Identification and validation of Biomarkers for Breast Cancer from human White Blood Cells

Jayakumar Mani

A thesis submitted for the degree of Doctor of Philosophy

Department of Biological Sciences

University of Essex

Acknowledgments

Firstly I would like to thank my supervisor Professor Elena Klenova for her continued support, training and faith in me through the course of this PhD. Without her guidance and constant feedback this work would not have been possible. I also thank my second supervisor Professor Nelson Fernandez for his expert and moral support during the time of my PhD. I would also like to thank Dr Metodi Metodiev for his help in performing mass spectrometry experiments and data analysis. Special thanks to Dr Dawn Farrar from our lab for her valuable guidance and practical advice in the lab providing support in the most needed times.

I would like to thank Adele Angel of our lab for her support in purchasing materials and help in organising the lab. I would like to thank Dr Anu Apte for her support and help in the clinical end of the project. I am thankful to the members of our lab Dr Yukti Hari, Dr Nikhil Pawa, Dr Zubair Cheema, Dr Svetlana Gretton, Dr Hulkar Mamayusupova, Dr Olayinka Oloko, Krista McHugh, Ioanna Pavlaki and Rosie Bryan for sharing troubles and successes during our time in the lab. I am thankful to the contribution of many other staff and faculty members of the School of Biological Sciences at the University of Essex.

I gratefully acknowledge the funding received towards my PhD from the University of Essex scholarships. I would also like to thank the staff at Colchester General Hospital breast surgery unit for providing the patients samples.

Finally I would like to thank my parents for their love, continued support, and motivation which made me progress this far in life and career.

Abstract

There is a great need for the identification of non-invasive biomarkers for early detection, prognosis and treatment efficacy of breast cancer. Peripheral white blood cells (WBCs) carrying the information related to the presence of cancer, represent an attractive source for novel biomarkers. The main aims of the study were to develop the pipeline to discover and validate novel biomarkers in WBCs of breast cancer patients using proteomic and genomic approaches, and assess these biomarkers' utility. Using the highthroughput mass spectrometry and 2D-gel-electrophoresis, the protein profiles of the WBCs from breast cancer patients and healthy individuals were generated and compared with publicly available gene expression data from the WBCs of breast cancer patients and the information on protein profiles of the WBCs from the metastatic breast cancer patients. The shortlisted 15 genes were then validated using Real-Time Quantitative Reverse Transcription *PCR (RT-qPCR)*. The mRNA levels of *ITGA4*, *LCN2*, *CPNE3* and *SERPINB1* were found to be altered significantly in the WBCs of breast cancer patients. The levels of SERPINB1 (Serpin B1, neutrophil elastase inhibitor) and CPNE3 (Copine 3, phospholipid binding protein) were assessed using Western blotting. These analyses demonstrated the association of SERPINB1 with breast cancer metastases, and suggested its potential utility as a biomarker of poor prognosis and treatment efficacy. Further quantitative validation of SERPINB1 in a larger panel of WBCs by ELISA will be required for a clinical phase in the biomarker development pipeline. No conclusive results were obtained for CPNE3 and, together with ITGA4, LCN2 and other additional candidate biomarkers (to be selected from the initial list), they will be tested further. The data generated for this study has also given insight into differences in the molecular portraits of the cells of immune system associated with breast cancer, which will need to be validated by laboratory based functional assays.

Abbreviations

Amp	Ampicillin
APP	Amyloid Protein Precursor
APS	Ammonium persulfate
BCR	B cell receptor
BRCA-1	Breast cancer – 1 gene
BRCA-2	Breast cancer - 2 gene
BORIS	Brother of the regulator of imprinting sites
Bp	Base pairs
BSA	Bovine serum albumin
$CaCl_2$	Calcium chloride
CKII	Casein kinase II
cDNA	Complementary DNA
CTS	CTCF target sequences
CTCF	CCCTC sequence binding factor
C-terminal	Carboxy-terminal domain
DAPI	4',6-diamidino-2-phenylindole dilactate
DM1	Myotonic dystrophy type 1

DMD	Differentially methylated region
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DNMT1	DNA (cytosine-5)-methyltransferase 1
ECL	Enhanced chemiluminescence immunodetection
EDTA	Ethylenediaminetetraacetic acid
EGFP	Enhanced green fluorescence protein
FBS	Fetal bovine serum
FISH	Fluorescence in situ hybridisation
FITC	Fluorescin isothiocyanate
FOG	Friend of GATA-1
HCE	Human corneal epithelial cells
HDAC	Histone deacetylase
HRP	Horseradish peroxidase
ICR	Imprinting control region
IHC	Immunohistochemistry
IF	Immunofluorescence
kDa	Kilodalton

LB	Luria broth
LOH	Loss of heterozygosity
LOI	Loss of imprinting
mES	mouse embryonic stem cells
MI	Mitotic index
mRNA	Messenger RNA
NaB	Sodium butyrate
NLS	Nuclear localisation signal
ORF	Open reading frame
RNA	Ribonucleic acid
RT	Room temperature
rpm	Revolutions per minute
PAR	Poly(ADP-ribose)
PARG	Poly(ADP-ribose)glycohydrolase
PARP	Poly(ADP-ribose)polymerase
PCNA	proliferating cell nuclear antigen
PBS	phosphate buffered saline
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl SulfatePolyacrylamide Gel Electrophoresis

siRNA	Small interference RNA
SUMO	Small ubiquitin-like modifier
TEMED	N,N,N’,N’,-tetramethylehylenediamine
TRITC	Tetramethyl Rhodamine Isothiocyanate
TSS	Transcription start site
WB	Western Blot
ZF	Zinc finger
3C	Chromosome conformation capture

Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Breast Cancer- Incidence and statistics

Cancer can be defined as the uncontrollable proliferation of cells which then become invasive and metastasise by spreading to various regions through blood or lymphatic vessels. Cancer can be classified based on their type of origin as follows: carcinomas (epithelial cells), sarcomas (mesenchymal connective tissues), leukaemia and lymphomas (haematopoietic cells). Breast cancer is the most common cancer in women who have 1 in 8 lifetime risk of developing breast cancer (Cancer Research UK, 2013). Around 49000 women were diagnosed with breast cancer in 2011 and breast cancer incidence has increased by 72% since the 1970s. Worldwide, 1.68 million women were diagnosed with breast cancer in 2012. It is estimated that ~522,000 women died of breast cancer in 2012 (Cancer Research UK, 2013). However, breast cancer survival rates have been increasing with ~ 80% of women surviving primary breast cancer for more than 5 years.

1.2 Mammary gland and breast cancer

The mammary gland is a modified sweat gland that originates from the ectoderm. The breast has two main parts: supporting stroma and parenchyma. The parenchyma consists of ducts, lobules and alveoli, and the stroma includes connective and fatty tissues. The parenchyma tissue has 15-20 lobes, and within each lobe there are smaller lobules which produce milk. The lobules are connected by thin ducts and all the ducts lead to the nipple. Epithelial cells line both the ducts and the lobules. The epithelial cells are surrounded by myoepithelial cells which are attached to the basal membrane. The lobules and ducts are surrounded by the stroma consisting of fibroblasts and adipocytes (Torácica and Mama, 2006). The mammary epithelium undergoes several cycles of remodelling, proliferation, differentiation and apoptosis at various stages of life

such as puberty, pregnancy, lactation and menopause. These changes of the breast are induced by hormones such as oestrogen, progesterone and prolactin (Russo and Hu, 2000).

1.2.1 Types of Breast cancer

Breast cancer is classified morphologically in two major categories based on their histology: Invasive and non-invasive (or *in situ*) cancer. Breast cancer is also classified based on the cells of origin: Ductal, lobular and tubular cancers. Non-invasive cancer cells do not penetrate the basal layer and do not spread to others parts of the breast. Invasive cancer spreads to the surrounding ducts, lobules, lymph nodes and to distant organs. Medullary, mucinous and pappilary cancers of the nipple are few of the rarer morphological types accounting for 2-5% of all cancers (Fentiman and D'Arrigo, 2004). Invasive ductal carcinoma (IDC) accounts to up to 80% of breast cancers. Invasive lobular carcinoma accounts to 5-15% of all diagnosed breast cancers.

Breast cancers are clinically classified into primary breast cancer and metastatic breast cancer. Primary breast cancer are small to medium sized tumours (<50mm) which are localised to only the primary site of tumour development. Metastatic breast cancers are spread to different parts of the body such as bone, brain, lungs etc.

Breast tumours are also described based on their size, spread and differentiation. Tumour staging (1-4) depends on the size of the tumour and lymph node metastasis. Tumour grading (1-3) depends on how different cancer cells look when compared to normal cells when viewed under the microscope. Estrogen receptors (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) statuses in the breast tumour are also important to decide the treatment protocol and assess response to therapy. ER and PR status are currently diagnosed using Immunohistochemistry (IHC). HER2 status is identified by nucleic acid hybridisation (Cianfrocca and Gradishar, 2009).

Over the last 15 years due to the development of highthroughput genomics and proteomics the molecular classification breast cancer has become clearer, although there is no common standard categorization of breast cancer based on this concept. For example breast tumours were classified by Perou et al based on gene expression of 456 genes in 65 tissue samples (Perou et al. 2000). This study resulted in the elucidation of four different molecular types of breast cancers: luminal-like (ER-positive, PR-positive, HER2-negative), basal-like (ER negative, PR negative, HER2 negative), HER2-positive, and normal breast. The most recent work revealed novel molecular subgroups of breast cancer based on genomic and transcriptomic experiments (Curtis et al., 2012; Dawson et al., 2013). In these studies, copy number, single nucleotide polymorphisms, and gene-transcription rates in 2,000 breast cancers were analysed and 10 different molecular subtypes of breast cancer based on patient treatment outcomes were identified. Furthermore, 45 regions in the genome were discovered which are either deleted or amplified abnormally and might contribute to the pathophysiology of breast cancer. These findings however are not yet strong enough to be applied to clinical research.

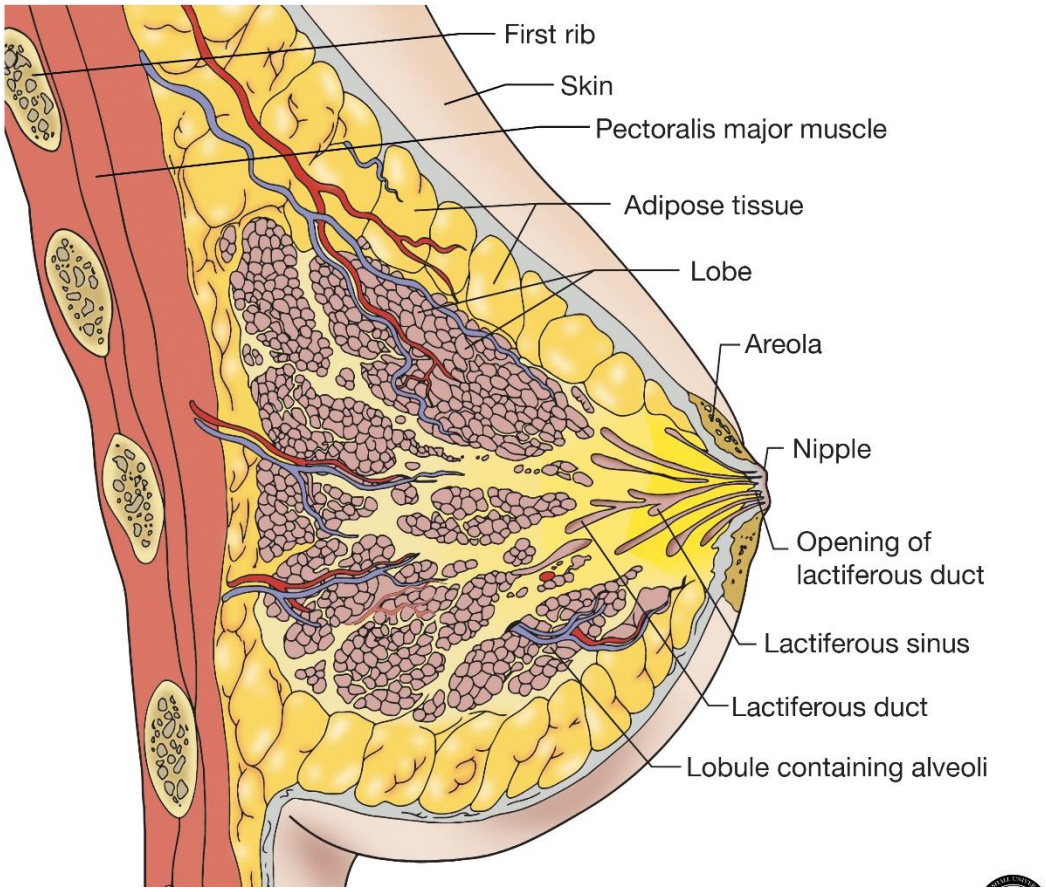


Figure 1.1 Anatomy of Breast. Front view and side view of a normal breast showing ducts, lobules, adipose tissue, muscle and lymph nodes. The breast lies superficial to the pectoralis muscle and drains to the regional nodes (Source: http://musom.marshall.edu/graphicdesign/ibooks/Reproductive%20Normal.html Last Accessed 09/07/2015).

1.3 Risk factors for breast cancer development

The exact cause of breast cancer is unknown and it can be caused by a number of risk factors. The risk factors include age (over 50 years), geographical variation (higher in western countries), first child bearing age (after 40 years), menarche age (before 11 years), menopause (after 54 years), family history of cancer incidence and genetic predisposition. Lifestyle factors such as obesity, alcohol intake, tobacco smoking, oral contraceptives and hormone replacement therapy increases susceptibility to breast cancer (Mcpherson et al., 2000).

1.4 Molecular pathology of Breast Cancer

The mutations in the genes controlling the cell cycle lead to defects in the regulation of cell cycle which eventually lead to tumour formation. The genes which are involved in tumour formation can broadly be classified into Oncogenes and Tumour suppressor genes. Oncogenes are genes which are mutated are highly expressed in tumour cells and tumour suppressor genes prevent the cell from entering into tumour development stage. When DNA damage occurs in a cell an immediate cell reaction called the DNA damage response (DDR) is triggered. This response prevents the cell from undergoing DNA replication and separation of damaged DNA into daughter cells and thereby preventing the propagation of corrupt genetic information and finally acts with other cellular components to repair the damaged DNA. If the damaged DNA is repaired the cells will then undergo normal proliferation but if the DNA damage is so severe that it cannot be repaired easily the cells undergo apoptosis (Lord and Ashworth, 2012). However an additional outcome is also possible that the cells may go into cellular senescence. Cellular senescence can be defined as an irreversible state of cell cycle arrest with the survival of the cell in absence of any further cell division. It is still not clear what determines the cell to make its choice between apoptosis and senescence (Sulli et al., 2012). The link between senescence and

cancer development is still debatable as to whether senescent cells lead to formation of cancer (Rodier and Campisi, 2011).

1.4.1 Genes involved in breast tumourigenesis

Genes which exhibit in hereditary mutations leading to breast cancer are the most extensively studied genes. Examples include *BRCA1, BRCA2, TP53, ATM, CHK2, PTEN* and *CDH*-1 which are hereditary breast cancer causing genes which are mutated to cause the formation of tumour. BReastCAncer-1 gene (BRCA1) and BReastCAncer-2 gene (BRCA2) are estimated to account for between 5%-10% of breast cancer. BRCA1 and BRCA2 are tumour suppressors with the main functions in homologous recombination. BRCA1 binds directly to the single stranded DNA and localises along with RAD51 and BARD1 to the site of the DNA damage which is already in complex with histones (Scully and Livingston, 2000). BRCA2 can bind with RAD51 while the binding nature of BRCA1 and RAD51 is unknown. BRCA1 is phosphorylated in this process by ataxia-telangiectasia mutated (ATM). Regardless of the presence BRCA1 the cellular localisation and the activity of RAD51 is controlled by BRCA2 (Cousineau et al., 2005). The levels of these three proteins also increase when the cell enters S-phase suggesting the role of these proteins in maintaining genome integrity (Narod and Foulkes, 2004). In BRCA1 and BRCA2 knockout mice p21 (a p53 dependent cell cycle inhibitor) was induced to delay the cell cycle growth arrest only to leading to delayed death (Scully and Livingston, 2000). These findings imply that if BRCA genes are mutated the cell becomes abnormal and the cell survives for a prolonged period and driven towards tumourigenesis. Figure 1.1 illustrates the functions of BRCA1 and BRCA2. The DNA breaks may lead to three possible outcomes: cell cycle arrest and repair, cell cycle arrest and apoptotic cell death, or the errors will be compromised leading to tumour development.

BRCA1 mutations responsible for the protein's malfunction are caused by the introduction of a stop codon or a substitution of base pair, and also insertions and deletions which leads to a truncated or mutated BRCA1 protein. BRCA1 and BRCA2 null mice cells were directed to apoptosis while the cells in which there were mutated BRCA proteins p53 and p21 were inactivated. The breast epithelial cells are clonal due to the rapid growth rate during puberty. Hence if one cell carries a mutation the progeny of that cell will carry that mutation and be localised in one part of the breast (Narod and Foulkes, 2004; Scully and Livingston, 2000). Data also suggests that hypermethylation of the promoter region of BRCA1 and p53 is involved in breast cancers without the hereditary factor (Alkam et al., 2013; Baylin and Herman, 2000; Krasteva et al., 2012). Other genes which are hereditary involved with BRCA genes are ATM and CHK2 genes. CHK2 also codes for RAD3 which is involved in BRCA1-RAD51-BRCA2 complex during double stranded breaks; Both CHK2 and ATM phosphorylate BRCA1 and p53 thereby regulating their function.

Tumour suppressor gene TP53 responsible for the expression of the protein, p53,a transcription factor involved in the regulation of many functions including cell cycle and apoptosis, and is functionally inactivated in 70% of human tumours. On encountering DNA damage, a cell will either activate a DNA damage checkpoint to slow down or arrest cell division and allow repair of the lesions, or apoptosis will take place to remove cells with potentially lethal mutations. P53, which levels rapidly increase after DNA damage, plays the key role in these processes. There are various mechanisms employed by p53 to induce apoptosis in a stage-, tissue- and stress-signal-specific manner. One of them is direct regulation of apoptosis-related genes such as DR5, Fas, Bax, Noxa and Bcl-2. However, p53 can also promote apoptosis via transcription-independent mechanisms (Haupt et al., 2003; Möröy and Zörnig, 1996).

The presence of functional p53 is important in senescence (Herbig and Sedivy, 2006; Itahana et al., 2001). The p53 communications with the multitude of proteins such as the Wnt-

beta-catenin, IGF-1-AKT, Rb-E2F, p38 MAP kinase, cyclin-cdk, p14/19 ARF and the cyclin G-PP2A, and p73 and others have been shown to be involved in the regulation of senescence and apoptosis in different cell types (Harris and Levine, 2005).

The link between p53 and breast cancer has been well established. For example, 30% of breast tumours have been reported to have mutations in p53, although it is higher in whereas 70% of colon tumours (70%) and generally in all cancers (50%) (Slee et al., 2004). However, other mechanisms, for example, nuclear exclusion can also lead to p53 malfunction and development of breast tumours(Moll et al., 1992). Overexpression of p53 protein was reported to be a tissue biomarker for disease recurrence and reduced survival rates in patients with breast cancer (Friedrichs et al., 1993).

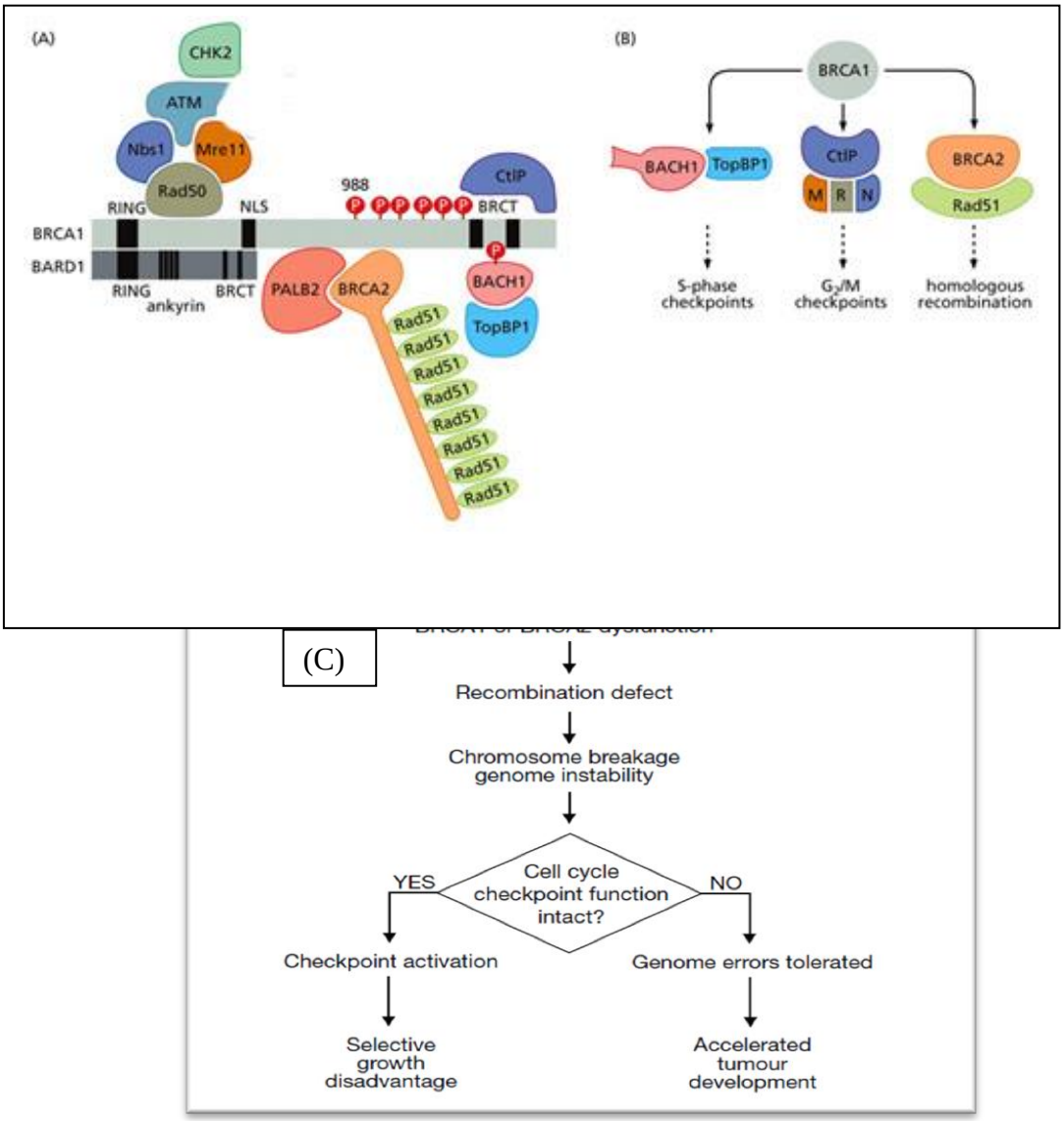


Figure 1.2 BRCA1 and BRCA2 are important for DNA repair processes, in particular homologous recombination: BRCA1 and BRCA2 proteins act as scaffold to assemble DNA repair proteins (A); Loss of function of BRCA1 and BRCA2 may result in defects in repair, accumulation of mutation and, ultimately, cancer (B). Taken from Weinberg, 2006. Mutated BRCA protein function may lead to tumour development or the cell growth is arrested depending on the cell cycle checkpoint proteins such as CHK1 and PLK1 (C)**.** (Taken from Scully & Livingston, 2000)

1.5 Metastasis of breast cancer

Metastatic breast cancers (MBC) also known as advanced breast cancers (ABS) are invasive breast cancers of higher grade and stage which spread to distant organs such as lymph nodes, bones, lungs, brain, liver. Approximately 20-30% of the primary breast cancer patients develop metastatic disease. Metastatic breast cancer is major cause of breast cancer related deaths. The median survival range after treatment of metastatic breast cancer is 24-36 months. Metastasis occurs in different steps (Fidler, 2003) that are illustrated in Figure 1.2. Several cellular processes and molecules are involved in cancer metastasis at each stage (Leber and Efferth, 2009).The subsequent metastatic events and the molecules involved are described below:

i. Division and growth of tumour at the primary site. Receptors such as Vascular endothelial growth factor receptor (VEGF) and epidermal growth factor receptor (EGFR) are activated tumour secreted growth factors such as HER2 and epidermal growth factor (EGF).

ii. Development of blood vessels (angiogenesis) in the tumour mass; when the tumour grows more than 2 cm it needs blood vessels to survive. Hypoxia induces the tumour cells to secrete angiogenic factors such as VEGF, platelet-derived growth factor (PDGF), fibroblast growth factor (FGF), angiostatin, endostatin and thrombospondins in a concentration gradient which drives angiogenesis.

iii. Invasion of tumour cells into the surrounding basal tissues or stroma; the tumour cells need various cell adhesion molecules such asintegrins, CD44, E-cadherin etc to adhere to the extracellular matrix. Matrix metalloproteinases, serine proteinases and cysteine proteinases help the tumour cells dissolve the extracellular matrix.

iv. Dissemination and aggregation of cancer cells in the blood stream. Survival of the disseminated cells in the blood stream

v. Adherence to blood vessel wall and invasion through the vessel wall into the basement membrane of the metastasis target tissue. Autocrine motility factors such as hepatocyte growth factor (HGF), insulin-like growth factor II (IGF-II) and autotoxin (ATX) stimulate the cells to start extravasation through blood vessel walls.

vi. Proliferation and angiogenesis of secondary tumour at the metastasis site.

1.6 Diagnosis of Breast cancer

Breast cancer is usually detected on the basis of particular symptoms or through the National Health Service breast screening program for women between the age ranges 47-73 (currently every three years). During 2009/2010, 15500 cases of breast cancer were diagnosed because the screening program (Cancer Research UK, 2012). Currently the process of breast cancer diagnosis involves “triple assessment”;

a) Physical examination
b) Radiological investigation- includes Mammography and Ultrasound
c) Pathological investigation- involves tumour biopsy collection through fine needle aspiration or excision biopsy.

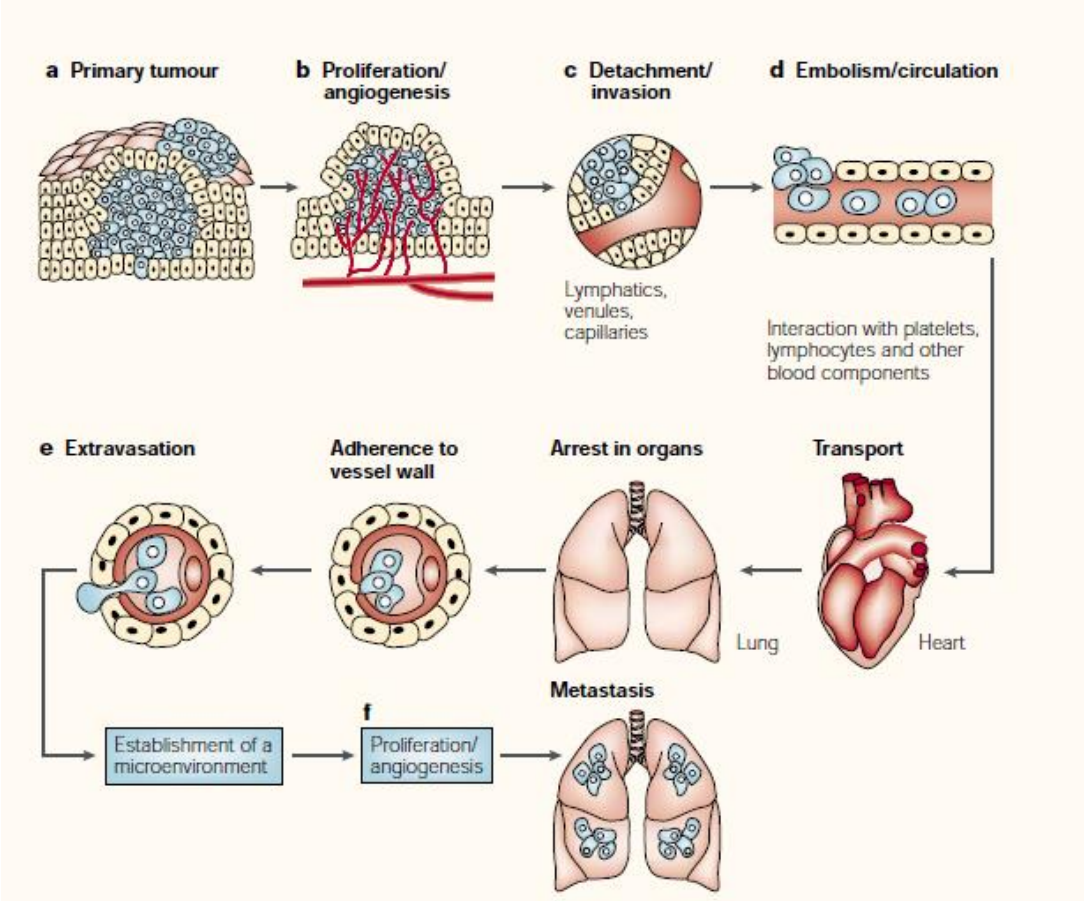


Figure 1.3 Development of cancer metastasis: Cancer invasion and metastasis takes place in different steps. The most important processes are formation of blood vessels in the primary tumour and extravasation through the blood vessel wall at the site of metastasis (taken from Fidler 2003).

