cathepsin B. Cysteine cathepsins, such as Cathepsin B, play a fundamental role in proteolytic cleavage that decreases cellular adhesion, promotes cancer cell spread and stimulates other MMPs involved in ECM remodelling (Fonović and Turk, 2014; Vizovišek *et al.*, 2019). Therefore, the upregulation of Cathepsin B suggests that this protease may be further investigated in ECM remodelling studies aimed at counteracting processes that promote tumour invasion and metastasis.

Tumour angiogenesis, controlled by both activator and suppressor mediators, is characterised by the formation of tumour vasculature that is vital for cancer cells to obtain adequate oxygen and nutrient supply to sustain tumour growth and cancer cell spread (Lugano *et al.*, 2020). Tumour angiogenesis seemed to be favoured by the downregulation of collagen alpha-1 (XVIII) chain, a precursor to anti-angiogenic mediator endostatin and the upregulation of periostin, an ECM protein that is associated with the enhancement of tumour angiogenesis by upregulating angiogenic proteins such as vascular endothelial growth factor receptor 2 (VEGFR-2). ECM associated proteins such as tenascin-C (TNC), versican core protein (VCAN) and calreticulin were also found to be upregulated in breast tumour tissue and provided additional evidence that tumour cells are able to upregulate ECM components reported to promote malignancy and metastatic processes. TNC is involved in cellular differentiation and proliferation through interactions with cytokines, is associated with pulmonary metastasis of breast cancer and involved in TNC-induced neovascularisation (Orend and Chiquet-Ehrismann, 2006; Sun *et al.*, 2019). Calreticulin is linked to tumour formation, promoting cancer-associated fibroblasts as well as cancer cell migration particularly through its interaction with thrombospondin-1 (Wang *et al.*, 1996; Pallero *et al.*, 2008; Kalluri, 2016). Finally, reduced protein levels of three tumour suppressors, lumican, decorin and mimecan, in breast tumour samples offered additional evidence that cancer cells are able to alter the expression of ECM components to create a well vascularised and sustainable stromal niche for cancer cell survival and growth in IDC breast cancer.

Despite the small number of samples tested and compared, the overall conclusion is that the differential expression of these ECM proteins in breast tumour tissue and the significant role they play in tumour growth and progression suggests that these proteins may be considered as putative biological markers or targets for chemotherapeutic treatments for breast cancer.

4.2 Study limitations and recommendations

Study limitations and weaknesses associated with either experimental design or methodology occur in all research studies, and the impact of these scientific constraints on the final outcome of a study need to be acknowledged and taken into consideration when analysing large datasets. While efforts were made to automate as many experimental methods and

workflows as possible to reduce technical variance and improve reliability of the results to enable the achievement of the research aims and objectives of this study, there are aspects that introduce unavoidable limitations. Identifying and addressing these limitations could have improved the study outcomes and therefore the following recommendations are suggested.

Choosing a suitable sample size for any research study is of critical importance in order to ensure that clinically relevant differences can be reliably detected and quantified and to ensure statistical relevance. While a power analysis was be used to calculate the number of samples needed to ensure sufficient statistical power, additional factors such as access to qualifying and consenting study participants, available resources and time constraints also need to be taken into consideration when deciding on the final sample number. Certain limitations with regards to sample number and logistics with respect to appropriate sampling of the tissue that was resected arose in this study and should be improved upon for further studies. A vital part of this study was to differentiate between similar non-tumorous breast tissue and tumour sections before any further experiments were carried out. This component of the study required use of snap frozen resected tissue followed by cryo-sectioning of tissue samples. The adjacent sections of the same samples were then formalin fixed followed by H&E staining to confirmation of patient diagnosis by a qualified pathologist before any experimental work could be initiated. The histological screening process was labour intensive, time consuming and because it was done simultaneously with cryo-sectioning of tissue samples for proteomic analysis, it resulted in the exclusion of multiple tissue samples that had already been collected from patients, where post-histological screening indicated that samples had to be excluded due to either not matching the cancer diagnosis or where non-tumorous tissue proved to show cancerous characteristics. This part of the experimental workflow ultimately slowed progress with downstream proteomic analysis of samples significantly. Therefore, it is suggested that an alternative source of samples, such as formalin-fixed paraffin-embedded (FFPE) sections, be considered for similar studies in future. Advantages of FFPE samples include the fact that they can be preserved at room temperature for extended periods of time with structural and morphological integrity of samples remaining intact due the formalin fixation and wax embedding process (Grillo *et al.*, 2017). Furthermore,

FFPE blocks as well as patient clinical history, histology reports and patient outcomes that can be used to improve patient phenotyping and classification are stored in pathology laboratories for extended periods of time thus providing an expansive and invaluable archive of tissue material and clinical data that is easily accessible. It is important to note that protein cross-linking induced by formalin fixation has impeded the common use and acceptance of FFPE samples for proteomic- based studies. However this is changing as more validation studies that include advanced mass spectrometry techniques are being done to investigate protein expression levels and proteome coverage overlap between FFPE and matched fresh frozen breast tissue sections in order to establish whether FFPE-based proteomic experiments will be able to provide accurate and comprehensive proteomic data that is on par with data extracted from fresh frozen samples (Gräntzdörffer *et al.*, 2010; Gustafsson *et al.*, 2015; Coscia *et al.*, 2020; Uchida *et al.*, 2020).

Breast cancer shows a wide range of morphological and histopathological features that are essentially used to classify breast carcinomas into specific categories and molecular subtypes, each having their own specific clinical outcome (Makki, 2015). Diagnostic procedures to classify tumour samples according to their molecular subtype using immunohistochemical (IHC) detection of unique tumour markers on cancer cells, were not performed on any samples used in this study. The inability to reliably differentiate between molecular subtypes was considered to be a study limitation as certain ECM proteins of interest in this study have been reported to be differentially expressed between molecular subtypes of breast cancer (Roomi *et al.*, 2009; Mehner *et al.*, 2014; Yousef *et al.*, 2014). Therefore, future studies should incorporate immunohistochemical techniques to accurately assess the expression levels of ECM proteins based on molecular subtypes of the tumours. Data from such studies would provide a more comprehensive analysis of the tumour microenvironment in breast cancer and would offer valuable information that could potentially improve existing clinical models for breast cancer classification which may ultimately lead to better patient prognosis and treatment outcomes.

As mentioned in the discussion, although a considerable number of ECM proteins were identified in this present study, a specific group of ECM proteins like the collagens, appeared to be under represented. It has to be acknowledged that certain analytical difficulties were identified when characterising the complex composition of primary breast tumour ECM

samples and these analytical constraints are considered to be limitations in the present study. Therefore, alternative methods may be considered in order to improve the proteomic coverage of the ECM in breast tissue samples. It is imperative that methods used for sample preparation preserve the structural integrity and biological components of the ECM in addition to being compatible with downstream proteomic applications (Naba *et al.*, 2016). The following ECM enrichment strategies, each with their own pros and cons, may be considered in an attempt to improve the proteomic coverage of ECM proteins to achieve a more targeted analysis of the ECM. Decellularisation methods make use of chemical reagents or detergents and physical methods such as temperature, force and pressure to remove abundant cellular components in order to effectively identify lower abundant ECM proteins. However, these strategies are considered to be relatively harsh and tend to disrupt the proteomic composition of the ECM. Proteases such as trypsin or collagenases can also be used to remove cells but it may lead to the loss of ECM proteins and peptides through the partial digestion of the ECM thus affecting downstream proteomic analysis (Crapo *et al.*, 2011; Keane *et al.*, 2015; Ma *et al.*, 2019). Sub-cellular fractionation methods involving the sequential extraction of intracellular components that result in a final ECM enriched fraction, using buffer reagents, that vary in pH, salt and detergent concentrations, have been relatively successful in characterising the ECM of normal and diseased tissue (Naba *et al.*, 2017)). However, this method may result in the loss of soluble ECM components such as growth factors or ECM remodelling enzymes, thus compromising a comprehensive analysis of the ECM proteome (Naba *et al.*, 2015). Laser-capture microdissection is an advanced technique that utilises laser excision coupled with high resolution microscopy for precise tissue, cell or organelle isolation from complex biological samples (Shapiro *et al.*, 2012; Datta *et al.*, 2015). Laser-capture microdissection has been used as a powerful and robust sample preparation technique in combination with proteomic-based mass spectrometry for cancer tissue studies (Dilillo *et al.*, 2017; Coscia *et al.*, 2020; Herrera *et al.*, 2020). Laser-capture microdissection is commonly used on FFPE samples and has an advantage over whole tissue analysis as it allows for the isolation of specific cells or tissue areas, such as cellular or stromal compartments, leading to reduced tissue heterogeneity that may influence differential protein expression in non-tumorous and tumour samples. Laser-capture microdissection also serves as an enrichment technique which could be a useful approach when characterising less abundant ECM protein components (Johann Jr *et al.*, 2009; Coscia *et al.*, 2020).

MS protein identification and quantification is directly linked to efficient and reproducible enzyme digestion of proteins prior to LC-MS/MS analysis. While trypsin is commonly used for protein digestion, certain ECM proteins such as collagens are resistant to trypsin proteolysis thus negatively affecting downstream peptide sequence coverage (Van Huizen *et al.*, 2020). A possible suggestion to circumvent this would be to use trypsin and Lys-C whereby studies have shown that proteolysis is enhanced and fewer missed cleavages occur when trypsin/Lys-C combination is used for peptide digestion thus leading to improved peptide sequence coverage of proteolytically resistant proteins (Saveliev *et al.*, 2013). Other proteolytic enzymes could be tested to determine their ability to render low solubility proteins, like the fibrous collagens, soluble enough to achieve cleavage into suitable peptides. Chemical cleavage could also be attempted where the long fibres may be cleaved to give more soluble and less cross-linked structures more suitable for enzymatic cleavage to achieve mass spectrometer friendly peptides. Despite the fact that a number of ECM enrichment methods have been developed, it is yet to be determined whether these methods are more efficient than whole tissue analysis when characterising ECM in solid tumours. Therefore, future studies should involve comparative data analysis based on the method used in this current study and the above-mentioned ECM extraction strategies to establish how well these methods perform for the proteomic characterisation of primary breast tumour ECM.

4.3 Concluding remarks

The main aim of this study was achieved by using semi-automated sample preparation methodology, advanced LC-MS/MS instrumentation with SWATH-based DIA data acquisition and a comprehensive project specific library for protein identification and relative quantification to provide a robust and comprehensive proteomic assessment of primary breast tumour proteome. Additionally, proteomic data confirmed that protein homeostasis including the extracellular matrix is measurably altered in solid invasive ductal breast cancer tumours of the breast in cancer patients. Protein synthesis by tumour cells is enhanced and altered in order to meet the demands of uncontrolled cell proliferation seen in tumour growth. SWATH-based mass spectrometric data collection enabled reliable peptide quantification using total peptide area under the curve normalisation with a coefficient of variation of less than 20% and could be successfully used to identify differentially expressed

proteins within the tumour microenvironment that are involved in ECM remodelling, tumour angiogenesis, cancer cell proliferation, invasion, and metastasis. This exploratory study has laid the foundation for prognostic and pharmacological based studies for cancer therapeutics by identifying putative ECM protein candidates that can be further assessed in independent verification studies consisting of a larger study cohort. With the above-mentioned recommendations and large- scale verification and validation studies, the optimised analytical approach utilised in this study has the ability to provide valuable proteomic data leading to the advancement of breast cancer research associated with the tumour microenvironment.

References

Abner, AL, Collins, L, Peiro, G, Recht, A, Come, S, Shulman, LN, Silver, B, Nixon, A, Harris, JR & Schnitt, SJ 1998. Correlation of tumor size and axillary lymph node involvement with prognosis in patients with T1 breast carcinoma. *Cancer: Interdisciplinary International Journal of the American Cancer Society,* 83**,** 2502-2508.

Abotaleb, M, Kubatka, P, Caprnda, M, Varghese, E, Zolakova, B, Zubor, P, Opatrilova, R, Kruzliak, P, Stefanicka, P & Büsselberg, D 2018. Chemotherapeutic agents for the treatment of metastatic breast cancer: An update. *Biomedicine and Pharmacotherapy,* 101**,** 458-477.

Abrams, JS 2001. Adjuvant therapy for breast cancer—results from the USA consensus conference. *Breast Cancer,* 8**,** 298-304.

Acerbi, I, Cassereau, L, Dean, I, Shi, Q, Au, A, Park, C, Chen, Y, Liphardt, J, Hwang, E & Weaver, V 2015. Human breast cancer invasion and aggression correlates with ECM stiffening and immune cell infiltration. *Integrative Biology,* 7**,** 1120-1134.

Aggarwal, S & Yadav, AK 2016. False discovery rate estimation in proteomics. *Statistical Analysis in Proteomics.* Springer.

Aghamaliyev, U, Hajiyeva, Y & Rückert, F 2016. Desmoplastic Reaction In Pancreatic Ductal Adenocarcinoma. *Pancreas-Open Journal,* 1**,** 22-29.

Ahadi, M, Heibatollahi, M & Zahedifard, S 2020. Comparison of ER, PR, Ki67 and HER-2/neu Reactivity Pattern with Patients' Age, Histologic Grade, Tumor Size and Lymph Node Status in Invasive Ductal Breast Cancer. *International Journal of Cancer Management,* 13.

Ahn, SG, Kim, SJ, Kim, C & Jeong, J 2016. Molecular Classification of Triple-Negative Breast Cancer. *Journal of Breast Cancer,* 19**,** 223-230.

Al-Thoubaity, FK 2020. Molecular classification of breast cancer: A retrospective cohort study. *Annals of Medicine and Surgery,* 49**,** 44-48.

Albain, KS, Barlow, WE, Ravdin, PM, Farrar, WB, Burton, GV, Ketchel, SJ, Cobau, CD, Levine, EG, Ingle, JN & Pritchard, KI 2009. Adjuvant chemotherapy and timing of tamoxifen in postmenopausal patients with endocrine-responsive, node-positive breast cancer: a phase 3, open-label, randomised controlled trial. *The Lancet,* 374**,** 2055-2063.

Albini, A & Sporn, MB 2007. The tumour microenvironment as a target for chemoprevention. *Nature Reviews Cancer,* 7**,** 139-147.

Amenta, PS, Hadad, S, Lee, MT, Barnard, N, Li, D & Myers, JC 2003. Loss of types XV and XIX collagen precedes basement membrane invasion in ductal carcinoma of the female breast. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland,* 199**,** 298-308.

Amodei, D, Egertson, J, Maclean, BX, Johnson, R, Merrihew, GE, Keller, A, Marsh, D, Vitek, O, Mallick, P & Maccoss, MJ 2019. Improving precursor selectivity in data-independent acquisition using overlapping windows. *Journal of the American Society for Mass Spectrometry,* 30**,** 669-684.

Anampa, J, Makower, D & Sparano, JA 2015. Progress in adjuvant chemotherapy for breast cancer: an overview. *BMC Medicine,* 13**,** 1-13.

Andruska, N, Zheng, X, Yang, X, Helferich, WG & Shapiro, DJ 2015. Anticipatory estrogen activation of the unfolded protein response is linked to cell proliferation and poor survival in estrogen receptor α-positive breast cancer. *Oncogene,* 34**,** 3760-3769.

Angel, TE, Aryal, UK, Hengel, SM, Baker, ES, Kelly, RT, Robinson, EW & Smith, RD 2012. Mass spectrometry-based proteomics: existing capabilities and future directions. *Chemical Society Reviews,* 41**,** 3912-3928.

Aquino, RGFD, Vasques, PHD, Cavalcante, DI, Oliveira, ALDS, Oliveira, B & Pinheiro, LGP 2017. Invasive ductal carcinoma: relationship between pathological characteristics and the presence of axillary metastasis in 220 cases. *Revista do Colégio Brasileiro de Cirurgiões,* 44**,** 163-170.

Arcamone, F 2012. *Doxorubicin: anticancer antibiotics*, Elsevier.

Arteaga, CL 2002. Epidermal growth factor receptor dependence in human tumors: more than just expression? *The Oncologist,* 7**,** 31-39.

Arthurs, C, Murtaza, BN, Thomson, C, Dickens, K, Henrique, R, Patel, HR, Beltran, M, Millar, M, Thrasivoulou, C & Ahmed, A 2017. Expression of ribosomal proteins in normal and cancerous human prostate tissue. *PloS One,* 12**,** e0186047.

Aslam, B, Basit, M, Nisar, MA, Khurshid, M & Rasool, MH 2017. Proteomics: technologies and their applications. *Journal of Chromatographic Science,* 55**,** 182-196.

Ataollahi, MR, Sharifi, J, Paknahad, MR & Paknahad, A 2015. Breast cancer and associated factors: a review. *Journal of Medicine and Life,* 8**,** 6-11.

Avril, T, Vauleon, E & Chevet, E 2017. Endoplasmic reticulum stress signaling and chemotherapy resistance in solid cancers. *Oncogenesis,* 6**,** e373-e373.

Bae, YK, Kim, A, Kim, MK, Choi, JE, Kang, SH & Lee, SJ 2013. Fibronectin expression in carcinoma cells correlates with tumor aggressiveness and poor clinical outcome in patients with invasive breast cancer. *Human Pathology,* 44**,** 2028-2037.

Barcus, CE, Keely, PJ, Eliceiri, KW & Schuler, LA 2013. Stiff collagen matrices increase tumorigenic prolactin signaling in breast cancer cells. *Journal of Biological Chemistry,* 288**,** 12722-12732.

Barnard, ME, Boeke, CE & Tamimi, RM 2015. Established breast cancer risk factors and risk of intrinsic tumor subtypes. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer,* 1856**,** 73-85.

Baskin, D 2014. Fixation and tissue processing in immunohistochemistry.

Baum, M 1998. Tamoxifen--the treatment of choice. Why look for alternatives? *British Journal of Cancer,* 78**,** 1-4.

Becker, S 2015. A historic and scientific review of breast cancer: The next global healthcare challenge. *International Journal of Gynecology & Obstetrics,* 131**,** S36-S39.

Begam, AJ, Jubie, S & Nanjan, M 2017. Estrogen receptor agonists/antagonists in breast cancer therapy: A critical review. *Bioorganic Chemistry,* 71**,** 257-274.

Belin, S, Beghin, A, Solano-Gonzàlez, E, Bezin, L, Brunet-Manquat, S, Textoris, J, Prats, A-C, Mertani, HC, Dumontet, C & Diaz, J-J 2009. Dysregulation of ribosome biogenesis and translational capacity is associated with tumor progression of human breast cancer cells. *PloS One,* 4**,** e7147.

Ben-Shem, A, De Loubresse, NG, Melnikov, S, Jenner, L, Yusupova, G & Yusupov, M 2011. The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science,* 334**,** 1524-1529.

Benjamini, Y & Hochberg, Y 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological),* 57**,** 289-300.

Bergamaschi, A, Tagliabue, E, Sørlie, T, Naume, B, Triulzi, T, Orlandi, R, Russnes, H, Nesland, J, Tammi, R & Auvinen, P 2008. Extracellular matrix signature identifies breast cancer subgroups with different clinical outcome. *The Journal of Pathology,* 214**,** 357-367.

Berns, K, Horlings, HM, Hennessy, BT, Madiredjo, M, Hijmans, EM, Beelen, K, Linn, SC, Gonzalez-Angulo, AM, Stemke-Hale, K & Hauptmann, M 2007. A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer. *Cancer Cell,* 12**,** 395-402.

Bertos, NR & Park, M 2011. Breast cancer—one term, many entities? *The Journal of Clinical Investigation,* 121**,** 3789-3796.

Bogdanow, B, Zauber, H & Selbach, M 2016. Systematic errors in peptide and protein identification and quantification by modified peptides. *Molecular & Cellular Proteomics,* 15**,** 2791-2801.

Bonadonna, G, Brusamolino, E, Valagussa, P, Rossi, A, Brugnatelli, L, Brambilla, C, De Lena, M, Tancini, G, Bajetta, E & Musumeci, R 1976. Combination chemotherapy as an adjuvant treatment in operable breast cancer. *New England Journal of Medicine,* 294**,** 405-410.

Bonnans, C, Chou, J & Werb, Z 2014. Remodelling the extracellular matrix in development and disease. *Nature Reviews Molecular Cell Biology,* 15**,** 786-801.

Borràs, E & Sabidó, E 2017. What is targeted proteomics? A concise revision of targeted acquisition and targeted data analysis in mass spectrometry. *Proteomics,* 17**,** 1700180.

Borst, MJ & Ingold, JA 1993. Metastatic patterns of invasive lobular versus invasive ductal carcinoma of the breast. *Surgery,* 114**,** 637-642.

Bray, F, Ferlay, J, Soerjomataram, I, Siegel, RL, Torre, LA & Jemal, A 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians,* 68**,** 394-424.

Buck, M, Karustis, DG, Day, N, Honn, K & Sloane, BF 1992. Degradation of extracellular-matrix proteins by human cathepsin B from normal and tumour tissues. *Biochemical Journal,* 282**,** 273-278.

Burstein, HJ, Sun, Y, Dirix, LY, Jiang, Z, Paridaens, R, Tan, AR, Awada, A, Ranade, A, Jiao, S & Schwartz, G 2010. Neratinib, an irreversible ErbB receptor tyrosine kinase inhibitor, in patients with advanced ErbB2-positive breast cancer. *Journal of Clinical Oncology,* 28**,** 1301-1307.

Byron, A, Humphries, JD & Humphries, MJ 2013. Defining the extracellular matrix using proteomics. *International Journal of Experimental Pathology,* 94**,** 75-92.

Cannon, A, Thompson, C, Hall, BR, Jain, M, Kumar, S & Batra, SK 2018. Desmoplasia in pancreatic ductal adenocarcinoma: insight into pathological function and therapeutic potential. *Genes & Cancer,* 9**,** 78.

Carpenter, PM, Dao, AV, Arain, ZS, Chang, MK, Nguyen, HP, Arain, S, Wang-Rodriguez, J, Kwon, S-Y & Wilczynski, SP 2009. Motility induction in breast carcinoma by mammary epithelial laminin 332 (laminin 5). *Molecular Cancer Research,* 7**,** 462-475.

Catteau, X, Simon, P, Jondet, M, Vanhaeverbeek, M & Noël, J-C 2019. Quantification of stromal reaction in breast carcinoma and its correlation with tumor grade and free progression survival. *PloS One,* 14**,** e0210263.

Cattley, RC & Radinsky, BR 2004. Cancer therapeutics: understanding the mechanism of action. *Toxicologic Pathology,* 32**,** 116-121.

Cavallaro, U & Christofori, G 2004. Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. *Nature Reviews Cancer,* 4**,** 118-132.

Chang, J & Chaudhuri, O 2019. Beyond proteases: Basement membrane mechanics and cancer invasion. *Journal of Cell Biology,* 218**,** 2456-2469.

Cheang, MC, Chia, SK, Voduc, D, Gao, D, Leung, S, Snider, J, Watson, M, Davies, S, Bernard, PS & Parker, JS 2009. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *JNCI: Journal of the National Cancer Institute,* 101**,** 736-750.

Chen, I-C, Hsiao, L-P, Huang, I-W, Yu, H-C, Yeh, L-C, Lin, C-H, Chen, TW-W, Cheng, A-L & Lu, Y-S 2017. Phosphatidylinositol-3 kinase inhibitors, buparlisib and alpelisib, sensitize estrogen receptor-positive breast cancer cells to tamoxifen. *Scientific Reports,* 7**,** 1-10.

Chen, S, Zhang, J, Duan, L, Zhang, Y, Li, C, Liu, D, Ouyang, C, Lu, F & Liu, X 2014. Identification of HnRNP M as a novel biomarker for colorectal carcinoma by quantitative proteomics. *American Journal of Physiology-Gastrointestinal and Liver Physiology,* 306**,** G394-G403.

Chen, X, Wei, S, Ji, Y, Guo, X & Yang, F 2015. Quantitative proteomics using SILAC: principles, applications, and developments. *Proteomics,* 15**,** 3175-3192.

Chia, J, Kusuma, N, Anderson, R, Parker, B, Bidwell, B, Zamurs, L, Nice, E & Pouliot, N 2007. Evidence for a role of tumor-derived laminin-511 in the metastatic progression of breast cancer. *The American Journal of Pathology,* 170**,** 2135-2148.

Choccalingam, C, Rao, L & Rao, S 2012. Clinico-pathological characteristics of triple negative and non triple negative high grade breast carcinomas with and without basal marker (CK5/6 and EGFR) expression at a rural tertiary hospital in India. *Breast cancer: basic and clinical research,* 6**,** BCBCR. S8611.

Chong, HC, Tan, CK, Huang, R-L & Tan, NS 2012. Matricellular proteins: a sticky affair with cancers. *Journal of Oncology,* 2012.

Chuanyu, S, Yuqing, Z, Chong, X, Guowei, X & Xiaojun, Z 2017. Periostin promotes migration and invasion of renal cell carcinoma through the integrin/focal adhesion kinase/c-Jun N-terminal kinase pathway. *Tumor Biology,* 39**,** 1010428317694549.