1.7 Breast Cancer Biomarkers

1.7.1 Biomarker overview

Breast cancer grows to a critical mass by the time of clinical diagnosis and becomes difficult to treat. Therefore there is a need for biomarkers to screen, predict and identify treatment as early as possible. Biomarker is a defined as a "cellular, biochemical, biological characteristics or molecular alterations that can be objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention" (Naylor, 2010).

There are three types of biomarkers 1) Diagnostic biomarker- identification of preclinical stage of disease and disease state. 2) Predictive biomarker- used for forecasting the likely response to treatment. 3) Prognostic biomarker- used for predicting the course of the disease irrespective of treatment (Buyse et al., 2010). The sources of cancer biomarkers can be classified into two types; 1) Cellular Biomarkers- markers that are associated with the cancer tissue, 2) Humoral biomarkers- found in body fluids such as blood, urine, saliva, nipple discharge etc.

For the past two decades there have been advances in genomics, proteomics and molecular pathology and tremendous efforts have been taken to identify novel cancer biomarkers to be used in clinical practice. These are exemplified in the number of the grants (4928) and publications (441,510) up to 2008 relating to cancer biomarkers (Ptolemy and Rifai, 2010). However there is a discrepancy in the effort to discover cancer biomarkers and the number of biomarkers actually approved by the US Food and Drug Administration (Füzéry et al., 2013). Table 1.1 gives the list of approved cancer biomarkers which are currently in clinical use. It can be observed from the table that there only 6 biomarkers used in monitoring treatment response and disease progression of breast cancer. There are no approved biomarkers to be used in

screening and diagnosis of the disease before clinical symptoms begin to manifest. This shows the need for new breast cancer biomarkers that could be used for screening and diagnosis of breast cancer.

Table 1.1 US Food and Drug Administration-Approved Cancer Biomarkers

Biomarker	Clinical use	Cancer type	Specimen	Methodology	Year approved
Pro2PSA	Discriminating cancer from benign disease	Prostate	Serum	Immunoassay	2012
ROMA (HE4+CA-125)	Prediction of malignancy	Ovarian	Serum	Immunoassay	2011
OVA1 (multiple proteins)	Prediction of malignancy	Ovarian	Serum	Immunoassay	2009
HE4	Monitoring recurrence or progression of disease	Ovarian	Serum	Immunoassay	2008
Fibrin/ fibrinogen degradation product (DR-70)	Monitoring progression of disease	Colorectal	Serum	Immunoassay	2008
AFP-L3	Risk assessment for development of disease	Hepatocellular	Serum	HPLC, microfluidic capillary electrophoresis	2005
Circulating Tumor Cells (EpCAM, CD45, cytokeratins 8, 18+, 19+)	Prediction of cancer progression and survival	Breast	Whole blood	Immunomagnetic capture/ immune-fluorescence	2005
p63 protein	Aid in differential diagnosis	Prostate	FFPE tissue	Immunohistochemistry	2005
c-Kit	Detection of tumors, aid in selection of	Gastrointestinal	FFPE tissue	Immunohistochemistry	2004

	patients	stromal tumors			
CA19-9	Monitoring disease status	Pancreatic	Serum, plasma	Immunoassay	2002
Estrogen receptor (ER)	Prognosis, response to therapy	Breast	FFPE tissue	Immunohistochemistry	1999
Progesterone receptor (PR)	Prognosis, response to therapy	Breast	FFPE tissue	Immunohistochemistry	1999
HER-2/neu	Assessment for therapy	Breast	FFPE tissue	Immunohistochemistry	1998
CA-125	Monitoring disease progression, response to therapy	Ovarian	Serum, plasma	Immunoassay	1997
CA15-3	Monitoring disease response to therapy	Breast	Serum, plasma	Immunoassay	1997
CA27.29	Monitoring disease response to therapy	Breast	Serum	Immunoassay	1997
Free PSA	Discriminating cancer from benign disease	Prostate	Serum	Immunoassay	1997
Thyroglobulin	Aid in monitoring	Thyroid	Serum, plasma	Immunoassay	1997
Nuclear Mitotic Apparatus protein (NuMA, NMP22)	Diagnosis and monitoring of disease (professional and home use)	Bladder	Urine	Lateral flow immunoassay	1996
Alpha-fetoprotein (AFP)	Management of cancer	Testicular	Serum,	Immunoassay	1992

			plasma, amniotic fluid		
Total PSA	Prostate cancer diagnosis and monitoring	Prostate	Serum	Immunoassay	1986
Carcino-embryonic antigen	Aid in management and prognosis	Not specified	Serum, plasma	Immunoassay	1985
Human hemoglobin (fecal occult blood)	Detection of fecal occult blood (home use)	Colorectal	Feces	Lateral flow immunoassay	1976

1.7.2 Blood based breast cancer biomarkers

Blood based biomarkers could be proteins in the serum/plasma, circulating tumour cells, circulating MicroRNAs (miRNAs), DNA, metabolites etc. Classical soluble breast cancer biomarkers include Cancer Antigen 15-3 (CA 15-3), Cancer Antigen 27-29 (CA 27-29) and carcinoembryonic antigen (CEA). Both CA 15-3 and CA 27-29 are members of the mucin-like membrane protein 1 (MUC1) family. The members of MUC1 family are large glycosylated molecules and their physiological functions are unclear but they have been implicated in cell adhesion and metastasis (Duffy, 2006). Both CA 15-3 and CA 27-29 are used for predicting prognosis after treatment in primary breast cancer. CA 15-3 along with alkaline phosphatase predicts breast cancer recurrence and metastasis (Keshaviah et al., 2007). The rise in concentrations of CA-15-3 occurs immediately after treatment. Decrease in the level of CA 15-3 by 50% indicates suppression of tumour growth (Duffy, 2006). CEA is a secreted glycoprotein which is a tumour associated antigen and serum tumour biomarker. CEA is elevated in 50-60% of patients with metastatic breast cancer and hence is used to predict tumour metastasis. It is also used to monitor the therapy in breast cancer (Ebeling et al., 1999).

Other well known soluble biomarkers in the serum/plasma include cytokeratin 18, cytokeratin 20 (Moll, 1994), mammoglobin (Bernstein et al., 2005), MicroRNAs (miR-10b, miR-21, miR-125b, miR-145, miR-155 miR-191 and miR-382) (George and Mittal, 2010; Mar-Aguilar et al., 2013) are in the development pipeline which are not yet applied in clinical diagnosis.

1.8 Microarray expression profiling and biomarker discovery

1.8.1 Overview of Microarray Technology

Microarray is a technique used to genotype cells for identifying mutations or expression profile different types of cells. DNA microarrays are created by robotic machines that arrange thousands of small DNA sequences on a microscopic slide. These DNA sequences are complementary to the gene sequences in the cell. When DNA is extracted from the cell it is fragmented and labelled with a fluorescent probe. These small gene sequences are then allowed to hybridise on the DNA-chip. When the labelled DNA sequences bind to the chip the fluorescent signal can be detected and quantified. Expression microarrays use the sample principle but the RNA from the cells are extracted and converted to cDNA using reverse transcriptase enzyme. The cDNA is then allowed to bind to the DNA-chip (Forster et al., 2003). There are several sequence probes corresponding to every gene known and this depends on the manufacturer. The fluorescent signal is then detected using a high resolution camera. The image is processed by quantifying and normalising the signal/noise ratio. The expression values for each probe corresponding to a single gene are obtained. Different DNA-chips are normalised by using a combination of housekeeping genes that usually include genes such as GAPDH, β-Actin, 18s rRNA, 28s rRNA etc. Expression microarrays can be used to identify changes in disease state and compare it with healthy or normal state. The first cDNA expression microarray was done by Augenlicht et al. 1987 and it started to gain popularity with the publication by Ron Davis and Pat Brown labs at Stanford University (Schena et al., 1995). Microarrays has several limitations and drawbacks such as; reliance on the existing knowledge of the genome and isoforms of transcripts, high background levels on the DNA-chip due to cross hybridisation, lack of rigorous standards for data analysis and quality/ integrity of RNA used.

1.8.2 Genomics in biomarker discovery

Expression microarray technology and RNA sequencing technology are conventionally used to identify the expression profiles of cancer tissues and cancer cell lines to classify tumours based on their genotype and expression. This approach is also used to discover cellular biomarkers from the only the cancer tissue, recently from circulating tumour cells (Moll, 1994) and MicroRNAs (George and Mittal, 2010; Mar-Aguilar et al., 2013). There have been a number of studies aimed at discovering biomarkers for breast cancer from tissues. Perou et al. 2000 used cDNA microarrays to test tissue samples from 42 breast cancer patients for the expression of around 8000 genes. This was the first study which was done to classify breast tumours into different subtypes based on expression patterns. Sørliea et al. 2001 confirmed the findings and demonstrated that gene expression signatures of these subtypes of cancer can be used as a prognostic marker with respect to overall and relapse free- survival. Five types of breast cancer were identified based on the expression patterns; Luminal A, Luminal B, Basal-like, ERBB2 positive/ER negative and normal breast-like. Various research groups have identified gene signatures which correlate with prognosis and outcome in breast cancer patients (Abba et al., 2010; Glinsky et al., 2004; Van't Veer et al., 2002; Yu et al., 2008b). Recent work done by Curtis et al. 2012 has revealed novel molecular subgroups of breast cancer based on genomic and transcriptomic experiments. The group analysed copy number, single nucleotide polymorphisms, and gene-transcription rates in 2,000 breast cancers and identified 10 different molecular subtypes of breast cancer based on patient treatment outcomes. The findings are yet to be implemented in clinical settings. The main disadvantage of genomic profiles to be used as prognostic biomarkers is the heterogeneity of cancers, expensive/time consuming method to be used in clinical settings, need for large randomised clinical trials, lack of specificity and sensitivity of individual genes and the RNA levels do not always correspond to protein levels.

Recently several groups have tried to discover biomarkers by expression profiling in the patients WBCs of breast cancer (Aarøe et al., 2010; Komatsu et al., 2012; LaBreche et al., 2011; Sharma et al., 2005; Zuckerman et al., 2013), pancreatic cancer (Baine et al., 2011), Non-small cell lung cancer (Showe et al., 2009), colorectal cancer (Xu et al., 2013a) and acute myeloid leukemia (Metzeler et al., 2008). All these groups have performed expression analysis and come up with panels or expression patterns to predict cancer but none except one group (Xu et al., 2013a) have performed validation by RT-qPCR. Since clinical tests for cancer use protein antibody based techniques such as ELISA, IHC and FISH, microarray data has to be validated so that it could be taken forward to the validation step. The difficulties in validation of biomarkers will be discussed later in Section 1.10.

1.9 Proteomic profiling using Mass Spectrometry and biomarker discovery

Mass spectrometry based proteomics can be used for the discovery step of biomarkers to identify a huge list of potential candidates which can be shortlisted in the following steps of development. Proteomics complements and further enlarges the information generated by genomics in breast cancer. But the mRNA levels do not necessarily correlate with corresponding protein abundance. The most commonly used technologies are tissue protein microarrays, 2D-gel electrophoresis and mass spectrometry based techniques such as matrix assisted laser desorption/ionisation (MALDI),Surface Enhanced Laser Desorption/Ionization time-of-flight mass spectrometry (SELDI) and liquid chromatography (LC-MS/MS). In this study we have used 2D-gel electrophoresis and mass spectrometry using the LTQ Orbitrap hybrid ion trap mass spectrometer.

1.9.1 Overview of 2-Dimensional Polyacrylamide Gel Electrophoresis (2D-PAGE)

2D-PAGE is a classical technique used to separate a cell lysate based on mass and charge, resulting in up to thousands of spots per gel. It can be used in tissue, plasma and serum

proteome analysis without prior enrichment or fractionation. The proteins are separated in two separate steps; in the first dimension proteins are solubilised in urea and are separated based on their isoelectric point on a pH gradient. Proteins will migrate until they have no net charge. After the proteins are separated based on their charge they are then loaded onto a polyacrylamide gel containing sodium dodecyl sulphate (SDS) under an electric field. The proteins get denatured and acquire a negative charge due to the presence of SDS. The proteins migrate under the electric field based on their molecular mass (Lopez, 1999; O'Farell, 1975). Thus proteins are separated in separate protein spots on an acrylamide gel with high resolution which can be stained. The protein spots are excised out of the gel and digested with proteases before mass spectrometric identification.

1.9.2 Overview of Mass spectrometry and LTQ Orbitrap hybrid ion trap mass spectrometer

Mass spectrometry has revolutionised the analysis of proteins and has become the tool of choice in biomarker discovery due to its speed, wide signal range, quantitative capability and the facility to interface with chromatographic methods. Time-of-flight mass spectrometry (ToF MS) is where the protein peptides are ionised and the ions are accelerated by an electric field of known strength. The mass analyser then separates ions according to their mass to charge ratio and by time taken for the ions to travel through the field (Boja et al., 2011). The data is then analysed by the application of algorithms to compare with the theoretical spectra of proteins in protein sequence databases. This mass analyser is usually coupled with MALDI, SELDI or liquid chromatography apparatus.

MALDI involves precipitation of proteins along with a matrix material (cyano-4-hydroxycinnamic acid or dihydroxybenzoic acid). The precipitated protein is then fragmented and ionized with laser pulses which are fed into a ToF MS. SELDI involves the protein sample to

be loaded onto a ProteinChip arrays that selectively bind to proteins in the lysate according to the protein chemistries based on adsorption or affinity chromatography (Srinivas et al., 2002). The chip is then washed to avoid unspecific binding. LC-MS/MS combines liquid chromatography and mass spectrometry. After enzymatic digestion the mixture of peptides are resolved in a chromatographic column. The peptides are eluted at different rates due to their interaction with the column material. The separated peptides are then fed into a mass spectrometer (Khadir et al., 2013).

The Orbitrap has an ion trap mass analyzer consisting of an outer barrel-like electrode and a co-axial inner spindle-like electrode that traps ions in an orbital motion around the spindle. The ions move inside in complex spiral patterns. The ions are then detected as an image current on the two halves of an electrode surrounding the orbitrap. Fourier transform is used obtain oscillation frequency for ions with different masses which gives the mass to charge ratio (Hu et al., 2005; Scigelova and Makarov, 2006). The Orbitrap mass spectrometer has very high resolution and high detection speeds of up to 5000 measurement per second per 5ppm. It also has very high sensitivity towards proteins of low concentrations in complex biological mixture.

1.9.3 Proteomics in breast cancer biomarker discovery

The advancement in new mass spectrometric techniques coupled with protein fractionation techniques has expanded the possibility of highthroughput protein identification and quantification in complex biological mixtures such as plasma, urine and cell lysates. The current FDA approved clinical tests are protein based antibody assays (Table 1.1). There are several candidate protein biomarkers such as cathepsin-D, 14-3-3 proteins, HSP60, PCNA, annexins, Calreticulin, matrix metalloproteinases etc that have been identified by mass spectrometry in breast cancer tissues (Hondermarck et al., 2001). There are no FDA approved biomarkers which was discovered by mass spectrometry based approaches (Drabovich et al.,

2014). Biomarkers have been searched for in the plasma of cancer patients based on the hypothesis that tumour specific proteins maybe found in blood circulation and would have use in the early detection of cancer.CA 15-3 and CA 27-29 are FDA approved biomarkers of breast cancer to monitor disease progression and recurrence in the plasma. The main utility of both of these molecules are to monitor plasma levels in metastatic breast cancer patients undergoing therapy (Misek and Kim, 2011; Nolen et al., 2008). Early periodic peaks in CA 15-3 and CA 27-29 levels have been reported during the therapy but these are not recommended for screening or diagnosis after primary therapy. Novel plasma biomarkers which have been discovered include apolipoprotein H (Chung et al., 2014) and 21-protein panel in metastatic breast cancer patients (Carlsson et al., 2011). There are several more plasma biomarkers which are in the development pipeline and have to be validated using clinical trials. The main drawbacks of protein biomarkers in plasma are dynamic changes in the blood, blood collection conditions, blood processing/ storage methods and detection of proteins of low concentrations.

1.9.4 Biomarker Development Pipeline

Similar to drug discovery and biomarker discovery is a multi-step process. The aim of the pipeline is to assess the maximum possible candidates and exclude the ineffective ones as early as possible. Figure 1.3 gives the steps in the protein biomarker development pipeline describing the timeline and the approach usually taken for each step. The first step can also be called the discovery phase involves using clinical samples and currently there are two types of approaches; Exploratory approach where highthroughput methods are used and targeted approach where usually molecules overexpressed in cancers are measured by targeted proteomics. The exploratory approach consumes more money and time. It took nearly 8 years since the discovery of human epididymis protein 4 for ovarian cancer and interleukin 1 receptor-like 1 protein for heart failure to be approved by the FDA for clinical use (Drabovich et al., 2014). The qualification step involves analysing data from the first step to identify association between

clinical outcome and protein abundance by using criteria filtering to obtain a manageable list of biomarkers. The verification or validation step includes measuring the promising candidates using semi-quantitative methods like RT-qPCR, western blotting and Immunohistochemistry where false positive candidates are excluded. The pre-clinical step typically involves protein expression and purification of the biomarkers to develop antibodies against them. The clinical stage usually involves validation on clinical samples in large sample numbers (usually 1000s). It is in this stage reference values, clinical endpoints and surrogate endpoints are defined. Clinical endpoint is the defined by how the patient feels or survives and surrogate endpoint is defined as the substitute for clinical endpoint to include it and prediction of effects of treatment on the clinical endpoint (Buyse et al., 2010). A number issues such as deciding the number of samples to be analysed, criteria for inclusion or exclusion for sample and candidates, appropriate statistical analysis to use for both discovery phase and clinical phase, sample collection methods, effects of sample preparation, limitations of validation methods, validation using independent datasets/investigators, time taken for biomarker discovery process causing clinicians dissatisfaction and the cost of the process should be considered while developing a biomarker.

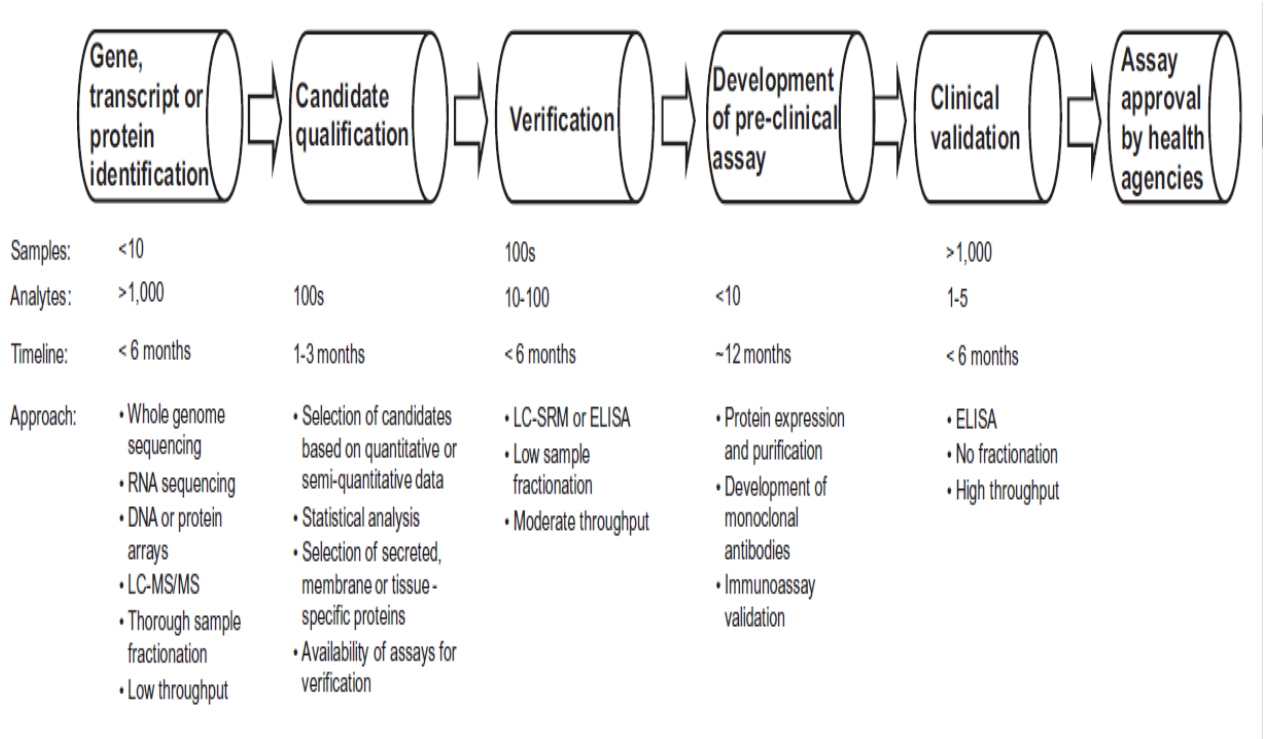


Figure 1.4 Integrated protein biomarker development pipeline: Both genomics and proteomics approach is combined and the biomarker candidates are validated using immunoassays. The numbers of candidates decrease and the number of samples increase along the pipeline (taken from Drabovich et al. 2014).

1.10 Cancer and the Immune system

1.10.1 Immune Surveillance and Immunoediting

The immune surveillance theory explains that cancer cells constantly originate in the body but are eliminated by the immune system as antigens. The tumour which manifests itself either has to escape this or impair the immune system. This theory then evolved into the cancer immunoediting hypothesis which suggests that the immune system sculpts the tumour during its development and, paradoxically, results in the generation of cancer cells resistant to the immune system. Cancer immunoediting consists of three stages: elimination, equilibrium and escape (Dunn et al., 2002). During the first phase the immune system complements the immune surveillance hypothesis where the immune system eliminates the tumour cells based on their antigenic properties. The equilibrium phase occurs when tumour cells with reduced immunogenicity and increased immunosuppressive properties evolve. This is achieved when tumours secrete immunosuppressive factors, which downregulate antigen presenting molecules or inhibit of immune cell homing (Chouaib et al., 1997). Tumour escape occurs when the tumour secretes immunosuppressive factors such asvascular endothelial growth factor (VEGF), IL-10, IL-6, TGF-b, prostaglandin E2, soluble phosphatidylserine, granulocyte macrophage colony stimulating factor, tumour necrosis factor-α, soluble Fas and soluble Fas ligand (Kim et al., 2007).

1.10.2 Role of the Innate Immune system in promoting tumour development and angiogenesis:

Approximately 15% of the cancers worldwide are caused by chronic infections. For example Hepatitis C increases the risk of liver carcinoma whereas chronic infection of *Heliobacter pylori* is a leading cause for stomach cancer (Cousens & Werb, 2002; De Visser, Eichten, & Coussens, 2006). There is a growing body of evidence suggesting the link between

inflammation and tumour development. The immune cells cause inflammation in the process of destroying an infection and produce cytokines which subject the surrounding cells to oxidative stress thereby damaging them. The damaged cells have a higher risk of developing into a neoplastic cell.

Cell mediated immunity or innate immunity is the first line of defence of the body against tumour cells and comprise of macrophages, dendritic cells, mast cells, granulocytes, natural killer cells (NK cells) and cytotoxic T-cells (CTLs). Reduced NK cells activity and reduced number of T-lymphocytes were observed in breast cancer patients prior to surgery or adjuvant therapy. One of the major inflammatory cytokine secreted by macrophages TNF-α has also been found to be secreted by breast cancer cells in high levels. Also the activity of TNFα was 100% higher in breast cancer patients (Standish et al., 2008). TNF-α expression positively correlated with increased tumour grade and lymph node metastasis (Leek et al., 1998). The expression of class I MHC is reduced in tumour cells therefore the cytotoxic T-cells cannot recognise the tumour cells as they are dependent on MHC I expression. Tumour tissues from breast cancer and prostate cancer have abundant infiltrations of immune cells and immunoglobulin deposits compared to normal tissue (De Visser et al., 2006). The density of tumour associated macrophages in the tumours has been shown to be correlated with poor prognosis and lymph node metastasis. Also tumour associated macrophages secrete other inflammatory cytokines such as IL-8, IL-10, VEGF and Fibroblast growth factor to inhibit T cells response (Crowther et al., 2001; Gabrilovich et al., 2012).

Neutrophils comprise of a large subset of the white blood cell population (~70%) and are important in the innate immune response. Neutrophils are activated by compounds such as lipopolysaccharide, Interleukin-8, Interferon-γ, and TNF-α. Activated neutrophils secrete a growth promoting factor Hepatocyte Growth Factor (HGF) (McCourt et al., 2001). Interleukin-8, one of the stimulators of neutrophils is secreted in the cancer microenvironment and increases

the population of tumour infiltrating neutrophils in the tumour site (Waugh and Wilson, 2008). HGF was found to be produced by Tumour infiltrating neutrophils in response to granulocyte macrophage colony stimulating factor and TNF-α in pulmonary adenocarcinoma microenvironment (Wislez et al., 2003). HGF promotes metastasis and is a growth factor which improves angiogenesis and cell proliferation. In another study *c-met* gene which codes for the receptor for HGF is upregulated in colorectal cancer cells (Di Renzo et al., 1995). Recently it was proved that HGF was necessary to disrupt cell to cell interactions by uncoupling the myosin VI and E-cadherin function accompanied by loss of F-actin at cell surface junctions of zonula adherens (Mangold et al., 2011). Oncostatin M, a cytokine belonging to the IL-6 family was secreted by tumour infiltrating neutrophils in breast cancer cells. It improves the invasive capacity of the tumour and metastasis. When neutrophils were co-cultured with human breast cancer cell lines T47D and MDA-MB-231 oncostatin-M was secreted and macrophage colony stimulating factor was necessary for it (Queen et al., 2005). Thus neutrophils play a very important part in the immune response to cancer and may promote tumour progression and metastasis. It has been showed that TGF-β polarises tumour associated neutrophils to a N2 pro-tumour phenotype. Blocking of TGF-β causes accumulation of N1 phenotype neutrophils which possess anti-tumour activity (Fridlender et al., 2009). The function of neutrophils in tumour microenvironment is still unclear as to whether they promote or suppress tumour development and growth.

1.10.3 Cancer and the Chemokine/cytokine Network

The chemokines and cytokines secreted by tumour cells have been implicated in the suppression of the cells if the immune system. Two cytokines, CCL2 and CCL5, secreted by various tumours have been studied extensively. CCL2 otherwise known as MCP-1, is secreted by tumour and stromal cells into the surrounding attracting monocytes, basophils and T-cells to sites of inflammation. CCL2 levels correlate with lymphocyte and macrophage numbers in the

epithelial areas of the tumour along with poor prognosis and relapse (Balkwill, 2004). CCL5 secreted by tumour cells, attracts T-cells, basophils and eosinophils; the CCL5 levels were high in advanced breast carcinomas (Ben-Baruch, 2003). Both CCL2 and CCL5 facilitate the production of matrix metalloproteinases such as MMP9 in macrophages which modify the extracellular matrix to aid tumour cell migration and invasion(Robinson et al., 2002). Another chemokine implicated in tumour development is CXCL12 (SDF1-α). CXCL12 is expressed by sites such as the bone marrow, liver, lung, brain etc which are metastasis sites for breast cancer. Its receptor, CXCR4, has been found to be up-regulated in many types of tumours including breast cancer, allowing cells to migrate to targeted distant organs containing cells that secrete its ligand (de Jong et al., 1998; Müller et al., 2001)

1.11 Leukocyte Transendothelial Migration

It is very important for the immune cells reach the site of infection or inflammation through the blood vessel wall. Leukocyte transendothelial migration is the process by which T-cell, neutrophils and monocytes move through the endothelial layer. Transendothelial migration has also been implicated in tumour escape mechanisms. The whole process, illustrated in Figure 1.4, happens in three stages: rolling and capture, adhesion on the endothelium and finally transmigration (Van Buul and Hordijk, 2004; Muller, 2009; Wagner and Roth, 2000). Monocytes, lymphocytes and neutrophils all migrate in a similar way into the endothelium but are activated by different signalling molecules. Many molecules such as TNF-α, IL-1, Leukotrienes, Platelet activating factor, Complement protein 5a, IL-8, SDF-1α etc mediate the interaction between the endothelium and the leukocytes (Wagner and Roth, 2000).

1.11.1 Rolling and Capture

Polymorphonuclear Leukocyte Neutrophils (PMNs) roll over the endothelium forming reversible interactions using molecules called selectins which are present on both PMNs and

endothelial cells. The PMNs, newly released from the bone marrow, express high levels of L-selectin and as they age the L-selectin molecules are lost from the surface of PMNs (Kuhns et al., 1995) .The ligand for L-selectin is CD34, it is highly expressed in lymph nodes and endothelial cells of the blood vessel (Krause et al., 1996). Two other selectins of the endothelium, P-selectin and E-selectin also have a role in this rolling step and are expressed when appropriate inflammatory signals are received. The ligand for P-selectin on the surface of leukocytes is PSGL1 which can bind to two P-selectin molecules at the same time (Wagner and Roth, 2000). The interaction with PSGL1 is longer than the L-selectin-CD34 interaction but if appropriate inflammatory signals are not received these selectin interactions are reversible and short-lived.

1.11.2 Adhesion and Activation

The next step in transendothelial migration is firm adhesion which happens due to either appropriate inflammatory signals or the activation of selectin interactions. Intracellular adhesion molecules and integrins play a very important part in the adhesion of PMNs onto the endothelium. Integrins are cells adhesion receptors which bind to a variety of ligands on the extracellular matrix such as collagen, fibronectin, fibrinogen, laminins, cadherins, vitronectins etc (Barczyk et al., 2010). Integrins are heterodimeric proteins found the membranes of most cells. There are 16 different α subunits and 8 β sub-units to form 23 different transmembrane Integrin receptors. Each Integrin is formed by one α and β sub-units with the cytoplasmic tail forming linkages with the cytoskeletal proteins and harbouring phosphorylation sites. The main integrins implicated in leukocyte migration are Mac-1(αMβ2), LFA-1(αLβ2) and VLA-4 (α4β1).VLA-4 and LFA-1 are responsible for the movement for T-cells and monocytes while Mac-1 is predominantly responsible for neutrophil and NK cell movement. The receptors for Mac-1, LFA-1 and VLA-4 are C4a:C4b complement, ICAM-1 and VCAM-1 respectively. Both ICAM-1 and VCAM-1 interactions are essential for neutrophil invasion. PECAM-1 and CD99 is also essential for the neutrophil movement.PEACAM-1 is its own ligand forming a homodimer

on the opposing cell perhaps essential for signalling the cell to inform the degree of extravasion (Wagner and Roth, 2000).

1.11.3 Transmigration

Leukocytes possess proteolytic enzymes such as neutrophil elastase, proteinase 3 and cathepsin G which can cleave components of the extracellular matrix such as collagen and laminins. This step involves the actin cytoskeletal rearrangement of the PMN cell. The neutrophil proteases and peroxidases disrupt the VE-cadherin junctions of the endothelial cells to loosen the cell junctions for the movement of PMNs. This involves the reorganisation of the cytoskeletal structure of the PMN so it has a leading edge that adheres and protrudes into the endothelium and the retracting end. In fibroblasts the integrin binding activates Cdc42 and Rac1 which intern activates the Arp2/3 complex. The molecules involved in the retracting end are CD44, CD43, ICAM-3 and PSLG-1 (Van Buul and Hordijk, 2004).

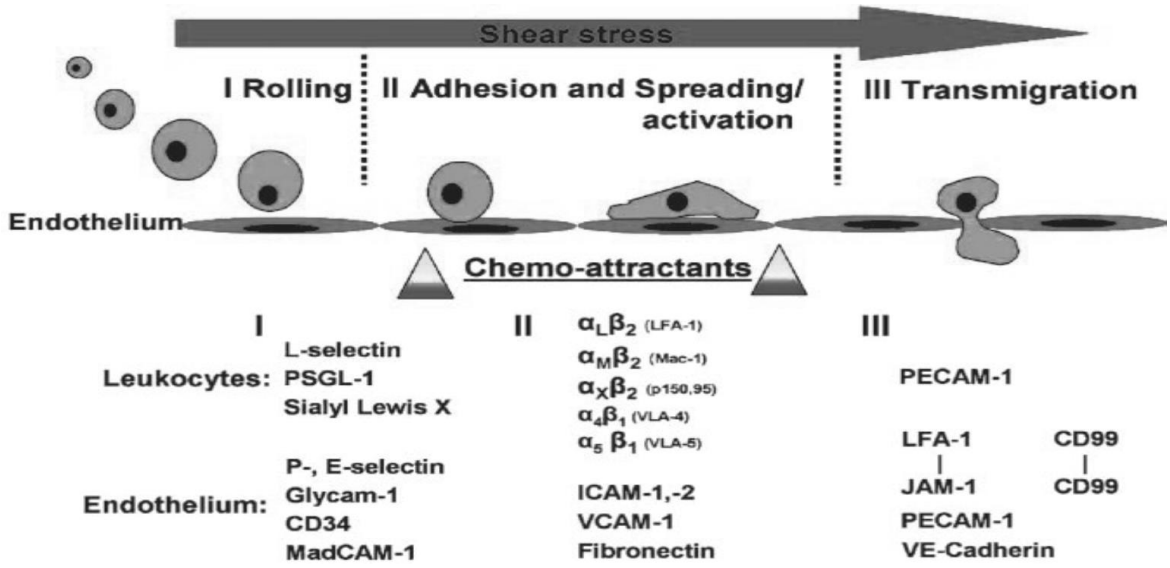


Figure 1.5 Overview of Leukocytes Transendothelial Migration: Three steps involved leukocyte migration along with the molecules involved. (Taken from Van Buul & Hordijk, 2004)

1.12 Aims of the project

The need for new biomarkers has been discussed in the previous sections. In this project the primary aim is to discover novel breast cancer biomarkers in the **White Blood Cells (WBCs)** fraction of blood of breast cancer patients using the exploratory biomarker discovery approach. The secondary aim is to use the findings to explain the effect of tumour on the immune system, specifically on the WBCs. To achieve this we divided the project into the following Objectives:

1. Identification of the difference in gene expression profiles of WBCs between healthy donors and primary breast cancer patients, using the existing data deposited in the Gene Expression Omnibus (GEO).
2. Identify the difference in protein profiles of WBCs of healthy donors and primary breast cancer patients by highthroughput proteomics and 2D-PAGE.
3. Integrate the data from both genomic and proteomics approaches to shortlist the biomarker candidates.
4. Validate shortlisted biomarker candidates using RT-qPCR and Western blotting.

The expected outcome of this study would be identification of new breast cancer biomarkers. In addition, this investigation will add to our knowledge of how the immune system is altered in the breast cancer patients.

Chapter 2: Materials and Methods

2.1 Isolation of White Blood Cells from whole blood

Ethical approval for this project was obtained from Essex Research Ethics Committee (Reference No: 10/H0301/60) and the Research & Development Department of Colchester Hospitals University NHS Foundation Trust. Blood samples of Breast cancer patients were obtained from Colchester general hospital in collections tubes with Ethylenediaminetetraacetic acid (EDTA) and were processed within 24 hours of collection. Blood was processed to separate White Blood Cells (WBCs) using the buoyancy density method (D'Arcy et al., 2006). Histopaque (Sigma) was used to separate the whole blood into a density gradient of plasma, white blood cells and red blood cells by spinning at 400 x g for 30 minutes. Red blood cells were discarded and the plasma was stored at -80^{0}C. The white blood cells layer was taken and washed with Hanks balance salt solution (HBSS) twice and was spun down at 250 x g for 5 minutes. The remaining pellet was treated with red blood cells lysis buffer (from Roche) to remove the red blood cells and washed with HBSS to obtain the WBC's which was then split into different fractions to analyse them.

2.2 Protein analysis methods

2.2.1 Overview of Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Western blotting

SDS-PAGE is a method used to separate proteins in a mixture under the influence of an electric field in a polyacrylamide gel. Cells are lysed in a buffer containing SDS. The proteins are denatured and bind to SDS to obtain a negative charge. The proteins are then separated based on their molecular weight in a polyacrylamide gel (Chrambach& Rodbard 1971). Western blotting is a technique where the proteins which were resolved by SDS-PAGE are transferred to

a nitrocellulose membrane using an electric field. The membrane is then probed with a primary antibody raised specifically against an antigen followed by probing with secondary antibody labelled either a fluorescent tag or an enzyme which produces a luminescent signal in the presence of a substrate (Towbin et al., 1979). All the buffer and solution compositions used are given in Table 2.1

2.2.1.1 Preparation of White Blood Cell lysates for gel electrophoresis

The protein lysates were prepared from $2x10^6$ whole white blood cells in SDS lysis buffer containing 0.1M Tris/HCl pH 6.8, 7 M urea, 4% SDS, 10% mercaptoethanol and phenol red. The mixture was then heated at 95^0C for 5 minutes to allow the proteins to be reduced.

2.2.1.2 Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Standard 8.1% SDS-PAGE gel was used to separate the proteins in the lysates. The separating gel was prepared and poured into the glass cassettes. After the separating gel was ready Stacking gel was prepared and poured on the separating gel with a well comb. Once the gels were set, protein samples were loaded onto the wells. SDS-PAGE gels were run with a protein standard marker with a molecular weight range of 7 - 175 kDa (New England Biolabs) under denaturing conditions at 125 V and 40 mA (80 mA for two gels run simultaneously) for just over two hours. The gel was incubated in running buffer with 1% methanol for 15 minutes on a gentle rocking platform. The PVDF membrane was treated with methanol (100%) for 10 seconds and washed thoroughly with RO water.

2.2.1.3 Western blotting analysis

Protein blotting was done from the acrylamide gel onto polyvinylidenedifluoride (PVDF) membrane by electrotransfer.Transfer was performed using transfer buffer in a semi-dry electro blotting apparatus for two hours at100 mA / 35 V. The PVDF membrane was blocked overnight at 4^0C in blocking buffer (PBS, 0.05% tween-20 and 5% milk). The membrane was then

incubated with primary antibody at a concentration at the recommended concentration in blocking buffer for two hours and then washed with washing buffer (1XPBS, 0.05%tween-20) three times for five minutes. The membrane was incubated with the horse radish peroxidase (HRP) conjugated secondary antibody for two hours in blocking buffer followed by washing thrice with washing buffer. The membrane was then soaked with the enhanced chemiluminescence (ECL) solution Luminata forte from Millipore (cat no: WBLUF0100) containing fluorescent peroxidase substrate and visualized. The signal was detected by exposing the membrane to an autoradiography film (Kodak Medical X-ray film). The film was then developed using GBX Developer and Fixer (Kodak, Japan).

2.2.1.4 Image J analysis

The image obtained from the PVDF membrane was then scanned as an image file and analysed with Image J analysis software which can downloaded freely from http://imagej.nih.gov/ij/. This software analyses the band intensity and gives values. The target protein molecule was normalised between samples using the signal obtained from β-Actin antibody probing. Protein density values from different gels were normalised based on the value obtained from a common sample used.

A.

Buffer	Composition
SDS Lysis Buffer	0.1M Tris/HCl pH 6.8, 7M Urea, 10% 2 β-mercaptoethanol (add just before use), 4% SDS and a pinch of Phenol red dye
Tissue Lysis buffer	20mM Tris/Hepes pH 8.0, 2mM EDTA, 0.5M NaCl, 0.5% Na deoxycholate, 0.5% Triton X-100, 0.25 M Sucrose, add directly before use 50mM 2-ME, 50µM PMSF, 1µM Pepstatin
Resolving Buffer*	2M Tris/HCl, 0.2% SDS (pH 8.9)
Stacking Buffer**	0.1M Tris/HCl, 0.1% SDS (pH 6.8)
Gel Running Buffer	0.025M Tris/HCl, 0.192M Glycine, 0.1% SDS
Transfer Buffer	20mM Na_2PO_4, 2% Methanol, 0.05% SDS
Phosphate Buffered Saline (PBS) (1x)	2mM KH_2PO_4, 10mM Na_2PO_4, 137mM NaCl, 2.7mM KCl, (pH 7.4)
Blocking buffer	0.1% Tween, 5% dried skimmed milk powder, 1x PBS
Washing Buffer	0.1% Tween, 1x PBS

B.

Resolving gel composition	8.1% gel (ml)
Acrylamide/Bis-acrylamide Solution (30%) (Bio Rad)	2.7
Resolving gel buffer*	5
10% APS (SIGMA)	0.05
ddH_20	2.23
TEMED (SIGMA)	0.02
Total	**10**

Stacking gel composition	4% gel (ml)
Acrylamide/Bis-acrylamide Solution (30%)	0.665
Stacking gel buffer**	1.75
10% APS	0.025
ddH_20	2.55
TEMED	0.01
Total	**5**

Table 2.1 Preparation of buffers and gels for Western Blotting. (A) Composition of general buffers used in the SDS-PAGE and Western Blotting.(B) Components of Resolving (left) and Stacking (Right) Gels used in SDS-PAGE

2.2.2 Protein Profiling by Mass spectrometry

2.2.2.1 Protein profiling of White blood cells using Orbitrap mass spectrometer

To identify the protein expression patterns specific for cancer patients' proteins from WBC cell pellets were fractionated into membrane and soluble fractions as discussed in Alldridge et al., 2008. The proteins from the leukocyte fractions were extracted and the soluble and membrane proteins were separated to generate two separate proteome fractions to maximize the coverage of the analysis. Both the fractions were then digested using trypsin and analysed on OrbitrapVelos instrument equipped with an Ultimate 3000 nano-scale HPLC (Olsen et al., 2009). The data generated by Orbitrap was processed by MaxQuant (a suite of algorithms and tools to perform automated analysis using the raw files from Orbitrap). The software has been accepted and adopted widely by the quantitative proteomics community (Cox and Mann, 2008).The above work was conducted by Dr.MetodiMetodiev in his lab. The output of the procedure was obtained as table with normalised label free quantitative intensity values for each protein. Only the data from only the membrane fraction was analysed. Mann Whitney U test was then performed to compare the protein profiles between the WBCs of cancer patients and healthy donors with a p values of <0.05. The proteins were then shortlisted based on their expression profiles manually.

2.2.3 Protein profiling using 2D gel electrophoresis

2.2.3.1 Protein profiling of White blood cells using 2D Gel electrophoresis

2.2.3.1.1 Sample preparation for 2D Gel electrophoresis

The protocol used for protein lysate preparation and focusing the samples on pH strips was developed in our lab by Dr. Dawn Farrar. Frozen cell pellets were dissolved in a lysis buffer

containing 1%SDS and 100mM Tris-HCl pH7 preheated to 95^0C. The sample was then sonicated using a bioruptor and 5% 2β-mercaptoethanol was added. The samples was dissolved in solubilisation buffer (8M urea, 2.5M thiourea, 4% 3-[(3-cholamidopropyl) dimethylammonio]-1 –propanesulfonate CHAPS, 50mM DTT and 24mM sperminetetrachloride. The proteins were solubilised for 1 hour at room temperature following the further addition of 24mM of spermine. The sample was centrifuged at 12000g for 30 minutes. The solubilised protein sample was divided into two and 1ml ice cold acetone was added and incubated overnight at 20^0C followed by centrifugation at 12000 g for 30 minutes. The protein pellet was then washed with ice cold methanol-chloroform mixture and resuspended in isoelectric focusing (IEF) buffer (8M urea, 2.5M thiourea and 4% 3-[(3-cholamidopropyl) dimethylammonio]-1 –propanesulfonate CHAPS). The sample protein concentration was quantified using CooAssay.

2.2.3.1.2 Separation in first dimension using Isoelectric focussing (IEF)

55µg of protein lysate was dissolved with IEF buffer containing 1µl IPG buffer and 2.4µl destreak reagent to a total volume of 202µl. IPG strips (pH3-10) from BioRad were rehydrated with the above sample mixture overnight according to manufacturer's protocol. The IPG strip was then focused on IPGphor system with the following conditions: 250V-30 minutes, 500V-15 minutes, 500V-3.25hrs, 1000V-1hr, 6000V-2hrs, 6000V-2hrs, and 200V-12hrs. The strips were then incubated in equilibrium buffer (6 mol/L urea, 20% w/v glycerol, 4% w/v SDS, and 0.375M/L Tris-HCl (pH 8.8)) with 5% 2β-mercaptoethanol and then with equilibrium buffer with 2.5% iodoacetamide for 15 minutes each.

2.2.3.1.3 Second dimension separation using acryl amide gel electrophoresis and silver staining of proteins

The second dimensional electrophoresis was carried out in Criterion gel system (BioRad). 12% criterion precast gels (cat no: 345-0121) were used to perform the electrophoresis. The running buffer used was. The gels were fixed overnight in 10% methanol

and 10% acetic acid. The gel was stained using the silver staining protocol from Shevchenko, et al., 1996. Each sample was prepared and run as triplicates. The 2D-gels were scanned and analysed using SameSpots analysis software.

2.2.3.1.4 Protein spot analysis using Samespots Software

Each gel was scanned using the ImageScanner scanner from GE Healthcare. The scanned images were then uploaded into the Samespots software. The spots on the gel were then identified by the software. Background subtraction and normalization of the spots were then performed. The gels were then grouped for differential analysis based on tumour grade or disease state. The spots were compared between gels based on volume of the spot. ANOVA test (p-value< 0.05) was performed was used to identify spots that were differentially present in primary breast cancer patients. The spots which were differentially present were cut from the gel. The spots were then destained and prepared using a compatible protocol for mass spectrometry (Shevchenko et al., 1996). The spots were sent to Dr. Metodi Metodiev's lab for identification by mass spectrometry.

2.3 Methods for RNA extraction and analysis

2.3.1 RNA extraction and quality control

Total cell RNA was extracted from WBC's for microarray analysis using trisure phenol chloroform extraction method. Due to difficulties experienced like low final RNA concentration and low purity of RNA samples when extracting RNA from blood samples the conventional phenol-chloroform RNA extraction protocol was modified to obtain high quality RNA without any degradation. This protocol has been described in figure 2.1. The DNA digestion proved to be essential since the contaminating DNA in the RNA sample would interfere the microarray process and which was seen when the residual DNA was amplified when PCR was done on the cDNA produced from the RNA sample. Cells were stored in RNAlater at -80^0C until they were

processed. The cells were lysed with 1 ml of trisure per 10 $X10^6$ cells. 200µl of chloroform was the added and the tubes were vortexed for 15 seconds. The samples were then centrifuged at 12000 xg for 15 minutes and the top aqueous phase was then transferred to a new tube. The chloroform step was then repeated again by adding 200 µl of chloroform. RNA was precipitated from the aqueous phase for 10 minutes using 500µl of isopropanol per 1 ml of trisure used. The samples were then centrifuged at 12000 xg for 10 minutes. The supernatant was then discarded and pellet was washed twice with 1ml of 75% ethanol at 7500 xg for 5 minutes each. The resultant RNA pellet was then dissolved in sterile water by heating it at 55^0C for 5 minutes.

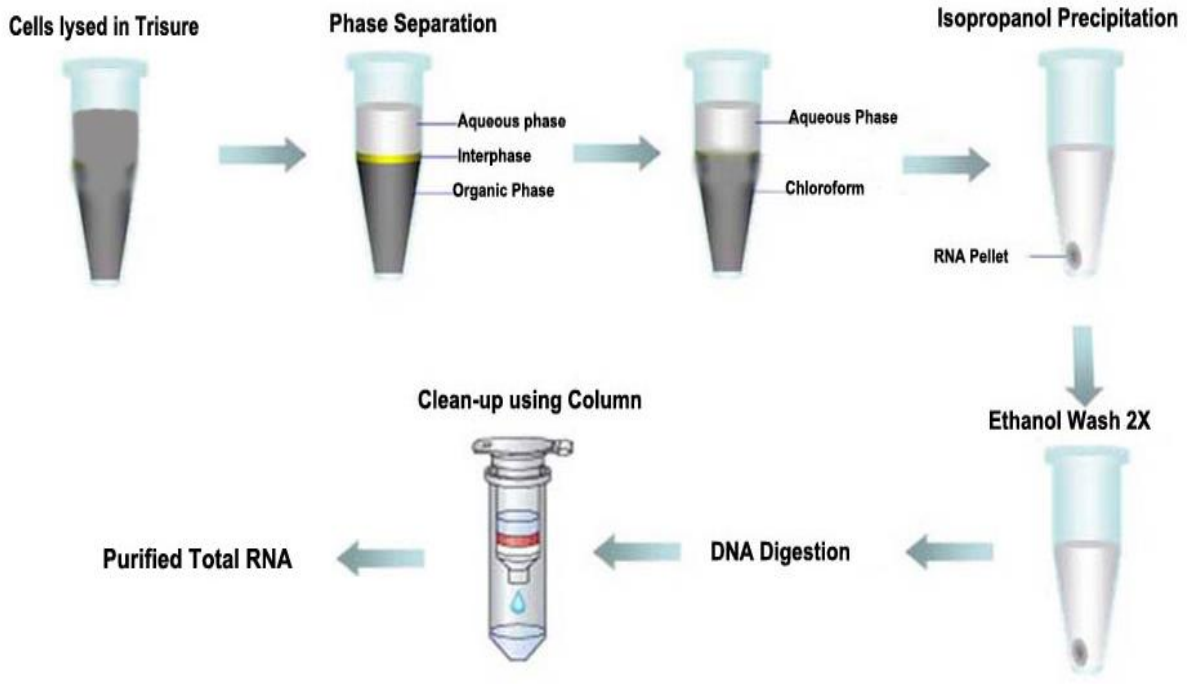


Figure 2.1 Modified RNA extraction protocol for White Blood Cells: The conventional Trisure phenol chloroform protocol was modified to extract RNA form white blood cells. The modified protocol includes an extra chloroform step and ethanol washing step. DNA digestion was also done to ensure the purity of RNA obtained.

2.3.2 DNase treatment of RNA extracted from cells and tissues

The purified RNA samples were treated with DNase to remove any DNA contamination carried over from RNA extraction process. TURBO DNA-*free*™ Kit (Ambion) was used and the digestion was performed according to manufacturer's instruction. Up to 10µg of RNA was treated with 1µl of TURBO DNase and 0.1 volume of 10X TURBO DNase buffer. Digestion was done at 37^0C for 30 minutes followed by deactivation of enzyme by incubating with 0.1 volume of DNase inactivation reagent for 5 minutes at room temperature. The DNA-free RNA was then obtained after centrifugation at 9000g for 2 minutes. The RNA quality and integrity was determined which are described in the following sections.

2.3.3 Quantification of RNA using Nanodrop® ND-1000 UV/VIS Spectrophotometer

Purified RNA samples were quantified using NanoDrop® ND-1000 UV/VIS Spectrophotometer (LabTech International Ltd, UK) following the manufacturer's guidelines. The nucleic acid concentration was determined by Beer's law and the absorbances were measured at 260, 280 and 230nm. The ratios of 260nm/280nm and 260nm/230nm were obtained to assess the purity of RNA. Both ratios had to be above 1.8 to be considered pure. Although Nanodrop measurements indicate the purity of RNA it does not indicate the integrity of RNA or indicate presence of DNA in the RNA sample.

2.3.4 Analysis of RNA quality using Agilent 2100 Bioanalyzer

Quality of all RNA samples was assessed using Agilent 2100 Bioanalyzer instrument according to the manufacturer's instruction. The Aglient Bioanalyzer performs capillary electrophoresis to analyze RNA, DNA and proteins (Sodowich et al., 2007). It is a very efficient method used to perform absolute RNA quality control before downstream applications such as gene expression analysis or microarrays. All the reagents of Aglient RNA 6000 Nano kit were incubated at RT for 30 minutes before the start of the procedure. The Dye was covered with foil

as it is light sensitive. The Gel matrix was filtered through the column provided in the kit. The gel-dye mix was prepared by mixing 1µl of the RNA 6000 Nano Dye with 65µl of the Agilent RNA 6000 Nano Gel matrix according to the manufacturer‘s instructions. The 6000 NanoChip was primed by placing a 9µl of this mixture into the specified well on the NanoChip. The gel was dispersed across the chip with a help of a plunger. The RNA samples were denatured for 2minutes at $+70^0$C and loaded into the NanoChip alongside 5µl of the RNA Nano Marker. The NanoChip was then vortexed for 60 seconds at 2499xg on a vortex mixer (IKA)and applied for analysis by the Bioanalyser. To start the experiment, Eukaryote Total RNA assay was selected on the program and run. After the run was complete the data was saved as a pdf file and the RNA quality was assessed from the electropherograms obtained. Only RNA of acceptable quality was used in downstream applications.

2.3.5 cDNA Synthesis (Reverse transcription)

RNA prepared and analyzed as described above was reverse-transcribed to from cDNA using VersoTM cDNA Synthesis Kit (Thermo Scientific) according to the manufacturer’s instructions. 500ng of total RNA was diluted in autoclaved ddH2O and heated for 2 min at 70°C to remove secondary structures. The cDNA reaction was prepared in sterile PCR tube by adding cDNA synthesis buffer, dNTPs, random hexamer RNA primers, RT enhancer, RT-polymerase as per the manufacture’s recommendations. The reaction mix (20µl) was run using following PCR cycle: 10 min 25°C, 120 min 42°C; 2 min at 95°C. In parallel, reactions with “no template” and “no enzyme” were performed as negative controls. Freshly synthesized cDNA was diluted 10 times and 1µl of the diluted cDNA was used in subsequent PCR or qPCR reaction.

To assess the quality of cDNA prepared, a PCR reaction usinga housekeeping gene *β-Actin* (Beta Actin) primers was performed. These are also known as reference genes and they help in normalizing the expression data when comparing different samples. As negative control,

a PCR reaction with 1µl of ddH2O instead of the cDNA template was run. The PCR-amplified samples were then resolved on an agarose gel.

2.3.6 Polymerase Chain Reaction (PCR)

Polymerase Chain Reaction or PCR method was originally developed by Kary Mullis (Mullis et al., 1986)who received the Nobel Prize in 1993 for this invention (Nobelprize.org., 1993). PCR uses the DNA polymerase function to synthesise a new strand of DNA. Short complementary strands of DNA called primers are used in PCR. The primers bind to the DNA sequence and add nucleotides to the 3'-OH end. Two primers are always used for a reaction; forward primer and reverse primer to mark the start and end of the region to be amplified. This reaction is performed in a series of DNA denaturation (95^0C), primer annealing ($50\text{-}60^0$C) and DNA polymerase extension (72^0C). Depending on the number of cycles the DNA is replicated to several thousand copies. This amplified DNA can be visualised on an agarose gel.

cDNA was amplified using KOD Hot Start Polymerase (Novagen) using the TECHNE TC-312 thermal cycler. Each 25µl PCR reaction contained 5µl of KOD buffer, 1mM of Magnesium, 0.1µM of each forward and reverse primers, 0.2mM of dNTPs, 100ng of DNA template and 1µl of KOD polymerase diluted in ddH_2O. A reaction where DNA template was substituted with ddH_2O was used as a background control.

2.3.7 Agarose gel electrophoresis

Agarose gel electrophoresis was used in this study to resolve DNA fragments to assess the size of PCR products. The nucleic acid samples were visualised with Ethidium Bromide (Sigma) in a 1% agarose gel using 300nm UV illumination (Alpha Innotech). Before loading 6 xDNA loading buffer (0.25% Bromophenol blue, 0.25% Xylene, 30% glycerol) was added to DNA preparations to increase density of the samples and also help track DNA migration. Agarose gels were run in 1xTAE running buffer (40mM Tris, 20mM acetic acid, 1mM EDTA).

The DNA marker, GeneRuler™ DNA Ladder Mix (Lab Aid, Fermentas), was run alongside samples to act as reference to establish the size of bands produced.

2.3.8 Gene Expression Analysis using Reverse Transcriptase Real Time PCR (RT-qPCR)

Quantitative PCR (qPCR) is a modification of the standard PCT technique where fluorescent dyes such as SYBR green are used. SYBR green binds to the minor grove of the DNA double helix. Thus the dye binds to amplified DNA fragments and the level of fluorescence corresponds to amount of DNA fragments present. The cycle threshold (Ct) is defined as the point where the signal becomes detectable (Bustin and Nolan, 2004). Melt curve analysis are performed to ensure that there is only one PCR product formed. When the DNA denatures the SYBR green dye disassociates and the signal is lost gradually which is measured. Melting temperature (Tm) is the point where the maximum signal SYBR green dye is lost.

Reverse Transcriptase Real Time (RT-qPCR) is qPCR using cDNA to measure expression levels of specific genes. To compare gene expression amongst different cell lines or tissue samples, it is very important that equal amounts of cDNA are used. Since there is no method to evaluate the efficiency of reverse transcription reaction, number of cDNA copies formed for each sample cannot be calculated. Therefore the selection of appropriate housekeeping or reference gene becomes an essential criterion for reliable gene expression analysis. The levels of the gene of interest are always compared between samples by normalising the target gene levels with the levels of a reference gene.