Clemons, M & Goss, P 2001. Estrogen and the risk of breast cancer. *New England Journal of Medicine,* 344**,** 276-285.
Colleoni, M, Rotmensz, N, Peruzzotti, G, Maisonneuve, P, Mazzarol, G, Pruneri, G, Luini, A, Intra, M, Veronesi, P & Galimberti, V 2005. Size of breast cancer metastases in axillary lymph nodes: clinical relevance of minimal lymph node involvement. *Journal of Clinical Oncology,* 23**,** 1379-1389.
Coltri, PP, Dos Santos, MG & Da Silva, GH 2019. Splicing and cancer: Challenges and opportunities. *Wiley Interdisciplinary Reviews: RNA,* 10**,** e1527.
Conklin, MW & Keely, PJ 2012. Why the stroma matters in breast cancer: insights into breast cancer patient outcomes through the examination of stromal biomarkers. *Cell adhesion & migration,* 6**,** 249-260.
Corazzari, M, Gagliardi, M, Fimia, GM & Piacentini, M 2017. Endoplasmic reticulum stress, unfolded protein response, and cancer cell fate. *Frontiers in Oncology,* 7**,** 78.
Coscia, F, Doll, S, Bech, JM, Schweizer, L, Mund, A, Lengyel, E, Lindebjerg, J, Madsen, GI, Moreira, JM & Mann, M 2020. A streamlined mass spectrometry–based proteomics workflow for large-scale FFPE tissue analysis. *The Journal of Pathology,* 251**,** 100-112.
Cotlar, AM, Dubose, JJ & Rose, DM 2003. History of surgery for breast cancer: radical to the sublime. *Current Surgery,* 60**,** 329-337.
Cowell, CF, Weigelt, B, Sakr, RA, Ng, CK, Hicks, J, King, TA & Reis-Filho, JS 2013. Progression from ductal carcinoma in situ to invasive breast cancer: revisited. *Molecular Oncology,* 7**,** 859-869.
Crapo, PM, Gilbert, TW & Badylak, SF 2011. An overview of tissue and whole organ decellularization processes. *Biomaterials,* 32**,** 3233-3243.
Dai, X, Li, T, Bai, Z, Yang, Y, Liu, X, Zhan, J & Shi, B 2015. Breast cancer intrinsic subtype classification, clinical use and future trends. *American Journal of Cancer Research,* 5**,** 2929.
Datta, S, Malhotra, L, Dickerson, R, Chaffee, S, Sen, CK & Roy, S 2015. Laser capture microdissection: Big data from small samples. *Histology and Histopathology,* 30**,** 1255.
De Lucca Camargo, L, Babelova, A, Mieth, A, Weigert, A, Mooz, J, Rajalingam, K, Heide, H, Wittig, I, Lopes, LR & Brandes, RP 2013. Endo-PDI is required for TNFα-induced angiogenesis. *Free Radical Biology and Medicine,* 65**,** 1398-1407.
De Santis, R, Albertoni, C, Petronzelli, F, Campo, S, D'alessio, V, Rosi, A, Anastasi, AM, Lindstedt, R, Caroni, N & Arseni, B 2006. Low and high tenascin-expressing tumors are efficiently targeted by ST2146 monoclonal antibody. *Clinical cancer research,* 12**,** 2191-2196.
Declerck, YA 2012. Desmoplasia: a response or a niche? *Cancer Discovery,* 2**,** 772-774.
Delpech, B, Chevallier, B, Reinhardt, N, Julien, JP, Duval, C, Maingonnat, C, Bastit, P & Asselain, B 1990. Serum hyaluronan (hyaluronic acid) in breast cancer patients. *International Journal of Cancer,* 46**,** 388-390.
Deng, J, Erdjument-Bromage, H & Neubert, TA 2019. Quantitative comparison of proteomes using SILAC. *Current protocols in protein science,* 95**,** e74.
Denoix, PF, Schutzenberger & Viollet 1952. [Records to aid in the study of different types of breast cancer]. *Acta Unio Int Contra Cancrum,* 8**,** 136-139.
Depuydt, M, Messens, J & Collet, J-F 2011. How proteins form disulfide bonds. *Antioxidants & Redox Signaling,* 15**,** 49-66.
Desgrosellier, JS & Cheresh, DA 2010. Integrins in cancer: biological implications and therapeutic opportunities. *Nature Reviews Cancer,* 10**,** 9-22.
Diaz, LK, Cristofanilli, M, Zhou, X, Welch, KL, Smith, TL, Yang, Y, Sneige, N, Sahin, AA & Gilcrease, MZ 2005. β4 integrin subunit gene expression correlates with tumor size and nuclear grade in early breast cancer. *Modern Pathology,* 18**,** 1165-1175.
Dilillo, M, Pellegrini, D, Ait-Belkacem, R, De Graaf, EL, Caleo, M & Mcdonnell, LA 2017. Mass spectrometry imaging, laser capture microdissection, and LC-MS/MS of the same tissue section. *Journal of Proteome Research,* 16**,** 2993-3001.

Doncheva, NT, Morris, JH, Gorodkin, J & Jensen, LJ 2018. Cytoscape StringApp: network analysis and visualization of proteomics data. *Journal of Proteome Research,* 18**,** 623-632.

Duffy, MJ, Maguire, TM, Hill, A, Mcdermott, E & O'higgins, N 2000. Metalloproteinases: role in breast carcinogenesis, invasion and metastasis. *Breast Cancer Research,* 2**,** 252.

Dumartin, L, Whiteman, HJ, Weeks, ME, Hariharan, D, Dmitrovic, B, Iacobuzio-Donahue, CA, Brentnall, TA, Bronner, MP, Feakins, RM & Timms, JF 2011. AGR2 is a novel surface antigen that promotes the dissemination of pancreatic cancer cells through regulation of cathepsins B and D. *Cancer Research,* 71**,** 7091-7102.

Dunn, MJ 1993. *Gel Electrophoresis: Proteins*, BIOS Scientific Publishers Limited.

Dupree, EJ, Jayathirtha, M, Yorkey, H, Mihasan, M, Petre, BA & Darie, CC 2020. A critical review of bottom-up proteomics: The good, the bad, and the future of this field. *Proteomes,* 8**,** 14.

Easton, D, Ford, D & Peto, J 1993a. Inherited susceptibility to breast cancer. *Cancer Surveys,* 18**,** 95-113.

Easton, DF, Bishop, D, Ford, D & Crockford, G 1993b. Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The Breast Cancer Linkage Consortium. *American Journal of Human Genetics,* 52**,** 678.

Edge, J, Buccimazza, I, Cubasch, H & Panieri, E 2014. The challenges of managing breast cancer in the developing world-a perspective from sub-Saharan Africa. *SAMJ: South African Medical Journal,* 104**,** 377-379.

Efthymiou, G, Saint, A, Ruff, M, Rekad, Z, Ciais, D & Obberghen-Schilling, V 2020. Shaping up the tumor microenvironment with cellular fibronectin. *Frontiers in Oncology,* 10**,** 641.

Egeblad, M, Nakasone, ES & Werb, Z 2010a. Tumors as organs: complex tissues that interface with the entire organism. *Developmental Cell,* 18**,** 884-901.

Egeblad, M, Rasch, MG & Weaver, VM 2010b. Dynamic interplay between the collagen scaffold and tumor evolution. *Current Opinion in Cell Biology,* 22**,** 697-706.

Elias, JE & Gygi, SP 2007. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods,* 4**,** 207-214.

Elias, JE & Gygi, SP 2010. Target-decoy search strategy for mass spectrometry-based proteomics. *Proteome bioinformatics.* Springer.

Eliyatkın, N, Yalçın, E, Zengel, B, Aktaş, S & Vardar, E 2015. Molecular Classification of Breast Carcinoma: From Traditional, Old-Fashioned Way to A New Age, and A New Way. *The journal of breast health,* 11**,** 59-66.

Ellis, H & Mahadevan, V 2013. Anatomy and physiology of the breast. *Surgery (oxford),* 31**,** 11-14.

Escobar-Hoyos, L, Knorr, K & Abdel-Wahab, O 2019. Aberrant RNA splicing in cancer. *Annual Review of Cancer Biology,* 3**,** 167-185.

Eshchenko, TY, Rykova, V, Chernakov, A, Sidorov, S & Grigorieva, E 2007. Expression of different proteoglycans in human breast tumors. *Biochemistry (Moscow),* 72**,** 1016-1020.

Ezhilarasan, S 2016. Wax embedding technique-innovative and cost effective method of preservation of human cadaveric specimens. *Indian Journal of Clinical Anatomy and Physiology,* 3**,** 423-426.

Fang, M, Yuan, J, Peng, C & Li, Y 2014. Collagen as a double-edged sword in tumor progression. *Tumor Biology,* 35**,** 2871-2882.

Feldman, AT & Wolfe, D 2014. Tissue processing and hematoxylin and eosin staining. *Histopathology.* Springer.

Fernandez-Garcia, B, Eiró, N, Marín, L, González-Reyes, S, González, LO, Lamelas, ML & Vizoso, FJ 2014. Expression and prognostic significance of fibronectin and matrix metalloproteases in breast cancer metastasis. *Histopathology,* 64**,** 512-522.

Fernandez, PM, Tabbara, SO, Jacobs, LK, Manning, FC, Tsangaris, TN, Schwartz, AM, Kennedy, KA & Patierno, SR 2000. Overexpression of the glucose-regulated stress gene GRP78 in malignant but not benign human breast lesions. *Breast Cancer Research and Treatment,* 59**,** 15-26.

Fidler, IJ 2003. The pathogenesis of cancer metastasis: the'seed and soil'hypothesis revisited. *Nature Reviews Cancer,* 3**,** 453-458.

Fischer, AH, Jacobson, KA, Rose, J & Zeller, R 2008. Hematoxylin and eosin staining of tissue and cell sections. *Cold spring harbor protocols,* 2008**,** pdb. prot4986.

Fisher, B, Brown, A, Mamounas, E, Wieand, S, Robidoux, A, Margolese, RG, Cruz Jr, AB, Fisher, ER, Wickerham, DL & Wolmark, N 1997. Effect of preoperative chemotherapy on local-regional disease in women with operable breast cancer: findings from National Surgical Adjuvant Breast and Bowel Project B-18. *Journal of Clinical Oncology,* 15**,** 2483-2493.

Fisher, B, Carbone, P, Economou, SG, Frelick, R, Glass, A, Lerner, H, Redmond, C, Zelen, M, Band, P & Katrych, DL 1975. L-Phenylalanine mustard (L-PAM) in the management of primary breast cancer: a report of early findings. *New England Journal of Medicine,* 292**,** 117-122.

Fisher, B, Ravdin, RG, Ausman, RK, Slack, NH, Moore, GE & Noer, RJ 1968. Surgical adjuvant chemotherapy in cancer of the breast: results of a decade of cooperative investigation. *Annals of Surgery,* 168**,** 337.

Fleming, RA 1997. An overview of cyclophosphamide and ifosfamide pharmacology. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy,* 17**,** 146S-154S.

Folkman, J 2006. Antiangiogenesis in cancer therapy—endostatin and its mechanisms of action. *Experimental Cell Research,* 312**,** 594-607.

Fonović, M & Turk, B 2014. Cysteine cathepsins and extracellular matrix degradation. *Biochimica et Biophysica Acta (BBA)-general subjects,* 1840**,** 2560-2570.

Frank, TS, Deffenbaugh, AM, Reid, JE, Hulick, M, Ward, BE, Lingenfelter, B, Gumpper, KL, Scholl, T, Tavtigian, SV & Pruss, DR 2002. Clinical characteristics of individuals with germline mutations in BRCA1 and BRCA2: analysis of 10,000 individuals. *Journal of Clinical Oncology,* 20**,** 1480-1490.

Frantz, C, Stewart, KM & Weaver, VM 2010. The extracellular matrix at a glance. *Journal of Cell Science,* 123**,** 4195-4200.

Fu, Q, Kowalski, MP, Mastali, M, Parker, SJ, Sobhani, K, Van Den Broek, I, Hunter, CL & Van Eyk, JE 2018. Highly reproducible automated proteomics sample preparation workflow for quantitative mass spectrometry. *Journal of Proteome Research,* 17**,** 420-428.

Gabrijelcic, D, Svetic, B, Spaić, D, Skrk, J, Budihna, M, Dolenc, I, Popovic, T, Cotic, V & Turk, V. Cathepsins B, H and L in human breast carcinoma. European journal of clinical chemistry and clinical biochemistry: journal of the Forum of European Clinical Chemistry Societies, 1992. 69-74.

Gajdos, C, Tartter, PI & Bleiweiss, IJ 1999. Lymphatic invasion, tumor size, and age are independent predictors of axillary lymph node metastases in women with T1 breast cancers. *Annals of Surgery,* 230**,** 692.

Galea, MH, Wilson, A, Roebuck, EJ, Elston, CW, Ellis, IO & Blamey, RW 1992. Breast cancer screening: A benefit can be anticipated. *The Breast,* 1**,** 155.

Garfin, DE 2009. One-dimensional gel electrophoresis. *Methods in Enzymology,* 463**,** 497-513.

Geiger, T, Madden, SF, Gallagher, WM, Cox, J & Mann, M 2012. Proteomic portrait of human breast cancer progression identifies novel prognostic markers. *Cancer Research,* 72**,** 2428-2439.

Geuens, T, Bouhy, D & Timmerman, V 2016. The hnRNP family: insights into their role in health and disease. *Human Genetics,* 135**,** 851-867.

Gifford, JB & Hill, R 2018. GRP78 influences chemoresistance and prognosis in cancer. *Current Drug Targets,* 19**,** 701-708.

Gilkes, DM, Chaturvedi, P, Bajpai, S, Wong, CC, Wei, H, Pitcairn, S, Hubbi, ME, Wirtz, D & Semenza, GL 2013. Collagen prolyl hydroxylases are essential for breast cancer metastasis. *Cancer Research,* 73**,** 3285-3296.

Gill, JK, Maskarinec, G, Pagano, I & Kolonel, LN 2006. The association of mammographic density with ductal carcinoma in situ of the breast: the Multiethnic Cohort. *Breast Cancer Research,* 8**,** R30.

Gillet, LC, Navarro, P, Tate, S, Röst, H, Selevsek, N, Reiter, L, Bonner, R & Aebersold, R 2012. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new

concept for consistent and accurate proteome analysis. *Molecular & Cellular Proteomics,* 11**,** O111. 016717.

Giussani, M, Merlino, G, Cappelletti, V, Tagliabue, E & Daidone, MG. Tumor-extracellular matrix interactions: Identification of tools associated with breast cancer progression. Seminars in Cancer Biology, 2015. Elsevier, 3-10.

Giussani, M, Triulzi, T, Sozzi, G & Tagliabue, E 2019. Tumor extracellular matrix remodeling: new perspectives as a circulating tool in the diagnosis and prognosis of solid tumors. *Cells,* 8**,** 81.

Glover, D, Lipton, A, Keller, A, Miller, AA, Browning, S, Fram, RJ, George, S, Zelenakas, K, Macerata, RS & Seaman, JJ 1994. Intravenous pamidronate disodium treatment of bone metastases in patients with breast cancer. A dose-seeking study. *Cancer: Interdisciplinary International Journal of the American Cancer Society,* 74**,** 2949-2955.

Gocheva, V, Zeng, W, Ke, D, Klimstra, D, Reinheckel, T, Peters, C, Hanahan, D & Joyce, JA 2006. Distinct roles for cysteine cathepsin genes in multistage tumorigenesis. *Genes & Development,* 20**,** 543-556.

Godwin, AR, Singh, M, Lockhart-Cairns, MP, Alanazi, YF, Cain, SA & Baldock, C 2019. The role of fibrillin and microfibril binding proteins in elastin and elastic fibre assembly. *Matrix Biology,* 84**,** 17-30.

Goicoechea, S, Orr, AW, Pallero, MA, Eggleton, P & Murphy-Ullrich, JE 2000. Thrombospondin mediates focal adhesion disassembly through interactions with cell surface calreticulin. *Journal of Biological Chemistry,* 275**,** 36358-36368.

Goldfarb, D, Wang, W & Major, MB 2016. MSAcquisitionSimulator: data-dependent acquisition simulator for LC-MS shotgun proteomics. *Bioinformatics,* 32**,** 1269-1271.

Goldhirsch, A, Winer, EP, Coates, A, Gelber, R, Piccart-Gebhart, M, Thürlimann, B, Senn, H-J, Members, P, Albain, KS & André, F 2013. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Annals of Oncology,* 24**,** 2206-2223.

Goldhirsch, A, Wood, WC, Coates, AS, Gelber, RD, Thürlimann, B, Senn, H-J & Members, P 2011. Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Annals of Oncology,* 22**,** 1736-1747.

Gong, J, Chehrazi-Raffle, A, Reddi, S & Salgia, R 2018. Development of PD-1 and PD-L1 inhibitors as a form of cancer immunotherapy: a comprehensive review of registration trials and future considerations. *Journal for Immunotherapy of Cancer,* 6**,** 1-18.

González-González, L & Alonso, J 2018. Periostin: a matricellular protein with multiple functions in cancer development and progression. *Frontiers in Oncology,* 8**,** 225.

Goss, PE, Ingle, JN, Martino, S, Robert, NJ, Muss, HB, Piccart, MJ, Castiglione, M, Tu, D, Shepherd, LE & Pritchard, KI 2003. A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer. *New England Journal of Medicine,* 349**,** 1793-1802.

Govaert, E, Van Steendam, K, Willems, S, Vossaert, L, Dhaenens, M & Deforce, D 2017. Comparison of fractionation proteomics for local SWATH library building. *Proteomics,* 17**,** 1700052.

Gräntzdörffer, I, Yumlu, S, Gioeva, Z, Von Wasielewski, R, Ebert, MP & Röcken, C 2010. Comparison of different tissue sampling methods for protein extraction from formalin-fixed and paraffin-embedded tissue specimens. *Experimental and Molecular Pathology,* 88**,** 190-196.

Graves, PR & Haystead, TA 2002. Molecular biologist's guide to proteomics. *Microbiology and Molecular Biology Reviews,* 66**,** 39-63.

Grillo, F, Bruzzone, M, Pigozzi, S, Prosapio, S, Migliora, P, Fiocca, R & Mastracci, L 2017. Immunohistochemistry on old archival paraffin blocks: is there an expiry date? *Journal of Clinical Pathology,* 70**,** 988-993.

Griss, J 2016. Spectral library searching in proteomics. *Proteomics,* 16**,** 729-740.

Gross, V, Carlson, G, Kwan, AT, Smejkal, G, Freeman, E, Ivanov, AR & Lazarev, A 2008. Tissue fractionation by hydrostatic pressure cycling technology: the unified sample preparation technique for systems biology studies. *Journal of biomolecular techniques: JBT,* 19**,** 189.

Guo, W, Huai, Q, Zhang, G, Guo, L, Song, P, Xue, X, Tan, F, Xue, Q, Gao, S & He, J 2021. Elevated heterogeneous nuclear ribonucleoprotein C expression correlates with poor prognosis in patients with surgically resected lung adenocarcinoma. *Frontiers in Oncology,* 10**,** 3145.

Gustafsson, OJ, Arentz, G & Hoffmann, P 2015. Proteomic developments in the analysis of formalin-fixed tissue. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics,* 1854**,** 559-580.

Halsted, WS 1894. I. The results of operations for the cure of cancer of the breast performed at the Johns Hopkins Hospital from June, 1889, to January, 1894. *Annals of Surgery,* 20**,** 497.

Hamadeh, IS, Patel, JN, Rusin, S & Tan, AR 2018. Personalizing aromatase inhibitor therapy in patients with breast cancer. *Cancer Treatment Reviews,* 70**,** 47-55.

Henson, DE, Ries, L, Freedman, LS & Carriaga, M 1991. Relationship among outcome, stage of disease, and histologic grade for 22,616 cases of breast cancer. The basis for a prognostic index. *Cancer,* 68**,** 2142-2149.

Herrera, JA, Mallikarjun, V, Rosini, S, Montero, MA, Lawless, C, Warwood, S, O'cualain, R, Knight, D, Schwartz, MA & Swift, J 2020. Laser capture microdissection coupled mass spectrometry (LCM-MS) for spatially resolved analysis of formalin-fixed and stained human lung tissues. *Clinical Proteomics,* 17**,** 1-12.

Hetz, C, Chevet, E & Oakes, SA 2015. Proteostasis control by the unfolded protein response. *Nature Cell Biology,* 17**,** 829-838.

Hewitt, R, Powe, D, Morrell, K, Balley, E, Leach, I, Ellis, I & Turner, D 1997. Laminin and collagen IV subunit distribution in normal and neoplastic tissues of colorectum and breast. *British Journal of Cancer,* 75**,** 221-229.

Higgins, MJ & Baselga, J 2011. Targeted therapies for breast cancer. *The Journal of Clinical Investigation,* 121**,** 3797-3803.

Ho, CS, Lam, C, Chan, M, Cheung, R, Law, L, Lit, L, Ng, K, Suen, M & Tai, H 2003. Electrospray ionisation mass spectrometry: principles and clinical applications. *The Clinical Biochemist Reviews,* 24**,** 3.

Howell, SJ, Johnston, SRD & Howell, A 2004. The use of selective estrogen receptor modulators and selective estrogen receptor down-regulators in breast cancer. *Best Practice & Research Clinical Endocrinology & Metabolism,* 18**,** 47-66.

Hu, A, Noble, WS & Wolf-Yadlin, A 2016. Technical advances in proteomics: new developments in data-independent acquisition. *F1000Research,* 5.

Hu, G, Li, L & Xu, W 2017. Extracellular matrix in mammary gland development and breast cancer progression. *Frontiers in Laboratory Medicine,* 1**,** 36-39.

Huang, Q, Yang, L, Luo, J, Guo, L, Wang, Z, Yang, X, Jin, W, Fang, Y, Ye, J & Shan, B 2015. SWATH enables precise label-free quantification on proteome scale. *Proteomics,* 15**,** 1215-1223.

Hubler, SL, Kumar, P, Mehta, S, Easterly, C, Johnson, JE, Jagtap, PD & Griffin, TJ 2019. Challenges in Peptide-Spectrum Matching: a Robust and Reproducible Statistical Framework for Removing Low-Accuracy, High-Scoring Hits. *Journal of Proteome Research,* 19**,** 161-173.

Hudis, CA 2007. Trastuzumab—mechanism of action and use in clinical practice. *New England Journal of Medicine,* 357**,** 39-51.

Huitema, A, Spaander, M, Mathôt, R, Tibben, M, Holtkamp, M, Beijnen, J & Rodenhuis, S 2002. Relationship between exposure and toxicity in high-dose chemotherapy with cyclophosphamide, thiotepa and carboplatin. *Annals of Oncology,* 13**,** 374-384.

Hynes, RO 2009. The extracellular matrix: not just pretty fibrils. *Science,* 326**,** 1216-1219.

Iacobuzio-Donahue, CA, Argani, P, Hempen, PM, Jones, J & Kern, SE 2002a. The desmoplastic response to infiltrating breast carcinoma: gene expression at the site of primary invasion and implications for comparisons between tumor types. *Cancer Research,* 62**,** 5351-5357.

Iacobuzio-Donahue, CA, Ryu, B, Hruban, RH & Kern, SE 2002b. Exploring the host desmoplastic response to pancreatic carcinoma: gene expression of stromal and neoplastic cells at the site of primary invasion. *The American Journal of Pathology,* 160**,** 91-99.

Inic, Z, Zegarac, M, Inic, M, Markovic, I, Kozomara, Z, Djurisic, I, Inic, I, Pupic, G & Jancic, S 2014. Difference between luminal A and luminal B subtypes according to Ki-67, tumor size, and progesterone receptor negativity providing prognostic information. *Clinical Medicine Insights: Oncology,* 8**,** CMO. S18006.

Insua-Rodríguez, J & Oskarsson, T 2016. The extracellular matrix in breast cancer. *Advanced Drug Delivery Reviews,* 97**,** 41-55.

Ioachim, E, Charchanti, A, Briasoulis, E, Karavasilis, V, Tsanou, H, Arvanitis, D, Agnantis, N & Pavlidis, N 2002. Immunohistochemical expression of extracellular matrix components tenascin, fibronectin, collagen type IV and laminin in breast cancer: their prognostic value and role in tumour invasion and progression. *European Journal of Cancer,* 38**,** 2362-2370.

Iozzo, RV 2005. Basement membrane proteoglycans: from cellar to ceiling. *Nature Reviews Molecular Cell Biology,* 6**,** 646-656.

Ishihara, A, Yoshida, T, Tamaki, H & Sakakura, T 1995. Tenascin expression in cancer cells and stroma of human breast cancer and its prognostic significance. *Clinical Cancer Research,* 1**,** 1035-1041.

Johann Jr, DJ, Rodriguez-Canales, J, Mukherjee, S, Prieto, DA, Hanson, JC, Emmert-Buck, M & Blonder, J 2009. Approaching solid tumor heterogeneity on a cellular basis by tissue proteomics using laser capture microdissection and biological mass spectrometry. *Journal of Proteome Research,* 8**,** 2310-2318.

Johnson, J, Decker, S, Zaharevitz, D, Rubinstein, L, Venditti, J, Schepartz, S, Kalyandrug, S, Christian, M, Arbuck, S & Hollingshead, M 2001. Relationships between drug activity in NCI preclinical in vitro and in vivo models and early clinical trials. *British Journal of Cancer,* 84**,** 1424.

Jordan, MA & Wilson, L 1998. Microtubules and actin filaments: dynamic targets for cancer chemotherapy. *Current Opinion in Cell Biology,* 10**,** 123-130.

Jordan, VC & Brodie, AM 2007. Development and evolution of therapies targeted to the estrogen receptor for the treatment and prevention of breast cancer. *Steroids,* 72**,** 7-25.