The RT-qPCR reactions were normally carried out in 96-well plates (Bio-Rad) using 2x SensiFAST™ SYBR No-ROX mix (Bioline) according to manufacturer's manual. The real time qPCR reactions were prepared in triplicates using cDNA (≈100ng), forward primer (400nM), reverse primer (400nM) and 2x SensiFAST™ SYBR No-ROX mix and autoclaved ddH2O to a final volume of 20µl per reaction. Negative control reactions without the DNA were also run

alongside. The target genes were always measured alongside reference genes to normalise and compare expression profiles. Primer sequences and their PCR cycling conditions used for specific experiments are described in Table 2.2 and Table 2.3. CFX96 Real-Time Thermal Cycler C1000 (Bio-Rad) was used and the data obtained was analysed using Bio-Rad CFX System Test software. The software calculates the relative quantities of gene expression based on the delta-delta Ct (ΔΔCt) method which assumes efficiency of each RT-qPCR reaction to be 100%. A modified formula based on ΔΔCt method that takes into account the actual efficiency of the qPCR reaction is given below

$$\text{Relative expression ratio (R)} = \frac{(E_{target})^{\Delta Ct_{target}\,(\text{Control-sample})}}{(E_{reference})^{\Delta Ct_{reference}\,(\text{Control-sample})}}$$

(E=Efficiency of PCR reaction; Ct=Threshold Cycle)

Reference Gene	For Primer	Rev Primer	Real-Time PCR conditions (initial denaturation: 95°C 2 min)
HPRT1	GACCAGTCAACAGGGGACAT	AACACTTCGTGGGGTCCTTTTC	Cycle(x40): 95°C 5 sec; 58°C 10 sec; 72°C 10 sec.
RPL32	CATCTCCTTCTCGGCATCA	AACCCTGTTGTCAATGCCTC	Cycle(x40): 95°C 5 sec; 60°C 10 sec; 72°C 10 sec.
B2M	ACTGAATTCACCCCCACTGA	CCTCCATGATGCTGCTTACA	Cycle(x40): 95°C 5 sec; 59°C 10 sec; 72°C 10 sec.
18s rRNA	CAGCCACCCGAGATTGAGCA	TAGTAGCGACGGGCGGTGTG	Cycle(x40): 95°C 5 sec; 61°C 10 sec; 72°C 10 sec.
CycloB	TGGCACAGGAGGAAAGAGCATC	AAAGGGCTTCTCCACATCGAT	Cycle(x40): 95°C 5 sec; 60°C 10 sec; 72°C 10 sec.
TBP	GCCCGAAACGCCGAATATA	CGTGGCTCTCTTATCCTCATGA	Cycle(x40): 95°C 5 sec; 58°C 10 sec; 72°C 10 sec.
HuP0	GCAGCATCTACAACCCTGAAG	CACTGGCAACATTGCGGAC	Cycle(x40): 95°C 5 sec; 58°C 10 sec; 72°C 10 sec.

Table 2.2 Gene primers used for finding suitable internal control housekeeping gene and cycling conditions: HuPO, human acidic ribosomal protein; CycloB, cyclophylinB; B2M, β2-microglobulin; HPRT, hypoxanthine phosphoribosyltransferase; 18sRNA, 18S ribosomal RNA; RPL32, ribosomal protein 32; TBP, Tata box Binding Protein.

Gene Name	Sequence	Annealing temperature (X^0C)	cDNA product size
SERPINB1 Forward	TGAGTGAGAACAATCCGGCT	62	128
SERPINB1 Reverse	GTTGAAATGGAAAGTCTTGGACAG		
RHOA Forward	CCCTCTCCTACCCAGATACC	60	167
RHOA reverse	TGCTCATCATTCCGAAGATCCT		
ITGA4 Forward	TCACACTTTCCAGACAGCCA	61	119
ITGA4 Reverse	CACTCCATAGCAACCACCAG		
CALR Forward	CAACTTCCTCATCACCAACGA	62	188
CALR Reverse	TGTCCTCATCATCCTCCTTGTC		
FGL2 Forward	CCAGCCAAGAACAAATACAGTC	60	115
FGL2 Reverse	GGATCAGGTGTAACTCTGTAGG		
ALOX5 Forward	CTACATCTACCTCAGCCTCGT	61	188
*ALOX5*Reverse	GTACCAGTCGTCATTCAGCC		
ANXA1 Forward	TTCAATACCATCCTTACCACCA	61	124
*ANXA1*Reverse	CTCAATGTCACCTTTCAACTCC		
WDR1 Forward	AGATCCAAGATGCACACCG	60	122
WDR1 Reverse	GCTCCTCAGTAGGTGATTGTC		
YWHAE Forward	CGATACGACGAAATGGTGGA	60	121
YWHAE Reverse	GGCTCTTCTAGCTCCAATCAC		
OSTF1 Forward	AGTGGGTGTTAATGGCTTAGAC	59	137
OSTF1 Reverse	TATCTCCCAACTTGTTCTGCTG		
LTF Forward	CAAATGTGCCTTCTCCTCCC	61	153
LTF Reverse	GTAACTCATACTCGTCCCTTTCAG		
NONO Forward	AGAGCAGGAGATTCGGATGG	61	172
NONO Reverse	AAGCGTTCAGTTGTTGGTGG		
ANXA3 Forward	TCTTAACTACCAGGACAAGCAG	59	111
ANXA3 Reverse	GAAGTCACCAGATGTTTCGG		
CPNE3 Forward	AGAGATATTGTCCAGTTTGTGCC	60	164
CPNE3 Reverse	ACTGCTTCTGTTGTTTCGTGG		

Table 2.3 Primer sequences used to validate candidate biomarkers using RT-qPCR: Primer sequences used to validate a biomarker panel by RT-qPCR. Real time PCR conditions: {Initial denaturation 95°C 2 min [Cycle(x40): 95°C 5 sec; X°C 10 sec; 72°C 10 sec]}.

2.4 High throughput data analysis and bioinformatics methods

2.4.1 Public microarray databases and Gene Expression Omnibus

Public microarray databases are repositories of raw/processed gene expression data that are peer reviewed and produced by adhering to industrial standards. These databases include often include tools for data analysis. There are two types of these repositories; peer reviewed public databases (eg. Gene Expression Omnibus (GEO) from NCBI and ArrayExpress from European bioinformatics institute) and specialised repository which might be associated with research groups, labs, disease types, cell types etc. In this study we used GEO to mine for microarray datasets. GEO provides a flexible and open design that facilitates submission, storage and retrieval of heterogeneous data sets from high-throughput gene expression and genomic hybridization experiments (Edgar et al., 2002). Both raw data and processed data for microarray datasets can be downloaded from GEO and data analysis can be independently performed.

2.4.2 Statistical analysis methods used in the study

The statistical hypothesis tests used in this study were Student's T-test and one- way Analysis of Variance (ANOVA). Both T-test and ANOVA assume the data is normally distributed. The T-test looks at the t-statistic, t-distribution and degrees of freedom to determine a p value (probability) that can be used to determine whether the population means differ. To compare three or more variables ANOVA is used. Other tests used in this study were Mann-Whitney U-test and Kruskal-Wallis test. Both are non-parametric tests and is used when the data is not normally distributed, if the variances for the two conditions are markedly different or if the data are measurements on an ordinal scale.

2.4.3 Tools for Molecular and Functional analysis of highthroughput data

2.4.3.1 Multi Experiment Viewer (MeV)

Multi Experiment Viewer is freely available software which can be used to analyse microarray data. It incorporates algorithms for normalisation, clustering, visualisation statistical analysis and annotation based meta-analysis. MeV is one of the components of the TM4 microarray suite and can be freely downloaded from www.tm4.org.

2.4.3.2 Gene Ontology (GO)

Understanding the biological meaning of the results obtained through microarray or RNA-Seq can be difficult. The conventional way to achieve this is to classify the genes based on their function and their cellular location. Gene ontology is used to annotate genes with respect to their function. The gene ontology project is an initiative which evolved out of the need to consolidate the descriptions of each gene and its products across different species for consistence to enable functional interpretation of experimental data; the ontology database is maintained by the gene ontology consortium (http://geneontology.org/). The Gene Ontology analysis was performed using a tool called Gene Ontology for Functional Analysis (GOFFA) developed by the USA Food and Drug Administration (FDA) Department which is available as a standalone tool at http://www.fda.gov/ScienceResearch/BioinformaticsTools/ucm233315.htm.

2.4.3.3 Pathway enrichment analysis

Pathway analysis is another approach to interpret the biological meaning of a microarray dataset. This approach can be used to observe subtle and consistent changes in the pathways using functional annotations. Pathway analysis was done using the KEGG database (Kyoto Encyclopedia of Genes and Genomes) (Weblink: http://www.genome.jp/kegg) and the software used to view to apply and view the data was GenMapp (Weblink: http://www.genmapp.org). KEGG pathway database is a collection of manually drawn pathway maps representing molecular

interactions and networks for metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems, human diseases and drug development (Kanehisa et al., 2012).

2.4.3.4 Protein Protein Interaction analysis

Protein-Protein interactions are crucial for all biological processes. With the rapid development of genomics and proteomics technologies the knowledge about cellular processes has increased. In practice information about molecular interactions is dispersed in the scientific literature and difficult to retrieve in a structured format. Therefore organised, user-friendly and consolidated databases about molecular interactions are needed to interpret complex biological data. Currently several of these databases exist; Biomolecular Interaction Network Database (BIND), Biological General Repository for Interaction Datasets (BIOGRID), IntAct molecular interaction database, Search Tool for the Retrieval of Interacting Genes/Proteins (STRING), Molecular INTeraction database (MINT) etc. These databases use information from yeast hybrid experiments, immunoprecipitation, user submitted information etc. Osprey is a tool for graphical visualisation of complex biological interaction networks using the gene ontology annotated dataset maintained by BIOGRID. Osprey was used in this study and it can freely downloaded from http://www.cs.duke.edu/donaldlab/osprey.php.

Chapter 3: Analysis of Protein profiles of White blood cells in breast cancer patients using Proteomic Approaches

3.1 Introduction

Early detection of breast cancer allows the possibility to determine the best treatment options which may be surgical resection, hormone therapy, chemotherapy, radiotherapy or combination of those. Our understanding of the molecular pathology of breast cancer has recently advanced with the help of new proteome- and genome-wide technologies. The conventional classification based on traits such as invasiveness, tissue of origin, histopathology, degree of cellular differentiation etc has now been complemented by novel molecular stratification which derived from integrated analyses of gene expression and copy number (Curtis et al., 2012; Dawson et al., 2013). These studies also created an opportunity for discovering biomarkers to predict the prognosis of the disease thereby deciding the course of treatment.

Proteomics approaches have been used in a number of studies to discover biomarkers in the blood plasma. Examples include biomarkers for invasive breast cancer in a mouse model (Pitteri et al., 2008, 2011), diagnostic and prognostic markers for gastric cancer (Qiu et al.,2009), non-small cell lung cancer (Izbicka et al., 2012) and breast cancer (Nolen et al., 2008) in patients' plasma. Circulating biomarkers such as CA 125, CA 19-9 and carcinoembryonic antigen in ovarian, pancreatic and colon cancer, respectively, have been used to monitor response to therapy and recurrence (Ludwig and Weinstein, 2005). There are a number FDA approved biomarkers (Ludwig and Weinstein, 2005) for breast cancer including CA15-3, CA27-29, Oestrogen receptor, progesterone receptor, HER2/NEU for prognosis and monitoring but none of them can be used for diagnosis. Therefore, there is a continuing quest to discover new

biomarkers from various sources, in particular from biological fluids (blood, urine, nipple aspirates) because of the less invasive procedures are involved to obtain those specimens.

Proteomics technologies have now been widely used to study proteomes of the cells. The term "proteome" refers to the entire set of proteins synthesised by the cell or tissue and the large scale study of it is proteomics. This approach has advantages because (i) proteomics provide more physiologically accurate information about processes in the cell; (ii) changes in protein activities associated with the disease can be identified and (iii) proteomics platforms are more applicable from therapeutics point of view. A range of methods including 2D-PAGE (Gharbi, 2001; Somiari et al., 2003), protein arrays (Sreekumar et al., 2001) coupled with mass spectrometric techniques, for example, SELDI-ToF, MALDI-ToF, etc have been used to identify biomarkers for cancer using tissue, cancer cell lines and blood plasma.

The evolution of mass spectrometry in the last decade has made it more sensitive, although it is not yet possible to resolve accurately all the proteins found in complex biological mixtures such as serum, cell lysates etc. Nevertheless, this approach has many advantages since protein biomarkers are more desirable in clinical settings. These methods will be used in this study to identify breast cancer biomarkers in the white blood cells (WBCs), a relatively unexplored source for biomarker discovery.

3.2 Aim of the Chapter

The main aim of this chapter is to investigate protein profiles specific for the WBCs of breast cancer patients and identify new breast cancer biomarkers. To achieve this, two approaches will be applied: the high throughput mass spectrometry using the Velos Orbitrap instrument and 2D-PAGE.

High throughput proteomics was used to identify the differences in protein profiles of white blood cells (WBCs) of breast cancer patients and healthy donors. For these analyses, samples from three participant groups will be processed: the healthy donors (Group C, n=5) and breast cancer patients (Group 1) stratified into two distinct groups: Group 1A with more favourable prognosis (Grade 1, Stage 1, n=5) and Group 1B with less favourable prognosis (Grade 3, Stage 3, n=5). We reasoned that using the groups with opposing characteristics will facilitate the analysis and simplify the identification of the candidates showing significant differences in protein profiles. Clinical information linked to the specimens used for these analyses is given in the Table 3.1.

The 2D-PAGE method was employed to independently identify proteins differentially present in the WBCs of breast cancer patients and to compare the data with high throughput proteomics analysis. Clinical information linked to the specimens used in this investigation is presented in Table 3.2.

Table 3.1 Sample Patient data for the High-throughput Proteomics study: Breast cancer patient WBC samples used for high throughput analysis by OrbitrapVelos instrument. T size- Tumour size, T Grade- Tumour grade, T stage- Tumour stage, ER- Estrogen Receptor, PR- Progestrone Receptor, DCIS- Ductal Carcinoma in situ, IDC- Invasive Ductal Carcinoma, ITC- Invasive tubular carcinoma and ILC- Invasive Lobular Carcinoma.

Patient No	AGE	T SIZE	NODE STATUS	T STAGE	T GRADE	ER	PR	HER2	DIAGNOSIS
1462	65	11	NEG	I	1	8	ND	-ve	ITC+DCIS
1466	56	10	NEG	I	1	8	ND	-ve	IDC+ILC
1477	46	8	NEG	I	1	8	ND	-ve	IDC,DCIS(LOW GRADE)
1483	52	17	NEG	I	1	8	ND	-ve	IDC+ILC
1565	51	11	NEG	I	1	8	NA	-ve	Tubular+DCIS
453	26	50	POS	III	3	3	0	-ve	IDC
1261	44	80	POS	III	3	4	NA	-ve	IDC+DCIS
1503	75	20	POS	III	3	8	NA	-ve	IDC+DCIS
1516	46	14	NEG	III	3	3	NA	-ve	IDC+DCIS
1540	48	28	POS	III	3	8	NA	-ve	IDC+Chondrosarcoma

Table 3.2 Sample Patient data for the 2D-PAGE: Breast cancer patient WBC samples used for 2D-PAGE. T size- Tumour size, T Grade- Tumour grade, T stage- Tumour stage, ER- Estrogen Receptor, PR-Progestrone Receptor, DCIS- Ductal Carcinoma in situ, IDC- Invasive Ductal Carcinoma, ITC- Invasive tubular carcinoma and ILC- Invasive Lobular Carcinoma.

Patient No	**AGE**	**T SIZE**	**NODES**	**T STAGE**	**T GRADE**	**ER**	**PR**	**HER2**	**DIAGNOSIS**
1310	69	8	NEG	I	1	8	ND	-ve	ILC
1455	49	18	NEG	I	1	8	ND	-ve	ILC+LCIS
1462	65	11	NEG	I	1	8	ND	-ve	ITC+DCIS
1499	67	15	NEG	I	1	8	NA	+ve	IDC+DCIS+Tubulo-lobular
1565	51	11	NEG	I	1	8	NA	-ve	Tubular+DCIS
1261	44	80	POS	III	3	4	NA	-ve	IDC+DCIS
1436	60	20	NEG	II	3	8	ND	-ve	IDC+DCIS
1528	84	20	NEG	II	3	8	NA	-ve	IDC+DCIS
1560	82	20	NEG	II	3	0	NA	-ve	IDC
1540	48	28	POS	III	3	8	NA	-ve	IDC+Chondrosarcoma

3.3 Results

3.3.1 High throughput proteomic profiling of WBCs using the Velos Orbitrap mass spectrometer.

3.3.1.1 High throughput proteomic profiling and comparative analysis of WBCs from the Healthy donors and all Breast cancer patients groups.

High throughput proteomics was used to identify the differences in protein profiles of white blood cells (WBCs) of breast cancer patients and healthy donors. The WBCs were fractionated into membrane and soluble fractions as described in Alldridge et al., 2008. Both fractions were digested and analysed using the OrbitrapVelos instrument (see section 2.2.2.1) equipped with an Ultimate 3000 nano-scale HPLC (Olsen et al., 2009). Dr Metodi Metodiev (University of Essex) had kindly performed the mass spectrometric analysis using this instrument. Only the data for the membrane fraction was analysed for this study. In total, 2577 proteins were identified in the membrane fraction of the WBCs. Since the data were not normally distributed, the Mann Whitney U-Test was performed with a P-value cutoff of <0.05 to compare the protein profiles between healthy donor and breast cancer patients. The test yielded 136 proteins which were significantly different between the healthy donor and the breast cancer group. Out of the significant 136 proteins, 83 proteins were overexpressed and 53 proteins were under expressed in the breast cancer patient category. Hierarchical clustering was performed with the normalised Label free quantification (LFQ) intensity values of significant 136 proteins which are given in Figure 3.1. For this analyses the Multiexperiment viewer (see section 2.4.3.1) from the TM4 microarray software suite available at http://www.tm4.org/ was used.

The hierarchical clustering revealed that the data had a high degree of variation especially among the breast cancer patient cohort. This can be due to the fact that WBCs are a mixture of different types of cells and the dynamic nature of the immune system.

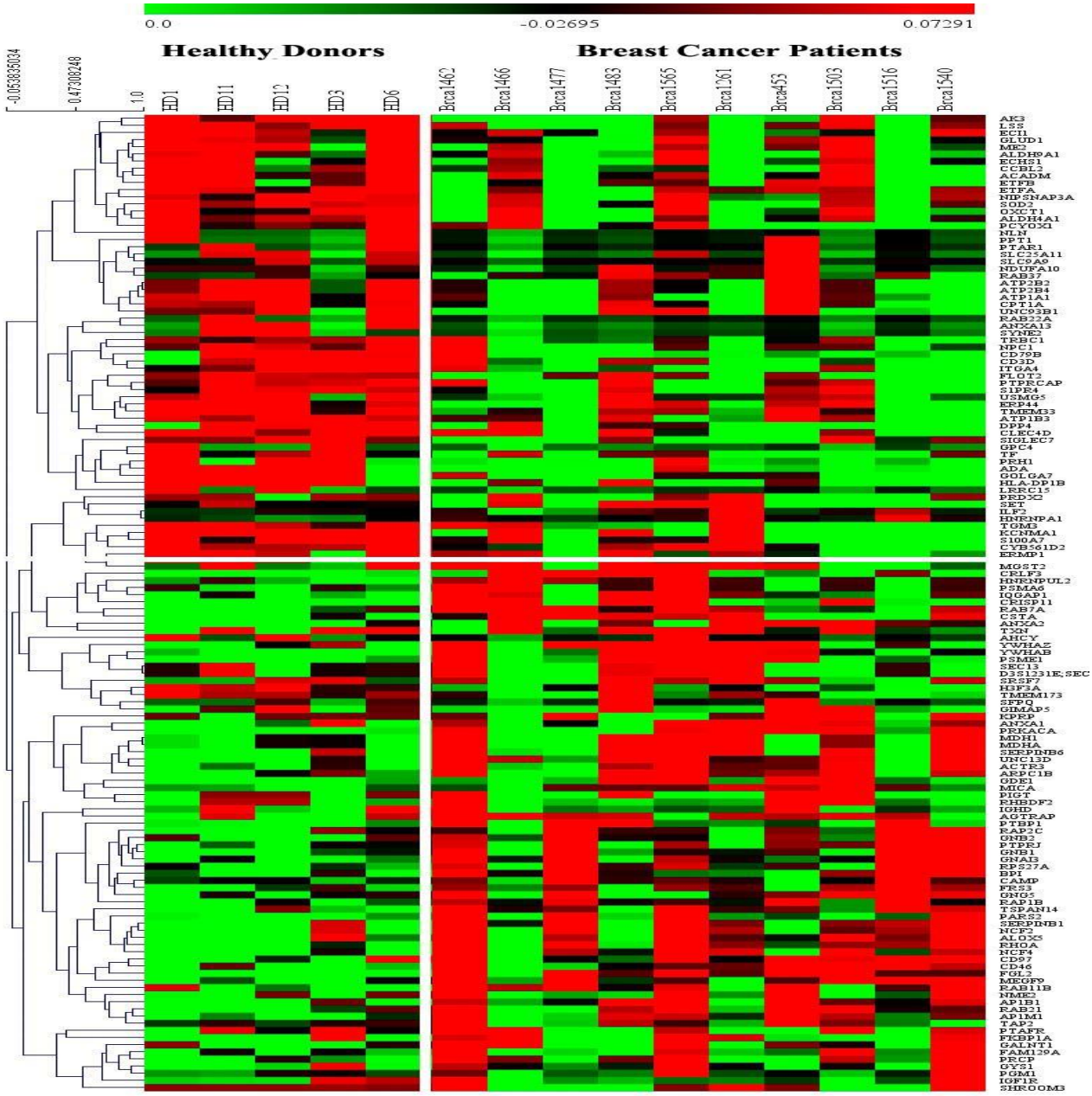


Figure 3.1 Hierarchical Clustering of significant proteins between the WBCs of Healthy Donors and Breast Cancer Patients: Differentially present proteins from the membrane fraction were identified using the Mann Whitney test between breast cancer patients and healthy donors. Hierarchical clustering was performed using the Multiexperiment viewer. In the breast cancer patient category 83 proteins were overexpressed and 53 proteins were underexpressed. Red indicates that the protein is relatively overexpressed and green the vice versa in breast cancer patient samples.

3.3.1.2 Comparison between the samples with different Tumour grades: Healthy Donors vs Low grade Breast Cancer patients and Healthy Donors vs High grade Breast Cancer Patients.

To identify the differences the between different grades and stages samples were chosen in patients with either Grade1-Stage1 tumours (favourable prognosis) or Grade3-Stage3 tumours (less favourable prognosis). We reasoned that by choosing samples with these characteristics the differences in proteomes of breast cancer will be more apparent. The WBCs protein profiles obtained using the high throughput mass spectrometry was compared as follows: Healthy Donors vs Low grade Breast Cancer patients and Healthy Donors vs High grade Breast Cancer Patients. Kruskal-Wallis teast (dunn's correction, p value<0.05) was used as described in section 2.4.3.1 to compare between the groups; this resulted in 44 proteins as significantly different between the categories.

Hierarchical clustering was performed with the normalised Label free quantification (LFQ) intensity values (section 2.4.3.1) of significant 44 proteins (Figure 3.2). The Multiexperiment viewer from the TM4 microarray software was used for this analysis. There were three major groups of proteins observed from clustering: Proteins which were overexpressed in healthy donors (n=10), proteins overexpressed in both breast cancer cohorts (n=14) and proteins overexpressed in Low grade patients (n=20). The protein profiles which follow the trend of the first two groups were chosen and the mean for each group was taken to plot graphs with standard deviation (Figure 3.3).

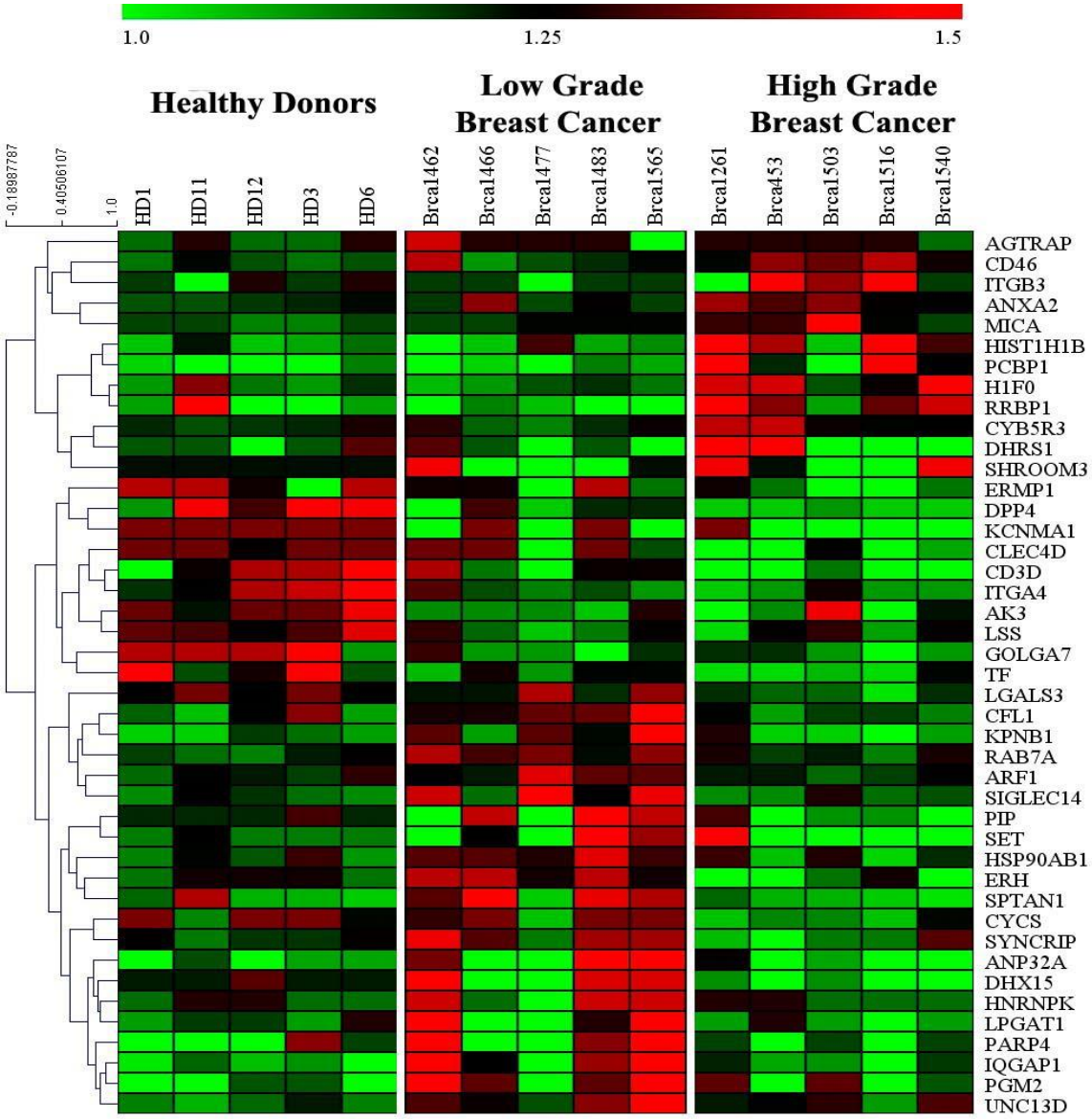


Figure 3.2 Hierarchical clustering of significant proteins between the WBCs of Healthy Donors, Grade1-Stage1 and Grade3-Stage3 Breast Cancer Patients: Mann-Whitney test was performed to obtain the significantly expressed proteins between the WBCs of healthy donors, low grade breast cancer patients and high grade breast cancer patients. Hierarchical Clustering was done using the proteins which had significant differences using Multiexperiment viewer. Red indicates that the protein is relatively overexpressed and green the vice versa in breast cancer patient samples.

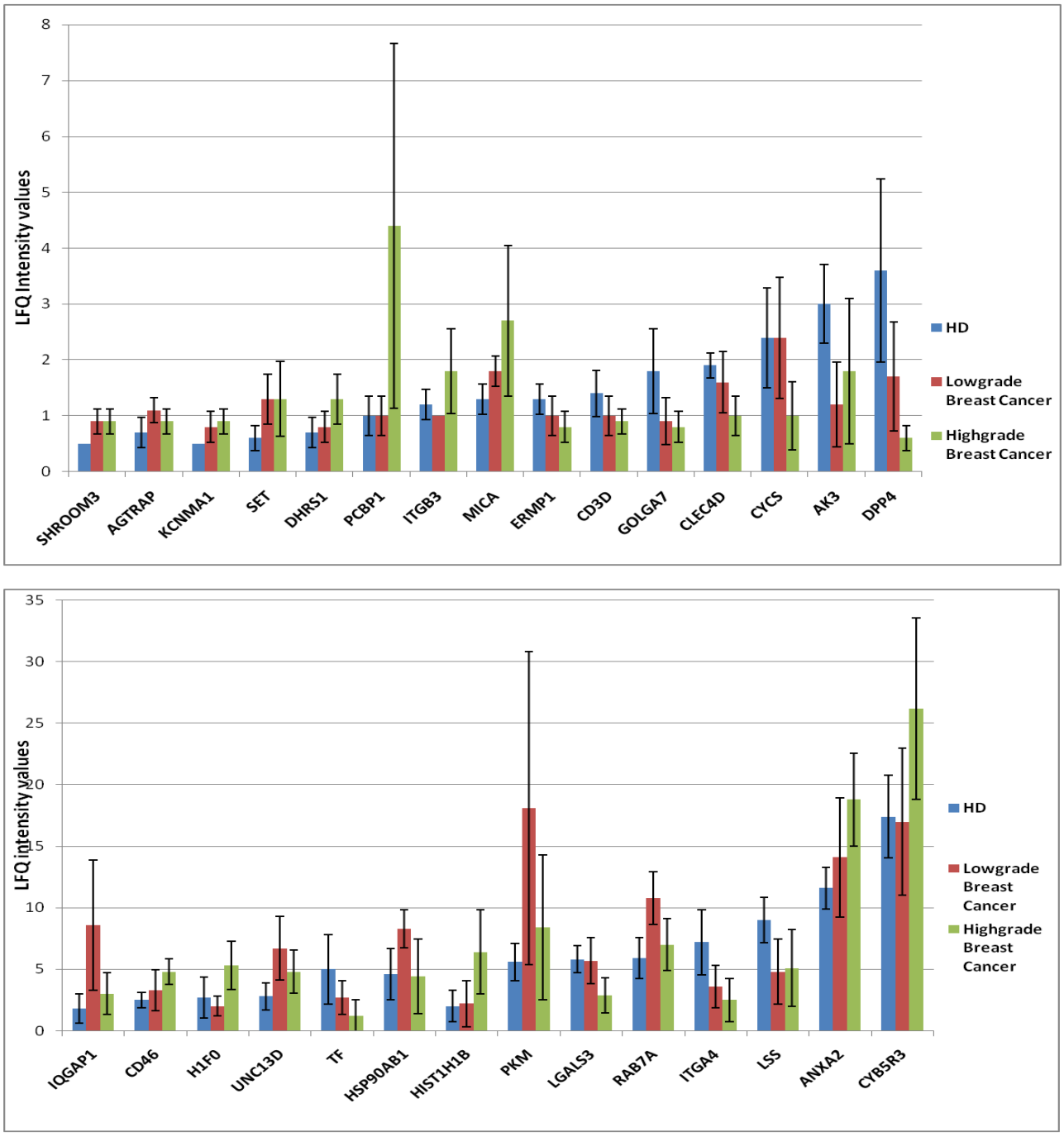


Figure 3.3 Protein expression profiles of significant proteins obtained by comparison between the WBCs of Healthy Donor, Grade1-Stage1 and Grade3-Stage3 Breast cancer Patients: Kruskal-Wallis was used to compare membrane fraction protein profiles of WBCs between healthy donors, low grade-low stage and high grade-high stage breast cancer patients. The profiles of proteins which showed significant changes in healthy donors and breast cancer patients (both low grade and high grade) were chosen and their mean was taken to plot graphs.