Joseph, C, Alsaleem, M, Orah, N, Narasimha, PL, Miligy, IM, Kurozumi, S, Ellis, IO, Mongan, NP, Green, AR & Rakha, EA 2020. Elevated MMP9 expression in breast cancer is a predictor of shorter patient survival. *Breast Cancer Research and Treatment,* 182**,** 267-282.

Jylhä, A, Nättinen, J, Aapola, U, Mikhailova, A, Nykter, M, Zhou, L, Beuerman, R & Uusitalo, H 2018. Comparison of iTRAQ and SWATH in a clinical study with multiple time points. *Clinical Proteomics,* 15**,** 1-11.

Kadler, KE, Baldock, C, Bella, J & Boot-Handford, RP 2007. Collagens at a glance. *Journal of Cell Science,* 120**,** 1955-1958.

Kalluri, R 2016. The biology and function of fibroblasts in cancer. *Nature Reviews Cancer,* 16**,** 582.

Kaplan, RN, Riba, RD, Zacharoulis, S, Bramley, AH, Vincent, L, Costa, C, Macdonald, DD, Jin, DK, Shido, K & Kerns, SA 2005. VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. *Nature,* 438**,** 820-827.

Karousou, E, D'angelo, ML, Kouvidi, K, Vigetti, D, Viola, M, Nikitovic, D, De Luca, G & Passi, A 2014. Collagen VI and hyaluronan: the common role in breast cancer. *BioMed research international,* 2014.

Katari, SK, Pasala, C, Nalamolu, RM, Vankadoth, UN, Alexander, SP, Pakala, SR, Bitla, AR & Umamaheswari, A 2019. Pathophysiology of matrix metalloproteinases in breast cancer progression. *Journal of Clinical and Scientific Research,* 8**,** 145.

Katayama, M & Sekiguchi, K 2004. Laminin-5 in epithelial tumour invasion. *Journal of Molecular Histology,* 35**,** 277-286.

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. *The Journal of Pathology,* 186**,** 262-268.

Keane, TJ, Swinehart, IT & Badylak, SF 2015. Methods of tissue decellularization used for preparation of biologic scaffolds and in vivo relevance. *Methods,* 84**,** 25-34.

Keihanian, S, Koochaki, N, Pouya, M & Zakerihamidi, M 2019. Factors Affecting axillary lymph node involvement in patients with breast cancer. *Tehran University Medical Journal TUMS Publications,* 77**,** 484-490.

Kelemen, O, Convertini, P, Zhang, Z, Wen, Y, Shen, M, Falaleeva, M & Stamm, S 2013. Function of alternative splicing. *Gene,* 514**,** 1-30.

Khamis, ZI, Sahab, ZJ & Sang, Q-XA 2012. Active roles of tumor stroma in breast cancer metastasis. *International Journal of Breast Cancer,* 2012.

Khan, MM, Simizu, S, Suzuki, T, Masuda, A, Kawatani, M, Muroi, M, Dohmae, N & Osada, H 2012. Protein disulfide isomerase-mediated disulfide bonds regulate the gelatinolytic activity and secretion of matrix metalloproteinase-9. *Experimental Cell Research,* 318**,** 904-914.

Khurana, S & George, SP 2008. Regulation of cell structure and function by actin-binding proteins: villin's perspective. *FEBS Letters,* 582**,** 2128-2139.

Kielty, CM & Grant, ME 2002. The collagen family: structure, assembly, and organization in the extracellular matrix. *In:* Royce, PMaS, B., (ed.) *Connective tissue and its heritable disorders: molecular, genetic, and medical aspects.* New York: Wiley-Liss, Inc.

Kiernan, JA 2000. Formaldehyde, formalin, paraformaldehyde and glutaraldehyde: what they are and what they do. *Microscopy Today,* 8**,** 8-13.

Kim, BG, An, HJ, Kang, S, Choi, YP, Gao, M-Q, Park, H & Cho, NH 2011. Laminin-332-rich tumor microenvironment for tumor invasion in the interface zone of breast cancer. *The American Journal of Pathology,* 178**,** 373-381.

Kim, G-E, Lee, JS, Park, MH & Yoon, JH 2017. Epithelial periostin expression is correlated with poor survival in patients with invasive breast carcinoma. *PloS One,* 12**,** e0187635.

Kim, Y-I & Cho, J-Y 2019. Gel-based proteomics in disease research: Is it still valuable? *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics,* 1867**,** 9-16.

Kim, Y-M, Hwang, S, Kim, Y-M, Pyun, B-J, Kim, T-Y, Lee, S-T, Gho, YS & Kwon, Y-G 2002. Endostatin blocks vascular endothelial growth factor-mediated signaling via direct interaction with KDR/Flk-1. *Journal of Biological Chemistry,* 277**,** 27872-27879.

Kim, Y-M, Jang, J-W, Lee, O-H, Yeon, J, Choi, E-Y, Kim, K-W, Lee, S-T & Kwon, Y-G 2000. Endostatin inhibits endothelial and tumor cellular invasion by blocking the activation and catalytic activity of matrix metalloproteinase 2. *Cancer Research,* 60**,** 5410-5413.

Kimura, K, Shimazu, K, Toki, T, Misawa, M, Fukuda, K, Yoshida, T, Taguchi, D, Fukuda, S, Iijima, K & Takahashi, N 2020. Outcome of colorectal cancer in Diamond–Blackfan syndrome with a ribosomal protein S19 mutation. *Clinical Journal of Gastroenterology,* 13**,** 1173-1177.

Kischel, P, Waltregny, D, Dumont, B, Turtoi, A, Greffe, Y, Kirsch, S, De Pauw, E & Castronovo, V 2010. Versican overexpression in human breast cancer lesions: known and new isoforms for stromal tumor targeting. *International Journal of Cancer,* 126**,** 640-650.

Kleemann, M, Schneider, H, Unger, K, Sander, P, Schneider, EM, Fischer-Posovszky, P, Handrick, R & Otte, K 2018. MiR-744-5p inducing cell death by directly targeting HNRNPC and NFIX in ovarian cancer cells. *Scientific Reports,* 8**,** 1-15.

Klont, F, Bras, L, Wolters, JC, Ongay, S, Bischoff, R, Halmos, GB & Horvatovich, PT 2018. Assessment of sample preparation bias in mass spectrometry-based proteomics. *Analytical Chemistry,* 90**,** 5405-5413.

Koedoot, E, Wolters, L, Van De Water, B & Le Dévédec, SE 2019. Splicing regulatory factors in breast cancer hallmarks and disease progression. *Oncotarget,* 10**,** 6021.

Koh, CM, Bezzi, M, Low, DH, Ang, WX, Teo, SX, Gay, FP, Al-Haddawi, M, Tan, SY, Osato, M & Sabo, A 2015. MYC regulates the core pre-mRNA splicing machinery as an essential step in lymphomagenesis. *Nature,* 523**,** 96-100.

Koopmans, F, Ho, JT, Smit, AB & Li, KW 2018. Comparative analyses of data independent acquisition mass spectrometric approaches: DIA, WiSIM-DIA, and untargeted DIA. *Proteomics,* 18**,** 1700304.

Kwapisz, D 2017. Cyclin-dependent kinase 4/6 inhibitors in breast cancer: palbociclib, ribociclib, and abemaciclib. *Breast Cancer Research and Treatment,* 166**,** 41-54.

Kwon, S-Y, Chae, SW, Wilczynski, SP, Arain, A & Carpenter, PM 2012. Laminin 332 expression in breast carcinoma. *Applied immunohistochemistry & molecular morphology: AIMM/official publication of the Society for Applied Immunohistochemistry,* 20**,** 159.

Laemmli, UK 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* 227**,** 680-685.

Lee, E 2017. Emerging roles of protein disulfide isomerase in cancer. *BMB Reports,* 50**,** 401.

Lee, E, Nichols, P, Spicer, D, Groshen, S, Mimi, CY & Lee, AS 2006. GRP78 as a novel predictor of responsiveness to chemotherapy in breast cancer. *Cancer Research,* 66**,** 7849-7853.

Lee, H-H, Lim, C-A, Cheong, Y-T, Singh, M & Gam, L-H 2012. Comparison of protein expression profiles of different stages of lymph nodes metastasis in breast cancer. *International Journal of Biological Sciences,* 8**,** 353.

Lehmann-Che, J, André, F, Desmedt, C, Mazouni, C, Giacchetti, S, Turpin, E, Espié, M, Plassa, L-F, Marty, M & Bertheau, P 2010. Cyclophosphamide dose intensification may circumvent anthracycline resistance of p53 mutant breast cancers. *The oncologist,* 15**,** 246.

Leppä, S, Saarto, T, Vehmanen, L, Blomqvist, C & Elomaa, I 2004. A high serum matrix metalloproteinase-2 level is associated with an adverse prognosis in node-positive breast carcinoma. *Clinical cancer research,* 10**,** 1057-1063.

Levental, KR, Yu, H, Kass, L, Lakins, JN, Egeblad, M, Erler, JT, Fong, SF, Csiszar, K, Giaccia, A & Weninger, W 2009. Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell,* 139**,** 891-906.

Levine, MN, Bramwell, VH, Pritchard, KI, Norris, BD, Shepherd, LE, Abu-Zahra, H, Findlay, B, Warr, D, Bowman, D & Myles, J 1998. Randomized trial of intensive cyclophosphamide, epirubicin, and fluorouracil chemotherapy compared with cyclophosphamide, methotrexate, and fluorouracil in premenopausal women with node-positive breast cancer. National Cancer Institute of Canada Clinical Trials Group. *Journal of Clinical Oncology,* 16**,** 2651-2658.

Levitsky, LI, Ivanov, MV, Lobas, AA & Gorshkov, MV 2017. Unbiased false discovery rate estimation for shotgun proteomics based on the target-decoy approach. *Journal of Proteome Research,* 16**,** 393-397.

Leygue, E, Snell, L, Dotzlaw, H, Hole, K, Hiller-Hitchcock, T, Roughley, PJ, Watson, PH & Murphy, LC 1998. Expression of lumican in human breast carcinoma. *Cancer Research,* 58**,** 1348-1352.

Leygue, E, Snell, L, Dotzlaw, H, Troup, S, Hiller-Hitchcock, T, Murphy, LC, Roughley, PJ & Watson, PH 2000. Lumican and decorin are differentially expressed in human breast carcinoma. *The Journal of Pathology,* 192**,** 313-320.

Li, C, Uribe, D & Daling, J 2005. Clinical characteristics of different histologic types of breast cancer. *British Journal of Cancer,* 93**,** 1046-1052.

Li, H, Qiu, Z, Li, F & Wang, C 2017. The relationship between MMP-2 and MMP-9 expression levels with breast cancer incidence and prognosis. *Oncology Letters,* 14**,** 5865-5870.

Lindemann, C, Thomanek, N, Hundt, F, Lerari, T, Meyer, HE, Wolters, D & Marcus, K 2017. Strategies in relative and absolute quantitative mass spectrometry based proteomics. *Biological Chemistry,* 398**,** 687-699.

Lingasamy, P, Tobi, A, Haugas, M, Hunt, H, Paiste, P, Asser, T, Rätsep, T, Kotamraju, VR, Bjerkvig, R & Teesalu, T 2019. Bi-specific tenascin-C and fibronectin targeted peptide for solid tumor delivery. *Biomaterials,* 219**,** 119373.

Liot, S, Aubert, A, Hervieu, V, El Kholti, N, Schalkwijk, J, Verrier, B, Valcourt, U & Lambert, E 2020. Loss of Tenascin-X expression during tumor progression: A new pan-cancer marker. *Matrix Biology Plus***,** 100021.

Liu, S & Cheng, C 2013. Alternative RNA splicing and cancer. *Wiley Interdisciplinary Reviews: RNA,* 4**,** 547-566.

Lowy, CM & Oskarsson, T 2015. Tenascin C in metastasis: A view from the invasive front. *Cell adhesion & migration,* 9**,** 112-124.

Lu, J & Gao, FH 2016. Role and molecular mechanism of heterogeneous nuclear ribonucleoprotein K in tumor development and progression. *Biomedical reports,* 4**,** 657-663.

Lu, P, Weaver, VM & Werb, Z 2012. The extracellular matrix: a dynamic niche in cancer progression. *The Journal of Cell Biology,* 196**,** 395-406.

Lucas, N, Robinson, AB, Marcker Espersen, M, Mahboob, S, Xavier, D, Xue, J, Balleine, RL, Defazio, A, Hains, PG & Robinson, PJ 2018. Accelerated barocycler lysis and extraction sample preparation for clinical proteomics by mass spectrometry. *Journal of Proteome Research,* 18**,** 399-405.

Ludwig, C, Gillet, L, Rosenberger, G, Amon, S, Collins, BC & Aebersold, R 2018. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Molecular Systems Biology,* 14**,** e8126.

Lugano, R, Ramachandran, M & Dimberg, A 2020. Tumor angiogenesis: causes, consequences, challenges and opportunities. *Cellular and Molecular Life Sciences,* 77**,** 1745-1770.

Lynch, KL 2016. CLSI C62-A: a new standard for clinical mass spectrometry. *Clinical Chemistry,* 62**,** 24-29.

Ma, F, Tremmel, DM, Li, Z, Lietz, CB, Sackett, SD, Odorico, JS & Li, L 2019. In depth quantification of extracellular matrix proteins from human pancreas. *Journal of Proteome Research,* 18**,** 3156-3165.

Määttä, M, Soini, Y, Pääkkö, P, Salo, S, Tryggvason, K & Autio-Harmainen, H 1999. Expression of the laminin γ2 chain in different histological types of lung carcinoma. A study by immunohistochemistry and in situ hybridization. *The Journal of Pathology,* 188**,** 361-368.

Macéa, JR & Fregnani, JH 2006. Anatomy of the Thoracic Wall, Axilla and Breast. *International Journal of Morphology,* 24**,** 691-704.

Mai, J, Sameni, M, Mikkelsen, T & Sloane, BF 2002. Degradation of extracellular matrix protein tenascin-C by cathepsin B: an interaction involved in the progression of gliomas. *Biological Chemistry,* 383**,** 1407-1413.

Majumder, B, Baraneedharan, U, Thiyagarajan, S, Radhakrishnan, P, Narasimhan, H, Dhandapani, M, Brijwani, N, Pinto, DD, Prasath, A & Shanthappa, BU 2015. Predicting clinical response to anticancer drugs using an ex vivo platform that captures tumour heterogeneity. *Nature Communications,* 6.

Makki, J 2015. Diversity of breast carcinoma: histological subtypes and clinical relevance. *Clinical Medicine Insights: Pathology,* 8**,** CPath. S31563.

Malhotra, V & Perry, MC 2003. Classical chemotherapy: mechanisms, toxicities and the therapeutc window. *Cancer Biology & Therapy,* 2**,** 1-3.

Mansour, EG, Gray, R, Shatila, AH, Osborne, C, Tormey, DC, Gilchrist, KW, Cooper, MR & Falkson, G 1989. Efficacy of adjuvant chemotherapy in high-risk node-negative breast cancer. *New England Journal of Medicine,* 320**,** 485-490.

Mansour, EG, Gray, R, Shatila, AH, Tormey, DC, Cooper, MR, Osborne, CK & Falkson, G 1998. Survival advantage of adjuvant chemotherapy in high-risk node-negative breast cancer: ten-year analysis--an intergroup study. *Journal of Clinical Oncology,* 16**,** 3486-3492.

Markiewski, MM, Vadrevu, SK, Sharma, SK, Chintala, NK, Ghouse, S, Cho, J-H, Fairlie, DP, Paterson, Y, Astrinidis, A & Karbowniczek, M 2017. The ribosomal protein S19 suppresses antitumor immune responses via the complement C5a receptor 1. *The Journal of Immunology,* 198**,** 2989-2999.

Martin, A-M & Weber, BL 2000. Genetic and hormonal risk factors in breast cancer. *Journal of the National Cancer Institute,* 92**,** 1126-1135.

Massacesi, C, Di Tomaso, E, Urban, P, Germa, C, Quadt, C, Trandafir, L, Aimone, P, Fretault, N, Dharan, B & Tavorath, R 2016. PI3K inhibitors as new cancer therapeutics: implications for clinical trial design. *OncoTargets and Therapy,* 9**,** 203.

Matei, I, Ghajar, CM & Lyden, D 2011. A TeNaCious foundation for the metastatic niche. *Cancer Cell,* 20**,** 139-141.

Mayrhofer, R, Ng, H, Putti, T & Kuchel, P 2013. Magnetic Resonance in the Detection of Breast Cancers of Different Histological Types. *Magnetic Resonance Insights,* 6**,** 33-49.

Mccafferty, MP, Healy, NA & Kerin, MJ 2009. Breast cancer subtypes and molecular biomarkers. *Diagnostic Histopathology,* 15**,** 485-489.

Mcgrath, EP, Logue, SE, Mnich, K, Deegan, S, Jäger, R, Gorman, AM & Samali, A 2018. The unfolded protein response in breast cancer. *Cancers,* 10**,** 344.

Mcgrogan, BT, Gilmartin, B, Carney, DN & Mccann, A 2008. Taxanes, microtubules and chemoresistant breast cancer. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer,* 1785**,** 96-132.

Mcpherson, K, Steel, C & Dixon, J 2000. Breast cancer—epidemiology, risk factors, and genetics. *British Medical Journal,* 321**,** 624-628.

Mcqueen, P & Krokhin, O 2012. Optimal selection of 2D reversed-phase–reversed-phase HPLC separation techniques in bottom-up proteomics. *Expert review of proteomics,* 9**,** 125-128.

Megger, DA, Bracht, T, Meyer, HE & Sitek, B 2013. Label-free quantification in clinical proteomics. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics,* 1834**,** 1581-1590.

Mehner, C, Hockla, A, Miller, E, Ran, S, Radisky, DC & Radisky, ES 2014. Tumor cell-produced matrix metalloproteinase 9 (MMP-9) drives malignant progression and metastasis of basal-like triple negative breast cancer. *Oncotarget,* 5**,** 2736.

Merdad, A, Karim, S, Schulten, H-J, Dallol, A, Buhmeida, A, Al-Thubaity, F, Gari, MA, Chaudhary, AG, Abuzenadah, AM & Al-Qahtani, MH 2014. Expression of matrix metalloproteinases (MMPs) in primary human breast cancer: MMP-9 as a potential biomarker for cancer invasion and metastasis. *Anticancer Research,* 34**,** 1355-1366.

Midwood, K, Sacre, S, Piccinini, AM, Inglis, J, Trebaul, A, Chan, E, Drexler, S, Sofat, N, Kashiwagi, M & Orend, G 2009. Tenascin-C is an endogenous activator of Toll-like receptor 4 that is essential for maintaining inflammation in arthritic joint disease. *Nature Medicine,* 15**,** 774-780.

Miki, Y, Swensen, J, Shattuck-Eidens, D, Futreal, PA, Harshman, K, Tavtigian, S, Liu, Q, Cochran, C, Bennett, LM & Ding, W 1994. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science,* 266**,** 66-71.

Miller, WR, Bartlett, J, Brodie, AMH, Brueggemeier, RW, Di Salle, E, Lønning, PE, Llombart, A, Maass, N, Maudelonde, T, Sasano, H & Goss, PE 2008. Aromatase Inhibitors: Are There Differences Between Steroidal and Nonsteroidal Aromatase Inhibitors and Do They Matter? *The Oncologist,* 13**,** 829-837.

Minchinton, AI & Tannock, IF 2006. Drug penetration in solid tumours. *Nature Reviews Cancer,* 6**,** 583-592.

Mittendorf, EA, Wu, Y, Scaltriti, M, Meric-Bernstam, F, Hunt, KK, Dawood, S, Esteva, FJ, Buzdar, AU, Chen, H & Eksambi, S 2009. Loss of HER2 amplification following trastuzumab-based neoadjuvant systemic therapy and survival outcomes. *Clinical Cancer Research,* 15**,** 7381-7388.

Mohan, V, Das, A & Sagi, I. Emerging roles of ECM remodeling processes in cancer. Seminars in Cancer Biology, 2020. Elsevier, 192-200.

Molloy, MP, Brzezinski, EE, Hang, J, Mcdowell, MT & Vanbogelen, RA 2003. Overcoming technical variation and biological variation in quantitative proteomics. *Proteomics,* 3**,** 1912-1919.

Moon, HW, Han, HG & Jeon, YJ 2018. Protein quality control in the endoplasmic reticulum and cancer. *International Journal of Molecular Sciences,* 19**,** 3020.

Moreno, L & Pearson, AD 2013. How can attrition rates be reduced in cancer drug discovery? *Expert opinion on drug discovery,* 8**,** 363-368.

Morhason-Bello, IO, Odedina, F, Rebbeck, TR, Harford, J, Dangou, JM, Denny, L & Adewole, IF 2013. Challenges and opportunities in cancer control in Africa: a perspective from the African Organisation for Research and Training in Cancer. *Lancet Oncology,* 14**,** e142-151.

Moscatello, DK, Santra, M, Mann, DM, Mcquillan, DJ, Wong, AJ & Iozzo, RV 1998. Decorin suppresses tumor cell growth by activating the epidermal growth factor receptor. *The Journal of Clinical Investigation,* 101**,** 406-412.

Moudi, M, Go, R, Yien, CYS & Nazre, M 2013. Vinca alkaloids. *International Journal of Preventive Medicine,* 4**,** 1231.

Mutolo, MJ, Morris, KJ, Leir, S-H, Caffrey, TC, Lewandowska, MA, Hollingsworth, MA & Harris, A 2012. Tumor suppression by collagen XV is independent of the restin domain. *Matrix Biology,* 31**,** 285-289.

Naba, A, Clauser, KR, Ding, H, Whittaker, CA, Carr, SA & Hynes, RO 2016. The extracellular matrix: Tools and insights for the "omics" era. *Matrix Biology,* 49**,** 10-24.

Naba, A, Clauser, KR & Hynes, RO 2015. Enrichment of extracellular matrix proteins from tissues and digestion into peptides for mass spectrometry analysis. *Journal of Visualized Experiments: JoVE*.

Naba, A, Pearce, OM, Del Rosario, A, Ma, D, Ding, H, Rajeeve, V, Cutillas, PR, Balkwill, FR & Hynes, RO 2017. Characterization of the extracellular matrix of normal and diseased tissues using proteomics. *Journal of Proteome Research,* 16**,** 3083-3091.

Nahnsen, S, Bielow, C, Reinert, K & Kohlbacher, O 2013. Tools for label-free peptide quantification. *Molecular & Cellular Proteomics,* 12**,** 549-556.

Narla, A & Ebert, BL 2010. Ribosomopathies: human disorders of ribosome dysfunction. *Blood, The Journal of the American Society of Hematology,* 115**,** 3196-3205.

Nesvizhskii, AI 2010. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *Journal of Proteomics,* 73**,** 2092-2123.

Nielsen, M & Karsdal, MA 2016. Type III collagen. *Biochemistry of Collagens, Laminins and Elastin.* Elsevier.

Nikitovic, D, Papoutsidakis, A, Karamanos, NK & Tzanakakis, GN 2014. Lumican affects tumor cell functions, tumor–ECM interactions, angiogenesis and inflammatory response. *Matrix Biology,* 35**,** 206-214.

Nishimura, R & Arima, N 2008. Is triple negative a prognostic factor in breast cancer? *Breast Cancer,* 15**,** 303-308.

Nweke, EE, Naicker, P, Aron, S, Stoychev, S, Devar, J, Tabb, DL, Omoshoro-Jones, J, Smith, M & Candy, G 2020. SWATH-MS based proteomic profiling of Pancreatic Ductal Adenocarcinoma tumours reveals the interplay between the extracellular matrix and related intracellular pathways. *PloS One,* 15**,** e0240453.

O'shaughnessy, JA, Clark, RS, Blum, JL, Mennel, RG, Snyder, D, Ye, Z, Liepa, AM, Melemed, AS & Yardley, DA 2005. Phase II study of pemetrexed in patients pretreated with an anthracycline, a taxane, and capecitabine for advanced breast cancer. *Clinical Breast Cancer,* 6**,** 143-149.

O'rourke, MB, Town, SE, Dalla, PV, Bicknell, F, Koh Belic, N, Violi, JP, Steele, JR & Padula, MP 2019. What is Normalization? The Strategies Employed in Top-Down and Bottom-Up Proteome Analysis Workflows. *Proteomes,* 7**,** 29.

Oakes, SA 2017. Endoplasmic reticulum proteostasis: a key checkpoint in cancer. *American Journal of Physiology-Cell Physiology,* 312**,** C93-C102.

Oakes, SA 2020. Endoplasmic reticulum stress signaling in cancer cells. *The American Journal of Pathology,* 190**,** 934-946.

Oakes, SA & Papa, FR 2015. The role of endoplasmic reticulum stress in human pathology. *Annual Review of Pathology: Mechanisms of Disease,* 10**,** 173-194.

Oberg, AL & Vitek, O 2009. Statistical design of quantitative mass spectrometry-based proteomic experiments. *Journal of Proteome Research,* 8**,** 2144-2156.

Orend, G & Chiquet-Ehrismann, R 2006. Tenascin-C induced signaling in cancer. *Cancer Letters,* 244**,** 143-163.

Oskarsson, T 2013. Extracellular matrix components in breast cancer progression and metastasis. *The Breast,* 22**,** S66-S72.

Oskarsson, T, Acharyya, S, Zhang, XH, Vanharanta, S, Tavazoie, SF, Morris, PG, Downey, RJ, Manova-Todorova, K, Brogi, E & Massagué, J 2011. Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs. *Nature Medicine,* 17**,** 867-874.