3.3.1.3 Molecular and Functional analysis of proteins obtained from the high throughput proteomics analysis

It is important to biologically interpret the data obtained from the high throughput analyses and there are various approaches to achieve this. Commonly used approaches include: gene ontology, pathway enrichment and protein- protein interaction network analyses (see section 2.4.3). The ontology analysis was done based on three categories immune response, cell death and cell migration. The proteins which were differentially expressed in the WBCs of breast cancer patients obtained by both tests mentioned in sections 3.3.1.1 and 3.3.1.2 were analysed for functional significance.

3.3.1.3.1 The Gene Ontology analysis: classification of Biological Processes

The differentially present proteins in the breast cancer patient cohort were obtained by comparing the WBC membrane fraction profiles of breast cancer patients and healthy donors. These proteins were subjected to ontology analysis to identify enriched biological processes. Gene ontology is used to functionally interpret data obtained from high throughput techniques such as microarray, proteomics etc (Ashburner et al., 2000). The gene ontology database is maintained by the Gene Ontology Consortium and can be accessed freely at www.geneontology.org.

Gene ontology analysis based on classification of Biological Processes was performed using GOFFA (Gene Ontology for Functional Analysis) as described in Section 2.4.3.2.The results for the classification are given in Figure 3.4.The categories which had the most number of proteins were cellular processes (n=73), metabolic processes (n=59), response to stimulus and reproduction (n=53). It was observed that 41 proteins were involved in localisation and establishment of localisation. This shows that around 25% of proteins which changed were involved in cellular localisation.

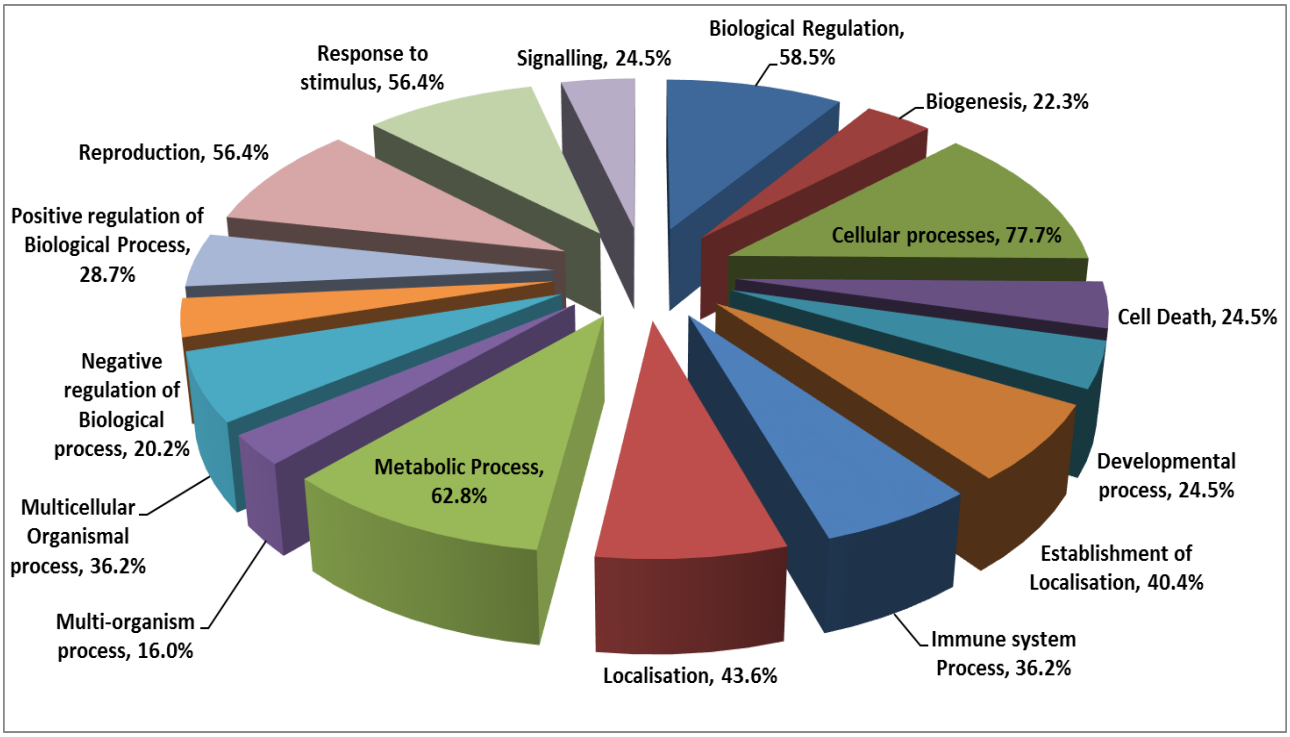


Figure 3.4 Gene Ontology of significantly different proteins between the WBCs of Healthy donors and Breast cancer Patients obtained through high throughput Proteomics: Proteins differentially present in the WBCs of breast cancer patient and healthy donors by high throughput proteomics were statistically significant by the Mann Whitney U test. Gene Ontology was performed using GOFFA and based on biological processes. The total number of proteins involved was 136.

3.3.1.3.2 Ontology analysis of proteins involved in immune system processes

There were 34 proteins involved in the immune system processes; the gene ontology classification of these processes and proteins involved in each subcategory are given in Table 3.4. Among those, 21 proteins were involved in regulation of the immune system processes, with 14 activators. In the latter category, 8 proteins were overexpressed and 6 proteins were underexpressed in the breast cancer cohort. The number of genes involved in negative regulation of immune system was 5 and all the proteins except ADA (Adenosine deaminase) were overexpressed. These data provided inconclusive results as to how the immune response was affected by breast cancer. Therefore, the proteins involved in leukocyte and lymphocyte activation were analysed. The proteins involved in both categories were the same except for BPI (Bactericidal/Permeability-Increasing Protein) and YWHAZ (tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta). The proteins involved in positive regulation of leukocyte and lymphocyte activity (ADA, CD3D, DPP4 and TRBC1) were underexpressed. The proteins which negatively regulated leukocyte and lymphocyte activity, BPI and GIMAP5, were overexpressed. It can be concluded from the data that activation of leukocytes and lymphocytes are negatively affected by the breast tumour.

The proteins involved in the inflammatory response (n=10) are shown in Table 3.4; except for ADA and PRDX2, all the proteins were overexpressed. ADA negatively regulates the inflammatory response (GO:0050728). Peroxiredoxin-2 (PRDX2) is an enzyme which has antioxidant activity reducing hydrogen peroxide alkyl hydroperoxides in the cell and hence has an anti-inflammatory effect. This shows that inflammatory response is increased in the WBCs of breast cancer patients.

Table 3.3 List of Proteins involved in the immune response: Proteins identified by high-throughput proteomics to be differentially present between healthy donors and breast cancer patients by comparing the membrane fraction of the WBCs. Gene ontology was to choose proteins related to immune response. Proteins in red are overexpressed and green are underexpressed in breast cancer patient WBCs.

Gene Symbol	Gene name	Fold change
ADA	Adenosine deaminase	
CD79B	CD79b molecule, immunoglobulin-associated beta	
DPP4	Dipeptidyl-peptidase 4 (CD26, adenosine deaminase complexing protein 2)	
S1PR4	Endothelial differentiation, lysophosphatidic acid G-protein-coupled receptor, 6	
CD3D	CD3d molecule, delta (CD3-TCR complex)	
S100A7	S100 calcium binding protein A7	
ITGA4	Integrin, alpha 4 (antigen CD49D, alpha 4 subunit of VLA-4 receptor)	
CLEC4D	C-type lectin domain family 4, member D	
TRBC1	T cell receptor beta constant 1	
ATP1B3	ATPase, Na+/K+ transporting, beta 3 polypeptide	
SOD2	Superoxide dismutase 2, mitochondrial	
UNC93B1	Unc-93 homolog B1 (C. elegans)	
TMEM173	Transmembrane protein 173	
PRDX2	Peroxiredoxin 2	
PTAFR	Platelet-Activating Factor Receptor	
GIMAP5	GTPase, IMAP family member 5	
CD97	CD97 molecule	
PTPRJ	Protein tyrosine phosphatase, receptor type, J	
YWHAZ	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	
ANXA1	Annexin A1	

S100A9	S100 calcium binding protein A7;Calgranulin B	
FKBP1A	FK506 binding protein 1A, 12kDa	
IGF1R	Insulin-like growth factor 1 receptor	
RPS27A	Ribosomal protein S27a	
CD46	CD46 molecule, complement regulatory protein	
BPI	Bactericidal/permeability-increasing protein	
MICA	MHC Class I Polypeptide-Related Sequence A	
TAP2	Transporter 2, ATP-binding cassette, sub-family B (MDR/TAP)	
UNC13D	Unc-13 homolog D (C. elegans)	
ILF2	Interleukin enhancer binding factor 2, 45kDa	
AP1B1	Adaptor-related protein complex 1, beta 1 subunit	
NCF2	Neutrophil cytosolic factor 2	
NCF4	Neutrophil cytosolic factor 4, 40kDa	
AP1M1	Adaptor-related protein complex 1, mu 1 subunit	
CD200R1	CD200 receptor 1	

Table 3.4 Gene ontology of the Proteins involved in immune system processes: Immune system proteins identified to be differentially present in the WBCs of breast cancer patient and healthy donors by high throughput proteomics were obtained by Mann Whitney U test. Gene Ontology was performed using GOFFA and based on Immune system process. The total genes involved in immune system processes were 34.

Gene Ontology Term	Gene Ontology ID	Genes involved	P value
Regulation of Immune system process	GO:0002682	ADA,AP1B1,AP1M1,BPI,CD200R1,CD3D, CD46,DPP4,FKBP1A,GIMAP5,ITGA4,MICA, PRDX2,PTPRJ,RPS27A,S100A7,TAP2,TMEM173, TRBC1,UNC13D,UNC93B1	0.0002
Positive regulation of the immune system	GO:002684	ADA,CD3D,CD46,DPP4,GIMAP5,MICA, PTPRJ,RPS27A,S100A7,TAP2,TMEM173,TRBC1, UNC13D,UNC93B1	0.0009
Negative regulation of the immune system	GO:002683	ADA,BPI,GIMAP5,PRDX2,PTPRJ	0.03
Adaptive Immune Response	GO:0002250	ADA,CD46,DPP4,GIMAP5,TAP2,UNC13D	0.02
Innate Immune Response	GO:0045087	CD46,CLEC4D,GIMAP5,MICA,NCF2,PTAFR,RPS27A, S100A7,TMEM173,UNC13D,UNC93B1	0.002
Positive regulation of Innate immune response	GO:0045089	GIMAP5,MICA,RPS27A,TMEM173,UNC93B1	0.041
Leukocyte Activation	GO:0045321	ADA,ANXA1,BPI,CD3D,DPP4, FKBP1A,GIMAP5,ITGA4,MICA,PRDX2,TRBC1, UNC13D,YWHAZ	0.0001
Positive Regulation of Leukocyte activation	GO:0002696	ADA,CD3D,DPP4,TRBC1	0.004
Negative Regulation of Leukocyte activation	GO:0002695	BPI,GIMAP5,PRDX2	0.05

Lymphocyte activation	GO:0046649	ADA,ANXA1,CD3D,DPP4,FKBP1A,GIMAP5,ITGA4, MICA,PRDX2,TRBC1,UNC13D	0.002
Positive Regulation of Lymphocyte activation	GO:0051251	ADA,CD3D,DPP4,TRBC1	0.003
Negative Regulation of Lymphocyte activation	GO:0051250	GIMAP5,PRDX2	0.046
Inflammatory response	GO:0006954	ADA,ALOX5,ANXA1,CD97,MGST2,PRDX2, PSMA6,PTAFR,UNC13D,YWHAZ	0.0002

3.3.1.3.3 Ontology analysis of proteins involved in Cell death and Cell migration

The number of proteins in involved in the cell death in the 136 significant proteins obtained by gene ontology analysis by GOFFA was23 (17%). The genes involved in cell death (GO:0008219) are given in Table 3.5. Of the proteins involved in negative regulation of cell death (n=13) except for ADA, PPT1, SOD2 and PRDX2 were overexpressed in the breast cancer cohort. The proteins involved in the positive regulation of cell death (n=8) were also overexpressed. The analysis for cell proliferation also gave similar inconclusive results. The effect of breast cancer on WBC survival is ambiguous taking this data into consideration.

Ontology analysis was done for proteins involved in cell migration. It was found that 10 proteins involved in cell migration were differentially present in the WBCs of breast cancer patients. The genes involved in the positive regulation of cell migration are S100A7, S100A9, IGF1R, ITGB3 and SYNE2. Only S100 calcium binding protein A7 (S100A7) is underexpressed. The S100 family of proteins are involved in several cellular processes like proliferation, differentiation, migration etc. S100A7 has been implicated to aid in breast cancer progression and invasion (Emberley et al., 2004). S100A9 (Calgranulin B) is implicated in promoting leukocyte recruitment to sites of inflammation and abnormal differentiation of myeloid cell in the stroma of cancer (Cheng et al., 2008; Hiratsuka et al., 2006). The proteins which were involved in cell adhesion were CD97, FLOT2, ITGA4, NME2 and SIGLEC7 are shown in Table 3.5. Only ITGA4 and SIGLEC7 were underexpressed in the breast cancer cohort. ITGA4 (Integrin alpha 4) binds to cell surface adhesion molecules on endothelial cells enabling WBCs to transmigrate. SIGLEC7 is also known as adhesion inhibitory receptor molecule 1 is expressed on different tumour cells and inhibits leukocyte adhesion to endothelial cells (Jandus et al., 2014). The cell migration and cell adhesion ontology analysis show that the WBCs of breast cancer patients have increased cell adhesion and migratory functions.

Table 3.5 Gene ontology of the Proteins involved in Cell migration and Cell death: Ontology was done with the differentially expressed proteins, obtained by comparison between healthy donors and breast cancer cohorts using GOFFA. The proteins involved in Cell migration processes and Cell proliferation processes were obtained along with their gene ontology IDs.

Gene Ontology Term	Gene Ontology ID	Genes Involved	P value
Cell Migration	GO:0016477	ADA,ATP1B3,DPP4,IGF1R,ITGA4,PTPRJ, RAP2C,RHOA,S100A7,S100A9,SYNE2	0.017
Positive regulation of cell migration	GO:0030335	S100A7,S100A9,IGF1R,ITGB3,SYNE2	0.03
Negative regulation of cell migration	GO:0030336	ADA,PTPRJ,RAP2C	0.06
Cell Adhesion	GO:0007155	CD97,FLOT2,ITGA4,NME2,SIGLEC7	0.018
Cell proliferation	GO:0008383	ADA,ANXA1,DPP4,GNB1,GPC4,IGF1R, NME2,PRDX2,PTPRJ,RAP1B,SOD2,TXN	0.04
Negative regulation of cell proliferation	GO:0008285	PTPRJ,SOD2	0.009
Positive regulation of cell	GO:0008284	ADA,DPP4,IGF1R,ITGB3,NME2	0.017

proliferation			
Cell death	GO:0008219	CFL1,CYCS,GIMAP5,H1F0,IGF1R,KCNMA1, KPNB1,MICA,NME2,PARP4,PIGT,PPT1, PRDX2,PSMA6,RHOA,RPS27A,SET, SOD2,SPTAN1,TAP2,TMEM13 UNC13D,YWHAB	0.003
Negative regulation of Cell death	GO:0060548	ADA,ANXA1,GIMAP5,IGF1R,NME2, PPT1,PRDX2,PTPRJ,RHOA,RPS27A,SET,SOD2,YWHAZ	0.0001
Positive regulation of Cell death	GO:0010942	CYCS,GIMAP5,KCNMA1,NCF2,RPS27A, TAP2,UNC13D,YWHAB	0.011

3.3.1.3.4 Pathway enrichment analysis

The pathway enrichment analysis is another approach to interpret the biological meaning of high throughput data. This approach can be used to observe subtle and consistent changes in the pathways using functional annotations. Pathway analysis was performed on the significant genes obtained by comparing the protein profiles between the WBCs of breast cancer patients and healthy donors using GenMAPP software as described in section 2.4.3.3. The pathway database used was KEGG (Kyoto Encyclopedia of Genes and Genomes). The top 10 list of pathways with the proteins involved are given in Table 3.6. The majority of the pathways are related to metabolism. The only pathways related to immune function was leukocyte transendothelial migration (Figure 3.5) and PI3K-AKT signalling pathway (Figure 3.6). All the proteins (RHOA, NCF2, NCF4, GNAI3 and RAP1B) in the leukocyte transendothelial migration pathway except ITGA4 were overexpressed in the breast cancer cohort. It can be observed from Figure 3.5 that the leukocyte transendothelial pathway is promoted in the WBCs of breast cancer patients. It was previously reported that metastasis in melanoma patients was enhanced in the presence of neutrophils within the alveolar capillaries. Melanoma cells release interleukin 8 inducing neutrophils to overexpress β2 integrins and adhere to the vascular endothelium. The ICAM-1 on melanoma cells and the β2 integrins on neutrophils then used for these cells to bind and melanoma cells transverse and emigrate into lung tissue (Gregory and Houghton, 2011). It should be noted that the PI3K-AKT pathway is involved in the leukocyte transendothelial signalling (Figure3.5). The altered proteins involved in PI3K-AKT pathway were ITGA4, GNG5, GYS1, YWHAB, YWHAZ and IGF1R.GNG5, GYS1, YWHAZ and YWHAB were overexpressed in the WBCs of breast cancer patients. The PI3K-AKT pathway is a combination of insulin, MAPK and VEGF signaling pathways leading to cell growth progression, immune cell activation and increased cell migration (Koyasu, 2003).

Table 3.6 Top 10 Pathways obtained by comparison of Protein profiles of healthy donors and breast cancer patients: List of pathways and proteins involved in immune response to tumour obtained by high throughput proteomics by comparison of WBCs from healthy donor and breast cancer patients.

KEGG pathway	Proteins involved	Adj P value
Salivary secretion	ATP1B3, ATP1A1, CAMP, PRH1, ATP2B4, KCNMA1	1.33e-06
Fatty acid metabolism	ACADM, CPT1A, ECI1, ECHS1, ALDH9A1	1.33e-06
PI3K-AKTsignaling pathway	ITGA4, GNG5, GYS1, YWHAB, YWHAZIGF1R	1.89e-06
Valine, leucine and isoleucine degradation	ACADM, OXCT1, ECHS1, ALDH9A1	3.72e-05
Leukocyte transendothelial migration	RHOA, NCF4, GNAI3, ITGA4, RAP1B,NCF2	6.62e-05
Metabolic pathways	ALDH4A1, LSS, PGM1, PIGT, PPT1, ECHS1, NME2, MDH1, AHCY, ACADM, NDUFA10, ALDH9A1	6.94e-05
beta-Alanine metabolism	ACADM, ECHS1, ALDH9A1	0.0001
Proximal tubule bicarbonate reclamation	MDH1,ATP1B3, ATP1A1	0.0001
Propanoate metabolism	ALDH9A1, ECHS1, ACADM	0.0003
Pyruvate metabolism	MDH1, ME2, ALDH9A1	0.0005

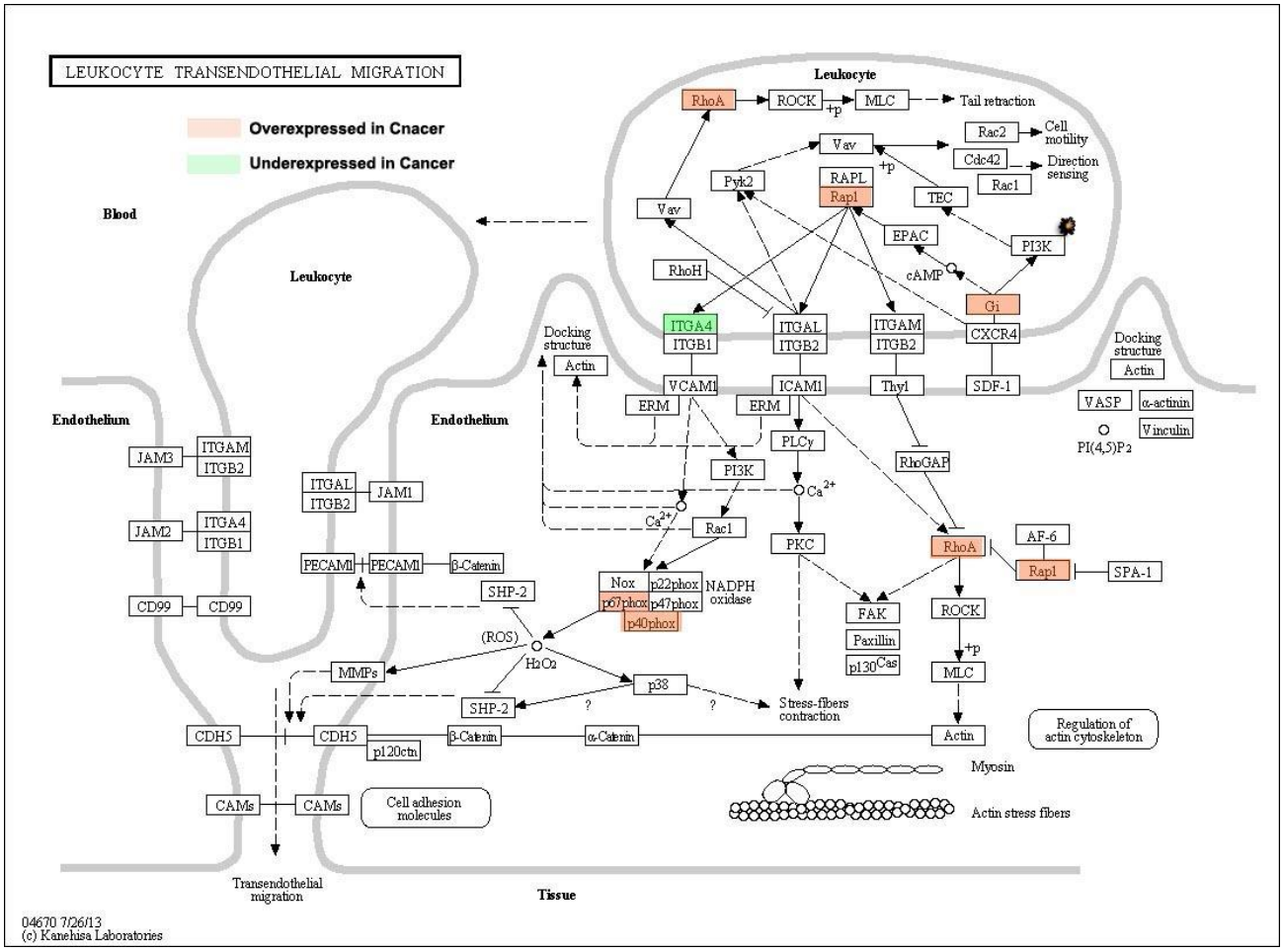


Figure 3.5 Regulation of the Leukocyte Transendothelial Migration Pathway in the WBCs of Breast Cancer Patients: Differentially expressed genes were obtained from the comparison between protein profiles of breast cancer patients and healthy donors. Pathway analysis was performed using the KEGG database. Red indicated overexpression and green underexpression in the breast cancer patients.

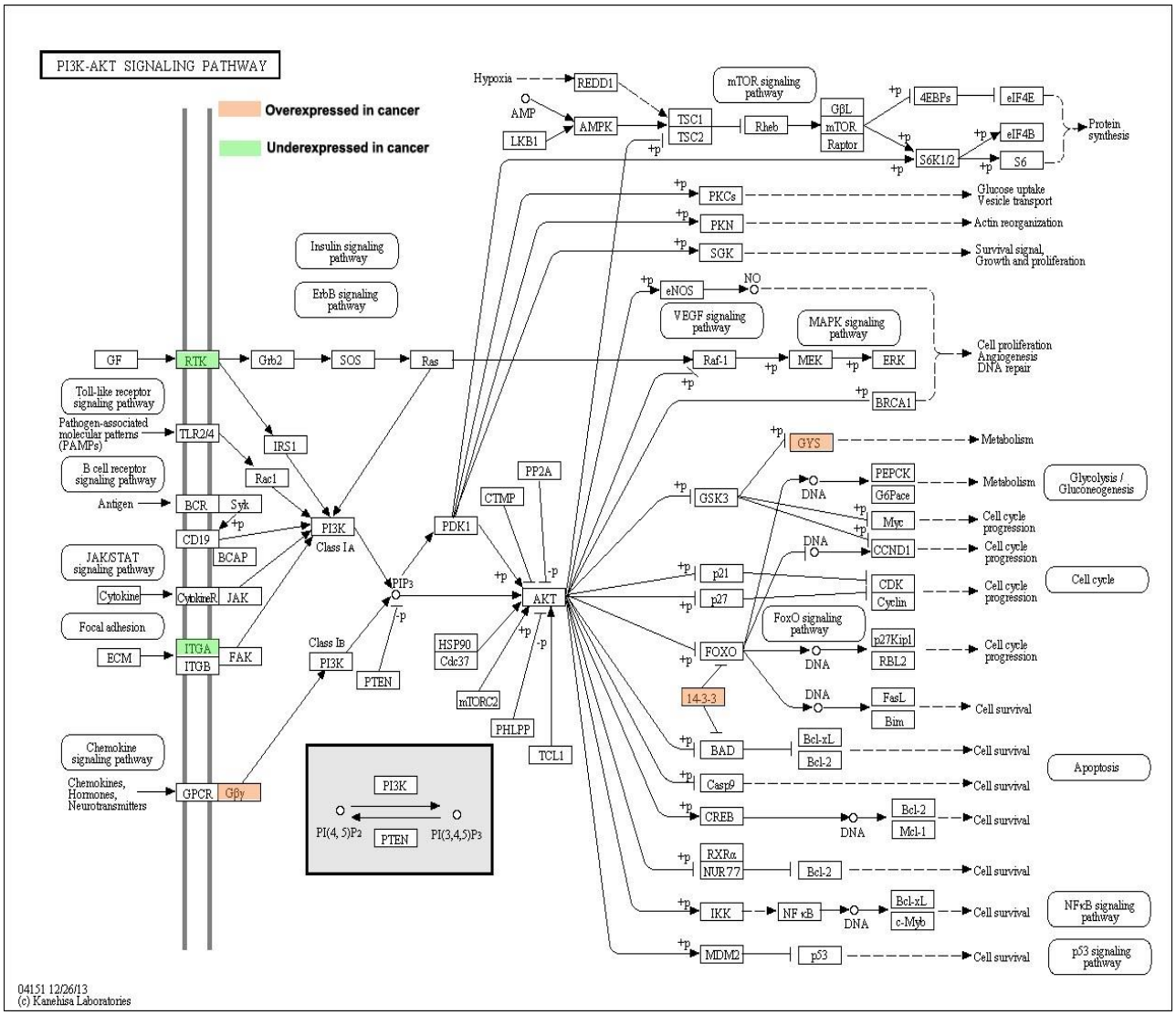


Figure 3.6 Regulation of the PI3K-Akt Signalling Pathway in the WBCs of Breast Cancer Patients: Differentially expressed genes were obtained from the comparison between protein profiles of breast cancer patients and healthy donors. Pathway analysis was performed using GENMAPP and the KEGG database. Red indicated overexpression and green underexpression in the breast cancer patients.