Özdemir, BC, Pentcheva-Hoang, T, Carstens, JL, Zheng, X, Wu, C-C, Simpson, TR, Laklai, H, Sugimoto, H, Kahlert, C & Novitskiy, SV 2014. Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival. *Cancer Cell,* 25**,** 719-734.

Paffenbarger Jr, RS, Kampert, JB & Chang, H 1980. Characteristics that predict risk of breast cancer before and after the menopause. *American Journal of Epidemiology,* 112**,** 258-268.

Paget, S 1889. The distribution of secondary growths in cancer of the breast. *The Lancet,* 133**,** 571-573.

Pallero, MA, Elzie, CA, Chen, J, Mosher, DF & Murphy-Ullrich, JE 2008. Thrombospondin 1 binding to calreticulin-LRP1 signals resistance to anoikis. *The FASEB Journal,* 22**,** 3968-3979.

Pandya, K, Meeke, K, Clementz, A, Rogowski, A, Roberts, J, Miele, L, Albain, K & Osipo, C 2011. Targeting both Notch and ErbB-2 signalling pathways is required for prevention of ErbB-2-positive breast tumour recurrence. *British Journal of Cancer,* 105**,** 796-806.

Pankov, R & Yamada, KM 2002. Fibronectin at a glance. *Journal of Cell Science,* 115**,** 3861-3863.

Park, YM, Hwang, SJ, Masuda, K, Choi, K-M, Jeong, M-R, Nam, D-H, Gorospe, M & Kim, HH 2012. Heterogeneous nuclear ribonucleoprotein C1/C2 controls the metastatic potential of glioblastoma by regulating PDCD4. *Molecular and Cellular Biology,* 32**,** 4237-4244.

Patterson, SD 2003. Data analysis—the Achilles heel of proteomics. *Nature Biotechnology,* 21**,** 221-222.

Paulo, JA 2016. Sample preparation for proteomic analysis using a GeLC-MS/MS strategy. *Journal of Biological Methods,* 3**,** e45.

Pelletier, J, Thomas, G & Volarević, S 2018. Ribosome biogenesis in cancer: new players and therapeutic avenues. *Nature Reviews Cancer,* 18**,** 51-63.

Penzo, M, Montanaro, L, Treré, D & Derenzini, M 2019. The ribosome biogenesis—cancer connection. *Cells,* 8**,** 55.

Perez, EA 2007. Safety profiles of tamoxifen and the aromatase inhibitors in adjuvant therapy of hormone-responsive early breast cancer. *Annals of Oncology,* 18**,** viii26-viii35.

Perez, EA, Romond, EH, Suman, VJ, Jeong, J-H, Davidson, NE, Geyer Jr, CE, Martino, S, Mamounas, EP, Kaufman, PA & Wolmark, N 2011. Four-year follow-up of trastuzumab plus adjuvant chemotherapy for operable human epidermal growth factor receptor 2–positive breast cancer: joint analysis of data from NCCTG N9831 and NSABP B-31. *Journal of Clinical Oncology,* 29**,** 3366.

Perou, CM, Sørlie, T, Eisen, MB, Van De Rijn, M, Jeffrey, SS, Rees, CA, Pollack, JR, Ross, DT, Johnsen, H & Akslen, LA 2000. Molecular portraits of human breast tumours. *Nature,* 406**,** 747-752.

Petronzelli, F, Pelliccia, A, Anastasi, AM, D'alessio, V, Albertoni, C, Rosi, A, Leoni, B, De Angelis, C, Paganelli, G & Palombo, G 2005. Improved tumor targeting by combined use of two antitenascin antibodies. *Clinical cancer research,* 11**,** 7137s-7145s.

Pham, TV, Piersma, SR, Oudgenoeg, G & Jimenez, CR 2012. Label-free mass spectrometry-based proteomics for biomarker discovery and validation. *Expert Review of Molecular Diagnostics,* 12**,** 343-359.

Piccart-Gebhart, MJ, Procter, M, Leyland-Jones, B, Goldhirsch, A, Untch, M, Smith, I, Gianni, L, Baselga, J, Bell, R & Jackisch, C 2005. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *New England Journal of Medicine,* 353**,** 1659-1672.

Pickup, MW, Mouw, JK & Weaver, VM 2014. The extracellular matrix modulates the hallmarks of cancer. *EMBO reports***,** e201439246.

Pike, M, Gerkins, V, Casagrande, J, Gray, G, Brown, J & Henderson, B 1979. The hormonal basis of breast cancer. *National Cancer Institute Monograph***,** 187-193.

Pino, LK, Rose, J, O'broin, A, Shah, S & Schilling, B 2020. Emerging mass spectrometry-based proteomics methodologies for novel biomedical applications. *Biochemical Society Transactions,* 48**,** 1953-1966.

Pobre, KFR, Poet, GJ & Hendershot, LM 2019. The endoplasmic reticulum (ER) chaperone BiP is a master regulator of ER functions: Getting by with a little help from ERdj friends. *Journal of Biological Chemistry,* 294**,** 2098-2108.

Polyak, K 2011. Heterogeneity in breast cancer. *The Journal of Clinical Investigation,* 121**,** 3786-3788.

Poratti, M & Marzaro, G 2019. Third-generation CDK inhibitors: A review on the synthesis and binding modes of Palbociclib, Ribociclib and Abemaciclib. *European Journal of Medicinal Chemistry,* 172**,** 143-153.

Pozniak, Y, Balint-Lahat, N, Rudolph, JD, Lindskog, C, Katzir, R, Avivi, C, Pontén, F, Ruppin, E, Barshack, I & Geiger, T 2016. System-wide clinical proteomics of breast cancer reveals global remodeling of tissue homeostasis. *Cell systems,* 2**,** 172-184.

Prat, A, Pineda, E, Adamo, B, Galván, P, Fernández, A, Gaba, L, Díez, M, Viladot, M, Arance, A & Muñoz, M 2015. Clinical implications of the intrinsic molecular subtypes of breast cancer. *Breast,* 24 Suppl 2**,** S26-35.

Provenzano, PP, Inman, DR, Eliceiri, KW & Keely, PJ 2009. Matrix density-induced mechanoregulation of breast cell phenotype, signaling and gene expression through a FAK–ERK linkage. *Oncogene,* 28**,** 4326-4343.

Provenzano, PP, Inman, DR, Eliceiri, KW, Knittel, JG, Yan, L, Rueden, CT, White, JG & Keely, PJ 2008. Collagen density promotes mammary tumor initiation and progression. *BMC Medicine,* 6**,** 11.

Przybylowska, K, Kluczna, A, Zadrozny, M, Krawczyk, T, Kulig, A, Rykala, J, Kolacinska, A, Morawiec, Z, Drzewoski, J & Blasiak, J 2006. Polymorphisms of the promoter regions of matrix metalloproteinases genes MMP-1 and MMP-9 in breast cancer. *Breast Cancer Research and Treatment,* 95**,** 65-72.

Pupa, SM, Ménard, S, Forti, S & Tagliabue, E 2002. New insights into the role of extracellular matrix during tumor onset and progression. *Journal of Cellular Physiology,* 192**,** 259-267.

Rakha, EA, El-Sayed, ME, Powe, DG, Green, AR, Habashy, H, Grainge, MJ, Robertson, JF, Blamey, R, Gee, J, Nicholson, RI, Lee, AH & Ellis, IO 2008. Invasive lobular carcinoma of the breast: response to hormonal therapy and outcomes. *European Journal of Cancer,* 44**,** 73-83.

Rakha, EA, Reis-Filho, JS, Baehner, F, Dabbs, DJ, Decker, T, Eusebi, V, Fox, SB, Ichihara, S, Jacquemier, J, Lakhani, SR, Palacios, J, Richardson, AL, Schnitt, SJ, Schmitt, FC, Tan, PH, Tse, GM, Badve, S & Ellis, IO 2010. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Research,* 12**,** 207.

Ramakrishnan, V 2011. The eukaryotic ribosome. *Science,* 331**,** 681-682.

Ramos, F, Serino, L, Carvalho, C, Lima, R, Urban, C, Cavalli, I & Ribeiro, E 2015. PDIA3 and PDIA6 gene expression as an aggressiveness marker in primary ductal breast cancer. *Genetics and Molecular Research,* 14**,** 6960-6967.

Ratajczak-Wielgomas, K, Grzegrzolka, J, Piotrowska, A, Gomulkiewicz, A, Witkiewicz, W & Dziegiel, P 2016. Periostin expression in cancer-associated fibroblasts of invasive ductal breast carcinoma. *Oncology Reports,* 36**,** 2745-2754.

Ratajczak-Wielgomas, K, Grzegrzolka, J, Piotrowska, A, Matkowski, R, Wojnar, A, Rys, J, Ugorski, M & Dziegiel, P 2017. Expression of periostin in breast cancer cells. *International Journal of Oncology,* 51**,** 1300-1310.

Reddy, RK, Mao, C, Baumeister, P, Austin, RC, Kaufman, RJ & Lee, AS 2003. Endoplasmic reticulum chaperone protein GRP78 protects cells from apoptosis induced by topoisomerase inhibitors:

role of ATP binding site in suppression of caspase-7 activation. *Journal of Biological Chemistry,* 278**,** 20915-20924.

Ricard-Blum, S 2011. The collagen family. *Cold Spring Harbor Perspectives in Biology,* 3**,** a004978.

Ricard-Blum, S, Ruggiero, F & Van Der Rest, M 2005. The collagen superfamily. *Collagen***,** 35-84.

Ricciardelli, C, Brooks, JH, Suwiwat, S, Sakko, AJ, Mayne, K, Raymond, WA, Seshadri, R, Lebaron, RG & Horsfall, DJ 2002. Regulation of stromal versican expression by breast cancer cells and importance to relapse-free survival in patients with node-negative primary breast cancer. *Clinical Cancer Research,* 8**,** 1054-1060.

Rimawi, M, Ferrero, J-M, De La Haba-Rodriguez, J, Poole, C, De Placido, S, Osborne, CK, Hegg, R, Easton, V, Wohlfarth, C & Arpino, G 2018. First-line trastuzumab plus an aromatase inhibitor, with or without pertuzumab, in human epidermal growth factor receptor 2–positive and hormone receptor–positive metastatic or locally advanced breast cancer (PERTAIN): a randomized, open-label phase II trial. *Journal of Clinical Oncology,* 36**,** 2826-2835.

Ringnér, M 2008. What is principal component analysis? *Nature Biotechnology,* 26**,** 303-304.

Rivenbark, AG, O'connor, SM & Coleman, WB 2013. Molecular and cellular heterogeneity in breast cancer: challenges for personalized medicine. *The American Journal of Pathology,* 183**,** 1113-1124.

Robertson, C 2016. The extracellular matrix in breast cancer predicts prognosis through composition, splicing, and crosslinking. *Experimental Cell Research,* 343**,** 73-81.

Robinson, GW 2007. Cooperation of signalling pathways in embryonic mammary gland development. *Nature Reviews Genetics,* 8**,** 963-972.

Roller, C & Maddalo, D 2013. The molecular chaperone GRP78/BiP in the development of chemoresistance: mechanism and possible treatment. *Frontiers in Pharmacology,* 4**,** 10.

Romond, EH, Perez, EA, Bryant, J, Suman, VJ, Geyer Jr, CE, Davidson, NE, Tan-Chiu, E, Martino, S, Paik, S & Kaufman, PA 2005. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *New England Journal of Medicine,* 353**,** 1673-1684.

Roomi, M, Monterrey, J, Kalinovsky, T, Rath, M & Niedzwiecki, A 2009. Patterns of MMP-2 and MMP-9 expression in human cancer cell lines. *Oncology Reports,* 21**,** 1323-1333.

Roshy, S, Sloane, BF & Moin, K 2003. Pericellular cathepsin B and malignant progression. *Cancer and Metastasis Reviews,* 22**,** 271-286.

Röst, HL, Rosenberger, G, Navarro, P, Gillet, L, Miladinović, SM, Schubert, OT, Wolski, W, Collins, BC, Malmström, J & Malmström, L 2014. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nature Biotechnology,* 32**,** 219-223.

Safaeian, M, Hildesheim, A, Gonzalez, P, Yu, K, Porras, C, Li, Q, Rodriguez, AC, Sherman, ME, Schiffman, M & Wacholder, S 2012. Single nucleotide polymorphisms in the PRDX3 and RPS19 and risk of HPV persistence and cervical precancer/cancer. *PloS One,* 7**,** e33619.

Sandhu, R, Chollet-Hinton, L, Kirk, EL, Midkiff, B & Troester, MA 2016. Digital histologic analysis reveals morphometric patterns of age-related involution in breast epithelium and stroma. *Human Pathology,* 48**,** 60-68.

Sarbanes, SL, Le Pen, J & Rice, CM 2018. Friend and foe, HNRNPC takes on immunostimulatory RNA s in breast cancer cells. *The EMBO Journal,* 37**,** e100923.

Saveliev, S, Bratz, M, Zubarev, R, Szapacs, M, Budamgunta, H & Urh, M 2013. Trypsin/Lys-C protease mix for enhanced protein mass spectrometry analysis. *Nature Methods,* 10**,** i-ii.

Schedin, P 2006. Pregnancy-associated breast cancer and metastasis. *Nature Reviews Cancer,* 6**,** 281-291.

Schickli, MA, Berger, MJ, Lustberg, M, Palettas, M & Vargo, CA 2019. Time to treatment failure of palbociclib and letrozole as second-line therapy or beyond in hormone receptor-positive advanced breast cancer. *Journal of Oncology Pharmacy Practice,* 25**,** 1374-1380.

Schilling, B, Gibson, BW & Hunter, CL 2017. Generation of high-quality SWATH® acquisition data for label-free quantitative proteomics studies using TripleTOF® mass spectrometers. *Proteomics.* Springer.

Segovia-Mendoza, M, González-González, ME, Barrera, D, Díaz, L & García-Becerra, R 2015. Efficacy and mechanism of action of the tyrosine kinase inhibitors gefitinib, lapatinib and neratinib in the treatment of HER2-positive breast cancer: preclinical and clinical evidence. *American Journal of Cancer Research,* 5**,** 2531.

Seidler, DG, Goldoni, S, Agnew, C, Cardi, C, Thakur, ML, Owens, RT, Mcquillan, DJ & Iozzo, RV 2006. Decorin protein core inhibits in vivo cancer growth and metabolism by hindering epidermal growth factor receptor function and triggering apoptosis via caspase-3 activation. *Journal of Biological Chemistry,* 281**,** 26408-26418.

Serra, V, Scaltriti, M, Prudkin, L, Eichhorn, PJ, Ibrahim, YH, Chandarlapaty, S, Markman, B, Rodriguez, O, Guzman, M & Rodriguez, S 2011. PI3K inhibition results in enhanced HER signaling and acquired ERK dependency in HER2-overexpressing breast cancer. *Oncogene,* 30**,** 2547-2557.

Seymour, SLH, C. L. High Quality 2017. High quality in-depth protein identification and protein Expression analysis. *ProteinPilotTM software overview*

Shao, R, Bao, S, Bai, X, Blanchette, C, Anderson, RM, Dang, T, Gishizky, ML, Marks, JR & Wang, X-F 2004. Acquired expression of periostin by human breast cancers promotes tumor angiogenesis through up-regulation of vascular endothelial growth factor receptor 2 expression. *Molecular and Cellular Biology,* 24**,** 3992-4003.

Shao, S, Guo, T, Gross, V, Lazarev, A, Koh, CC, Gillessen, S, Joerger, M, Jochum, W & Aebersold, R 2016. Reproducible tissue homogenization and protein extraction for quantitative proteomics using micropestle-assisted pressure-cycling technology. *Journal of Proteome Research,* 15**,** 1821-1829.

Shapiro, JP, Biswas, S, Merchant, AS, Satoskar, A, Taslim, C, Lin, S, Rovin, BH, Sen, CK, Roy, S & Freitas, MA 2012. A quantitative proteomic workflow for characterization of frozen clinical biopsies: laser capture microdissection coupled with label-free mass spectrometry. *Journal of Proteomics,* 77**,** 433-440.

Singh, V, Saunders, C, Wylie, L & Bourke, A 2008. New diagnostic techniques for breast cancer detection. *Future Oncology,* 4**,** 501-513.

Sinha, A & Mann, M 2020. A beginner's guide to mass spectrometry–based proteomics. *The Biochemist,* 42**,** 64-69.

Sinha, B & Politi, P 1990. Anthracyclines. *Cancer Chemotherapy and Biological Response Modifiers,* 11**,** 45-57.

Sinn, HP & Kreipe, H 2013. A Brief Overview of the WHO Classification of Breast Tumors, 4th Edition, Focusing on Issues and Updates from the 3rd Edition. *Breast Care (Basel),* 8**,** 149-154.

Slamon, D, Eiermann, W, Robert, N, Pienkowski, T, Martin, M, Press, M, Mackey, J, Glaspy, J, Chan, A & Pawlicki, M 2011. Adjuvant trastuzumab in HER2-positive breast cancer. *New England Journal of Medicine,* 365**,** 1273-1283.

Smith, I, Procter, M, Gelber, RD, Guillaume, S, Feyereislova, A, Dowsett, M, Goldhirsch, A, Untch, M, Mariani, G & Baselga, J 2007. 2-year follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer: a randomised controlled trial. *The Lancet,* 369**,** 29-36.

Soliman, NA & Yussif, SM 2016. Ki-67 as a prognostic marker according to breast cancer molecular subtype. *Cancer Biology & Medicine,* 13**,** 496.

Sørlie, T, Perou, CM, Tibshirani, R, Aas, T, Geisler, S, Johnsen, H, Hastie, T, Eisen, MB, Van De Rijn, M & Jeffrey, SS 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences,* 98**,** 10869-10874.

Sorlie, T, Tibshirani, R, Parker, J, Hastie, T, Marron, JS, Nobel, A, Deng, S, Johnsen, H, Pesich, R, Geisler, S, Demeter, J, Perou, CM, Lønning, PE, Brown, PO, Børresen-Dale, AL & Botstein, D 2003. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences,* 100**,** 8418-8423.

Srebrow, A & Kornblihtt, AR 2006. The connection between splicing and cancer. *Journal of Cell Science,* 119**,** 2635-2641.

Storey, JD 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 64**,** 479-498.

Stoychev, S, Gerber, I, Jordaan, J, Pty, RB, Naicker, P, Khumalo, F, Mamputha, S & Van Der, C. Universal unbiased sample clean-up pre-mass spectrometry using MagReSyn® HILIC for SPE. ASMS, 2012.

Sulima, SO, Hofman, IJ, De Keersmaecker, K & Dinman, JD 2017. How ribosomes translate cancer. *Cancer Discovery,* 7**,** 1069-1087.

Sun, W, Xing, B, Sun, Y, Du, X, Lu, M, Hao, C, Lu, Z, Mi, W, Wu, S & Wei, H 2007. Proteome analysis of hepatocellular carcinoma by two-dimensional difference gel electrophoresis: novel protein markers in hepatocellular carcinoma tissues. *Molecular & Cellular Proteomics,* 6**,** 1798-1808.

Sun, X, Sandhu, R, Figueroa, JD, Gierach, GL, Sherman, ME & Troester, MA 2014. Benign breast tissue composition in breast cancer patients: association with risk factors, clinical variables, and gene expression. *Cancer Epidemiology and Prevention Biomarkers,* 23**,** 2810-2818.

Sun, Y-S, Zhao, Z, Yang, Z-N, Xu, F, Lu, H-J, Zhu, Z-Y, Shi, W, Jiang, J, Yao, P-P & Zhu, H-P 2017. Risk factors and preventions of breast cancer. *International Journal of Biological Sciences,* 13**,** 1387.

Sun, Z, Velázquez-Quesada, I, Murdamoothoo, D, Ahowesso, C, Yilmaz, A, Spenlé, C, Averous, G, Erne, W, Oberndorfer, F & Oszwald, A 2019. Tenascin-C increases lung metastasis by impacting blood vessel invasions. *Matrix Biology,* 83**,** 26-47.

Sung, H, Ferlay, J, Siegel, RL, Laversanne, M, Soerjomataram, I, Jemal, A & Bray, F 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians,* 71**,** 209-249.

Suwiwat, S, Ricciardelli, C, Tammi, R, Tammi, M, Auvinen, P, Kosma, V-M, Lebaron, RG, Raymond, WA, Tilley, WD & Horsfall, DJ 2004. Expression of extracellular matrix components versican, chondroitin sulfate, tenascin, and hyaluronan, and their association with disease outcome in node-negative breast cancer. *Clinical Cancer Research,* 10**,** 2491-2498.

Tang, P & Tse, GM 2016. Immunohistochemical surrogates for molecular classification of breast carcinoma: a 2015 update. *Archives of Pathology and Laboratory Medicine,* 140**,** 806-814.

Tariq, A, Mateen, RM, Fatima, I & Akhtar, MW 2019. Calreticulin is Differentially Expressed in Invasive Ductal Carcinoma: A Comparative Study. *Current Proteomics,* 16**,** 148-155.

Taroni, P, Quarto, G, Pifferi, A, Abbate, F, Balestreri, N, Menna, S, Cassano, E & Cubeddu, R 2015. Breast tissue composition and its dependence on demographic risk factors for breast cancer: non-invasive assessment by time domain diffuse optical spectroscopy. *PloS One,* 10**,** e0128941.

Teo, K, Kreuzaler, P, Staniszewska, A & Watson, C 2012. Defining the Role of Lysosomal Cathepsins in Breast Cancer. *European Journal of Surgical Oncology,* 38**,** 448.

Thavarajah, R, Mudimbaimannar, VK, Elizabeth, J, Rao, UK & Ranganathan, K 2012. Chemical and physical basics of routine formaldehyde fixation. *Journal of Oral and Maxillofacial Pathology: JOMFP,* 16**,** 400.

Theocharis, AD, Manou, D & Karamanos, NK 2019. The extracellular matrix as a multitasking player in disease. *The FEBS Journal,* 286**,** 2830-2869.

Thoma, CR, Zimmermann, M, Agarkova, I, Kelm, JM & Krek, W 2014. 3D cell culture systems modeling tumor growth determinants in cancer target discovery. *Advanced Drug Delivery Reviews,* 69**,** 29-41.

Thomas, R, Rowell, R, Crichton, S & Cain, H 2018. An observational study investigating failure of primary endocrine therapy for operable breast cancer in the elderly. *Breast Cancer Research and Treatment,* 167**,** 73-80.

Thompson, A, Schäfer, J, Kuhn, K, Kienle, S, Schwarz, J, Schmidt, G, Neumann, T & Hamon, C 2003. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical Chemistry,* 75**,** 1895-1904.

Timpl, R, Rohde, H, Robey, PG, Rennard, SI, Foidart, J-M & Martin, GR 1979. Laminin–A glycoprotein from basement membranes. *Journal of Biological Chemistry,* 254**,** 9933-9937.

Trédan, O, Galmarini, CM, Patel, K & Tannock, IF 2007. Drug resistance and the solid tumor microenvironment. *Journal of the National Cancer Institute,* 99**,** 1441-1454.

Troup, S, Njue, C, Kliewer, EV, Parisien, M, Roskelley, C, Chakravarti, S, Roughley, PJ, Murphy, LC & Watson, PH 2003. Reduced expression of the small leucine-rich proteoglycans, lumican, and decorin is associated with poor outcome in node-negative invasive breast cancer. *Clinical Cancer Research,* 9**,** 207-214.

Tsai, M-J, Chang, W-A, Huang, M-S & Kuo, P-L 2014. Tumor microenvironment: A new treatment target for cancer. *ISRN Biochemistry,* 2014.

Tsang, J & Tse, GM 2020. Molecular classification of breast cancer. *Advances in Anatomic Pathology,* 27**,** 27-35.

Turashvili, G & Brogi, E 2017. Tumor heterogeneity in breast cancer. *Frontiers in Medicine,* 4**,** 227.

Tzu, J & Marinkovich, MP 2008. Bridging structure with function: structural, regulatory, and developmental role of laminins. *The International Journal of Biochemistry & Cell Biology,* 40**,** 199-214.

Uchida, Y, Sasaki, H & Terasaki, T 2020. Establishment and validation of highly accurate formalin-fixed paraffin-embedded quantitative proteomics by heat-compatible pressure cycling technology using phase-transfer surfactant and SWATH-MS. *Scientific Reports,* 10**,** 1-17.

Unger, FT, Witte, I & David, KA 2015. Prediction of individual response to anticancer therapy: historical and future perspectives. *Cellular and Molecular Life Sciences,* 72**,** 729-757.

Urakawa, H, Nishida, Y, Wasa, J, Arai, E, Zhuo, L, Kimata, K, Kozawa, E, Futamura, N & Ishiguro, N 2012. Inhibition of hyaluronan synthesis in breast cancer cells by 4-methylumbelliferone suppresses tumorigenicity in vitro and metastatic lesions of bone in vivo. *International Journal of Cancer,* 130**,** 454-466.

Urruticoechea, A, Smith, IE & Dowsett, M 2005. Proliferation marker Ki-67 in early breast cancer. *Journal of Clinical Oncology,* 23**,** 7212-7220.

Välikangas, T, Suomi, T & Elo, LL 2018. A systematic evaluation of normalization methods in quantitative label-free proteomics. *Briefings in Bioinformatics,* 19**,** 1-11.

Van Huizen, NA, Ijzermans, JN, Burgers, PC & Luider, TM 2020. Collagen analysis with mass spectrometry. *Mass Spectrometry Reviews,* 39**,** 309-335.

Varner, JA & Cheresh, DA 1996. Integrins and cancer. *Current Opinion in Cell Biology,* 8**,** 724-730.

Veronesi, U, Paganelli, G, Viale, G, Luini, A, Zurrida, S, Galimberti, V, Intra, M, Veronesi, P, Maisonneuve, P, Gatti, G, Mazzarol, G, De Cicco, C, Manfredi, G & Fernández, JR 2006. Sentinel-lymph-node biopsy as a staging procedure in breast cancer: update of a randomised controlled study. *Lancet Oncology,* 7**,** 983-990.

Vizovišek, M, Fonović, M & Turk, B 2019. Cysteine cathepsins in extracellular matrix remodeling: Extracellular matrix degradation and beyond. *Matrix Biology,* 75**,** 141-159.

Wahl, MC, Will, CL & Lührmann, R 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell,* 136**,** 701-718.

Walker, C, Mojares, E & Del Río Hernández, A 2018. Role of extracellular matrix in development and cancer progression. *International Journal of Molecular Sciences,* 19**,** 3028.

Walker, RA 2001. The complexities of breast cancer desmoplasia. *Breast Cancer Research,* 3**,** 143.

Walko, CM & Lindley, C 2005. Capecitabine: a review. *Clinical Therapeutics,* 27**,** 23-44.

Wang, E & Aifantis, I 2020. RNA splicing and cancer. *Trends in Cancer,* 6**,** 631-644.

Wang, JP & Hielscher, A 2017. Fibronectin: how its aberrant expression in tumors may improve therapeutic targeting. *Journal of Cancer,* 8**,** 674.

Wang, M, Hu, Y & Stearns, ME 2009. RPS2: a novel therapeutic target in prostate cancer. *Journal of Experimental and Clinical Cancer Research,* 28**,** 1-12.

Wang, M & Kaufman, RJ 2014. The impact of the endoplasmic reticulum protein-folding environment on cancer development. *Nature Reviews Cancer,* 14**,** 581-597.

Wang, TN, Qian, X-H, Granick, MS, Solomon, MP, Rothman, VL, Berger, DH & Tuszynski, GP 1996. Thrombospondin-1 (TSP-1) promotes the invasive properties of human breast cancer. *Journal of Surgical Research,* 63**,** 39-43.

Wang, X, He, Y, Ye, Y, Zhao, X, Deng, S, He, G, Zhu, H, Xu, N & Liang, S 2018. SILAC–based quantitative MS approach for real-time recording protein-mediated cell-cell interactions. *Scientific Reports,* 8**,** 1-9.

Wang, X, Liu, J, Wang, Z, Huang, Y, Liu, W, Zhu, X, Cai, Y, Fang, X, Lin, S & Yuan, L 2013. Periostin contributes to the acquisition of multipotent stem cell-like properties in human mammary epithelial cells and breast cancer cells. *PloS One,* 8**,** e72962.

Wasinger, VC, Cordwell, SJ, Cerpa-Poljak, A, Yan, JX, Gooley, AA, Wilkins, MR, Duncan, MW, Harris, R, Williams, KL & Humphery-Smith, I 1995. Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. *Electrophoresis,* 16**,** 1090-1094.

Watson, CJ & Khaled, WT 2008. Mammary development in the embryo and adult: a journey of morphogenesis and commitment. *Development,* 135**,** 995-1003.

Weidle, UH, Maisel, D, Klostermann, S, Weiss, EH & Schmitt, M 2011. Differential splicing generates new transmembrane receptor and extracellular matrix-related targets for antibody-based therapy of cancer. *Cancer Genomics & Proteomics,* 8**,** 211-226.

Wellings, SR 1980. A Hypothesis of the Origin of Human Breast Cancer from the Terminal Ductal Lobular Unit. *Pathology - Research and Practice,* 166**,** 515-535.

Wellings, SR & Jensen, HM 1973. On the origin and progression of ductal carcinoma in the human breast. *Journal of the National Cancer Institute,* 50**,** 1111-1118.

White, RJ 2005. RNA polymerases I and III, growth control and cancer. *Nature Reviews Molecular Cell Biology,* 6**,** 69-78.

Wiese, S, Reidegeld, KA, Meyer, HE & Warscheid, B 2007. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics,* 7**,** 340-350.

Wilkins, MR, Sanchez, J-C, Gooley, AA, Appel, RD, Humphery-Smith, I, Hochstrasser, DF & Williams, KL 1996. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology and Genetic Engineering Reviews,* 13**,** 19-50.

Will, CL & Lührmann, R 2011. Spliceosome structure and function. *Cold Spring Harbor Perspectives in Biology,* 3**,** a003707.

Wiseman, BS & Werb, Z 2002. Stromal effects on mammary gland development and breast cancer. *Science,* 296**,** 1046-1049.

Wishart, AL, Conner, SJ, Guarin, JR, Fatherree, JP, Peng, Y, Mcginn, RA, Crews, R, Naber, SP, Hunter, M & Greenberg, AS 2020. Decellularized extracellular matrix scaffolds identify full-length collagen VI as a driver of breast cancer cell invasion in obesity and metastasis. *Science Advances,* 6**,** eabc3175.

Wolff, A, Hammond, M, Hicks, D, Dowsett, M, Mcshane, L, Allison, K, Allred, D, Bartlett, J, Bilous, M & Fitzgibbons, P 2013. American Society of Clinical Oncology; College of American Pathologists. 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. *Journal of Clinical Oncology,* 31**,** 3997-4013.

Wooster, R, Bignell, G, Lancaster, J, Swift, S, Seal, S, Mangion, J, Collins, N, Gregory, S, Gumbs, C & Micklem, G 1995. Identification of the breast cancer susceptibility gene BRCA2. *Nature,* 378**,** 789-792.

Wu, X, Chen, G, Qiu, J, Lu, J, Zhu, W, Chen, J, Zhuo, S & Yan, J 2016. Visualization of basement membranes in normal breast and breast cancer tissues using multiphoton microscopy. *Oncology Letters,* 11**,** 3785-3789.

Wu, ZS, Wu, Q, Yang, JH, Wang, HQ, Ding, XD, Yang, F & Xu, XC 2008. Prognostic significance of MMP-9 and TIMP-1 serum and tissue expression in breast cancer. *International Journal of Cancer,* 122**,** 2050-2056.

Wullkopf, L, West, A-KV, Leijnse, N, Cox, TR, Madsen, CD, Oddershede, LB & Erler, JT 2018. Cancer cells' ability to mechanically adjust to extracellular matrix stiffness correlates with their invasive potential. *Molecular Biology of the Cell,* 29**,** 2378-2385.

Xiong, G, Deng, L, Zhu, J, Rychahou, PG & Xu, R 2014. Prolyl-4-hydroxylase α subunit 2 promotes breast cancer progression and metastasis by regulating collagen deposition. *BMC Cancer,* 14**,** 1-12.

Xu, D, Xu, H, Ren, Y, Liu, C, Wang, X, Zhang, H & Lu, P 2012. Cancer stem cell-related gene periostin: a novel prognostic marker for breast cancer. *PloS One,* 7**,** e46670.

Xu, S, Sankar, S & Neamati, N 2014a. Protein disulfide isomerase: a promising target for cancer therapy. *Drug Discovery Today,* 19**,** 222-240.

Xu, T, Zhang, R, Dong, M, Zhang, Z, Li, H, Zhan, C & Li, X 2019. Osteoglycin (OGN) Inhibits Cell Proliferation and Invasiveness in Breast Cancer via PI3K/Akt/mTOR Signaling Pathway. *OncoTargets and Therapy,* 12**,** 10639.

Xu, X, Xiong, X & Sun, Y 2016. The role of ribosomal proteins in the regulation of cell proliferation, tumorigenesis, and genomic integrity. *Science China Life Sciences,* 59**,** 656-672.

Xu, Y, Gao, XD, Lee, J-H, Huang, H, Tan, H, Ahn, J, Reinke, LM, Peter, ME, Feng, Y & Gius, D 2014b. Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. *Genes & Development,* 28**,** 1191-1203.

Yamashita, M & Fenn, JB 1984. Electrospray ion source. Another variation on the free-jet theme. *The Journal of Physical Chemistry,* 88**,** 4451-4459.

Yang, F, Shen, Y, Camp, DG & Smith, RD 2012. High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert review of proteomics,* 9**,** 129-134.

Yang, J, Song, H, Chen, L, Cao, K, Zhang, Y, Li, Y & Hao, X 2019. Integrated analysis of microfibrillar-associated proteins reveals MFAP4 as a novel biomarker in human cancers. *Epigenomics,* 11**,** 5-21.

Yang, W-H, Ding, M-J, Cui, G-Z, Yang, M & Dai, D-L 2018. Heterogeneous nuclear ribonucleoprotein M promotes the progression of breast cancer by regulating the axin/β-catenin signaling pathway. *Biomedicine and Pharmacotherapy,* 105**,** 848-855.

Yang, Z, Ni, W, Cui, C, Fang, L & Xuan, Y 2017. Tenascin C is a prognostic determinant and potential cancer-associated fibroblasts marker for breast ductal carcinoma. *Experimental and Molecular Pathology,* 102**,** 262-267.

Yap, TA, Sandhu, SK, Carden, CP & De Bono, JS 2011. Poly (ADP-Ribose) polymerase (PARP) inhibitors: Exploiting a synthetic lethal strategy in the clinic. *CA: A Cancer Journal for Clinicians,* 61**,** 31-49.

Yasuo Kunugiza, KI, Koibuchi, N & Fumihiro, S 2011. Role of periostin in cancer progression and metastasis: inhibition of breast cancer progression and metastasis by anti-periostin antibody in a murine model. *International Journal of Molecular Medicine,* 28**,** 181-186.

Yates, JR 1998. Mass spectrometry and the age of the proteome. *Journal of Mass Spectrometry,* 33**,** 1-19.

Yousef, EM, Tahir, MR, St-Pierre, Y & Gaboury, LA 2014. MMP-9 expression varies according to molecular subtypes of breast cancer. *BMC Cancer,* 14**,** 1-12.

Yurchenco, PD 2011. Basement membranes: cell scaffoldings and signaling platforms. *Cold Spring Harbor Perspectives in Biology,* 3**,** a004911.

Yurchenco, PD, Amenta, PS & Patton, BL 2004. Basement membrane assembly, stability and activities observed through a developmental lens. *Matrix Biology,* 22**,** 521-538.

Zahir, N, Lakins, JN, Russell, A, Ming, W, Chatterjee, C, Rozenberg, GI, Marinkovich, MP & Weaver, VM 2003. Autocrine laminin-5 ligates α6β4 integrin and activates RAC and NFκB to mediate anchorage-independent survival of mammary tumors. *The Journal of Cell Biology,* 163**,** 1397-1407.

Zamanian, M, Hamadneh, LaQ, Veerakumarasivam, A, Rahman, SA, Shohaimi, S & Rosli, R 2016. Calreticulin mediates an invasive breast cancer phenotype through the transcriptional dysregulation of p53 and MAPK pathways. *Cancer Cell International,* 16**,** 1-13.

Zecha, J, Satpathy, S, Kanashova, T, Avanessian, SC, Kane, MH, Clauser, KR, Mertins, P, Carr, SA & Kuster, B 2019. TMT labeling for the masses: a robust and cost-efficient, in-solution labeling approach. *Molecular & Cellular Proteomics,* 18**,** 1468-1478.

Zelek, L, Barthier, S, Riofrio, M, Fizazi, K, Rixe, O, Delord, J, Le Cesne, A & Spielmann, M 2001. Weekly vinorelbine is an effective palliative regimen after failure with anthracyclines and taxanes in metastatic breast carcinoma. *Cancer,* 92**,** 2267-2272.

Zhang, G & Neubert, TA 2009. Use of stable isotope labeling by amino acids in cell culture (SILAC) for phosphotyrosine protein identification and quantitation. *Phospho-Proteomics***,** 79-92.

Zhang, Y, Zhang, G, Li, J, Tao, Q & Tang, W 2010. The expression analysis of periostin in human breast cancer. *Journal of Surgical Research,* 160**,** 102-106.

Zhao, L, Cong, X, Zhai, L, Hu, H, Xu, J-Y, Zhao, W, Zhu, M, Tan, M & Ye, B-C 2020. Comparative evaluation of label-free quantification strategies. *Journal of Proteomics,* 215**,** 103669.

Zhao, S, Xu, L, Liu, W, Lv, C, Zhang, K, Gao, H, Wang, J & Ma, R 2015. Comparison of the expression of prognostic biomarkers between primary tumor and axillary lymph node metastases in breast cancer. *International Journal of Clinical and Experimental Pathology,* 8**,** 5744.

Zi, J, Zhang, S, Zhou, R, Zhou, B, Xu, S, Hou, G, Tan, F, Wen, B, Wang, Q & Lin, L 2014. Expansion of the ion library for mining SWATH-MS data through fractionation proteomics. *Analytical Chemistry,* 86**,** 7242-7246.

Zong, J, Guo, C, Liu, S, Sun, M-Z & Tang, J 2012. Proteomic research progress in lymphatic metastases of cancers. *Clinical and Translational Oncology,* 14**,** 21-30.

Appendix I: Ethical approvals

The Research Ethics Committee, Faculty Health Sciences, University of Pretoria complies with ICH-GCP guidelines and has US Federal wide Assurance.
- FWA 00002567, Approved dd 22 May 2002 and Expires 03/20/2022.
- IRB 0000 2235 IORG0001762 Approved dd 22/04/2014 and Expires 03/14/2020.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Health Sciences Research Ethics Committee

30/03/2017

Approval Certificate
New Application

Ethics Reference No.: 86/2017

Title: Proteomic characterisation of primary breast tumour extracellular matrix

Dear Miss Chanelle Pillay

The **New Application** as supported by documents specified in your cover letter dated 24/02/2017 for your research received on the 24/02/2017, was approved by the Faculty of Health Sciences Research Ethics Committee on its quorate meeting of 29/03/2017.

Please note the following about your ethics approval:
- Ethics Approval is valid for 2 years
- Please remember to use your protocol number (**86/2017**) on any documents or correspondence with the Research Ethics Committee regarding your research.
- Please note that the Research Ethics Committee may ask further questions, seek additional information, require further modification, or monitor the conduct of your research.

Ethics approval is subject to the following:
- The ethics approval is conditional on the receipt of **6 monthly written Progress Reports**, and
- The ethics approval is conditional on the research being conducted as stipulated by the details of all documents submitted to the Committee. In the event that a further need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.

We wish you the best with your research.

Yours sincerely

Dr R Sommers; MBChB; MMed (Int); MPharMed,PhD
Deputy Chairperson of the Faculty of Health Sciences Research Ethics Committee, University of Pretoria

The Faculty of Health Sciences Research Ethics Committee complies with the SA National Act 61 of 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 and 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes, Second Edition 2015 (Department of Health).

012 356 3084 deepeka.behari@up.ac.za / fhsethics@up.ac.za http://www.up.ac.za/healthethics
Private Bag X323, Arcadia, 0007 - Tswelopele Building, Level 4, Room 60, Gezina, Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Health Sciences

The Research Ethics Committee, Faculty Health Sciences, University of Pretoria complies with ICH-GCP guidelines and has US Federal wide Assurance.

- FWA 00002567, Approved dd 22 May 2002 and Expires 03/20/2022.
- IRB 0000 2235 IORG0001762 Approved dd 22/04/2014 and Expires 03/14/2020.

6 November 2019

Approval Certificate
Annual Renewal

Ethics Reference No.: 86/2017
Title: Proteomic characterisation of primary breast tumour extracellular matrix

Dear Miss CM Pillay

The **Annual Renewal** as supported by documents received between 2019-10-10 and 2019-11-06 for your research, was approved by the Faculty of Health Sciences Research Ethics Committee on its quorate meeting of 2019-11-06.

Please note the following about your ethics approval:

- Renewal of ethics approval is valid for 1 year, subsequent annual renewal will become due on 2020-11-06.
- Please remember to use your protocol number (86/2017) on any documents or correspondence with the Research Ethics Committee regarding your research.
- Please note that the Research Ethics Committee may ask further questions, seek additional information, require further modification, monitor the conduct of your research, or suspend or withdraw ethics approval.

Ethics approval is subject to the following:

- The ethics approval is conditional on the research being conducted as stipulated by the details of all documents submitted to the Committee. In the event that a further need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.

We wish you the best with your research.

Yours sincerely

Dr R Sommers
MBChB MMed (Int) MPharmMed PhD
Deputy Chairperson of the Faculty of Health Sciences Research Ethics Committee, University of Pretoria

The Faculty of Health Sciences Research Ethics Committee complies with the SA National Act 61 of 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 and 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes, Second Edition 2015 (Department of Health)

Research Ethics Committee
Room 4-60, Level 4, Tswelopele Building
University of Pretoria, Private Bag x323
Gezina 0031, South Africa
Tel: +27 (0)12 356 3084
Email: deepeka.behari@up.ac.za
www.up.ac.za

Fakulteit Gesondheidswetenskappe
Lefapha la Disaense tša Maphelo

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Health Sciences

Institution: The Research Ethics Committee, Faculty Health Sciences, University of Pretoria complies with ICH-GCP guidelines and has US Federal wide Assurance.

- FWA 00002567, Approved dd 22 May 2002 and Expires 03/20/2022.
- IORG #: IORG0001762 OMB No. 0990-0279 Approved for use through February 28, 2022 and Expires: 03/04/2023.

6 November 2020

Approval Certificate
Annual Renewal

Ethics Reference No.: 86/2017
Title: Proteomic characterisation of primary breast tumour extracellular matrix

Dear Miss CM Pillay

The **Annual Renewal** as supported by documents received between 2020-10-15 and 2020-11-04 for your research, was approved by the Faculty of Health Sciences Research Ethics Committee on 2020-11-04 as resolved by its quorate meeting.

Please note the following about your ethics approval:

- Renewal of ethics approval is valid for 1 year, subsequent annual renewal will become due on 2021-11-06.
- Please remember to use your protocol number (86/2017) on any documents or correspondence with the Research Ethics Committee regarding your research.
- Please note that the Research Ethics Committee may ask further questions, seek additional information, require further modification, monitor the conduct of your research, or suspend or withdraw ethics approval.

Ethics approval is subject to the following:

- The ethics approval is conditional on the research being conducted as stipulated by the details of all documents submitted to the Committee. In the event that a further need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.

We wish you the best with your research.

Yours sincerely

Dr R Sommers
MBChB MMed (Int) MPharmMed PhD
Deputy Chairperson of the Faculty of Health Sciences Research Ethics Committee, University of Pretoria

The Faculty of Health Sciences Research Ethics Committee complies with the SA National Act 61 of 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 and 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes, Second Edition 2015 (Department of Health)

Research Ethics Committee
Room 4-60, Level 4, Tswelopele Building
University of Pretoria, Private Bag x323
Gezina 0031, South Africa
Tel +27 (0)12 356 3084
Email: deepika.behari@up.ac.za
www.up.ac.za

Fakulteit Gesondheidswetenskappe
Lefapha la Disaense Sa Maphelo

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Health Sciences

Institution: The Research Ethics Committee, Faculty Health Sciences, University of Pretoria complies with ICH-GCP guidelines and has US Federal wide Assurance.
- FWA 00002567, Approved dd 22 May 2002 and Expires 03/20/2022.
- IORG #: IORG0001762 OMB No. 0990-0279 Approved for use through February 28, 2022 and Expires: 03/04/2023.

Faculty of Health Sciences **Research Ethics Committee**

11 November 2021

Approval Certificate
Annual Renewal

Dear Miss CM Pilay

Ethics Reference No.: 86/2017
Title: Proteomic characterisation of primary breast tumour extracellular matrix

The **Annual Renewal** as supported by documents received between 2021-11-03 and 2021-11-10 for your research, was approved by the Faculty of Health Sciences Research Ethics Committee on 2021-11-10 as resolved by its quorate meeting.

Please note the following about your ethics approval:

- Renewal of ethics approval is valid for 1 year, subsequent annual renewal will become due on 2022-11-11.
- Please remember to use your protocol number (86/2017) on any documents or correspondence with the Research Ethics Committee regarding your research.
- Please note that the Research Ethics Committee may ask further questions, seek additional information, require further modification, monitor the conduct of your research, or suspend or withdraw ethics approval.

Ethics approval is subject to the following:

- The ethics approval is conditional on the research being conducted as stipulated by the details of all documents submitted to the Committee. In the event that a further need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.

We wish you the best with your research.

Yours sincerely

On behalf of the FHS REC, Dr R Sommers
MBChB, MMed (Int), MPharmMed, PhD
***Deputy Chairperson** of the Faculty of Health Sciences Research Ethics Committee, University of Pretoria*

The Faculty of Health Sciences Research Ethics Committee complies with the SA National Act 61 of 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 and 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes, Second Edition 2015 (Department of Health)

Research Ethics Committee
Room 4-60, Level 4, Tswelopele Building
University of Pretoria, Private Bag x323
Gezina 0031, South Africa
Tel: +27 (0)12 356 3084
Email: deepika.behari@up.ac.za
www.up.ac.za

Fakulteit Gesondheidswetenskappe
Lefapha la Disaense tša Maphelo

Appendix II: Permission letter from Steve Biko

Permission to access Records / Files / Data base at the Steve Biko Academic Hospital (SBAH)

To: Chief Executive Officer/Information Officer
Steve Biko Academic Hospital

From: The Investigator
University of Pretoria

Dr E Kenoshi

Dr XX Stander

Re: Permission to do research at SBAH Hospital

Dr BA Stander, Prof T Mokoena and I are researchers working at the Cell culture Unit, Department of Surgery at SBAH Hospital and the University of Pretoria respectively. I am requesting permission on behalf of all of us to conduct a study on the SBAH Hospital grounds that involves access to patient records.

The request is lodged with you in terms of the requirements of the Promotion of Access to Information Act. No. 2 of 2000.

The title of the study is: Ex vivo tumour heterogenetic model to predict clinical response to anticancer drugs

The researchers request access to the following information:

Access to the clinical files, record book and the data base.

We intend to publish the findings of the study in a professional journal and/ or at professional meeting like symposia, congresses, or other meetings of such a nature.

We intend to protect the personal identity of the patients by assigning each patient a random code number.

We undertake not to proceed with the study until we have received approval from the Faculty of Health Sciences Research Ethics Committee, University of Pretoria.

Yours sincerely

Signature of the Principle Investigator

24/02/2016

Permission to do the research study at this hospital and to access the information as requested, is hereby approved.

Chief Executive Officer

Steve Biko Academic Hospital
Dr E Kenoshi

Signature of the CEO

10/3/2016

DR ME KENOSHI
CHIEF EXECUTIVE OFFICER
STEVE BIKO
ACADEMIC HOSPITAL

Hospital Official Stamp

Appendix III: CSIR permission letter

CSIR Research Ethics Committee
PO Box 395 Pretoria 0001 South Africa
Tel: +27 12 841 4060
Fax: +27 12 841 2476
Email:R&DEthics@csir.co.za

Permission certificate

14 June 2017

Dear: Dr Stoyan Stoychev

Title: Proteomic characterisation of primary breast tumour extracellular matrix (Ref: 213/2017).

Thank you for submitting your application to the CSIR Research Ethics Committee (REC). The submission was reviewed by the REC on the 08th of June 2017. The CSIR REC notes the University of Pretoria Faculty of Health Sciences Research Ethics Committee clearance certificate. Furthermore, that ethics related work will be performed at the University of Pretoria and at the Steve Biko Hospital as per approval, and that the CSIR's responsibility is the analysis of the outputs (samples). The CSIR REC therefore grants permission for the study to proceed with the following provisos:

- The PI is requested to submit a signed research collaboration agreement between the University of Pretoria and the CSIR;
- The PI is requested to submit a 2017 proof of registration for Ms Chanelle Pillay.

Kindly note that you are required to submit bi-annual progress reports to the CSIR REC and a final report on completion of the research in which you indicate (i) that the research has been completed; (ii) if any new or unexpected ethical issues emerged during the course of the study; and if so, (iii) how these ethical issues were addressed.

We wish you all of the best with your research project.

Kind regards

Dr Shenuka Singh
(CSIR REC Chair)

Appendix IV: Patient consent form

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Health Sciences
Research Ethics Committee

[Informed Consent Form for Participant and Patient for Clinical Non-intervention Study]

Title: Proteomic characterisation of primary breast tumour extracellular matrix

This Informed Consent Form is for participant and patient who attend the Steve Biko Academic Hospital/University of Pretoria, and who we are inviting to participate in research at the Department of Surgery/Pharmacology/Physiology.

Principal Investigator: Ms Chanelle Pillay, Department of Pharmacology, University of Pretoria
Co-Principal investigator: Prof AD Cromarty, Department of Pharmacology, University of Pretoria
Co-Principal investigator: Dr BA Stander, Department of Physiology, University of Pretoria
Co-Principal investigator: Dr SH Stoychev, Biosciences, CSIR

This informed Consent has two parts:
Part I: Information Sheet (to share information about the research with you)
Part II: Certificate of Consent (for signatures if you agree to take part in the study)

Part I: Information Sheet

Introduction

You are invited to volunteer for a research study. This information leaflet is to help you to decide if you would like to participate. Before you agree to take part in this study you should fully understand what is involved. If you have any questions, which are not fully explained in this leaflet, do not hesitate to ask the investigator. You should not agree to take part unless you are completely happy about all the procedures involved. In the best interests of your health, it is strongly recommended that you discuss with or inform your personal doctor of your possible participation in this study, wherever possible.