3.3.1.3.5 Protein-Protein interaction (PPI) analysis

Interactions between proteins are very important in many biological processes and the information about interacting protein partners can provide clues about potential protein functions. We therefore performed the protein-protein interaction (PPI) analysis of the identified proteins with the help of the Osprey tool used for visualization and manipulation of complex interaction networks (see section 2.4.3.4). The Osprey tool builds data-rich graphical representations from Gene Ontology (GO) annotated interaction data maintained by the BioGRID (available from http://thebiogrid.org/). The protein-protein interaction network was generated and the single nodes (proteins) which did not have any interaction were deleted (Figure 3.7).The hub genes were identified using this approach. The hub genes are those genes which are highly connected to other genes and when altered lead to drastic phenotypic changes in the cell which could determine the fate of the cell.

It was observed that 14-3-3 protein beta/alpha (YWHAB), insulin-like growth factor 1 receptor (IGF1R) and the Ras homolog gene family member A (RHOA) had the most number of interactions. These three proteins were overexpressed in the WBCs of breast cancer patients.

Other minor hub genes identified were Heterogeneous nuclear ribonucleoprotein A1 (HNRNPA1), AP-1 complex subunit beta-1 (AP1B1), SET Nuclear Oncogene (SET) and SEC13 Homolog (SEC13). YWHAB belongs to the 14-3-3 family of proteins involved in a wide range of cellular functions such as metabolism, protein trafficking, signal transduction, apoptosis and cell cycle regulation. YWHAB binds to RAF1 and CDC25 phosphatases suggesting it might play a role in cell proliferation (Conklin, 1995; Yuryev and Wennogle, 2003). IGF1R activates the PI3K-AKT signalling, JAK/STAT pathway and the Ras-MAPK pathway leading to cell growth/survival (Wilker et al., 2005). RHOA is a small GTPase protein which is involved in regulation of actin cytoskeleton rearrangement to form stress fibres during cell migration. RHOA

is also implicated in regulating the cell transformation and cell cycle progression (Klimov et al., 2013). The overexpression of all these proteins suggests that the WBCs of breast cancer patients are been driven towards cell proliferation and cell survival.

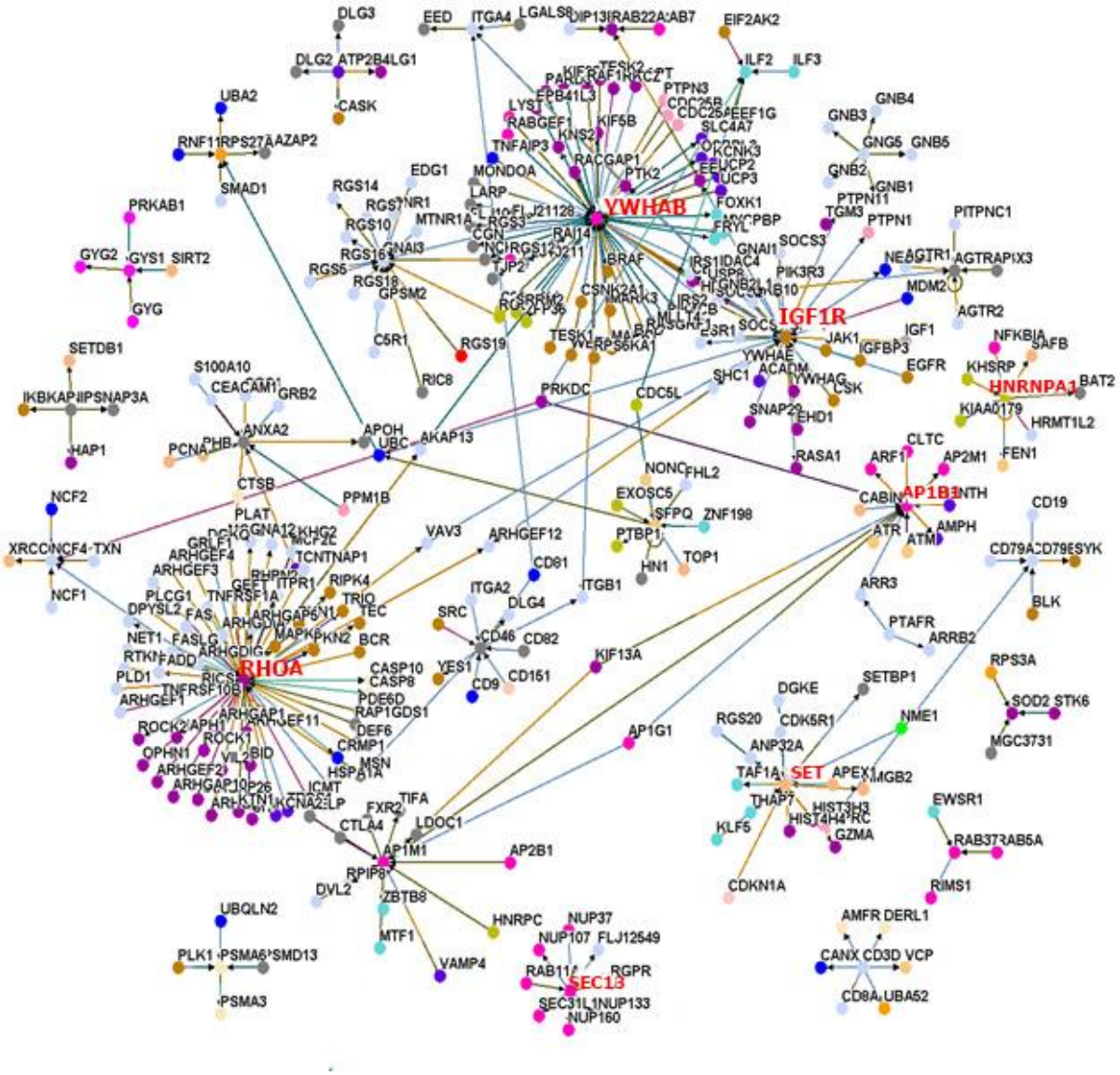
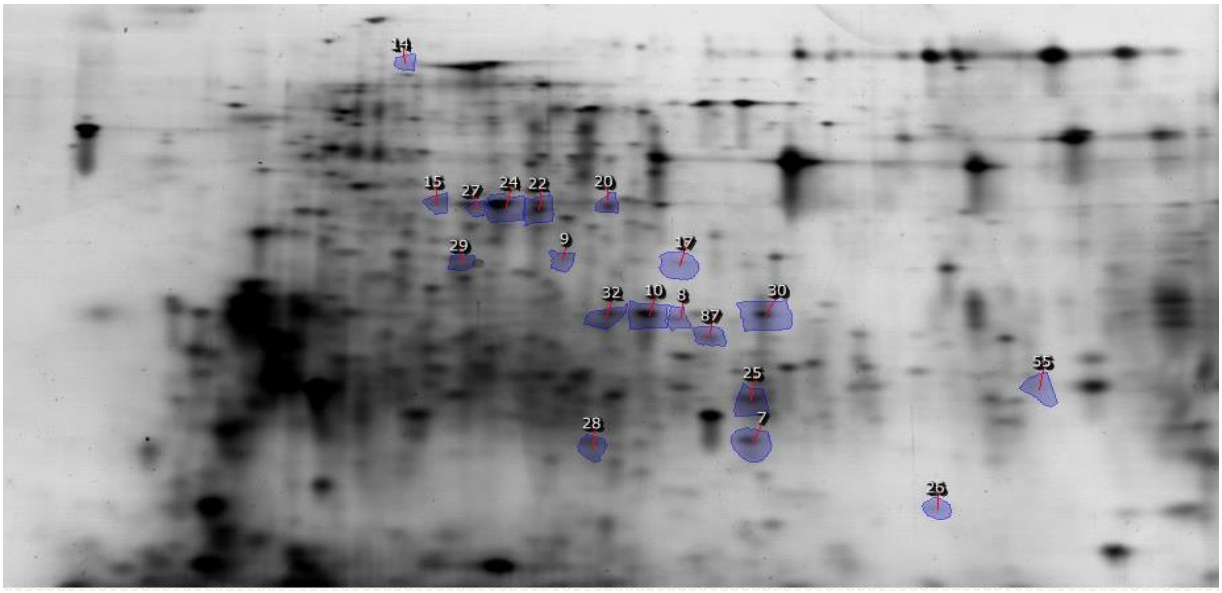


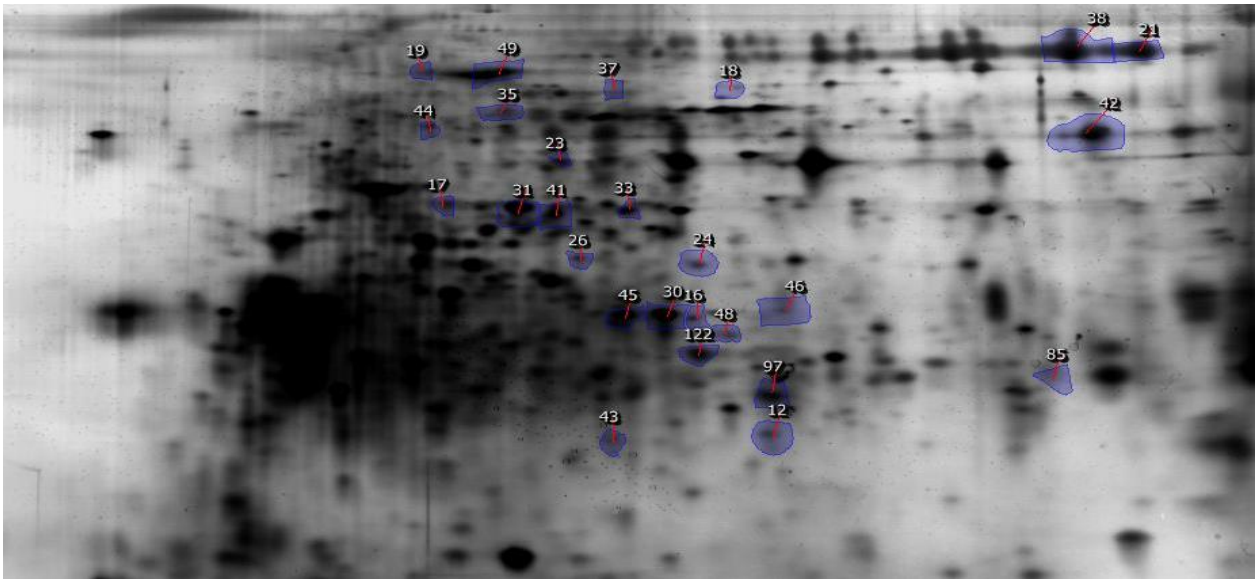
Figure 3.7 Identification Hub gene nodes changed in the WBCs of breast cancer patient cohort using high throughput protein data: Differentially expressed proteins in the WBCs of breast cancer patients were used to construct gene regulatory networks. Network was constructed using Osprey (http://biodata.mshri.on.ca/osprey/servlet/Index). The lines between gene nodes represent interactions and the genes with most interactions (hub genes) are independently with their own network.

3.3.2 Qualitative protein profiling using Two Dimensional (2D) Protein gel Electrophoresis

In this study an alternative method, the 2D-PAGE, was also used. To be resolved by 2D gels, the whole cell protein lysates from the WBCs of 10 breast cancer patients and 5 healthy donors were prepared according to procedure in Section 2.2.3.3. The sample details are given in Table 3.2. The protein lysates were then separated in the first dimension by the isoelectric point and in the second dimension by the molecular weight. All the samples were run in triplicates and the gels scanned after staining. The images were then analysed using the Samespots software (see section 2.2.3.2.4). Two types of analysis were performed on the 2D-gels based on the grouping of samples: (1) the disease state (Healthy donor vs Breast Cancer) and (2) tumour grade (Healthy donor, Low Grade and High Grade). The Samespot analysis based on disease state gave 18 spots which were significant and the analysis based on tumour grade gave 25 spots. The gel images showing the significant spots for both the analyses are presented in Figure 3.8. Between both analyses there was an overlap of 14 spots. The spots were excised out of the gels and given for mass spectrometric identification in Dr M. Metodiev's laboratory.



A



B

Figure 3.8 2D-PAGE gel separation of proteins identified with silver staining from WBCs of cancer patients and healthy donors. The proteins are separated by their isoelectric point in the first dimension and by molecular weight (*MW*) in the second dimension. Gels were run in triplicates and analysed with the Samespots software. The spots which were differentially present were obtained by comparing different conditions. The circled spots show proteins significantly different in the Disease State analysis **A** and Tumour Grade analysis **B.**

3.3.2.1 Protein Spot Identification and Functional analysis

The total numbers of unique protein identifications were14 out of 26 spots obtained from the Samespot analysis. There was more than one protein identification for certain spots. This can be attributed to post-translational modifications, the protein fragmentation during the sample preparation step or simply an error during mass spectrometric procedures. The protein ID which had the most number of peptide hits was taken as the spot identification. Some of the spots could not be identified due to low protein content in the spot. The results for the tumour grade analysis and diseased state analysis are given in Table 3.7 and Table 3.8 respectively. All the proteins picked up in the disease state analysis were present in the tumour grade analysis with the exception of 6-phosphogluconate dehydrogenase (PGD). In the tumour grade analysis only the following proteins showed a consistent trend with both low grade and high tumours (either increase or decrease); JUP, TKT, SERPINB1, CAPZA1, CORO1A and ARPC2 (Figure 3.9). The proteins which were overexpressed in low grade breast and underexpressed in high grade breast cancer cohort were ANXA1, PRH1, ITGAM and VCL. On the other hand the proteins which were underexpressed in the low grade breast cancer and overexpressed in high grade breast cancer cohort were CASP14 and CTNNG. The proteins which were common between both the types of analysis were TKT, SERPINB1, CAPZA1, CORO1A and ARPC2.

When gene ontology was performed as described in section2.4.3.2majority of the proteins identified are cytoskeletal proteins involved in actin rearrangement (Gene Ontology id:0005856); CASP14, ACTN1, CORO1A,ARPC2, VCL, ANXA1 and CAPZA1 (Table 3.9). The other categories obtained by gene ontology analysis were cell adhesion and cell migration. CAPZA and CORO1A are involved in the innate immune response (GO:0045087). ANXA1 and CORO1A are involved in leukocyte activation (GO:0045321). SERPINB1 is stored in the granules of neutrophils and it inhibits neutrophil elastase which has anti-microbial activity,

dissolves cellular junctions, increases cell migration etc (Chou et al., 2012). Except for ACTN1, VCL and CAPZA1 all the proteins which are involved in cell migration/adhesion were overexpressed in the WBCs of breast cancer patients. SERPINB1 and JUP affect cell migration in a negative manner (Chou et al., 2012; Lam et al., 2012).All the other proteins (CORO1A, ITGAM, VCL and ACTN1) involved in cell migration/adhesion enhance the process. Protein-Protein interaction network was constructed using the significant genes identified through 2D-PAGE analysis as described in Section 2.4.3.4. It was observed that Plakoglobin (JUP) had the most number of interactions. Plakoglobin is a tumour growth and metastasis suppressor (Aktary and Pasdar, 2012; Lam et al., 2012). From the data obtained in this analysis a conclusion could be drawn on the effect of tumour on WBC migration in cancer patients.

Table 3.7 Mass spectrometric Identification of 2D-PAGE spots from Tumour Grade analysis: List of Protein spots and their Protein Ids differentially present in the WBCs of breast cancer patients as identified through mass spectrometry obtained by comparing healthy donors, Low grade breast cancer and high grade breast cancer patients. The spot numbers correspond to the numbers in the spotpickingimageFigure3.8 B. FC-Fold Change in comparison with healthy donors. Red- FC>1.2, Pink-0<FC<1.2, Green FC<-1.2, Light green -1.2<FC<0.

Grade Analysis Spot number	Protein Identification	Protein Name	Average Normalised Spot Volumes			FC Low Grade	FC High Grade
			Healthy Donor	Low Grade	High Grade		
17	Junction Plakoglobin	JUP	2.136	2.83	3.398	+1.3	+1.6
19	Immunoglobulin J chain		14.64	16.57	21.6	+1.1	+1.3
21	Caspase 14	CASP14	1.836	1.349	2.142	-1.4	+1.2
23	Integrin Alpha M	ITGAM	1.723	2.503	1.641	+1.4	-1.04
30	Transketolase	TKT	9.738	14.3	14.42	+1.5	+1.5
31	Glycogen phosphorylase	PYGL	3.176	2.329	2.003	-1.4	-1.16
35	Vinculin	VCL	7.864	8.853	6.143	+1.12	-1.3
37	Gamma-catenin	CTNNG	2.595	1.851	2.655	-1.4	+1.02
38	Annexin A1	ANXA1	22.9	29.05	20.41	+1.3	-1.12
41	Neutrophil Elastase Inhibitor	SERPINB1	8.138	10.52	11.26	+1.3	+1.4
42	Parotid acidic protein	PRH1	26.81	32.2	23.39	+1.2	-1.15
43	F-actin-capping protein subunit alpha-1	CAPZA1	3.439	2.903	2.513	-1.4	-1.4
44	Annexin A1	ANXA1	2.222	2.836	2.083	+1.3	-1.06
45	Coronin 1A	CORO1A	5.813	6.18	7.797	+1.06	+1.3
46	Actin-related protein 2/3 complex subunit 2	ARPC2	18.22	15.69	13.62	-1.16	-1.3
49	Actinin alpha 1	ACTN1	25.38	25.32	19.87	1	-1.3
97	Immunoglobulin J chain		10.54	8.362	7.899	-1.3	-1.3
122	Phosphoglyceratemutase 1	PGAM1	9.551	9.258	7.534	1	-1.3

Table 3.8 Mass spectrometric Identification of 2D-PAGE spots from Disease state analysis; List of Protein spots and their Protein Ids differentially present in the WBCs of breast cancer patients as identified through mass spectrometry obtained by comparing healthy donors and breast cancer patients. The spot numbers correspond to the numbers in the spotpicking image Figure 3.8 A. FC-Fold Change in comparison with healthy donors. Red- FC>1.2, Green FC< -1.2.

Disease State Analysis Spot number	Protein Identification	Protein Name	Average Normalised Spot Volumes		Fold Change in Cancer
			Healthy Donor	Cancer	
14	Immunoglobulin J chain		3.17	2.16	-1.5
29	F-actin-capping protein subunit alpha-1	CAPZA1	3.53	4.52	-1.4
30	Actin-related protein 2/3 complex subunit 2	ARPC2	18.2	14.63	-1.2
10	Transketolase	TKT	9.72	14.34	+1.5
32	Coronin 1A	CORO1A	5.80	6.98	+1.2
22	Glycogen phosphorylase	PYGL	8.13	10.7	+1.3
24	Neutrophil Elastase Inhibitor	SERPINB1	14.6	19.05	+1.3
26	6-phosphogluconate dehydrogenase	PGD	1.33	1.71	+1.3

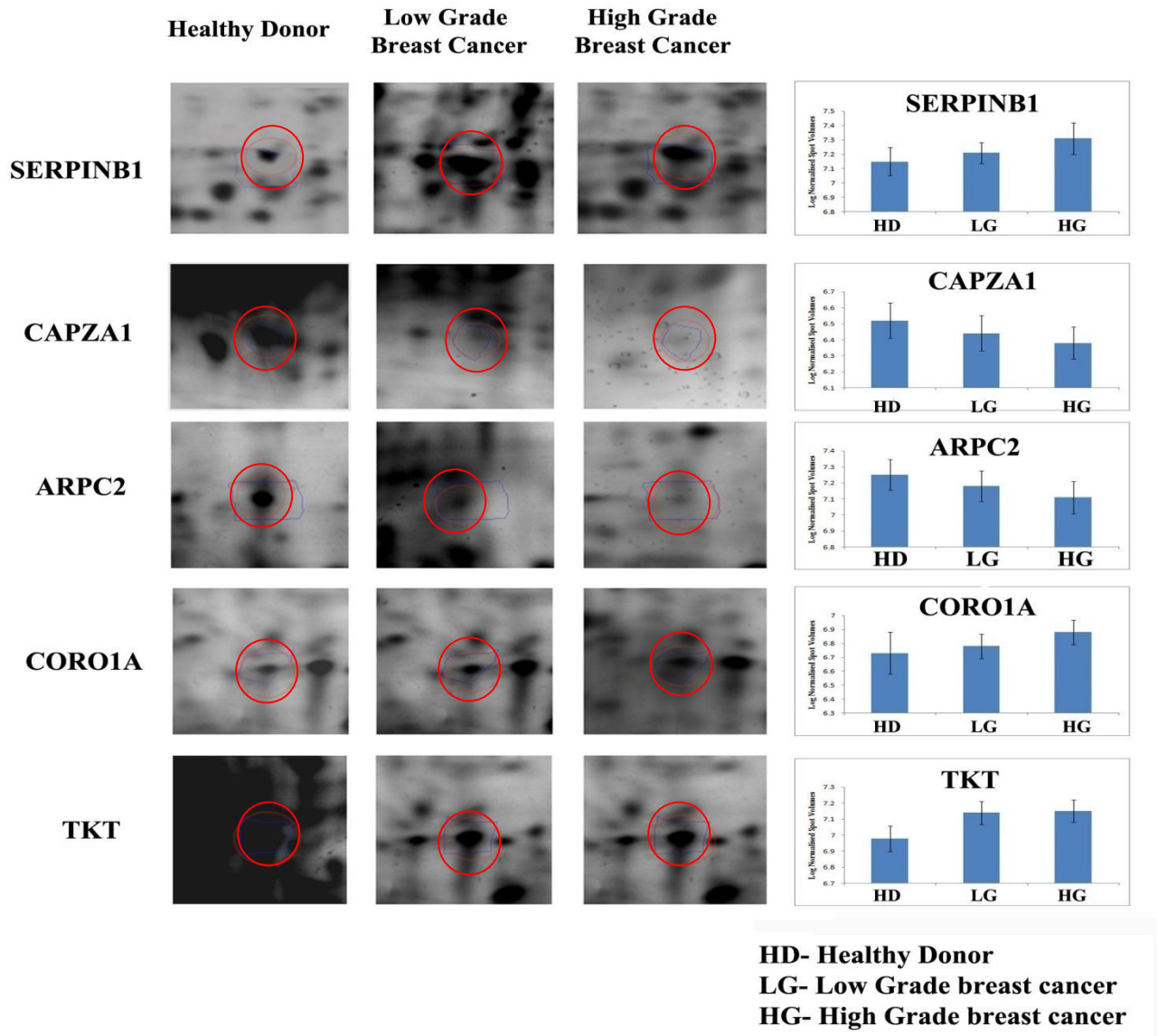


Figure 3.9 Protein Spots common between disease state and tumour grade analysis identified by 2D-PAGE: WBC sample lysates were run in triplicates to perform 2D-gel electrophoresis. Scanned protein gels were analysed using Samespots software. Two types of analysis was done; Disease state (Healthy donor vs Breast cancer) and Tumour grade (Healthy donor vs Low grade breast cancer vs High grade breast cancer). Proteins spots represented above were common between both analyses and also showed consistent trends in different samples between different grades in breast cancer patient and healthy donor cohorts.

Table 3.9 Gene ontology of the significant proteins identified through 2D-PAGE: White blood cell lysates were subjected to 2D-PAGE. Samples were run in triplicates and the scanned gels were analysed by Samespots software. Protein spots were identified using the Orbitrap mass spectrometer. Proteins identified to be differentially present in the WBCs of breast cancer patient and healthy donors were subjected to gene ontology analysis. Gene Ontology was performed using GOFFA and based on Biological process.

Gene ontology Term	Gene ontology ID	Proteins involved
Immune system Process	GO:0002376	ANXA1,CAPZA1,CORO1A,ITGAM
Cell death	GO:0016265	ACTN1,ANXA1,CASP14
Cell proliferation	GO:0008383	ANXA1,CORO1A
Actin cytoskeleton	GO:0015629	ACTN1,ANXA1,ARPC2,CAPZA1,CORO1A,JUP,VCL
Cell Adhesion	GO:0007155	ACTN1,CORO1A,ITGAM,JUP,VCL
Cell migration	GO:0016477	CORO1A,ITGAM,JUP,VCL

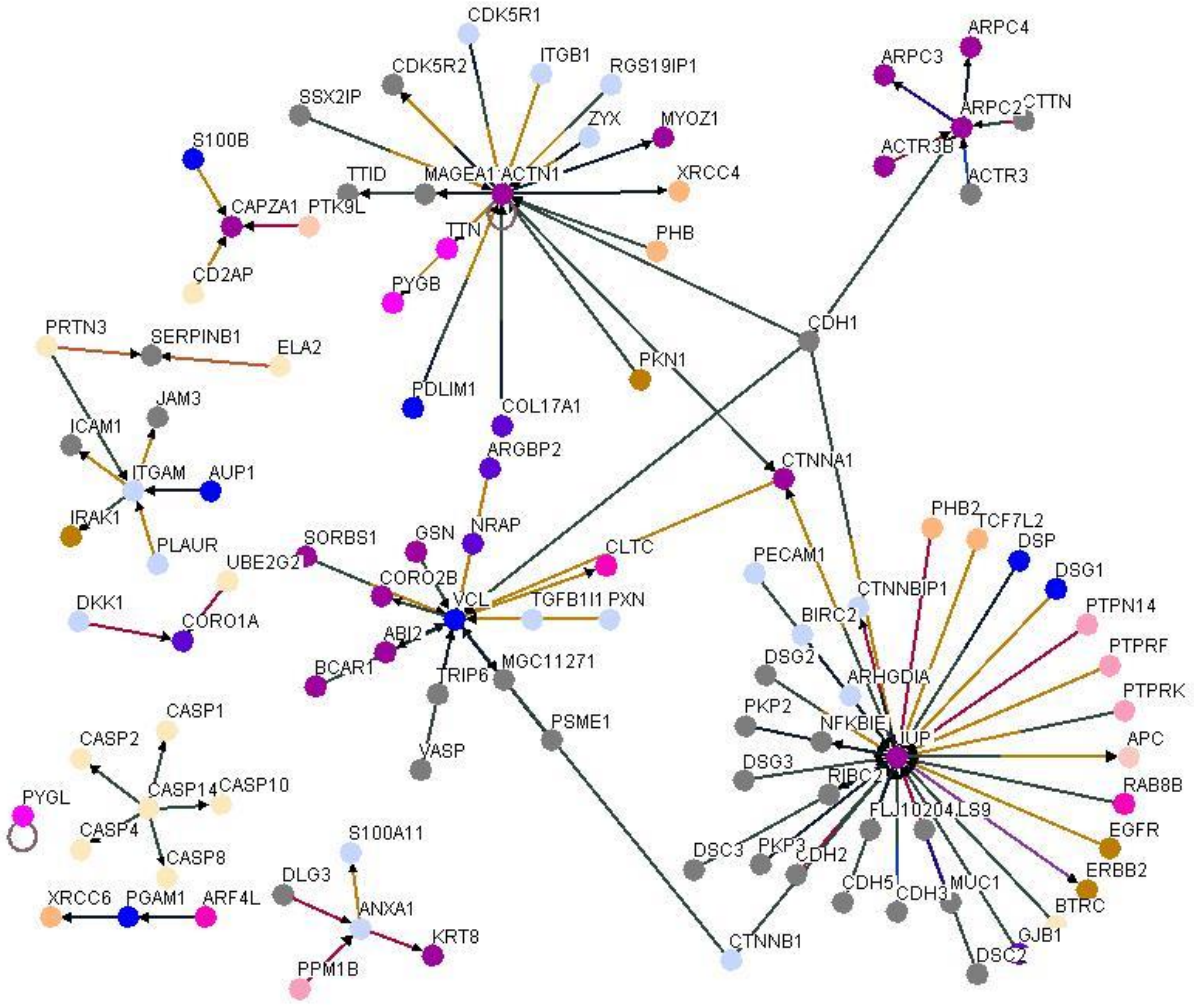


Figure 3.10 Identification Hub gene nodes changed in the WBCs of breast cancer patient cohort using 2D-PAGE: Differentially expressed proteins in the WBCs of breast cancer patients obtained by 2D-PAGE method were used to construct gene regulatory networks. Networks were constructed using the Osprey tool (http://biodata.mshri.on.ca/osprey/servlet/Index). The lines between gene nodes represent interactions and the genes with most interactions (hub genes) are independently with their own network.

3.4 Discussion

This study utilised proteomics approaches to identify differences in protein profiles of WBCs between breast cancer patients and healthy donors. Two types of technologies were used to identify differences in WBCs protein profiles: high throughput analysis using the Orbitrap mass spectrometer and the 2D-PAGE. These proteomic approaches are becoming very popular in finding novel cancer biomarkers in tissue and blood and give advantage over gene expression studies since protein biomarkers are preferred in clinical settings. They are also easier to detect due to well-developed antibody assays such as western blotting, enzyme-linked immunosorbent assay, antibody arrays etc. It should be noted that both methods have limitations, for example, the 2D-PAGE has reduced sensitivity and resolution, low throughput, requirement of the large amounts of starting material, more inter-gel variations and increased chances of experimental errors (Verrills, 2006). Compared with 2D-PAGE the high throughput proteomics has a better sensitivity and can be automated, however, the data mining, analysis and validation represents a challenge. The high throughput studies which were performed using the Orbitrap Velos instrument were very sensitive and even identified very specific changes. Around 2500 proteins were detected only in the membrane fraction of WBCs. For example ITGA4 (also known as CD49d) is present predominantly in the mononuclear cell fraction of the WBCs was found to be underexpressed in the breast cancer patient cohort. In our experiments, this analysis also showed variation in individual protein levels within the same cohort. This may be due to the heterogeneous composition of the WBCs and also the dynamic nature of the immune system. This can be solved by increasing the sample numbers which would provide more statistically significant proteins in cancer patients. Furthermore, the number of significant proteins identified through high throughput proteomics was much higher than 2D-PAGE analysis. However, these data will still need to be validated using other techniques such as western blotting, immunohistochemistry and others.