Purpose of the research

You are invited to take part in a research study. The aim of this study is to evaluate differences between cancer patients and non-cancer population in the analytical laboratory settings. We wish to take the surgical waste material (tumour samples) and keep it alive with the patient's serum (from blood) to test anticancer drugs in the laboratory. By doing so we would like to establish a model to help cancer patients moving towards personalised medicine.

Procedures to be followed

This study mainly takes place in the laboratory therefore patients will not be taking drugs. Primary tumours (surgical waste) from patients, who will be undergoing a mastectomy, from the Steve Biko Academic Hospital will be used in the study. In addition, this study involves the answering of a few

questions with regards to your illness as well as allowing us to access your hospital file for details of your condition, management and complications, if there are any.

Participation selection

We are inviting all adults (>18 years) who have been diagnosed with breast cancer and who attend the Steve Biko Academic Hospital to participate in this research.

Risk and discomfort involved

There are no risks involved in taking part in this study and no known side effects. You can refuse to take part in this study and this will not affect the treatment of your cancer in anyway. You will receive the prescribed treatment for your condition.

Side effects
There are no side effects from taking the surgical waste material after your mastectomy.

Confidentiality
The information that we collect from this research project will be kept confidential. It will not be shared with or given to anyone except your doctors, nurses and research staff.

Right to refuse or withdraw from the study

You do not have to take part in this research if you do not wish to do so and refusing to participate will not affect your treatment at this clinic in any way. You may stop participating in the research at any time that you wish without losing any of your rights as a patient here. Your treatment at the Steve Biko Academic Hospital will not be affected in any way.

Contact information

If you have any questions concerning this study, you should contact:
Ms Chanelle Pillay – Tel: 012 319 2558 or email: chanelle.pillay@up.ac.za
Prof Duncan Cromarty – Tel: 012 319 2622 or email: duncan.cromarty@up.ac.za
Dr BA Stander – Tel: 012 319 2241 or email: andre.stander@up.ac.za

This proposal has been reviewed and approved by the Research Ethics Committee, Faculty of Health Sciences, University of Pretoria, which is a committee whose task it is to make sure that research participants are protected from harm. If you wish to find out more about the committee, contact Ms Behari – Tel 012 354 1677

Part II: Certificate of Consent

I have read or had the following information read to me in a language that I understand before signing this consent form. The content and meaning of this information have been explained to me. I have been given an opportunity to ask questions which have been answered satisfactorily. I understand that if I do not participate it will not alter my treatment in any way. I hereby volunteer to take part in this study.

I have received a signed copy of this informed consent agreement.

Name of Patient	Signature	Date
Name of person obtaining consent	Signature	Date
Name of witness	Signature	Date

VERBAL PATIENT INFORMED CONSENT (applicable when patients cannot read or write)

I, the undersigned, ……, have read and have explained fully to the patient, named …… and/or/ his/her relative, the patient information leaflet, which has indicated the nature and purpose of the study in which I have asked the patient to participate in. The explanation I have given has mentioned both the possible risks and benefits of the study. The patient indicated that he/she understands that he/she will be free to withdraw from the study at any time for any reason and without jeopardizing his/her treatment.

I hereby certify that the patient has agreed to participate in this study.

Name of patient	Signature of patient	Date
Name of investigator	Signature of investigator	Date
Name of witness	Signature of witness	Date

(Witness – sign that he/she has witnessed the process of informed consent)

Appendix V: Patient clinical information

Pt no.	Age	Clinical diagnosis as stated on pathology report	Tumour Grade	ER (-/+)	PR (-/+)	Her-2 (-/+)	Ki-67 (%)	Lymph node involvement	Immunophenotype
P05	54	Poorly differentiated infiltrating ductal carcinoma	2	+	+	-	30	Yes	Luminal B
P09	45	infiltrating ductal carcinoma	3	-	-	-	80	No	Triple negative
P11	38	Moderately differentiated infiltrating ductal carcinoma	?	+	+	+	40	No	Luminal B
P13	73	Moderately differentiated infiltrating ductal carcinoma	2	+	+	-	40	Yes	Luminal B
P14	75	Moderately differentiated infiltrating ductal carcinoma	2	+	+	-	5	No	Luminal A
P18	88	Infiltrating ductal carcinoma	?	+	+	-	2	No	Luminal A
P19	38	Infiltrating ductal carcinoma	2	-	-	-	5	No	Triple negative
P21	66	Infiltrating ductal carcinoma	3	-	-	+	10	Yes	Her-2 enriched
P23	64	Moderately differentiated infiltrating ductal carcinoma	?	+	+	-	60	No	Luminal B
P24	47	Infiltrating ductal carcinoma	3	-	-	-	1	Yes	Triple negative
P25	47	Infiltrating tubular/micropapillary carcinoma	?	+	+	+	50	Yes	Luminal B
P26	34	Moderately differentiated infiltrating ductal carcinoma	?	?	?	?	?	Yes	Not specified
P29	54	Infiltrating ductal carcinoma	2	+	+	-	50	No	Luminal B
P30	46	Infiltrating ductal carcinoma	3	+	+	-	?	Yes	Not specified
P32	38	Infiltrating ductal carcinoma	?	-	-	-	1	Yes	Triple negative
P35	38	Infiltrating ductal carcinoma	?	+	+	+	10	Yes	Luminal B

P37	70	Infiltrating ductal carcinoma	3	-	-	-	90	No	Triple negative
P38	38	Infiltrating ductal carcinoma	3	+	+	-	?	Yes	Not specified
P39	59	Infiltrating ductal carcinoma	2	+	+	-	?	Yes	Not specified
P41	40	Poorly differentiated infiltrating ductal carcinoma	3	+	+	?	80	Yes	Not specified
P42	73	Infiltrating ductal carcinoma of no special type	2	+	+	-	30	Yes	Not specified
P43	41	Infiltrating ductal carcinoma	2	?	?	?	?	Yes	Not specified
P46	63	Moderately differentiated infiltrating ductal carcinoma	?	-	-	-	60	Yes	Triple negative
P49	66	Moderately differentiated infiltrating ductal carcinoma	?	?	?	?	?	Yes	Not specified
P51	60	Infiltrating ductal carcinoma	1	+	+	-	2	No	Luminal A
P52	72	Infiltrating ductal carcinoma	1	+	+	-	0	No	Luminal A
P53	51	Infiltrating ductal carcinoma	2	+	+	-	1	Yes	Luminal A
P57	53	Infiltrating ductal carcinoma	3	-	-	+	?	?	Her-2 enriched
P58	69	Infiltrating ductal carcinoma	1	?	?	?	?	Yes	Not specified

Appendix VI: Protein concentrations of baro-extracts

Table A.1 Protein concentrations of the barocycler extracts for matched non-tumorous breast tissue and tumour samples.

Sample no.	Average protein concentration µg/µl
NP05	3.26
TP05	4.44
NP21	3.18
TP21	3.85
NP09	3.59
TP09	3.92
NP11	3.85
TP11	4.38
NP26	1.97
TP26	4.08
NP57	3.88
TP57	4.30
NP19	3.01
TP19	3.71
NP13	3.00
TP13	3.17
NP24	2.31
TP24	2.99
NP41	2.38
TP41	2.81
NP14	2.14
TP14	2.57
NP32	2.67
TP32	2.51
NP25	2.38
TP25	2.78
NP43	2.77
TP43	2.45
NP18	3.05
TP18	3.77
NP37	2.39
TP37	3.47

NP35	3.57
TP35	3.21
NP49	3.44
TP49	3.66
NP23	2.73
TP23	3.26
NP46	3.51
TP46	3.69
NP58	3.27
TP58	3.44
NP29	2.89
TP29	2.85
NP30	2.36
TP30	4.30
NP38	3.56
TP38	3.57
NP39	3.46
TP39	2.68
NP42	2.66
TP42	3.98
NP51	3.41
TP51	2.97
NP52	3.25
TP52	3.66
NP53	3.81
TP53	3.57

NP: non-tumorous tissue sample
TP: tumour samples

Appendix VII: Post analysis overview

Run No.	Sample Name	Condition	Replicate	Precursors	Modified Sequences	Peptides	Protein Groups	Proteins
1	NP53	Normal	1	7264	6294	6180	1292	1394
2	NP52	Normal	2	6084	5352	5261	1144	1238
3	NP51	Normal	3	8220	7192	7045	1475	1584
4	NP42	Normal	4	4096	3641	3599	855	936
5	NP39	Normal	5	6658	5854	5730	1247	1351
6	NP38	Normal	6	5916	5188	5063	1084	1179
7	NP30	Normal	7	10236	8881	8680	1783	1901
8	NP29	Normal	8	8009	7004	6853	1445	1545
9	NP58	Normal	9	6770	5970	5839	1265	1361
10	NP46	Normal	10	8636	7481	7384	1504	1614
11	NP23	Normal	11	6538	5706	5570	1159	1256
12	NP49	Normal	12	6826	5980	5843	1270	1370
13	NP35	Normal	13	9168	7960	7809	1575	1684
14	NP37	Normal	14	8222	7149	6978	1457	1570
15	NP18	Normal	15	6602	5788	5683	1210	1308
16	NP43	Normal	16	6649	5803	5721	1253	1358
17	NP25	Normal	17	6561	5761	5641	1225	1323
18	NP32	Normal	18	7767	6819	6664	1435	1552
19	NP14	Normal	19	7672	6722	6496	1405	1498
20	NP41	Normal	20	7294	6377	6233	1315	1417
21	NP24	Normal	21	8583	7496	7274	1504	1611
22	NP13	Normal	22	6696	5877	5705	1196	1288
23	NP19	Normal	23	6380	5623	5531	1224	1323
24	NP57	Normal	24	11008	9577	9362	1941	2071
25	NP26	Normal	25	7585	6712	6554	1496	1606
26	NP11	Normal	26	8743	7572	7428	1523	1630

27	NP09	Normal	27	8097	7014	6832	1445	1565
28	NP21	Normal	28	8325	7255	7078	1512	1627
29	NP05	Normal	29	8258	7154	6988	1509	1619
30	TP53	Tumour	1	11532	10011	9843	2042	2171
31	TP52	Tumour	2	11389	9880	9673	1942	2072
32	TP51	Tumour	3	8833	7752	7574	1603	1715
33	TP42	Tumour	4	12780	10957	10749	2160	2292
34	TP39	Tumour	5	9296	8102	7910	1829	1958
35	TP38	Tumour	6	12670	10942	10735	2142	2261
36	TP30	Tumour	7	11125	9776	9495	1999	2130
37	TP29	Tumour	8	8069	7012	6933	1645	1735
38	TP58	Tumour	9	8681	7552	7454	1540	1652
39	TP46	Tumour	10	8628	7524	7396	1543	1650
40	TP23	Tumour	11	8947	7826	7719	1628	1735
41	TP49	Tumour	12	11867	10177	9987	2007	2131
42	TP35	Tumour	13	13797	11866	11724	2245	2382
43	TP37	Tumour	14	13730	11818	11600	2226	2358
44	TP18	Tumour	15	14301	11998	11752	2322	2459
45	TP43	Tumour	16	12226	10523	10299	2048	2172
46	TP25	Tumour	17	9698	8442	8290	1696	1809
47	TP32	Tumour	18	11401	9870	9697	2030	2160
48	TP14	Tumour	19	12292	10595	10288	2003	2134
49	TP41	Tumour	20	10996	9480	9296	1934	2052
50	TP24	Tumour	21	14091	12058	11728	2244	2380
51	TP13	Tumour	22	12962	11169	10934	2168	2297
52	TP19	Tumour	23	11572	10008	9718	1984	2113
53	TP57	Tumour	24	12777	10957	10710	2133	2263
54	TP26	Tumour	25	12508	10773	10542	2110	2240
55	TP11	Tumour	26	13002	11187	10959	2204	2335
56	TP09	Tumour	27	13706	11662	11250	2286	2430
57	TP21	Tumour	28	12344	10617	10350	2124	2261
58	TP05	Tumour	29	12760	10828	10633	2156	2295

Appendix VIII: Candidate list of proteins

Uniprot accession number	Protein descriptions	AVG Log_2 Ratio	Absolute AVG Log_2 Ratio	P- value	Q- value
Q9HD20	Manganese-transporting ATPase 13A1	1.00	1.00	3.67E-02	9.84E-03
Q9UNE7	E3 ubiquitin-protein ligase CHIP	1.00	1.00	2.12E-09	1.30E-09
P06748	Nucleophosmin	1.00	1.00	2.78E-09	1.70E-09
P61764	Syntaxin-binding protein 1	-1.01	1.01	1.90E-17	2.40E-17
P30566	Adenylosuccinate lyase	-1.01	1.01	5.76E-07	2.72E-07
Q9UI10	Translation initiation factor eIF-2B subunit delta	1.01	1.01	3.60E-03	1.12E-03
P13611	Versican core protein	1.01	1.01	1.69E-18	2.42E-18
P60228	Eukaryotic translation initiation factor 3 subunit E	1.01	1.01	6.76E-13	5.83E-13
P01042	Kininogen-1	-1.02	1.02	1.07E-21	2.33E-21
P08603	Complement factor H	-1.02	1.02	4.24E-23	1.16E-22
Q01469	Fatty acid-binding protein 5	-1.02	1.02	2.24E-17	2.78E-17
Q8N474	Secreted frizzled-related protein 1	-1.02	1.02	2.29E-02	6.33E-03
P31948	Stress-induced-phosphoprotein 1	1.02	1.02	6.01E-11	4.29E-11
P01008	Antithrombin-III	-1.02	1.02	1.70E-20	3.06E-20
Q9Y6U3	Adseverin	-1.02	1.02	2.32E-03	7.45E-04
P35611	Alpha-adducin	-1.02	1.02	5.13E-19	8.00E-19
P51512	Matrix metalloproteinase-16	-1.02	1.02	1.86E-16	2.16E-16
P49207	60S ribosomal protein L34	1.02	1.02	1.32E-03	4.36E-04
P01780	Immunoglobulin heavy variable 3-7	-1.02	1.02	4.77E-11	3.44E-11
Q8NBS9	Thioredoxin domain-containing protein 5	1.02	1.02	6.27E-07	2.94E-07
O15533	Tapasin	1.02	1.02	1.26E-02	3.64E-03
Q5SSJ5	Heterochromatin protein 1-binding protein 3	1.02	1.02	1.07E-15	1.16E-15

Q9Y2R5	28S ribosomal protein S17, mitochondrial	1.03	1.03	1.69E-07	8.50E-08
P35080	Profilin-2	1.03	1.03	6.91E-09	3.95E-09
Q16787	Laminin subunit alpha-3	-1.03	1.03	1.46E-07	7.36E-08
P62249	40S ribosomal protein S16	1.03	1.03	3.64E-19	5.73E-19
P12109	Collagen alpha-1(VI) chain	-1.03	1.03	6.61E-16	7.23E-16
Q86UP2	Kinectin	1.03	1.03	2.14E-12	1.74E-12
Q8IXB3	Trafficking regulator of GLUT4 1	-1.03	1.03	1.15E-09	7.26E-10
Q00688	Peptidyl-prolyl cis-trans isomerase FKBP3	1.03	1.03	1.80E-02	5.06E-03
P22102	Trifunctional purine biosynthetic protein adenosine-3	1.04	1.04	4.41E-04	1.54E-04
Q13813	Spectrin alpha chain, non-erythrocytic 1	-1.04	1.04	3.80E-34	2.15E-32
Q6YN16	Hydroxysteroid dehydrogenase-like protein 2	-1.04	1.04	1.61E-19	2.59E-19
P23229	Integrin alpha-6	-1.04	1.04	9.34E-17	1.11E-16
P62318	Small nuclear ribonucleoprotein Sm D3	1.04	1.04	1.96E-12	1.61E-12
P31150	Rab GDP dissociation inhibitor alpha	-1.04	1.04	2.66E-03	8.46E-04
P55795	Heterogeneous nuclear ribonucleoprotein H2	1.04	1.04	1.07E-17	1.39E-17
Q16822	Phosphoenolpyruvate carboxykinase [GTP], mitochondrial	1.04	1.04	1.74E-13	1.58E-13
Q13232	Nucleoside diphosphate kinase 3	1.04	1.04	8.06E-07	3.72E-07
Q15027	Arf-GAP with coiled-coil, ANK repeat and PH domain-containing protein 1	1.04	1.04	3.23E-11	2.36E-11
P14868	Aspartate--tRNA ligase, cytoplasmic	-1.04	1.04	4.34E-03	1.35E-03
P02760	Protein AMBP	-1.04	1.04	2.32E-10	1.56E-10
P08134	Rho-related GTP-binding protein RhoC	1.05	1.05	3.09E-07	1.50E-07
O94855	Protein transport protein Sec24D	1.05	1.05	1.02E-02	3.00E-03
P19367	Hexokinase-1	1.05	1.05	8.96E-19	1.34E-18
P22061	Protein-L-isoaspartate(D-aspartate) O-methyltransferase	-1.05	1.05	2.84E-22	6.81E-22
Q04323	UBX domain-containing protein 1	1.05	1.05	1.25E-02	3.62E-03
Q8TAQ2	SWI/SNF complex subunit SMARCC2	1.05	1.05	4.06E-21	8.04E-21
P29373	Cellular retinoic acid-binding protein 2	1.05	1.05	3.21E-04	1.13E-04
O75947	ATP synthase subunit d, mitochondrial	-1.05	1.05	5.71E-06	2.47E-06

P22087	rRNA 2'-O-methyltransferase fibrillarin	1.05	1.05	3.78E-10	2.53E-10
Q8NFU3	Thiosulfate: glutathione sulfurtransferase	1.05	1.05	6.28E-04	2.16E-04
Q9BSJ8	Extended synaptotagmin-1	-1.06	1.06	3.95E-23	1.09E-22
O60506	Heterogeneous nuclear ribonucleoprotein Q	1.06	1.06	5.60E-18	7.49E-18
P26639	Threonine--tRNA ligase, cytoplasmic	1.06	1.06	9.70E-10	6.16E-10
P23634	Plasma membrane calcium-transporting ATPase 4	-1.06	1.06	1.87E-10	1.28E-10
P07686	Beta-hexosaminidase subunit beta	1.06	1.06	2.53E-05	1.01E-05
Q6DD88	Atlastin-3	-1.06	1.06	1.04E-14	1.04E-14
Q6P1J9	Parafibromin	1.06	1.06	5.10E-06	2.22E-06
Q709F0	Acyl-CoA dehydrogenase family member 11	1.07	1.07	4.21E-03	1.31E-03
P04433	Immunoglobulin kappa variable 3D-11; Immunoglobulin kappa variable 3-11	-1.07	1.07	1.40E-18	2.03E-18
O75348	V-type proton ATPase subunit G 1	1.07	1.07	6.32E-13	5.46E-13
P35579	Myosin-9	1.07	1.07	4.63E-21	9.08E-21
Q13085	Acetyl-CoA carboxylase 1	-1.07	1.07	7.37E-09	4.20E-09
P50914	60S ribosomal protein L14	1.07	1.07	1.13E-23	3.43E-23
P25940	Collagen alpha-3(V) chain	-1.07	1.07	3.01E-05	1.20E-05
Q15287	RNA-binding protein with serine-rich domain 1	1.07	1.07	1.61E-03	5.26E-04
P0DOX8	Immunoglobulin lambda-like polypeptide 5; Immunoglobulin lambda-1 light chain	-1.08	1.08	2.34E-18	3.24E-18
Q14554	Protein disulfide-isomerase A5	1.08	1.08	4.20E-10	2.79E-10
O43747	AP-1 complex subunit gamma-1	1.08	1.08	1.85E-08	1.02E-08
P05455	Lupus La protein	1.08	1.08	9.75E-17	1.15E-16
P48163	NADP-dependent malic enzyme	-1.08	1.08	1.46E-18	2.11E-18
P08865	40S ribosomal protein SA	1.08	1.08	1.95E-22	4.89E-22
P49585	Choline-phosphate cytidylyltransferase A	1.08	1.08	1.77E-02	4.98E-03
O43491	Band 4.1-like protein 2	-1.08	1.08	1.60E-02	4.52E-03
P07099	Epoxide hydrolase 1	-1.09	1.09	4.61E-22	1.09E-21
Q9UHG3	Prenylcysteine oxidase 1	-1.09	1.09	6.14E-28	4.05E-27

P0C0L4	Complement C4-A	-1.09	1.09	1.34E-17	1.71E-17
P13861	cAMP-dependent protein kinase type II-alpha regulatory subunit	-1.09	1.09	1.26E-25	5.18E-25
P56556	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 6	1.09	1.09	1.09E-09	6.87E-10
P28838	Cytosol aminopeptidase	1.09	1.09	1.55E-12	1.29E-12
Q9H3U1	Protein unc-45 homolog A	1.09	1.09	2.44E-09	1.49E-09
Q9Y6I3	Epsin-1	1.10	1.10	8.62E-03	2.57E-03
O43615	Mitochondrial import inner membrane translocase subunit TIM44	1.10	1.10	1.23E-02	3.57E-03
P10909	Clusterin	-1.10	1.10	2.95E-32	8.35E-31
Q9HBI1	Beta-parvin	-1.10	1.10	4.67E-20	8.08E-20
Q6P2Q9	Pre-mRNA-processing-splicing factor 8	-1.10	1.10	1.76E-02	4.93E-03
O14791	Apolipoprotein L1	-1.10	1.10	1.16E-13	1.07E-13
P30050	60S ribosomal protein L12	1.10	1.10	1.20E-21	2.59E-21
Q99471	Prefoldin subunit 5	-1.10	1.10	3.12E-02	8.46E-03
P30038	Delta-1-pyrroline-5-carboxylate dehydrogenase, mitochondrial	-1.10	1.10	2.13E-22	5.31E-22
P67936	Tropomyosin alpha-4 chain	1.10	1.10	5.74E-15	5.91E-15
Q92629	Delta-sarcoglycan	-1.11	1.11	6.79E-17	8.12E-17
Q9NYL9	Tropomodulin-3	-1.11	1.11	5.83E-06	2.51E-06
P07305	Histone H1.0	1.11	1.11	5.93E-08	3.13E-08
P51858	Hepatoma-derived growth factor	1.11	1.11	1.67E-12	1.39E-12
P49591	Serine--tRNA ligase, cytoplasmic	1.11	1.11	1.86E-21	3.91E-21
P07900	Heat shock protein HSP 90-alpha	1.12	1.12	2.24E-18	3.16E-18
P53999	Activated RNA polymerase II transcriptional coactivator p15	1.12	1.12	6.97E-06	2.98E-06
Q8NCA5	Protein FAM98A	1.13	1.13	3.45E-19	5.45E-19
P42167	Lamina-associated polypeptide 2, isoforms beta/gamma	1.13	1.13	6.44E-08	3.39E-08
P40306	Proteasome subunit beta type-10	1.13	1.13	1.21E-05	5.03E-06
A0A0J9YXX1	Immunoglobulin heavy variable 5-10-1	-1.13	1.13	2.33E-05	9.37E-06
P21291	Cysteine and glycine-rich protein 1	-1.13	1.13	8.42E-08	4.37E-08
P22352	Glutathione peroxidase 3	-1.13	1.13	1.04E-21	2.28E-21
P61165	Transmembrane protein 258	1.13	1.13	2.91E-16	3.32E-16

Q9UL46	Proteasome activator complex subunit 2	1.13	1.13	8.97E-17	1.07E-16
Q6NUM9	All-trans-retinol 13,14-reductase	-1.13	1.13	2.28E-20	4.04E-20
Q969X5	Endoplasmic reticulum-Golgi intermediate compartment protein 1	1.14	1.14	3.38E-09	2.03E-09
Q16836	Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	-1.14	1.14	5.05E-25	1.87E-24
P21283	V-type proton ATPase subunit C 1	1.14	1.14	1.94E-05	7.85E-06
Q9Y4E8	Ubiquitin carboxyl-terminal hydrolase 15	-1.14	1.14	2.51E-05	1.01E-05
Q13724	Mannosyl-oligosaccharide glucosidase	1.14	1.14	8.95E-03	2.66E-03
P16144	Integrin beta-4	-1.14	1.14	2.59E-02	7.11E-03
P48059	LIM and senescent cell antigen-like-containing domain protein 1	-1.15	1.15	1.71E-08	9.51E-09
P22748	Carbonic anhydrase 4	-1.15	1.15	1.29E-02	3.72E-03
P30101	Protein disulfide-isomerase A3	1.15	1.15	3.31E-16	3.73E-16
P02774	Vitamin D-binding protein	-1.15	1.15	3.10E-25	1.22E-24
O15020	Spectrin beta chain, non-erythrocytic 2	-1.15	1.15	3.28E-02	8.85E-03
P27797	Calreticulin	1.15	1.15	5.18E-09	3.00E-09
Q07866	Kinesin light chain 1	1.15	1.15	2.84E-07	1.39E-07
P11532	Dystrophin	-1.16	1.16	5.15E-03	1.58E-03
Q9Y2X3	Nucleolar protein 58	1.16	1.16	2.01E-03	6.48E-04
P14678; P63162	Small nuclear ribonucleoprotein-associated proteins B and B'; Small nuclear ribonucleoprotein-associated protein N	1.16	1.16	3.27E-18	4.50E-18
Q01130	Serine/arginine-rich splicing factor 2	1.16	1.16	4.34E-12	3.43E-12
P10412	Histone H1.4	1.16	1.16	1.69E-07	8.48E-08
Q9H6V9	Lipid droplet-associated hydrolase	-1.16	1.16	2.59E-02	7.13E-03
Q13033	Striatin-3	1.16	1.16	2.09E-02	5.79E-03
O00142	Thymidine kinase 2, mitochondrial	-1.17	1.17	1.60E-09	9.97E-10
Q8WU39	Marginal zone B- and B1-cell-specific protein	1.17	1.17	3.30E-02	8.91E-03
P02655	Apolipoprotein C-II	-1.17	1.17	5.99E-14	5.62E-14
Q9Y5K8	V-type proton ATPase subunit D	1.17	1.17	5.54E-07	2.63E-07
Q15661	Tryptase beta-2; Tryptase alpha/beta-1	-1.17	1.17	4.46E-16	4.95E-16
P49257	Protein ERGIC-53	1.18	1.18	2.76E-12	2.21E-12

O60716	Catenin delta-1	1.18	1.18	3.50E-09	2.10E-09
P12956	X-ray repair cross-complementing protein 6	1.18	1.18	8.58E-21	1.59E-20
Q96HR9	Receptor expression-enhancing protein 6	1.18	1.18	1.80E-13	1.63E-13
P20929	Nebulin	-1.18	1.18	9.19E-03	2.72E-03
P49755	Transmembrane emp24 domain-containing protein 10	1.18	1.18	1.54E-20	2.80E-20
P23396	40S ribosomal protein S3	1.18	1.18	1.21E-25	5.08E-25
Q8WTS1	1-acylglycerol-3-phosphate O-acyltransferase ABHD5	-1.18	1.18	3.29E-09	1.98E-09
Q8TB22	Spermatogenesis-associated protein 20	-1.19	1.19	1.31E-04	4.85E-05
Q05682	Caldesmon	1.19	1.19	6.05E-12	4.73E-12
P54136	Arginine--tRNA ligase, cytoplasmic	1.19	1.19	8.35E-19	1.26E-18
P13693	Translationally-controlled tumor protein	-1.19	1.19	2.17E-07	1.08E-07
Q8WX93	Palladin	1.19	1.19	7.44E-11	5.29E-11
O95292	Vesicle-associated membrane protein-associated protein B/C	1.19	1.19	8.31E-11	5.88E-11
P01857	Immunoglobulin heavy constant gamma 1	-1.19	1.19	3.64E-09	2.17E-09
Q99829	Copine-1	1.19	1.19	2.14E-02	5.93E-03
Q00013	55 kDa erythrocyte membrane protein	-1.19	1.19	7.05E-15	7.22E-15
P11021	Endoplasmic reticulum chaperone BiP	1.20	1.20	6.08E-20	1.02E-19
A0A075B6S2	Immunoglobulin kappa variable 2D-29; Immunoglobulin kappa variable 2D-26; Immunoglobulin kappa variable 2-29	-1.20	1.20	1.01E-05	4.22E-06
A0A0A0MS15	Immunoglobulin heavy variable 3-49	-1.20	1.20	5.84E-20	9.97E-20
A0A0B4J1U7	Immunoglobulin heavy variable 6-1	-1.20	1.20	7.97E-20	1.33E-19
P61604	10 kDa heat shock protein, mitochondrial	1.20	1.20	2.85E-09	1.73E-09
P40121	Macrophage-capping protein	1.20	1.20	1.03E-14	1.04E-14
P13804	Electron transfer flavoprotein subunit alpha, mitochondrial	-1.20	1.20	2.27E-08	1.25E-08
O75051	Plexin-A2	-1.20	1.20	2.05E-13	1.84E-13
P08729	Keratin, type II cytoskeletal 7	-1.20	1.20	2.09E-12	1.70E-12
Q07955	Serine/arginine-rich splicing factor 1	1.20	1.20	2.56E-06	1.14E-06
P43155	Carnitine O-acetyltransferase	-1.21	1.21	9.08E-12	7.00E-12
Q9NQC3	Reticulon-4	-1.21	1.21	1.18E-18	1.73E-18

Q14BN4	Sarcolemmal membrane-associated protein	-1.21	1.21	3.62E-05	1.43E-05
P01709	Immunoglobulin lambda variable 2-8	-1.21	1.21	4.48E-03	1.38E-03
P36871	Phosphoglucomutase-1	-1.22	1.22	7.19E-22	1.62E-21
P19823	Inter-alpha-trypsin inhibitor heavy chain H2	-1.22	1.22	1.43E-20	2.64E-20
Q15149	Plectin	1.22	1.22	1.16E-21	2.51E-21
P05091	Aldehyde dehydrogenase, mitochondrial	-1.22	1.22	1.59E-24	5.33E-24
Q13586	Stromal interaction molecule 1	1.22	1.22	3.40E-04	1.19E-04
O43324	Eukaryotic translation elongation factor 1 epsilon-1	1.22	1.22	2.29E-18	3.20E-18
P11387	DNA topoisomerase 1	1.22	1.22	1.81E-04	6.56E-05
P09651	Heterogeneous nuclear ribonucleoprotein A1	1.22	1.22	1.92E-19	3.07E-19
Q13418	Integrin-linked protein kinase	-1.23	1.23	1.41E-29	1.37E-28
P63220	40S ribosomal protein S21	1.23	1.23	1.45E-22	3.75E-22
P01009	Alpha-1-antitrypsin	-1.23	1.23	5.18E-23	1.41E-22
P01860	Immunoglobulin heavy constant gamma 3	-1.24	1.24	2.80E-11	2.04E-11
P0DMM9	Sulfotransferase 1A3; Sulfotransferase 1A4	-1.24	1.24	1.92E-09	1.18E-09
P27348	14-3-3 protein theta	1.24	1.24	3.62E-16	4.05E-16
P62070	Ras-related protein R-Ras2	-1.24	1.24	1.89E-11	1.40E-11
P69892	Hemoglobin subunit gamma-2	-1.24	1.24	4.76E-05	1.86E-05
P46776	60S ribosomal protein L27a	1.24	1.24	2.05E-15	2.17E-15
P84090	Enhancer of rudimentary homolog	1.24	1.24	1.28E-11	9.70E-12
P06727	Apolipoprotein A-IV	-1.24	1.24	5.16E-26	2.32E-25
Q8NFV4	Protein ABHD11	1.24	1.24	8.38E-03	2.50E-03
Q06323	Proteasome activator complex subunit 1	1.24	1.24	4.93E-18	6.62E-18
P16219	Short-chain specific acyl-CoA dehydrogenase, mitochondrial	-1.24	1.24	5.21E-26	2.32E-25
P61513	60S ribosomal protein L37a	1.24	1.24	9.09E-11	6.38E-11
P01011	Alpha-1-antichymotrypsin	-1.25	1.25	4.86E-21	9.44E-21
O15075	Serine/threonine-protein kinase DCLK1	1.25	1.25	2.40E-02	6.64E-03
P20962	Parathymosin	1.25	1.25	1.05E-11	8.06E-12
Q92614	Unconventional myosin-XVIIIa	1.25	1.25	2.33E-03	7.48E-04

P04275	von Willebrand factor	-1.25	1.25	1.43E-31	2.71E-30
Q99816	Tumor susceptibility gene 101 protein	1.25	1.25	4.05E-13	3.54E-13
Q13310	Polyadenylate-binding protein 4	1.25	1.25	1.04E-05	4.36E-06
P04003	C4b-binding protein alpha chain	-1.25	1.25	5.37E-19	8.33E-19
P46783	40S ribosomal protein S10	1.25	1.25	7.05E-22	1.60E-21
P62266	40S ribosomal protein S23	1.26	1.26	7.92E-28	5.14E-27
P02671	Fibrinogen alpha chain	-1.26	1.26	1.87E-10	1.28E-10
P05452	Tetranectin	-1.26	1.26	7.83E-03	2.36E-03
P21399	Cytoplasmic aconitate hydratase	-1.26	1.26	1.37E-26	7.35E-26
P42765	3-ketoacyl-CoA thiolase, mitochondrial	-1.26	1.26	3.40E-26	1.60E-25
P04233	HLA class II histocompatibility antigen gamma chain	1.27	1.27	1.16E-04	4.32E-05
Q92530	Proteasome inhibitor PI31 subunit	-1.27	1.27	9.04E-20	1.49E-19
P07942	Laminin subunit beta-1	-1.27	1.27	1.56E-24	5.28E-24
P00966	Argininosuccinate synthase	-1.27	1.27	1.39E-23	4.11E-23
Q9Y6N5	Sulfide: quinone oxidoreductase, mitochondrial	1.27	1.27	3.26E-04	1.15E-04
Q16363	Laminin subunit alpha-4	-1.27	1.27	1.20E-25	5.08E-25
P11233	Ras-related protein Ral-A	-1.27	1.27	1.80E-24	5.98E-24
O75339	Cartilage intermediate layer protein 1	-1.27	1.27	5.86E-04	2.02E-04
P35858	Insulin-like growth factor-binding protein complex acid labile subunit	-1.27	1.27	8.86E-11	6.23E-11
Q16576	Histone-binding protein RBBP7	1.27	1.27	3.12E-09	1.88E-09
Q04446	1,4-alpha-glucan-branching enzyme	-1.27	1.27	1.90E-26	9.63E-26
P30046	D-dopachrome decarboxylase	-1.27	1.27	9.68E-15	9.78E-15
O14782	Kinesin-like protein KIF3C	-1.27	1.27	9.39E-09	5.30E-09
O43633	Charged multivesicular body protein 2a	1.28	1.28	5.17E-14	4.88E-14
P06702	Protein S100-A9	-1.28	1.28	8.57E-11	6.04E-11
O43390	Heterogeneous nuclear ribonucleoprotein R	1.28	1.28	3.57E-30	4.29E-29
P84103	Serine/arginine-rich splicing factor 3	1.29	1.29	1.89E-09	1.17E-09
Q96A72	Protein mago nashi homolog 2	1.29	1.29	1.47E-02	4.19E-03
Q99459	Cell division cycle 5-like protein	-1.29	1.29	8.07E-21	1.51E-20

P06310	Immunoglobulin kappa variable 2-30	-1.29	1.29	1.06E-18	1.57E-18
P62701	40S ribosomal protein S4, X isoform	1.29	1.29	4.36E-24	1.38E-23
O95445	Apolipoprotein M	-1.29	1.29	1.47E-16	1.72E-16
P62750	60S ribosomal protein L23a	1.29	1.29	6.43E-10	4.20E-10
P05543	Thyroxine-binding globulin	-1.30	1.30	4.97E-05	1.94E-05
P16083	Ribosyldihydronicotinamide dehydrogenase [quinone]	-1.30	1.30	1.53E-05	6.33E-06
Q5VW32	BRO1 domain-containing protein BROX	1.30	1.30	7.88E-04	2.65E-04
P13639	Elongation factor 2	1.31	1.31	2.58E-20	4.53E-20
Q7L014	Probable ATP-dependent RNA helicase DDX46	1.31	1.31	2.99E-03	9.44E-04
P02790	Hemopexin	-1.31	1.31	4.82E-28	3.29E-27
P12270	Nucleoprotein TPR	1.31	1.31	2.17E-22	5.36E-22
P15559	NAD(P)H dehydrogenase [quinone] 1	-1.31	1.31	2.07E-08	1.13E-08
Q9BW30	Tubulin polymerization-promoting protein family member 3	-1.31	1.31	7.49E-06	3.18E-06
P02794	Ferritin heavy chain	1.31	1.31	5.07E-06	2.21E-06
P01019	Angiotensinogen	-1.31	1.31	2.09E-21	4.39E-21
Q8N4P3	Guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase MESH1	-1.32	1.32	4.72E-03	1.46E-03
A0A075B6P5	Immunoglobulin kappa variable 2-28; Immunoglobulin kappa variable 2-40; Immunoglobulin kappa variable 2D-40; Immunoglobulin kappa variable 2D-28	-1.32	1.32	2.32E-13	2.08E-13
Q99442	Translocation protein SEC62	1.32	1.32	1.47E-16	1.72E-16
Q92785	Zinc finger protein ubi-d4	1.32	1.32	6.42E-07	3.01E-07
O75781	Paralemmin-1	-1.32	1.32	3.25E-30	4.02E-29
P14625	Endoplasmin	1.33	1.33	9.92E-14	9.23E-14
Q8N163	Cell cycle and apoptosis regulator protein 2	-1.33	1.33	4.19E-09	2.47E-09
P15880	40S ribosomal protein S2	1.33	1.33	3.01E-15	3.16E-15
P62841	40S ribosomal protein S15	1.34	1.34	9.91E-08	5.09E-08
Q86TX2	Acyl-coenzyme A thioesterase 1	-1.34	1.34	1.05E-14	1.05E-14
P54920	Alpha-soluble NSF attachment protein	1.34	1.34	3.15E-05	1.25E-05
P15502	Elastin	-1.34	1.34	1.45E-03	4.77E-04

P61978	Heterogeneous nuclear ribonucleoprotein K	1.34	1.34	1.70E-13	1.55E-13
P05556	Integrin beta-1	-1.34	1.34	1.68E-26	8.66E-26
P02461	Collagen alpha-1(III) chain	-1.35	1.35	2.14E-07	1.07E-07
P53618	Coatomer subunit beta	-1.35	1.35	1.06E-02	3.12E-03
P52597	Heterogeneous nuclear ribonucleoprotein F	1.35	1.35	1.22E-17	1.58E-17
A0A0B4J1X5	Immunoglobulin heavy variable 3-74	-1.35	1.35	1.33E-18	1.93E-18
Q8NE62	Choline dehydrogenase, mitochondrial	-1.35	1.35	3.65E-05	1.44E-05
P00352	Retinal dehydrogenase 1	-1.35	1.35	8.29E-27	4.63E-26
A0A075B6S5	Immunoglobulin kappa variable 1-27; Immunoglobulin kappa variable 1-8; Immunoglobulin kappa variable 1-9	-1.35	1.35	4.54E-14	4.31E-14
P30040	Endoplasmic reticulum resident protein 29	1.35	1.35	7.98E-25	2.85E-24
Q02818	Nucleobindin-1	1.35	1.35	1.25E-11	9.50E-12
P61254	60S ribosomal protein L26	1.35	1.35	2.97E-21	6.00E-21
P06753	Tropomyosin alpha-3 chain	1.35	1.35	1.11E-18	1.63E-18
P46779	60S ribosomal protein L28	1.35	1.35	8.78E-08	4.55E-08
O00567	Nucleolar protein 56	1.35	1.35	7.39E-21	1.39E-20
P39060	Collagen alpha-1(XVIII) chain	-1.36	1.36	1.11E-28	8.95E-28
P61313	60S ribosomal protein L15	1.36	1.36	4.62E-21	9.08E-21
P50851	Lipopolysaccharide-responsive and beige-like anchor protein	1.36	1.36	8.03E-08	4.19E-08
P53621	Coatomer subunit alpha	1.36	1.36	3.55E-25	1.36E-24
P04626	Receptor tyrosine-protein kinase erbB-2	1.36	1.36	4.41E-03	1.37E-03
P30613	Pyruvate kinase PKLR	1.36	1.36	2.41E-21	4.97E-21
P00450	Ceruloplasmin	-1.36	1.36	1.98E-26	9.91E-26
P07814	Bifunctional glutamate/proline--tRNA ligase	1.36	1.36	3.21E-17	3.94E-17
P21397	Amine oxidase [flavin-containing] A	-1.36	1.36	1.55E-21	3.31E-21
Q07507	Dermatopontin	-1.36	1.36	4.13E-12	3.27E-12
P62081	40S ribosomal protein S7	1.37	1.37	4.18E-22	9.90E-22
P08697	Alpha-2-antiplasmin	-1.37	1.37	3.49E-04	1.22E-04
P01859	Immunoglobulin heavy constant gamma 2	-1.37	1.37	2.21E-18	3.13E-18

P02753	Retinol-binding protein 4;	-1.37	1.37	2.54E-28	1.87E-27
P02750	Leucine-rich alpha-2-glycoprotein	-1.38	1.38	4.57E-18	6.22E-18
P39019	40S ribosomal protein S19	1.38	1.38	4.81E-21	9.37E-21
P36543	V-type proton ATPase subunit E 1	1.38	1.38	3.22E-18	4.44E-18
Q9Y3B7	39S ribosomal protein L11, mitochondrial	1.38	1.38	3.12E-04	1.10E-04
Q13838	Spliceosome RNA helicase DDX39B	1.38	1.38	1.85E-28	1.43E-27
Q92841	Probable ATP-dependent RNA helicase DDX17	1.38	1.38	4.88E-22	1.14E-21
O15230	Laminin subunit alpha-5	-1.38	1.38	2.40E-21	4.97E-21
P01871	Immunoglobulin heavy constant mu	-1.38	1.38	7.95E-16	8.65E-16
P08621	U1 small nuclear ribonucleoprotein 70 kDa	1.38	1.38	2.55E-08	1.39E-08
P18428	Lipopolysaccharide-binding protein	-1.38	1.38	1.42E-02	4.05E-03
P01861	Immunoglobulin heavy constant gamma 4	-1.38	1.38	6.19E-09	3.56E-09
P0DOX6	Immunoglobulin mu heavy chain	-1.39	1.39	2.67E-13	2.37E-13
Q9NQT5	Exosome complex component RRP40	-1.39	1.39	1.34E-02	3.85E-03
Q96GQ7	Probable ATP-dependent RNA helicase DDX27	1.39	1.39	3.05E-16	3.46E-16
Q8TE77	Protein phosphatase Slingshot homolog 3	1.39	1.39	1.06E-03	3.52E-04
P49406	39S ribosomal protein L19, mitochondrial	1.40	1.40	9.17E-04	3.07E-04
Q14839	Chromodomain-helicase-DNA-binding protein 4	1.40	1.40	2.48E-15	2.62E-15
Q9NTX5	Ethylmalonyl-CoA decarboxylase	-1.40	1.40	1.55E-24	5.27E-24
P06737	Glycogen phosphorylase, liver form	-1.40	1.40	3.09E-24	1.00E-23
O76070	Gamma-synuclein	-1.41	1.41	2.70E-12	2.18E-12
O75891	Cytosolic 10-formyltetrahydrofolate dehydrogenase	-1.41	1.41	7.77E-15	7.93E-15
P63313	Thymosin beta-10	1.41	1.41	5.65E-06	2.45E-06
P07996	Thrombospondin-1	1.41	1.41	3.22E-07	1.56E-07
Q9NYF8	Bcl-2-associated transcription factor 1	1.41	1.41	2.97E-02	8.09E-03
P18564	Integrin beta-6	-1.41	1.41	4.47E-32	1.04E-30
P35527	Keratin, type I cytoskeletal 9	-1.41	1.41	1.49E-22	3.83E-22
P08238	Heat shock protein HSP 90-beta	1.41	1.41	1.28E-14	1.27E-14
P84098	60S ribosomal protein L19	1.41	1.41	2.39E-07	1.18E-07

Q96AC1	Fermitin family homolog 2	-1.42	1.42	9.36E-20	1.54E-19
P11498	Pyruvate carboxylase, mitochondrial	-1.42	1.42	6.04E-20	1.02E-19
P16403	Histone H1.2	1.42	1.42	7.86E-10	5.07E-10
O43175	D-3-phosphoglycerate dehydrogenase	-1.42	1.42	5.76E-22	1.32E-21
P05109	Protein S100-A8	-1.42	1.42	5.75E-06	2.48E-06
P06899	Histone H2B type 1-J; Histone H2B type 1-O; Histone H2B type 1-B; Histone H2B type 2-E	1.43	1.43	1.30E-14	1.28E-14
P78385	Keratin, type II cuticular Hb6; Keratin, type II cuticular Hb3; Keratin, type II cuticular Hb5; Keratin, type II cuticular Hb1	-1.43	1.43	1.48E-03	4.87E-04
Q6DN03; Q6DRA6	Putative histone H2B type 2-C; Putative histone H2B type 2-D	1.44	1.44	8.16E-13	6.98E-13
Q96KP4	Cytosolic non-specific dipeptidase	1.44	1.44	3.84E-23	1.07E-22
P48444	Coatomer subunit delta	1.44	1.44	4.44E-15	4.62E-15
Q9Y6E2	Basic leucine zipper and W2 domain-containing protein 2	1.44	1.44	1.50E-15	1.62E-15
O76009	Keratin, type I cuticular Ha3-I	-1.44	1.44	8.60E-05	3.25E-05
P22676	Calretinin	-1.44	1.44	3.06E-14	2.93E-14
P11171	Protein 4.1	-1.44	1.44	5.67E-23	1.52E-22
P05546	Heparin cofactor 2	-1.44	1.44	2.53E-23	7.26E-23
P04114	Apolipoprotein B-100	-1.44	1.44	3.49E-17	4.26E-17
P02679	Fibrinogen gamma chain	-1.45	1.45	3.49E-16	3.93E-16
O60831	PRA1 family protein 2	1.45	1.45	1.69E-06	7.65E-07
Q06830	Peroxiredoxin-1	-1.45	1.45	3.58E-23	1.01E-22
P62854	40S ribosomal protein S26	1.45	1.45	9.09E-23	2.40E-22
P52746	Zinc finger protein 142	-1.45	1.45	4.77E-28	3.29E-27
Q8N1G4	Leucine-rich repeat-containing protein 47	1.45	1.45	4.81E-18	6.53E-18
P08397	Porphobilinogen deaminase	-1.46	1.46	4.57E-09	2.66E-09
O95865	N(G), N(G)-dimethylarginine dimethylaminohydrolase 2	-1.46	1.46	8.47E-11	5.98E-11
P49720	Proteasome subunit beta type-3	-1.46	1.46	6.91E-08	3.63E-08
P15086	Carboxypeptidase B	1.47	1.47	1.84E-03	5.96E-04
P62847	40S ribosomal protein S24	1.47	1.47	7.03E-21	1.33E-20

P01024	Complement C3	-1.47	1.47	4.51E-28	3.19E-27
Q15084	Protein disulfide-isomerase A6	1.47	1.47	9.18E-19	1.37E-18
Q13835	Plakophilin-1	-1.47	1.47	2.76E-02	7.56E-03
Q7Z7G0	Target of Nesh-SH3	-1.48	1.48	1.42E-12	1.18E-12
P62280	40S ribosomal protein S11	1.48	1.48	3.50E-13	3.08E-13
P14174	Macrophage migration inhibitory factor	1.48	1.48	1.90E-12	1.57E-12
P01876	Immunoglobulin heavy constant alpha 1	-1.48	1.48	1.16E-10	8.11E-11
O75937	DnaJ homolog subfamily C member 8	1.48	1.48	3.12E-21	6.28E-21
P05155	Plasma protease C1 inhibitor	-1.49	1.49	5.83E-35	4.62E-33
Q08211	ATP-dependent RNA helicase A	1.49	1.49	2.13E-28	1.59E-27
Q9H078	Caseinolytic peptidase B protein homolog	1.49	1.49	6.78E-17	8.12E-17
Q86UE4	Protein LYRIC	1.49	1.49	6.16E-19	9.45E-19
P02647	Apolipoprotein A-I	-1.50	1.50	1.42E-31	2.71E-30
Q9BPW8	Protein NipSnap homolog 1	1.50	1.50	7.51E-14	7.03E-14
P13674	Prolyl 4-hydroxylase subunit alpha-1	1.50	1.50	1.87E-02	5.23E-03
P13667	Protein disulfide-isomerase A4	1.50	1.50	2.47E-09	1.51E-09
P11940	Polyadenylate-binding protein 1	1.51	1.51	7.09E-07	3.29E-07
Q15276	Rab GTPase-binding effector protein 1	-1.51	1.51	1.20E-02	3.48E-03
Q14112	Nidogen-2	-1.51	1.51	6.18E-29	5.32E-28
Q8IUG5	Unconventional myosin-XVIIIb	-1.51	1.51	1.15E-04	4.28E-05
Q5TDH0	Protein DDI1 homolog 2	-1.51	1.51	6.01E-20	1.02E-19
P02452	Collagen alpha-1(I) chain	-1.51	1.51	1.13E-04	4.23E-05
P00746	Complement factor D	-1.52	1.52	2.95E-23	8.40E-23
Q7KZF4	Staphylococcal nuclease domain-containing protein 1	1.52	1.52	2.03E-12	1.65E-12
P43243	Matrin-3	1.52	1.52	4.00E-24	1.28E-23
P20700	Lamin-B1	1.52	1.52	2.17E-17	2.70E-17
Q15404	Ras suppressor protein 1	-1.52	1.52	3.33E-20	5.83E-20
P55145	Mesencephalic astrocyte-derived neurotrophic factor	1.52	1.52	1.62E-06	7.31E-07
P0DOX2	Immunoglobulin alpha-2 heavy chain	-1.52	1.52	1.17E-09	7.36E-10