One of the reasons the two techniques were used independently in our studies was to identify common proteins. Three candidates, SERPINB1, JUP and ANXA1, appeared in both analyses with SERPINB1 and ANXA1 overexpressed in the breast cancer cohort. All the three proteins are involved in cell adhesion and cell migration in immune cells. SERPINB1, also known as neutrophil elastase inhibitor, is a serine protease inhibitor which inhibits the neutrophil-derived proteinases neutrophil elastase, cathepsin G, and proteinase-3 whereby protecting tissues from damage at inflammatory sites. In this study SERPINB1 levels were found to be increased in the WBCs of breast cancer patients using both approaches (2D-PAGE and high throughput proteomics).The link between SERPINB1 and cancer has been reported previously. For example, decreased expression of SERPINB1 in hepatocellular carcinoma positively correlates with poor prognosis (Cui et al., 2014). SERPINB1 was also found to positively correlated with cell invasiveness in oral cancer cell lines (Tseng et al., 2009).

ANXA1 (Annexin A1) is a protein with the phospholipase A2 inhibitory activity. Phospholipase A2 is required for the biosynthesis of important mediators of inflammation such as prostaglandins and leukotrienes, hence, ANXA1 may have potential anti-inflammatory activity (Perretti and D'Acquisto, 2009). ANXA1 was also found to aberrantly overexpressed in highly invasive basal like and HER2 positive breast tumours (Yousef et al., 2013). Other proteins related to leukotriene synthesis which changed in the breast cancer patient cohort were ALOX5 (Arachidonate 5-lipoxygenase) and LTA4H (Leukotriene A4 hydrolase); were overexpressed in the breast cancer patient cohort especially in the high grade category. CORO1A (Coronin-1A) is essential for T-cell mediated immunity and was present in higher levels in the breast cancer patient cohort in the 2D-PAGE analysis. It was found that CORO1A plasma levels increase in mice with tumour which corresponds to experiments done in white blood cells using 2D-PAGE (Pitteri et al., 2008).

The high throughput proteomics has provided a list of statistically significant 136 proteins differentially present between the WBCs of breast cancer patients and healthy donors. It is very important to note that these data only include the membrane proteins; the data from the soluble fraction were not analysed. One of the reasons for this was the presence of abundant proteins which masked the signals from other proteins of low abundance in this fraction. The Gene ontology analysis of the identified proteins revealed that 34 of those were involved immune response and 41 in cellular localisation; generally, this analysis showed that inflammatory response was enhanced in the WBCs of breast cancer patients. This finding supports the theory that inflammation contributes to generation of tumour-related mutations and promotes tumour progression (Grivennikov et al., 2010; Jackson et al., 1997; Nowarski et al., 2013). The Gene ontology analysis on the high throughput data also demonstrated that around 11 proteins were involved in the cell adhesion and migration. The ability of the immune cells to probe tissues for infection monitoring is very important. This is achieved by the immune cells by movement into tissues called transendothelial migration (Luster et al., 2005). If this movement of cells is affected the immune function can be compromised.

The Pathway analysis using KEGG database revealed leukocyte transendothelial migration was positively regulated in the breast cancer cohort. The Leukocyte Transendothelial migration is the process by which WBCs invade tissue through the endothelial layer to routinely search for any inflammatory signals or infections (Van Buul and Hordijk, 2004). This analysis also revealed that the PI3K pathway which is associated with the transendothelial migration pathway was activated through the overexpression of chemokine receptors. Breast cancer cells overexpress SDF-1α which chemotactically attracts lymphocytes and macrophages to sites of tumour enhancing the tumour metastasis (Müller et al., 2001). Leukocytes and macrophages have been implicated in enhancing tumour angiogenesis/metastasis through their recruitment through tumour secreted chemokines such as IL6, granulocyte macrophage colony stimulating

factor, SDF-1α etc (Gabrilovich et al., 2012). The 2D-PAGE analysis also complemented the high throughput study where 9 out of 18 proteins identified were involved in actin cytoskeletal rearrangement (ACTN1, ITGAM, CORO1A, ARPC2, JUP, ANXA1 CAPZA1 and VCL) and 4 proteins (ACTN1, ITGAM, SERPINB1and VCL) were involved in leukocyte transendothelial migration. The effect of the tumour tissue on the movement of leukocytes into the endothelium is implicated in cancer. Leukocyte adherence inhibition (89%) was observed when leukocytes from cancer patients were treated with tumour extracts (Halliday et al., 1980). Also tumour cells mimic immune cells by expressing surface molecules found on WBCs when they invade the extracellular matrix to metastasise (Jöhrer et al., 2004; Onrust et al., 1996).

This proteomics analysis will yield better results if it can be repeated using the datasets as the training dataset with the increased sample size. Gene expression data will give the changes in all the gene mRNA levels which will include secreted, membrane bound and soluble proteins and the proteomics data has to be linked up with the gene expression data to map out the immune response to tumour. This will also help in short listing the markers for validation at both RNA levels and protein levels.

Chapter 4: Comparative analysis of gene expression profiles of white blood cells in breast cancer patients and healthy donor

4.1 Introduction

The use of gene expression microarray and RNA-Seq technologies in the effort to identify biomarkers has been employed for various types of cancers (See Section 1.8.2 and references therein). In these studies, tissue samples were often used to identify genes differentially expressed in cancers and, ultimately, discover novel tissue cancer biomarkers. However, tissue biomarkers have limitations which include the costs and invasiveness of the biopsy procedures; in addition tumour tissue is only guaranteed at diagnosis, when the tumour is biopsied or resected.

Blood can used as an alternative source for biomarkers, as blood collection is less invasive and cheap. However, current serum markers for breast cancer such as CA27.29 and CA15-3 lack sensitivity for early diagnosis of breast cancer. Despite of that the search for novel blood plasma cancer biomarkers continues, for example for breast cancer (Pitteri et al., 2008, 2011) and Non-small cell lung cancer (Izbicka et al., 2012). As an alternative to serum or plasma, white blood cells (WBC) can be used to identify new biomarkers for cancer. A number of studies have been undertaken to screen for biomarkers for the purpose of both diagnosis and prognosis for cancer using genomic approaches in the WBCs for breast cancer (Aarøe et al., 2010; Komatsu et al., 2012; LaBreche et al., 2011; Sharma et al., 2005; Zuckerman et al., 2013), pancreatic cancer (Baine et al., 2011), Non-small cell lung cancer (Showe et al., 2009) and acute myeloid leukemia (Metzeler et al., 2008). There also have been efforts to prove DNA methylation patterns in WBCs may be useful in detecting cancer (Brennan et al., 2012).

In the study reported by Aarøe et al., 2010 blood samples from 67 female patients with invasive breast cancer, 54 healthy donors and 9 other control samples (pregnant women and

benign breast cancer) were collected. The whole WBC fraction was isolated and RNA was extracted. Gene expression analysis was done using the single channel Applied Biosystems Human Genome Survey microarray V.2 chips. The group used partial least squares regression modelling and suggested 208 core regulated probes which discriminate breast cancer patients from healthy controls. The functional analysis done by the group was limited and showed very few functionally enriched categories. Furthermore the aim of their analysis was only to identify gene expression patterns in the WBCs of breast cancer patients without validation of the data. These gene expression data will be analysed in this chapter to identify a list of potential biomarkers and to perform a independent functional analysis on the data.

4.2 Aims

The main aim of this Chapter is to analyse the gene expression patterns of WBCs from breast cancer patients from the previously published study (Aarøe et al. 2010). The gene expression data were deposited in the Gene expression omnibus (GEO) database (the GEO identity number is GSE16443). These data will be used to analyse for difference in global gene expression between helathy donors and breast cancer patients. This list of genes differentially expressed in the breast cancer patient cohort will be used for further functional analysis to identify possible biological pathways involved in the response of WBCs to the breast cancer. The latter included gene ontology, pathway and protein-protein interaction analyses. The Gene ontology and pathway analyses will be primarily focussed on immune response processes. The differentially expressed genes will also be considered as potential biomarker candidates for further validation. The results from this analysis will also be combined with the results obtained in Chapter 3 for biomarker identification and maping out the changes in immune response in breast cancer patients.

4.3 Results

4.3.1 Comparison of gene expression profiles between peripheral Blood cells of breast cancer patients and healthy donors.

The gene expression data were obtained from the GEO using the GEO id GSE16443 in the form of log2 normalised gene expression value for each sample and gene in matrix. The numbers of breast cancer patient samples were 67 and healthy donors were 54. Students' paired T-test was done using Multiexperiment viewer (see Section 2.4.3.1) between the two groups to obtain significant probes with fold change greater than ±1.2 and p-value less than 0.05. Genes which had more than one probe were filtered and the most significant fold change taken for further analysis. The test yielded 506 genes being significant with 214 genes under expressed and 292 genes overexpressed in the breast cancer patient cohort. Figure 4.1 is the volcano plot showing significant probes in blue which are being overexpressed in breast cancer patients with the fold change greater than ±1.2 and pvalue>0.05. Table4.1 gives the list of all the genes along with their fold change that were changed in the WBCs of breast cancer patients and are arranged in alphabetical order of their gene symbols. Hierarchical clustering for the 506 significant probes (Figure 4.2) was done using Multiexperiment viewer as discussed in Section2.4.3.1. It was observed that intensity values had a high degree of variance. This can be explained due to the fact that the molecular pathology of breast cancer is different to each tumour and also the immune system is very dynamic.

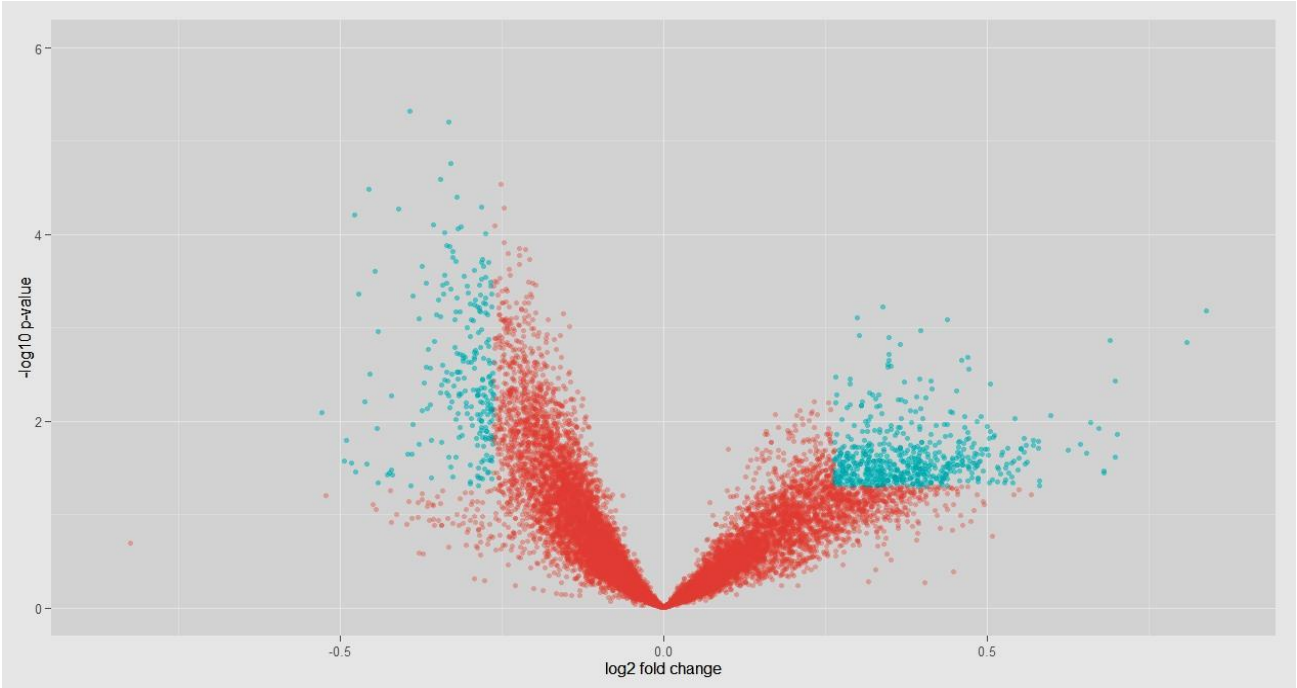


Figure 4.1 Volcano plot of the probes in GSE16443 dataset: Student's T test was performed to obtain probes which are significantly different between breast cancer patients and healthy donors. Individual probes are represented as dots. Blue dots represent probes which have p value <0.05 and fold change of ±1.2 which is relative to the breast cancer patient cohort.

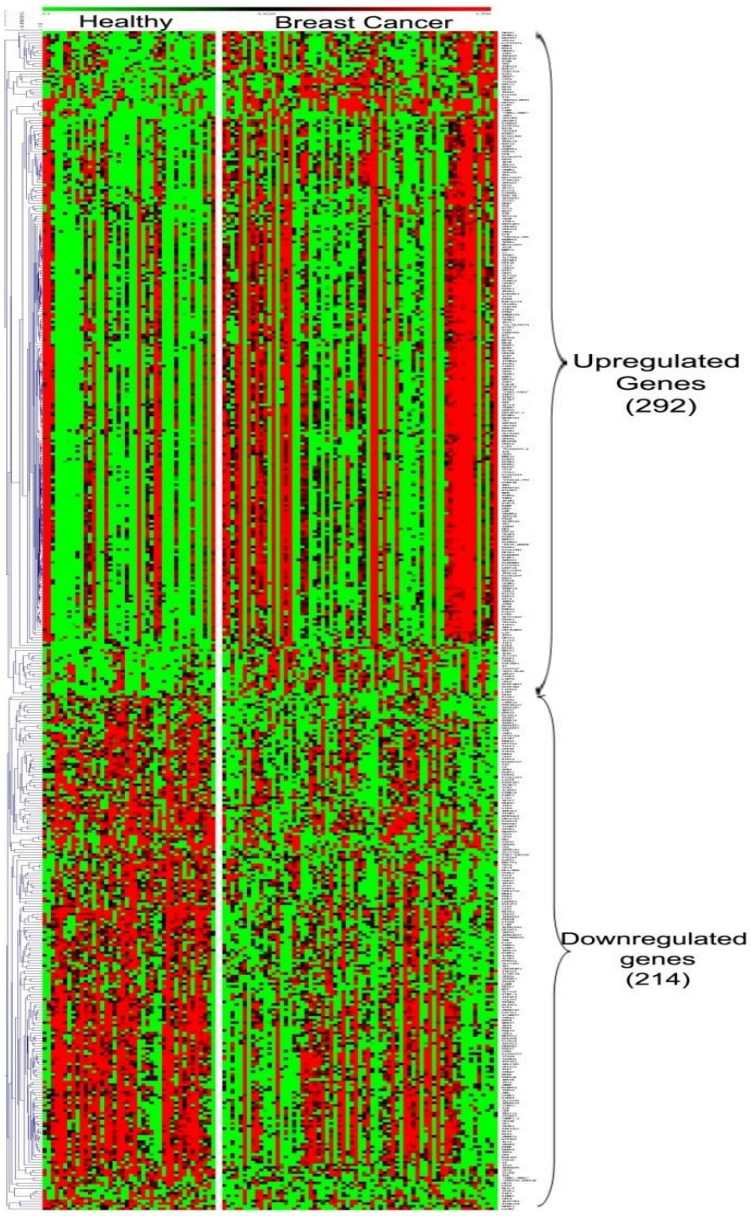


Figure 4.2 Hierarchical Clustering of significant genes differentially expressed in the WBCs of Breast Cancer Patients: Hierarchical Clustering was done using the 506 genes which had significant differences between Healthy donors and Breast cancer Patients as identified by T test. Mutiexperiment viewer was used to perform clustering. Red indicates that the gene is relatively overexpressed and green the vice versa in the breast cancer cohort.

Table 4.1 List of genes changed significantly in the WBCs of breast cancer patients: Students T-test was done on the GSE16443 dataset comparing between the expression profiles of WBCs of breast cancer patients and healthy donors. The p value cutoff for the test was 0.05 and fold change ±1.2. The test gave 506 genes being significantly different in breast cancer cohort.

Gene symbol	Gene name	FC
AANAT	arylalkylamine N-acetyltransferase	1.34
ABP1	amiloride binding protein 1 (amine oxidase (copper-containing))	1.21
ACAD8	"acyl-Coenzyme A dehydrogenase family, member 8"	-1.20
ACATE2	"likely ortholog of mouse acyl-Coenzyme A thioesterase 2, mitochondrial"	-1.21
ACBD6	acyl-Coenzyme A binding domain containing 6	-1.31
ACE	angiotensin I converting enzyme (peptidyl-dipeptidase A) 1	1.25
ACPT	"acid phosphatase, testicular"	1.26
ADCK2	aarF domain containing kinase 2	-1.21
ADRA2B	"adrenergic, alpha-2B-, receptor"	1.33
ADRA2C	"adrenergic, alpha-2C-, receptor"	1.34
ADRB1	"adrenergic, beta-1-, receptor"	1.38
ADSS	adenylosuccinate synthase	-1.28
AF1Q	ALL1-fused gene from chromosome 1q	-1.20
AFG3L1	AFG3 ATPase family gene 3-like 1 (yeast)	-1.25
AGPAT5	"1-acylglycerol-3-phosphate O-acyltransferase 5 (lysophosphatidic acid acyltransferase, epsilon)"	-1.23
AGR2	anterior gradient 2 homolog (Xenopuslaevis)	1.40
AGRP	agouti related protein homolog (mouse)	1.32
AIM1L	absent in melanoma 1-like	1.25
AIP1	atrophin-1 interacting protein 1	1.29
AK2	adenylate kinase 2	-1.23
ALDH18A1	"aldehyde dehydrogenase 18 family, member A1"	-1.23
ALPL	"alkaline phosphatase, liver/bone/kidney"	1.54
ALS2CR3	"amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 3"	-1.20
AMELX	"amelogenin (amelogenesisimperfecta 1, X-linked)"	1.31
AMELY	"amelogenin, Y-linked"	1.25
AMY2B	"amylase, alpha 2B; pancreatic"	-1.24
ANKRD10	ankyrin repeat domain 10	-1.26
ANXA3	annexin A3	1.97
AP1M2	"adaptor-related protein complex 1, mu 2 subunit"	1.32
AP3M2	"adaptor-related protein complex 3, mu 2 subunit"	1.26
APOC2	apolipoprotein C-II	1.35
APOC4	apolipoprotein C-IV	1.22
APOE	apolipoprotein E	1.36
APOL3	"apolipoprotein L, 3"	-1.22
ARFGEF2	ADP-ribosylation factor guanine nucleotide-exchange factor 2 (brefeldin A-inhibited)	1.27
ARHGAP17	Rho GTPase activating protein 17	-1.21
ARHGDIG	"protein disulfide isomerase, pancreatic,Rho GDP dissociation inhibitor	1.29

	(GDI) gamma"	
ARHGEF3	Rho guanine nucleotide exchange factor (GEF) 3	-1.20
ARHGEF9	Cdc42 guanine nucleotide exchange factor (GEF) 9	-1.21
ARL11	ADP-ribosylation factor-like 11	1.29
ARL6IP6	ADP-ribosylation-like factor 6 interacting protein 6	-1.21
ASIP	"agouti signaling protein, nonagouti homolog (mouse)"	1.29
B3GAT3	"beta-1,3-glucuronyltransferase 3 (glucuronosyltransferase I)"	1.31
BASP1	"brain abundant, membrane attached signal protein 1"	1.23
BAZ2A	"bromodomain adjacent to zinc finger domain, 2A"	-1.21
BCL7A	B-cell CLL/lymphoma 7A	1.25
BCNP1	B-cell novel protein 1	1.30
BM88	BM88 antigen	1.28
BNIP3L	BCL2/adenovirus E1B 19kDa interacting protein 3-like	1.29
BOMB	BH3-only member B protein	1.28
BRSK2	BR serine/threonine kinase 2	1.29
BTBD2	BTB (POZ) domain containing 2	1.27
BTC	betacellulin	1.20
C10orf33	chromosome 10 open reading frame 33	-1.28
C10orf47	chromosome 10 open reading frame 47	-1.22
C11orf16	chromosome 11 open reading frame 16	1.33
C19orf25	chromosome 19 open reading frame 25	1.39
C1orf19	chromosome 1 open reading frame 19	-1.27
C2	complement component 2	1.30
C20orf175	chromosome 20 open reading frame 175	1.22
C21orf106	chromosome 21 open reading frame 106	1.32
C21orf88	chromosome 21 open reading frame 88	1.30
C22orf3	chromosome 22 open reading frame 3	-1.22
C2orf22	chromosome 2 open reading frame 22	-1.29
C5orf13	chromosome 5 open reading frame 13	-1.21
C5orf15	chromosome 5 open reading frame 15	-1.31
C5orf18	chromosome 5 open reading frame 18	-1.22
C6orf110	chromosome 6 open reading frame 110	1.40
C6orf209	chromosome 6 open reading frame 209	-1.21
C8orf21	chromosome 8 open reading frame 21	1.25
CAMP	cathelicidin antimicrobial peptide	1.99
CARD15	"caspase recruitment domain family, member 15"	-1.36
CASP5	"caspase 5, apoptosis-related cysteine protease"	1.24
CCL1	chemokine (C-C motif) ligand 1	1.20
CCNG1	cyclin G1	-1.24
CD3Z	"CD3Z antigen, zeta polypeptide (TiT3 complex)"	-1.21
CD68	CD68 antigen	-1.21
CD74	"CD74 antigen (invariant polypeptide of major histocompatibility complex, class II antigen-associated)"	-1.22
CD9	CD9 antigen (p24)	-1.32
CDC2L1	"cell division cycle 2-like 1 (PITSLRE proteins),cell division cycle 2-like 2 (PITSLRE proteins)"	1.39
CDC42BPG	CDC42 binding protein kinase gamma (DMPK-like)	1.33

CDH15	"cadherin 15, M-cadherin (myotubule)"	1.23
CDK5R1	"cyclin-dependent kinase 5, regulatory subunit 1 (p35)"	1.31
CDK5RAP1	CDK5 regulatory subunit associated protein 1	-1.28
CDKL5	cyclin-dependent kinase-like 5	1.31
CEI	CEI protein	1.35
CEP1	centrosomal protein 1	-1.22
Cep192	centrosomal protein 192 kDa	-1.21
CG018	hypothetical gene CG018	-1.21
CGB8	"chorionic gonadotropin, beta polypeptide 8,chorionic gonadotropin, beta polypeptide 5"	1.42
CHGA	chromogranin A (parathyroid secretory protein 1)	1.31
CHMP1.5	CHMP1.5 protein	-1.21
CKLF	chemokine-like factor	1.28
CLDN11	claudin 11 (oligodendrocyte transmembrane protein)	1.22
CLEC12A	"C-type lectin domain family 12, member A"	1.60
CLEC4G	"C-type lectin superfamily 4, member G"	1.26
CNFN	cornifelin	1.21
COG5	component of oligomericgolgi complex 5	-1.26
CPA3	carboxypeptidase A3 (mast cell)	-1.34
CPNE3	copine III	-1.20
CREB5	cAMP responsive element binding protein 5	1.32
CRHR1	corticotropin releasing hormone receptor 1	1.35
CS	citrate synthase	-1.25
CSTA	cystatin A (stefin A)	1.32
CTAG1A	"cancer/testis antigen 1A,cancer/testis antigen 1B,cancer/testis antigen 2"	1.32
CTSC	cathepsin C	-1.23
CTSO	cathepsin O	-1.30
CXXC5	CXXC finger 5	-1.22
D4ST1	dermatan 4 sulfotransferase 1	-1.21
DATF1	death associated transcription factor 1	-1.23
DCN	decorin	1.24
DEFA3	"defensin, alpha 1, myeloid-related sequence,defensin, alpha 3, neutrophil-specific"	1.86
DERA	2-deoxyribose-5-phosphate aldolase homolog (C. elegans)	-1.21
DKFZp434I099	hypothetical protein DKFZp434I099	1.22
DMRTB1	"DMRT-like family B with proline-rich C-terminal, 1"	1.23
DOC2A	"double C2-like domains, alpha"	1.24
DOCK10	dedicator of cytokinesis 10	-1.37
DPAGT1	dolichyl-phosphate (UDP-N-acetylglucosamine) N-acetylglucosaminephosphotransferase 1 (GlcNAc-1-P transferase)	-1.22
DRD2	dopamine receptor D2	1.26
DYRK1B	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1B	1.25
EDG6	"endothelial differentiation, G-protein-coupled receptor 6"	1.28
EDG8	"endothelial differentiation, sphingolipid G-protein-coupled receptor, 8"	1.27
EIF3S6	"eukaryotic translation initiation factor 3, subunit 6 48kDa"	1.28
EIF4ENIF1	eukaryotic translation initiation factor 4E nuclear import factor 1	-1.21
EIF4G2	"eukaryotic translation initiation factor 4 gamma, 2"	-1.21

EIF5B	eukaryotic translation initiation factor 5B	-1.22
ELAVL1	"ELAV (embryonic lethal, abnormal vision, Drosophila)-like 1 (Hu antigen R)"	-1.26
ELF1	E74-like factor 1 (ets domain transcription factor)	-1.36
ELMO2	"engulfment and cell motility 2 (ced-12 homolog, C. elegans)"	-1.21
EME2	essential meiotic endonuclease 1 homolog 2 (S. pombe)	1.38
ENTH	enthoprotin	-1.28
EPB41L3	erythrocyte membrane protein band 4.1-like 3	-1.21
EVER2	epidermodysplasiaverruciformis 2	1.37
F5	"coagulation factor V (proaccelerin, labile factor)"	1.23
FAD158	factor for adipocyte differentiation 158	-1.22
FAM33A	"family with sequence similarity 33, member A"	1.30
FBS1	fibrosin 1	1.21
FBXL12	F-box and leucine-rich repeat protein 12	-1.22
FBXL14	F-box and leucine-rich repeat protein 14	1.20
FBXL18	F-box and leucine-rich repeat protein 18	1.36
FBXO11	F-box protein 11	-1.26
FER1L3	"fer-1-like 3, myoferlin (C. elegans)"	-1.26
FKBP14	"FK506 binding protein 14, 22 kDa"	1.26
FKBP5	FK506 binding protein 5	-1.23
FKSG2	apoptosis inhibitor	1.39
FLI1	Friend leukemia virus integration 1	-1.24
FLJ10925	hypothetical protein FLJ10925	1.41
FLJ11127	hypothetical protein FLJ11127	-1.20
FLJ12270	FLJ12270 protein	1.37
FLJ13841	hypothetical protein FLJ13841	1.23
FLJ14753	hypothetical protein FLJ14753	-1.26
FLJ20160	FLJ20160 protein	1.30
FLJ30277	hypothetical protein FLJ30277	1.35
FLJ32115	hypothetical protein FLJ32115	-1.39
FLJ32770	hypothetical protein FLJ32770	1.45
FLJ36268	FLJ36268 protein	1.31
FLJ37543	hypothetical protein FLJ37543	1.33
FLJ41131	FLJ41131 protein	1.23
FLJ43855	similar to sodium- and chloride-dependent creatine transporter	1.36
FN5	FN5 protein	1.28
FNTA	"farnesyltransferase, CAAX box, alpha"	-1.22
FRAS1	Fraser syndrome 1	1.31
FSCN1	"fascin homolog 1, actin-bundling protein (Strongylocentrotuspurpuratus)"	1.22
FUSIP1	FUS interacting protein (serine-arginine rich) 1	-1.20
FUT5	"fucosyltransferase 5 (alpha (1,3) fucosyltransferase)"	1.25
FXYD7	FXYD domain containing ion transport regulator 7	1.35
GADD45G	"growth arrest and DNA-damage-inducible, gamma"	1.25
GATA2	GATA binding protein 2	-1.38
GBGT1	"globoside alpha-1,3-N-acetylgalactosaminyltransferase 1"	1.49
GBP4	guanylate binding protein 4	-1.34