O75494	Serine/arginine-rich splicing factor 10	1.53	1.53	1.34E-02	3.86E-03
P62906	60S ribosomal protein L10a	1.53	1.53	1.53E-25	6.24E-25
Q9ULZ3	Apoptosis-associated speck-like protein containing a CARD	1.54	1.54	5.73E-05	2.21E-05
P00568	Adenylate kinase isoenzyme 1	-1.54	1.54	7.58E-18	1.00E-17
P07225	Vitamin K-dependent protein S	-1.54	1.54	2.65E-25	1.05E-24
Q13445	Transmembrane emp24 domain-containing protein 1	1.54	1.54	2.34E-03	7.48E-04
Q8IYB3	Serine/arginine repetitive matrix protein 1	1.55	1.55	5.97E-11	4.27E-11
P04217	Alpha-1B-glycoprotein	-1.55	1.55	3.32E-33	1.46E-31
P0C0S8	Histone H2A type 1; Histone H2A type 1-D; Histone H2A type 2-C; Histone H2A type 2-A; Histone H2A type 1-H; Histone H2A type 1-J; Histone H2A.J	1.55	1.55	8.28E-18	1.09E-17
P05164	Myeloperoxidase	-1.55	1.55	8.33E-13	7.11E-13
Q13683	Integrin alpha-7	-1.56	1.56	2.06E-18	2.92E-18
P62917	60S ribosomal protein L8	1.56	1.56	4.88E-19	7.64E-19
Q15746	Myosin light chain kinase, smooth muscle	-1.56	1.56	1.32E-08	7.41E-09
P40925	Malate dehydrogenase, cytoplasmic	-1.56	1.56	6.96E-31	1.10E-29
P51911	Calponin-1	-1.57	1.57	1.06E-03	3.53E-04
P47914	60S ribosomal protein L29	1.57	1.57	2.22E-12	1.80E-12
P07195	L-lactate dehydrogenase B chain	-1.58	1.58	2.10E-32	6.40E-31
Q9H307	Pinin	1.58	1.58	3.61E-16	4.05E-16
P06681	Complement C2	-1.58	1.58	2.10E-16	2.42E-16
P22570	NADPH: adrenodoxin oxidoreductase, mitochondrial	1.59	1.59	1.53E-20	2.80E-20
O14745	Na (+)/H (+) exchange regulatory cofactor NHE-RF1	1.59	1.59	5.04E-14	4.76E-14
Q13642	Four and a half LIM domains protein 1	-1.59	1.59	6.02E-20	1.02E-19
P00739	Haptoglobin-related protein	-1.59	1.59	1.06E-16	1.25E-16
Q00839	Heterogeneous nuclear ribonucleoprotein U	1.60	1.60	5.52E-20	9.46E-20
Q13228	Methanethiol oxidase	-1.60	1.60	6.43E-34	3.18E-32
P18124	60S ribosomal protein L7	1.60	1.60	9.94E-25	3.51E-24
Q9Y4L1	Hypoxia up-regulated protein 1	1.61	1.61	1.11E-14	1.11E-14

P15311	Ezrin	1.61	1.61	1.91E-14	1.87E-14
P34897	Serine hydroxymethyltransferase, mitochondrial	1.61	1.61	2.43E-22	5.97E-22
Q969G5	Caveolae-associated protein 3	-1.62	1.62	9.60E-29	7.92E-28
P24821	Tenascin	1.62	1.62	4.23E-09	2.49E-09
P10253	Lysosomal alpha-glucosidase	1.62	1.62	7.37E-05	2.81E-05
P46781	40S ribosomal protein S9	1.62	1.62	1.03E-26	5.62E-26
P55083	Microfibril-associated glycoprotein 4	-1.62	1.62	6.49E-15	6.66E-15
P61353	60S ribosomal protein L27	1.63	1.63	3.18E-23	9.00E-23
P60059	Protein transport protein Sec61 subunit gamma	1.63	1.63	4.50E-09	2.63E-09
P61009	Signal peptidase complex subunit 3	-1.64	1.64	9.17E-04	3.07E-04
Q9P2D6	Protein FAM135A	-1.65	1.65	2.21E-10	1.49E-10
P0C0L5	Complement C4-B	-1.65	1.65	1.92E-12	1.58E-12
P43121	Cell surface glycoprotein MUC18	-1.66	1.66	4.55E-30	5.15E-29
P55265	Double-stranded RNA-specific adenosine deaminase	-1.67	1.67	5.23E-08	2.77E-08
P46778	60S ribosomal protein L21	1.67	1.67	1.84E-23	5.39E-23
Q92781	Retinol dehydrogenase 5	-1.67	1.67	1.38E-16	1.62E-16
P19012	Keratin, type I cytoskeletal 15	1.67	1.67	5.38E-09	3.11E-09
Q5VT66	Mitochondrial amidoxime-reducing component 1	-1.67	1.67	7.32E-09	4.18E-09
Q9Y3U8	60S ribosomal protein L36	1.67	1.67	8.91E-20	1.48E-19
P19971	Thymidine phosphorylase	1.68	1.68	7.74E-04	2.61E-04
P83731	60S ribosomal protein L24	1.68	1.68	1.19E-24	4.18E-24
P35542	Serum amyloid A-4 protein	-1.68	1.68	6.80E-26	2.96E-25
P09874	Poly [ADP-ribose] polymerase 1	1.68	1.68	2.98E-14	2.86E-14
Q07020	60S ribosomal protein L18	1.69	1.69	3.93E-26	1.83E-25
Q16851	UTP--glucose-1-phosphate uridylyltransferase	-1.69	1.69	3.25E-29	2.93E-28
Q9NR28	Diablo homolog, mitochondrial	1.69	1.69	3.86E-16	4.30E-16
P12532	Creatine kinase U-type, mitochondrial	1.69	1.69	1.56E-09	9.77E-10
P24158	Myeloblastin	-1.69	1.69	2.05E-08	1.12E-08
P22626	Heterogeneous nuclear ribonucleoproteins A2/B1	1.70	1.70	8.85E-19	1.33E-18

P23610	Factor VIII intron 22 protein	-1.70	1.70	1.75E-05	7.15E-06
Q16629	Serine/arginine-rich splicing factor 7	1.70	1.70	7.58E-13	6.49E-13
P00738	Haptoglobin	-1.70	1.70	5.80E-19	8.98E-19
O00159	Unconventional myosin-Ic	-1.70	1.70	4.55E-31	7.50E-30
O43301	Heat shock 70 kDa protein 12A	-1.71	1.71	1.37E-30	1.93E-29
P51884	Lumican	-1.71	1.71	5.26E-13	4.57E-13
P11586	C-1-tetrahydrofolate synthase, cytoplasmic	-1.71	1.71	1.07E-05	4.48E-06
P30041	Peroxiredoxin-6	-1.71	1.71	4.30E-26	1.96E-25
P68431	Histone H3.1; Histone H3.3; Histone H3.1t; Histone H3.2	1.71	1.71	4.96E-15	5.15E-15
P18621	60S ribosomal protein L17	1.72	1.72	8.43E-18	1.10E-17
P27105	Erythrocyte band 7 integral membrane protein	-1.73	1.73	3.73E-25	1.41E-24
Q7Z4H3	HD domain-containing protein 2	-1.73	1.73	3.16E-02	8.58E-03
P23786	Carnitine O-palmitoyltransferase 2, mitochondrial	-1.73	1.73	1.32E-06	6.01E-07
P48506	Glutamate--cysteine ligase catalytic subunit	-1.73	1.73	2.93E-04	1.04E-04
P40429	60S ribosomal protein L13a	1.73	1.73	2.04E-15	2.17E-15
P23246	Splicing factor, proline- and glutamine-rich	1.74	1.74	2.80E-30	3.58E-29
P62269	40S ribosomal protein S18	1.74	1.74	4.75E-27	2.73E-26
Q12906	Interleukin enhancer-binding factor 3	1.74	1.74	1.45E-21	3.10E-21
P62899	60S ribosomal protein L31	1.74	1.74	9.44E-20	1.55E-19
P78527	DNA-dependent protein kinase catalytic subunit	1.74	1.74	1.38E-19	2.24E-19
A8MZA4	Golgin subfamily A member 6-like protein 6; Golgin subfamily A member 6-like protein 22; Golgin subfamily A member 6-like protein 1	-1.76	1.76	3.91E-10	2.61E-10
P02689	Myelin P2 protein	-1.77	1.77	1.28E-24	4.40E-24
Q9P1F3	Costars family protein ABRACL	1.77	1.77	2.42E-27	1.47E-26
P35908	Keratin, type II cytoskeletal 2 epidermal	-1.77	1.77	2.95E-08	1.60E-08
Q8N4H5	Mitochondrial import receptor subunit TOM5 homolog	1.77	1.77	6.71E-10	4.37E-10
Q14624	Inter-alpha-trypsin inhibitor heavy chain H4	-1.78	1.78	2.07E-28	1.58E-27
P62805	Histone H4	1.78	1.78	1.24E-25	5.18E-25
P05783	Keratin, type I cytoskeletal 18	1.78	1.78	1.20E-04	4.46E-05

P31947	14-3-3 protein sigma	-1.78	1.78	1.92E-02	5.36E-03
Q9BX66	Sorbin and SH3 domain-containing protein 1	-1.79	1.79	7.63E-29	6.43E-28
O75822	Eukaryotic translation initiation factor 3 subunit J	1.79	1.79	1.28E-22	3.35E-22
Q9HBL0	Tensin-1	-1.79	1.79	4.82E-33	1.73E-31
Q14699	Raftlin	1.79	1.79	3.21E-15	3.36E-15
P10620	Microsomal glutathione S-transferase 1	-1.79	1.79	1.04E-26	5.62E-26
P62263	40S ribosomal protein S14	1.79	1.79	4.58E-13	3.98E-13
P06454	Prothymosin alpha	1.80	1.80	2.22E-11	1.63E-11
P62851	40S ribosomal protein S25	1.80	1.80	2.74E-21	5.56E-21
Q96AZ6	Interferon-stimulated gene 20 kDa protein	1.80	1.80	1.36E-13	1.26E-13
P01023	Alpha-2-macroglobulin	-1.81	1.81	1.93E-30	2.64E-29
P16402	Histone H1.3	1.81	1.81	6.33E-07	2.97E-07
P51888	Prolargin	-1.81	1.81	1.55E-17	1.97E-17
P15289	Arylsulfatase A	-1.82	1.82	1.14E-10	7.94E-11
Q9P2E9	Ribosome-binding protein 1	1.82	1.82	1.87E-16	2.17E-16
Q86UX2	Inter-alpha-trypsin inhibitor heavy chain H5	-1.82	1.82	2.76E-08	1.50E-08
P04271	Protein S100-B	-1.84	1.84	5.45E-22	1.25E-21
P16930	Fumarylacetoacetase	-1.84	1.84	2.75E-26	1.34E-25
P26373	60S ribosomal protein L13	1.85	1.85	3.33E-22	7.94E-22
P17661	Desmin	-1.85	1.85	2.25E-07	1.12E-07
P07585	Decorin	-1.85	1.85	7.84E-16	8.55E-16
Q99685	Monoglyceride lipase	-1.86	1.86	3.77E-21	7.53E-21
P04216	Thy-1 membrane glycoprotein	1.86	1.86	1.63E-14	1.60E-14
P07910	Heterogeneous nuclear ribonucleoproteins C1/C2	1.86	1.86	9.90E-19	1.47E-18
P19525	Interferon-induced, double-stranded RNA-activated protein kinase	1.87	1.87	1.25E-14	1.24E-14
P02675	Fibrinogen beta chain	-1.87	1.87	9.70E-14	9.04E-14
Q8WZ42	Titin	-1.87	1.87	2.36E-20	4.18E-20
P16949	Stathmin	1.88	1.88	6.80E-06	2.92E-06
Q7Z794	Keratin, type II cytoskeletal 1b	-1.88	1.88	1.94E-10	1.32E-10

P51991	Heterogeneous nuclear ribonucleoprotein A3	1.88	1.88	7.32E-18	9.73E-18
Q15233	Non-POU domain-containing octamer-binding protein	1.88	1.88	8.83E-22	1.96E-21
Q9BRX8	Peroxiredoxin-like 2A	-1.90	1.90	6.29E-27	3.56E-26
P20774	Mimecan	-1.91	1.91	4.36E-15	4.54E-15
P52895	Aldo-keto reductase family 1 member C2	-1.91	1.91	3.56E-34	2.15E-32
Q9H6R3	Acyl-CoA synthetase short-chain family member 3, mitochondrial	-1.91	1.91	9.20E-05	3.46E-05
Q12765	Secernin-1	-1.92	1.92	2.38E-03	7.60E-04
Q02878	60S ribosomal protein L6	1.92	1.92	6.79E-22	1.54E-21
P98160	Basement membrane-specific heparan sulfate proteoglycan core protein	-1.92	1.92	1.09E-31	2.27E-30
O60814	Histone H2B type 1-K; Histone H2B type 1-D; Histone H2B type 1-C/E/F/G/I; Histone H2B type 2-F; Histone H2B type 1-H; Histone H2B type 1-N; Histone H2B type 1-M	1.94	1.94	2.30E-18	3.21E-18
O60568	Multifunctional procollagen lysine hydroxylase and glycosyltransferase LH3	1.94	1.94	2.21E-13	1.98E-13
Q08495	Dematin	-1.94	1.94	9.45E-22	2.09E-21
P22105	Tenascin-X	-1.94	1.94	1.74E-10	1.19E-10
Q9NQ50	39S ribosomal protein L40, mitochondrial	1.94	1.94	1.53E-09	9.58E-10
Q03135	Caveolin-1	-1.95	1.95	1.06E-30	1.62E-29
P63173	60S ribosomal protein L38	1.96	1.96	6.43E-24	1.98E-23
O75367	Core histone macro-H2A.1	1.96	1.96	9.44E-18	1.23E-17
Q16853	Membrane primary amine oxidase	-1.97	1.97	1.57E-26	8.18E-26
P31323	cAMP-dependent protein kinase type II-beta regulatory subunit	-1.98	1.98	2.02E-26	1.00E-25
Q14134	Tripartite motif-containing protein 29	-1.99	1.99	3.58E-02	9.62E-03
P14543	Nidogen-1	-1.99	1.99	1.71E-31	3.07E-30
P09467	Fructose-1,6-bisphosphatase 1	2.00	2.00	8.23E-19	1.25E-18
Q8TC05	Nuclear protein MDM1	-2.00	2.00	1.56E-26	8.18E-26
P13716	Delta-aminolevulinic acid dehydratase	-2.00	2.00	5.71E-25	2.05E-24
P13647	Keratin, type II cytoskeletal 5	-2.01	2.01	1.05E-03	3.51E-04
Q9NR12	PDZ and LIM domain protein 7	-2.02	2.02	7.13E-13	6.13E-13

P07858	Cathepsin B	2.02	2.02	6.86E-10	4.45E-10
P10301	Ras-related protein R-Ras	-2.02	2.02	4.47E-32	1.04E-30
Q9UMS6	Synaptopodin-2	-2.02	2.02	1.58E-05	6.52E-06
Q9Y6K5	2'-5'-oligoadenylate synthase 3	2.03	2.03	1.19E-23	3.56E-23
P09972	Fructose-bisphosphate aldolase C	-2.04	2.04	5.76E-21	1.10E-20
Q96CT7	Coiled-coil domain-containing protein 124	-2.04	2.04	4.42E-05	1.73E-05
P39059	Collagen alpha-1(XV) chain	-2.04	2.04	1.84E-25	7.38E-25
P17844	Probable ATP-dependent RNA helicase DDX5	2.05	2.05	2.90E-26	1.40E-25
Q6NZI2	Caveolae-associated protein 1	-2.07	2.07	3.69E-33	1.46E-31
P03952	Plasma kallikrein	-2.08	2.08	1.15E-03	3.81E-04
P08572	Collagen alpha-2(IV) chain	-2.08	2.08	2.23E-29	2.06E-28
Q07065	Cytoskeleton-associated protein 4	2.08	2.08	3.87E-24	1.25E-23
O75382	Tripartite motif-containing protein 3	2.08	2.08	3.01E-07	1.47E-07
P98095	Fibulin-2	-2.09	2.09	4.12E-08	2.21E-08
P25189	Myelin protein P0	-2.09	2.09	6.46E-06	2.77E-06
P30043	Flavin reductase (NADPH)	-2.11	2.11	6.05E-24	1.89E-23
Q9NZN4	EH domain-containing protein 2	-2.11	2.11	4.28E-27	2.50E-26
Q86V81	THO complex subunit 4	2.11	2.11	3.48E-08	1.87E-08
P52272	Heterogeneous nuclear ribonucleoprotein M	2.13	2.13	1.59E-11	1.19E-11
P02511	Alpha-crystallin B chain	-2.13	2.13	4.06E-26	1.87E-25
Q99715	Collagen alpha-1(XII) chain	2.13	2.13	4.39E-10	2.91E-10
Q9NWF9	E3 ubiquitin-protein ligase RNF216	-2.14	2.14	5.89E-29	5.18E-28
P55268	Laminin subunit beta-2	-2.14	2.14	2.36E-37	4.33E-35
O76062	Delta (14)-sterol reductase	-2.14	2.14	3.99E-21	7.93E-21
P11047	Laminin subunit gamma-1	-2.16	2.16	1.10E-35	1.08E-33
P02462	Collagen alpha-1(IV) chain	-2.19	2.19	2.98E-26	1.42E-25
Q96AG4	Leucine-rich repeat-containing protein 59	2.19	2.19	8.94E-18	1.17E-17
P53814	Smoothelin	-2.19	2.19	4.52E-12	3.57E-12
P33121	Long-chain-fatty-acid--CoA ligase 1	-2.21	2.21	7.97E-22	1.78E-21

P32119	Peroxiredoxin-2	-2.21	2.21	1.39E-22	3.62E-22
O14558	Heat shock protein beta-6	-2.21	2.21	7.85E-30	8.18E-29
Q14980	Nuclear mitotic apparatus protein 1	-2.22	2.22	4.80E-17	5.79E-17
O14776	Transcription elongation regulator 1	2.24	2.24	1.45E-02	4.12E-03
Q4G0F5	Vacuolar protein sorting-associated protein 26B	2.24	2.24	2.91E-13	2.57E-13
P08727	Keratin, type I cytoskeletal 19	2.27	2.27	4.31E-13	3.76E-13
Q9UKX2	Myosin-2	-2.27	2.27	4.64E-22	1.09E-21
Q92575	UBX domain-containing protein 4	-2.28	2.28	6.84E-08	3.60E-08
P36578	60S ribosomal protein L4	2.28	2.28	1.42E-27	8.81E-27
P42330	Aldo-keto reductase family 1 member C3	-2.29	2.29	3.35E-31	5.77E-30
P16157	Ankyrin-1	-2.29	2.29	5.04E-24	1.58E-23
P08579	U2 small nuclear ribonucleoprotein B''	-2.31	2.31	8.71E-06	3.67E-06
P49247	Ribose-5-phosphate isomerase	-2.31	2.31	1.64E-29	1.55E-28
P80303	Nucleobindin-2	2.32	2.32	4.39E-23	1.20E-22
P32455	Guanylate-binding protein 1	2.32	2.32	2.59E-14	2.51E-14
Q02790	Peptidyl-prolyl cis-trans isomerase FKBP4	2.32	2.32	2.27E-23	6.57E-23
O00534	von Willebrand factor A domain-containing protein 5A	-2.34	2.34	2.86E-02	7.81E-03
P23141	Liver carboxylesterase 1	-2.35	2.35	5.18E-22	1.20E-21
Q12955	Ankyrin-3	-2.36	2.36	2.65E-22	6.44E-22
Q05469	Hormone-sensitive lipase	-2.37	2.37	7.68E-33	2.53E-31
P16452	Erythrocyte membrane protein band 4.2	-2.37	2.37	5.33E-25	1.95E-24
P62861	40S ribosomal protein S30	2.38	2.38	2.51E-24	8.28E-24
P62277	40S ribosomal protein S13	2.38	2.38	1.37E-19	2.23E-19
Q8TC12	Retinol dehydrogenase 11	2.38	2.38	2.09E-19	3.32E-19
Q9BUL8	Programmed cell death protein 10	-2.38	2.38	4.66E-07	2.24E-07
P00325	Alcohol dehydrogenase 1B	-2.39	2.39	2.96E-27	1.78E-26
P69905	Hemoglobin subunit alpha	-2.39	2.39	2.71E-30	3.58E-29
P11217	Glycogen phosphorylase, muscle form	-2.40	2.40	1.80E-04	6.53E-05
P15090	Fatty acid-binding protein, adipocyte	-2.40	2.40	1.67E-21	3.53E-21

P42766	60S ribosomal protein L35	2.41	2.41	6.15E-24	1.90E-23
P08246	Neutrophil elastase	-2.42	2.42	4.87E-18	6.58E-18
P02458	Collagen alpha-1(II) chain	-2.43	2.43	2.91E-05	1.16E-05
O95678	Keratin, type II cytoskeletal 75	2.43	2.43	1.48E-13	1.36E-13
P08294	Extracellular superoxide dismutase [Cu-Zn]	-2.44	2.44	1.66E-37	4.33E-35
Q9HBH5	Retinol dehydrogenase 14	2.47	2.47	8.46E-10	5.42E-10
P02549	Spectrin alpha chain, erythrocytic 1	-2.49	2.49	1.07E-21	2.33E-21
O60240	Perilipin-1	-2.49	2.49	3.95E-30	4.60E-29
P11277	Spectrin beta chain, erythrocytic	-2.50	2.50	2.20E-23	6.41E-23
O95810	Caveolae-associated protein 2	-2.50	2.50	3.28E-37	4.33E-35
Q63HN8	E3 ubiquitin-protein ligase RNF213	-2.50	2.50	5.54E-23	1.49E-22
P20591	Interferon-induced GTP-binding protein Mx1	2.50	2.50	1.85E-12	1.53E-12
P16671	Platelet glycoprotein 4	-2.50	2.50	9.51E-30	9.66E-29
P16401	Histone H1.5	2.51	2.51	5.21E-15	5.39E-15
Q8TD06	Anterior gradient protein 3	2.52	2.52	3.24E-16	3.66E-16
Q9BYD6	39S ribosomal protein L1, mitochondrial	-2.53	2.53	4.92E-06	2.15E-06
P04040	Catalase	-2.53	2.53	6.11E-28	4.05E-27
Q04695	Keratin, type I cytoskeletal 17	-2.54	2.54	3.47E-06	1.53E-06
CASK_BOVIN	NaN	-2.54	2.54	9.82E-03	2.90E-03
Q15063	Periostin	2.57	2.57	1.70E-09	1.06E-09
P11166	Solute carrier family 2, facilitated glucose transporter member 1	-2.60	2.60	6.60E-26	2.90E-25
P21695	Glycerol-3-phosphate dehydrogenase [NAD (+)], cytoplasmic	-2.60	2.60	4.71E-30	5.18E-29
P04264	Keratin, type II cytoskeletal 1	-2.61	2.61	1.25E-05	5.20E-06
P45877	Peptidyl-prolyl cis-trans isomerase C	2.61	2.61	9.09E-26	3.91E-25
O75716	Serine/threonine-protein kinase 16	-2.72	2.72	4.16E-32	1.04E-30
P13645	Keratin, type I cytoskeletal 10	-2.75	2.75	5.79E-06	2.50E-06
P68871	Hemoglobin subunit beta	-2.75	2.75	4.33E-28	3.12E-27
P07738	Bisphosphoglycerate mutase	-2.76	2.76	1.27E-24	4.40E-24
Q96Q06	Perilipin-4	-2.78	2.78	5.41E-30	5.79E-29

P36269	Glutathione hydrolase 5 proenzyme	2.78	2.78	1.95E-02	5.42E-03
P06732	Creatine kinase M-type	-2.81	2.81	1.54E-02	4.36E-03
P02792	Ferritin light chain	2.83	2.83	9.02E-08	4.66E-08
P07203	Glutathione peroxidase 1	-2.86	2.86	1.01E-08	5.70E-09
Q13576	Ras GTPase-activating-like protein IQGAP2	-2.91	2.91	2.38E-10	1.60E-10
P02730	Band 3 anion transport protein	-2.95	2.95	5.52E-25	2.00E-24
Q59GN2	60S ribosomal protein L39; Putative 60S ribosomal protein L39-like 5	2.98	2.98	3.53E-25	1.36E-24
P31944	Caspase-14	-3.03	3.03	5.58E-03	1.70E-03
O60264	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	-3.12	3.12	1.32E-30	1.93E-29
Q16630	Cleavage and polyadenylation specificity factor subunit 6	3.12	3.12	1.42E-02	4.06E-03
P02008	Hemoglobin subunit zeta	-3.15	3.15	2.60E-04	9.28E-05
P00918	Carbonic anhydrase 2	-3.21	3.21	1.15E-28	9.12E-28
O75970	Multiple PDZ domain protein	-3.26	3.26	1.27E-23	3.78E-23
Q9Y3P9	Rab GTPase-activating protein 1	-3.29	3.29	2.76E-22	6.65E-22
P02042	Hemoglobin subunit delta	-3.33	3.33	8.33E-28	5.32E-27
P05787	Keratin, type II cytoskeletal 8	3.48	3.48	3.36E-15	3.51E-15
Q14232	Translation initiation factor eIF-2B subunit alpha	3.55	3.55	5.18E-05	2.01E-05
P00915	Carbonic anhydrase 1	-3.57	3.57	1.42E-27	8.81E-27
O95994	Anterior gradient protein 2 homolog	3.58	3.58	8.39E-18	1.10E-17
P14780	Matrix metalloproteinase-9	-3.58	3.58	3.57E-07	1.72E-07
P61221	ATP-binding cassette sub-family E member 1	-3.96	3.96	5.89E-32	1.30E-30
P02100	Hemoglobin subunit epsilon	-4.16	4.16	1.23E-06	5.62E-07
P15170	Eukaryotic peptide chain release factor GTP-binding subunit ERF3A	-4.22	4.22	2.36E-21	4.91E-21
P02144	Myoglobin	-4.56	4.56	5.01E-05	1.95E-05
P47929	Galectin-7	-5.87	5.87	1.44E-02	4.11E-03
O95154	Aflatoxin B1 aldehyde reductase member 3	6.02	6.02	1.89E-17	2.39E-17