GCH1	GTP cyclohydrolase 1 (dopa-responsive dystonia)	-1.30
GFOD1	glucose-fructose oxidoreductase domain containing 1	-1.25
GNE	glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase	-1.20
GOLGA1	"golgiautoantigen, golgin subfamily a, 1"	-1.25
GOT2	"glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)"	-1.21
GPR103	G protein-coupled receptor 103	1.26
GPR135	G protein-coupled receptor 135	1.25
GPR158	G protein-coupled receptor 158	1.32
GPR172A	G protein-coupled receptor 172A	-1.22
GPR25	G protein-coupled receptor 25	1.27
GPR35	G protein-coupled receptor 35	1.46
GPR44	G protein-coupled receptor 44	-1.31
GPR45	G protein-coupled receptor 45	1.38
GPR6	G protein-coupled receptor 6	1.25
GPR8	G protein-coupled receptor 8	1.49
GRP58	"glucose regulated protein, 58kDa"	-1.20
GTPBP3	GTP binding protein 3 (mitochondrial)	-1.21
GYPC	glycophorin C (Gerbich blood group)	1.25
GYPE	glycophorin E	1.37
HAMP	hepcidin antimicrobial peptide	1.31
HAVCR2	hepatitis A virus cellular receptor 2	-1.21
HBZ	"hemoglobin, zeta"	1.40
HDC	histidine decarboxylase	-1.44
HEBP1	heme binding protein 1	-1.21
HGFAC	HGF activator	1.33
HIRIP3	HIRA interacting protein 3	-1.20
HIST1H3I	"histone 1, H3i"	1.31
HIST3H3	"histone 3, H3"	1.29
HLA-DMA	"major histocompatibility complex, class II, DM alpha"	-1.20
HLA-F	"major histocompatibility complex, class I, F"	-1.29
HMBS	hydroxymethylbilane synthase	1.20
HMGA1	high mobility group AT-hook 1	1.26
HNRPH1	heterogeneous nuclear ribonucleoprotein H1 (H)	-1.33
HRH2	histamine receptor H2	1.27
HS6ST2	heparan sulfate 6-O-sulfotransferase 2	1.20
HSD-40	HSD-40 protein	1.24
HSPA9B	heat shock 70kDa protein 9B (mortalin-2)	-1.27
HSPB1	heat shock 27kDa protein 1	1.23
HTR1D	5-hydroxytryptamine (serotonin) receptor 1D	1.25
HUMPPA	paraneoplastic antigen	1.58
IDH3A	isocitrate dehydrogenase 3 (NAD+) alpha	-1.23
IDI1	isopentenyl-diphosphate delta isomerase	-1.21
IGFBP7	insulin-like growth factor binding protein 7	-1.23
IGFL1	insulin growth factor-like family member 1	1.21
IL20	interleukin 20	1.24
IL27RA	"interleukin 27 receptor, alpha"	1.47

IL28A	"interleukin 28A (interferon, lambda 2)"	1.23
IL2RB	"interleukin 2 receptor, beta"	-1.29
ISYNA1	myo-inositol 1-phosphate synthase A1	1.28
ITGA4	"integrin, alpha 4 (antigen CD49D, alpha 4 subunit of VLA-4 receptor)"	-1.29
ITPA	inosinetriphosphatase (nucleoside triphosphate pyrophosphatase)	-1.25
JAK1	Janus kinase 1 (a protein tyrosine kinase)	-1.25
KAAG1	kidney associated antigen 1	1.34
KCNF1	"potassium voltage-gated channel, subfamily F, member 1"	1.26
KCNQ4	"potassium voltage-gated channel, KQT-like subfamily, member 4"	1.20
KIAA0217	KIAA0217	-1.21
KIAA0528	KIAA0528 gene product	-1.22
KIAA1185	KIAA1185 protein	-1.20
KIAA1441	KIAA1441 protein	1.29
KIAA1446	brain-enriched guanylate kinase-associated protein	1.39
KIAA1539	KIAA1539	1.41
KIAA1875	hypothetical protein KIAA1875	1.42
KIAA1904	KIAA1904 protein	1.35
KIR2DL3	"killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 3"	1.35
KIRREL	kin of IRRE like (Drosophila)	1.27
KLHDC1	kelch domain containing 1	1.29
KPNA2	"karyopherin alpha 2 (RAG cohort 1, importin alpha 1)"	-1.24
KRT16	keratin 16 (focal non-epidermolyticpalmoplantarkeratoderma)	1.35
KRTAP12-3	keratin associated protein 12-3	1.38
KRTCAP3	keratinocyte associated protein 3	1.24
KRTHB3	"keratin, hair, basic, 3"	1.26
LACTB	"lactamase, beta"	-1.20
LAIR2	leukocyte-associated Ig-like receptor 2	-1.20
LCMT2	leucine carboxyl methyltransferase 2	-1.27
LCN2	lipocalin 2 (oncogene 24p3)	2.16
LCN6	lipocalin 6	1.26
LDLR	low density lipoprotein receptor (familial hypercholesterolemia)	-1.27
LEPRE1	leucine proline-enriched proteoglycan (leprecan) 1	-1.20
LGMN	legumain	-1.22
LHB	luteinizing hormone beta polypeptide	1.26
LIF	leukemia inhibitory factor (cholinergic differentiation factor)	1.26
LIG1	"ligase I, DNA, ATP-dependent"	-1.24
LILRA1	"leukocyte immunoglobulin-like receptor, subfamily A (with TM domain), member 1"	1.22
LIN7A	lin-7 homolog A (C. elegans)	1.37
LIPA	"lipase A, lysosomal acid, cholesterol esterase (Wolman disease)"	-1.27
LIPF	"lipase, gastric"	1.40
LIR9	leukocyte Ig-like receptor 9	1.24
LOC113655	hypothetical protein BC011982	1.29
LOC134548	hypothetical protein LOC134548	1.21
LOC149837	hypothetical protein LOC149837	1.21
LOC220070	hypothetical protein BC004224	1.26

LOC221091	similar to hypothetical protein	1.29
LOC283932	hypothetical protein LOC283932	1.26
LOC374969	hypothetical protein LOC374969	1.31
LOC55831	30 kDa protein	1.22
LONP	peroxisomallon protease	-1.23
LPIN2	lipin 2	-1.23
LRRC10	leucine rich repeat containing 10	1.30
LTF	lactotransferrin	2.38
LUC7L2	LUC7-like 2 (S. cerevisiae)	-1.25
MAP17	membrane-associated protein 17	1.32
MAP1LC3A	microtubule-associated protein 1 light chain 3 alpha	1.35
MAP3K8	mitogen-activated protein kinase kinasekinase 8	-1.26
MAT1A	"methionine adenosyltransferase I, alpha"	1.29
MC1R	melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor)	1.26
MCM7	MCM7 minichromosome maintenance deficient 7 (S. cerevisiae)	1.22
MEN1	multiple endocrine neoplasia I	-1.24
METTL3	methyltransferase like 3	-1.26
MGC11335	hypothetical protein MGC11335	1.21
MGC13057	hypothetical protein MGC13057	1.30
MGC15476	thymus expressed gene 3-like	1.34
MGC15619	hypothetical protein MGC15619	-1.27
MGC16635	hypothetical protein BC009980	1.25
MGC2574	hypothetical protein MGC2574	1.21
MGC29816	hypothetical protein MGC29816	-1.20
MGC3036	hypothetical protein MGC3036	1.44
MGC5178	hypothetical protein MGC5178	1.47
MGC61598	similar to ankyrin-repeat protein Nrarp	1.31
MGC70857	similar to RIKEN cDNA C030006K11 gene	1.21
MGC9712	hypothetical protein MGC9712	1.24
MHC2TA	MHC class II transactivator	-1.28
MINPP1	"multiple inositol polyphosphate histidine phosphatase, 1"	-1.25
MMP15	matrix metalloproteinase 15 (membrane-inserted)	1.24
MMP28	matrix metalloproteinase 28	1.29
MPN	pancreasin	1.31
MRAP	melanocortin 2 receptor accessory protein	1.28
MRGPRD	"MAS-related GPR, member D"	1.33
MRVI1	murine retrovirus integration site 1 homolog	2.18
MSX2	mshhomeo box homolog 2 (Drosophila)	1.23
MT1G	metallothionein 1G	1.22
MTSS1	metastasis suppressor 1	-1.24
MYC	v-mycmyelocytomatosis viral oncogene homolog (avian)	-1.22
MYCN	"v-mycmyelocytomatosis viral related oncogene, neuroblastoma derived (avian)"	1.26
MYL4	"myosin, light polypeptide 4, alkali; atrial, embryonic"	1.39
NARG1L	NMDA receptor regulated 1-like	-1.22
NCOR2	nuclear receptor co-repressor 2	1.37
NDUFV3	"NADH dehydrogenase (ubiquinone) flavoprotein 3, 10kDa"	-1.25

NELL2	NEL-like 2 (chicken)	-1.30
NEUROG1	neurogenin 1	1.33
NFAM1	NFAT activating protein with ITAM motif 1	1.33
NFKBIA	"nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha"	-1.20
NGB	neuroglobin	1.26
NKTR	natural killer-tumor recognition sequence	-1.22
NOXA1	NADPH oxidase activator 1	-1.22
NPAS1	neuronal PAS domain protein 1	1.27
NR1I2	"nuclear receptor subfamily 1, group I, member 2"	1.26
NUP50	nucleoporin 50kDa	-1.25
OR10H1	"olfactory receptor, family 10, subfamily H, member 1"	1.27
OR1D4	"olfactory receptor, family 1, subfamily D, member 4"	1.40
OR4D1	"olfactory receptor, family 4, subfamily D, member 1"	1.21
OR4M1	"olfactory receptor, family 4, subfamily M, member 1"	1.38
OR51G1	"olfactory receptor, family 51, subfamily G, member 1"	1.30
OR52E6	"olfactory receptor, family 52, subfamily E, member 6"	1.25
OR52N4	"olfactory receptor, family 52, subfamily N, member 4"	1.60
OR56B4	"olfactory receptor, family 56, subfamily B, member 4"	1.24
OR6B3	"olfactory receptor, family 6, subfamily B, member 3"	1.48
OR6N1	"olfactory receptor, family 6, subfamily N, member 1"	1.29
ORF1-FL49	putative nuclear protein ORF1-FL49	1.32
ORM1	"orosomucoid 1,orosomucoid 2"	2.22
OTOS	otospiralin	1.26
OVOL1	ovo-like 1(Drosophila)	1.30
P53AIP1	p53-regulated apoptosis-inducing protein 1	1.42
PAOX	polyamine oxidase (exo-N4-amino)	-1.25
PAPD4	PAP associated domain containing 4	-1.22
PBP	prostatic binding protein	1.29
PCDHA5	protocadherin alpha 5	1.21
PCNT1	pericentrin 1	-1.20
PCYT2	"phosphate cytidylyltransferase 2, ethanolamine"	1.56
PDE7A	phosphodiesterase 7A	-1.31
PDZK1	PDZ domain containing 1	1.25
PGLYRP1	peptidoglycan recognition protein 1	1.42
PHKG1	"phosphorylase kinase, gamma 1 (muscle)"	1.24
PIAS2	"protein inhibitor of activated STAT, 2"	-1.22
PIGV	"phosphatidylinositol glycan, class V"	-1.21
PKD1	polycystic kidney disease 1 (autosomal dominant)	-1.24
PKMYT1	"protein kinase, membrane associated tyrosine/threonine 1"	1.21
PLAGL1	pleiomorphic adenoma gene-like 1	-1.20
PLCB3	"phospholipase C, beta 3 (phosphatidylinositol-specific)"	1.33
PLEKHA9	"pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 9"	1.21
PLEKHH1	"pleckstrin homology domain containing, family H (with MyTH4 domain) member 1"	1.30
PLXNA1	plexin A1	1.33

PPFIA1	"protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 1"	-1.25
PPP1R12C	"protein phosphatase 1, regulatory (inhibitor) subunit 12C"	-1.22
PQLC2	PQ loop repeat containing 2	1.21
PRKCG	"protein kinase C, gamma"	1.37
PRKCH	"protein kinase C, eta"	-1.21
PRLH	prolactin releasing hormone	1.33
PRNP	"prion protein (p27-30) (Creutzfeld-Jakob disease, Gerstmann-Strausler-Scheinker syndrome, fatal familial insomnia)"	-1.27
PRPS2	phosphoribosyl pyrophosphate synthetase 2	-1.20
PSD4	pleckstrin and Sec7 domain containing 4	1.20
PSMB10	"proteasome (prosome, macropain) subunit, beta type, 10"	-1.20
PSMC3	"proteasome (prosome, macropain) 26S subunit, ATPase, 3"	-1.25
PSME2	"proteasome (prosome, macropain) activator subunit 2 (PA28 beta)"	-1.21
PSPH	phosphoserine phosphatase	1.24
PSTPIP2	proline-serine-threonine phosphatase interacting protein 2	-1.22
PTDSS1	phosphatidylserine synthase 1	-1.25
PTK2B	PTK2B protein tyrosine kinase 2 beta	1.44
PTMS	parathymosin	1.25
PUSL1	pseudouridylate synthase-like 1	1.34
RAB1B	"RAB1B, member RAS oncogene family"	1.32
RAB3-GAP150	"rab3 GTPase-activating protein, non-catalytic subunit (150kD)"	-1.20
RASSF4	Ras association (RalGDS/AF-6) domain family 4	-1.22
RBBP4	retinoblastoma binding protein 4	-1.25
RBM15	RNA binding motif protein 15	1.31
RBM16	RNA binding motif protein 16	-1.20
RBM19	RNA binding motif protein 19	1.26
RINT-1	Rad50-interacting protein 1	-1.20
RNASET2	ribonuclease T2	-1.21
ROBO4	"roundabout homolog 4, magic roundabout (Drosophila)"	1.35
ROD1	ROD1 regulator of differentiation 1 (S. pombe)	1.23
RHOA	Ras homolog gene family, member A	-1.22
RPA3	"replication protein A3, 14kDa"	1.20
RPL21	ribosomal protein L21	1.26
RPS6	ribosomal protein S6	1.22
RRN3	RRN3 RNA polymerase I transcription factor homolog (yeast)	-1.21
RTP2	receptor transporting protein 2	1.26
S100A12	S100 calcium binding protein A12 (calgranulin C)	1.61
S100A8	S100 calcium binding protein A8 (calgranulin A)	1.47
SACM1L	SAC1 suppressor of actin mutations 1-like (yeast)	-1.25
SCGB1A1	"secretoglobin, family 1A, member 1 (uteroglobin)"	1.37
SCO1	SCO cytochrome oxidase deficient homolog 1 (yeast)	1.22
SDFR1	stromal cell derived factor receptor 1	-1.20
SDS	serine dehydratase	1.21
SEH1L	SEH1-like (S. cerevisiae)	-1.21
SERF1A	"small EDRK-rich factor 1B (centromeric),small EDRK-rich factor 1A (telomeric)"	-1.20

SFRS1	"splicing factor, arginine/serine-rich 1 (splicing factor 2, alternate splicing factor)"	-1.22
SFRS2	"splicing factor, arginine/serine-rich 2"	-1.25
SHANK1	SH3 and multiple ankyrin repeat domains 1	1.31
SIRT6	sirtuin (silent mating type information regulation 2 homolog) 6 (S. cerevisiae)	1.36
SLC1A6	"solute carrier family 1 (high affinity aspartate/glutamate transporter), member 6"	1.21
SLC22A18	"solute carrier family 22 (organic cation transporter), member 18"	1.21
SLC2A3	"solute carrier family 2 (facilitated glucose transporter), member 3"	1.31
SLC31A1	"solute carrier family 31 (copper transporters), member 1"	-1.26
SLC35A5	"solute carrier family 35, member A5"	-1.21
SLC35E2	"solute carrier family 35, member E2"	-1.21
SLC41A1	"solute carrier family 41, member 1"	1.26
SLC7A6	"solute carrier family 7 (cationic amino acid transporter, y+ system), member 6"	-1.27
SMAD4	"SMAD, mothers against DPP homolog 4 (Drosophila)"	-1.24
SMARCA2	"SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2"	-1.22
SMBP	SM-11044 binding protein	-1.24
SMCR8	"Smith-Magenis syndrome chromosome region, candidate 8"	1.45
SMN1	"survival of motor neuron 1, telomeric,survival of motor neuron 2, centromeric"	-1.22
SMNDC1	survival motor neuron domain containing 1	-1.23
SMS	spermine synthase	-1.20
SOCS3	suppressor of cytokine signaling 3	1.29
SP3	Sp3 transcription factor	-1.26
SP6	Sp6 transcription factor	1.35
SRGAP3	SLIT-ROBO Rho GTPase activating protein 3	1.51
SRRM2	serine/arginine repetitive matrix 2	1.32
ST8SIA4	"ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 4"	-1.23
STAT1	"signal transducer and activator of transcription 1, 91kDa"	-1.27
STK39	"serine threonine kinase 39 (STE20/SPS1 homolog, yeast)"	-1.20
STX1A	syntaxin 1A (brain)	1.27
STXBP3	syntaxin binding protein 3	-1.22
SUMF1	sulfatase modifying factor 1	-1.22
SUSD2	sushi domain containing 2	1.42
SYK	spleen tyrosine kinase	-1.21
SYPL	synaptophysin-like protein	-1.20
TACC1	"transforming, acidic coiled-coil containing protein 1"	-1.25
TAP1	"transporter 1, ATP-binding cassette, sub-family B (MDR/TAP)"	-1.23
TCP1	t-complex 1	-1.22
TEKT2	tektin 2 (testicular)	1.39
TESK2	testis-specific kinase 2	-1.24
TFCP2L1	transcription factor CP2-like 1	1.24
TFG	TRK-fused gene	-1.23
TGFB1I4	transforming growth factor beta 1 induced transcript 4	-1.25
THEG	Theg homolog (mouse)	1.33

TITF1	thyroid transcription factor 1	1.21
TIZ	TRAF6-inhibitory zinc finger protein	1.21
TK1	"thymidine kinase 1, soluble"	1.32
TLR9	toll-like receptor 9	1.44
TM4SF11	transmembrane 4 superfamily member 11 (plasmolipin)	1.38
TM4SF9	transmembrane 4 superfamily member 9	1.30
TMEM38A	transmembrane protein 38A	1.26
TNFAIP2	"tumor necrosis factor, alpha-induced protein 2"	-1.21
TNNI2	"troponin I, skeletal, fast"	1.29
TOR3A	"torsin family 3, member A"	-1.27
TP53I5	tumor protein p53 inducible protein 5	1.24
TPM3	tropomyosin 3	-1.22
TPP2	tripeptidyl peptidase II	-1.26
TRAF4	TNF receptor-associated factor 4	-1.21
TRIM22	tripartite motif-containing 22	-1.30
TRIM4	tripartite motif-containing 4	-1.24
TRIM50A	tripartite motif-containing 50A	1.35
TSN	translin	-1.21
TTBK1	tau tubulin kinase 1	1.35
TTYH3	tweety homolog 3 (Drosophila)	-1.26
TUB	tubby homolog (mouse)	1.21
TXK	TXK tyrosine kinase	-1.36
UBA52	ubiquitin A-52 residue ribosomal protein fusion product 1	1.29
UBXD1	UBX domain containing 1	1.28
UCKL1	uridine-cytidine kinase 1-like 1	-1.22
UCN	urocortin	1.27
UCP2	"uncoupling protein 2 (mitochondrial, proton carrier)"	-1.21
UGT2A1	"UDP glycosyltransferase 2 family, polypeptide A1"	1.41
ULK2	unc-51-like kinase 2 (C. elegans)	1.31
UNQ5783	DTFT5783	-1.21
UPK2	uroplakin 2	1.28
UPK3B	uroplakin 3B	1.31
UROC1	urocanase domain containing 1	1.22
UROS	uroporphyrinogen III synthase (congenital erythropoietic porphyria)	-1.22
USP49	ubiquitin specific protease 49	-1.21
USP52	ubiquitin specific protease 52	-1.20
USP54	ubiquitin specific protease 54	1.25
UTX	"ubiquitously transcribed tetratricopeptide repeat, X chromosome"	-1.21
UVRAG	UV radiation resistance associated gene	-1.22
VDP	vesicle docking protein p115	-1.26
VPS13C	vacuolar protein sorting 13C (yeast)	-1.24
VPS4B	vacuolar protein sorting 4B (yeast)	-1.25
WDFY2	WD repeat and FYVE domain containing 2	-1.21
WDR1	WD repeat domain 1	1.24
WDR37	WD repeat domain 37	-1.26
XLKD1	extracellular link domain containing 1	1.44

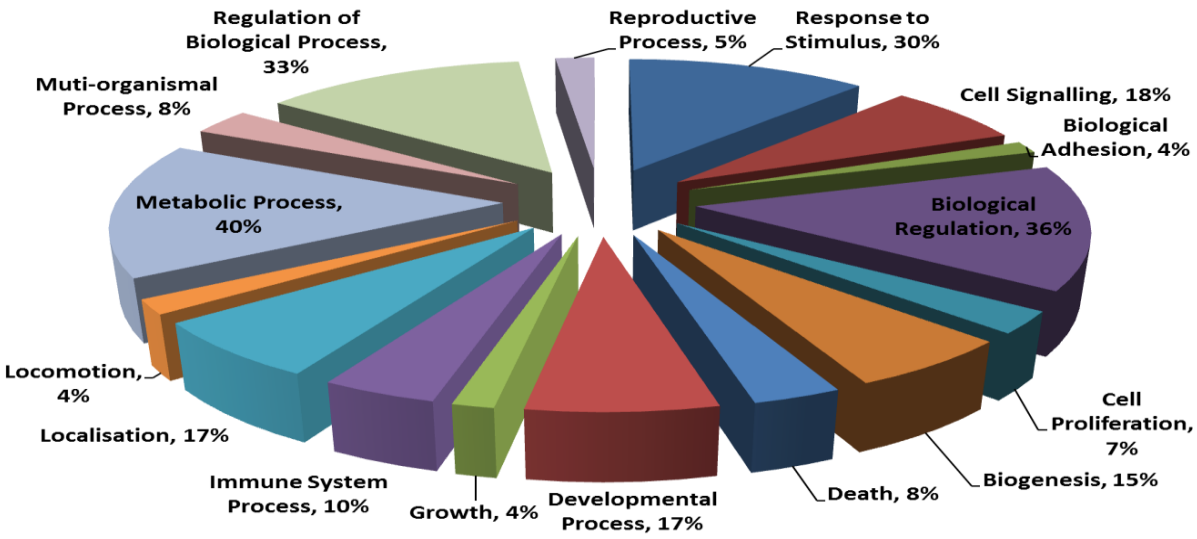
XPO4	exportin 4	1.28
ZC3HDC7	zinc finger CCCH type domain containing 7	-1.23
ZCCHC14	"zinc finger, CCHC domain containing 14"	-1.22
ZF	HCF-binding transcription factor Zhangfei	-1.22
ZFP36L2	"zinc finger protein 36, C3H type-like 2"	-1.37
ZMAT2	"zinc finger, matrin type 2"	1.27
ZNF3	zinc finger protein 3 (A8-51)	1.32
ZNF488	zinc finger protein 488	1.32
ZNF579	zinc finger protein 579	1.33
ZNFN1A1	"zinc finger protein, subfamily 1A, 1 (Ikaros)"	-1.41

4.3.2 Molecular and functional analysis of the significant genes obtained through the comparison of gene expression profiles of Healthy donors and breast cancer patients

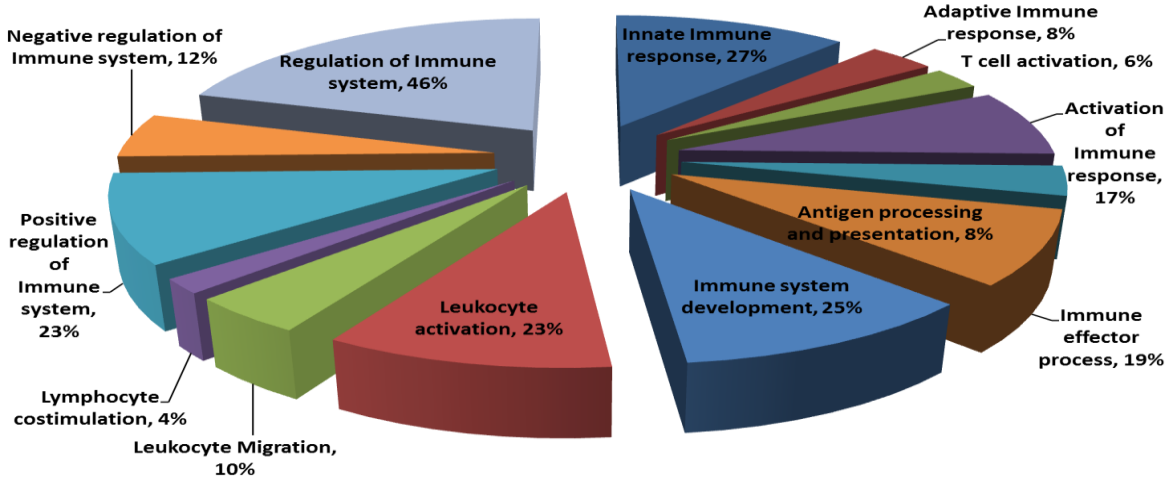
Understanding the biological meaning of the results obtained through microarray or RNA-Seq can be difficult. The conventional way to achieve this is to classify the genes based on their function and their cellular location. Gene ontology is used to annotate genes with respect to their function. The Gene ontology project is an initiative which evolved out of the need to consolidate the descriptions of each gene and its products across different species for consistence to enable functional interpretation of experimental data. The Gene Ontology analysis was performed using a tool called Gene Ontology for Functional Analysis (GOFFA) developed by the USA Food and Drug Administration (FDA) Department. This tool is further explained in Section2.4.3.2.Other methods to analyse microarray data include the KEGG pathway and protein-protein interaction cluster analyses using databases such as Biomolecular Interaction Network Database (BIND), Biological General Repository for Interaction Datasets (BioGRID) and others (discussed in Section2.4.3).

4.3.2.1 Gene ontology-Biological Processes

Gene ontology analysis was performed with the 506 significant probes obtained through the comparison of gene expression profiles between WBCs of breast cancer patients and healthy donors, with a particular purpose to highlight the changes in immune functions of WBCs in breast cancer patients. GOFFA was used to perform ontology as described in Section2.4.3.1. The software organises gene ontology into three main categories: Molecular function, Biological processes and Cellular components. Figure 4.3A shows the results for the gene ontology classification based on the biological function. The number of genes changed are related to metabolic processes (40%) followed by biological regulation (36%) and response to stimulus (30%).



A



B

Figure 4.3 Gene Ontology of significantly different genes between the PBCs of Healthy donors and Breast cancer Patients from GSE16443 dataset: Genes identified to be differentially present in the WBCs of breast cancer patient and healthy donors Gene Ontology was performed using GOFFA and based on biological processes (n=506). A) Ontology was done based on Biological processes and genes were grouped under different categories. B) Sub classification of immune system related genes (n=54).

4.3.2.2 Analysis of genes associated with immune system processes

The number of significant genes which were related to immune system processes (GO: 0002376) are 54 genes (10%). The genes related to the immune system were further sub classified based on different immune system functions (Figure 4.3B). The hierarchical clustering was performed using the Multiexperiment viewer on the significant probes which correspond to the immune system processes which shown in Figure 4.4. The clustering showed four distinct groups of genes in which the patterns of expression were same within the different cohorts of samples. The first and second group has genes which are under expressed in the breast cancer patient cohort. The genes in the last two groups are overexpressed in the breast cancer patient cohort. The category having most number of genes within the immune system process is the regulation of immune system (40%) followed by the innate immune response (27%). Table 4.2 gives the list of all the genes which are related to the immune system.

The categories involved in the immune system processes are given in Table 4.3 along with the significant genes involved. It can be observed that most of the genes have functions corresponding to the innate immune system processes such as leukocyte activation, leukocyte migration, antigen presentation etc. This might be due to the fact that leukocytes constitute a major fraction (~75%) of the WBCs. The genes involved in innate immune response which were down-regulated were *HLA-DMA, JAK1, STAT1, LCN2, TAP1, HLA-F, PGLYRP1* and up-regulated genes were *C2, UBA52, GCH1, S100A12, TLR9, SOCS3* and *NFKBIA*. The down-regulated genes in the innate immune response are mostly involved in antigen recognition, processing and presentation. The upregulated genes are involved in inflammatory responses. All the genes involved in lymphocyte activation were also involved with leukocyte activation (Table 4.3). Only three genes *ANXA3, STXBP3* and *ZNF3* were unique to leukocyte activation. All the genes involved in leukocyte activation were down regulated with the exception of *ANXA3, NFAM1* and *ZNF3*.Annexin A3 (ANXA3) is a calcium dependent phospholipid binding protein

which inhibits phospholipase A2. Phospholipase A2 releases arachidonic acid causes inflammation through the prostaglandin and leukotriene pathway(Hofmann et al., 2000). *SCGB1A1* (Secretoglobin Family 1A, Member 1) encodes a small disulfide linked dimeric protein also called as Uteroglobin has also been implicated to inhibit phospholipase A2 activity (Peri et al., 1993). Zinc finger protein 3 (*ZNF3*) is DNA binding transcription factor which involved in cell differentiation and proliferation. Syntaxin binding protein (STXBP3) binds to STX2 and STX4 which are involved in targeting and fusion of intracellular transport vesicles. NFAT Activating Protein with ITAM Motif 1 (*NFAM1*) encodes for a type 1 membrane receptor which regulates B cell development and signalling (Ohtsuka and Arase, 2004). Table 4.3 shows the list of genes which are involved in the inflammatory response and only *APOL3* and *LIPA* were downregulated which suggests that the inflammatory response is induced in the breast cancer patients. Only Lipase A (*LIPA*) and Apolipoprotein L3 (*APOL3*) were down regulated. LIPA catalyses the hydrolysis of cholesteryl esters and triglycerides in the lysosome and APOL3 affects the movements of lipids and facilitates their attachment to the organelles. Among the 506 significant genes the genes involved in T-cell receptor signalling (*ELF1, NKBIA, PRNP, HLA-DMA* and *SYK*) were down regulated in the breast cancer patient category.

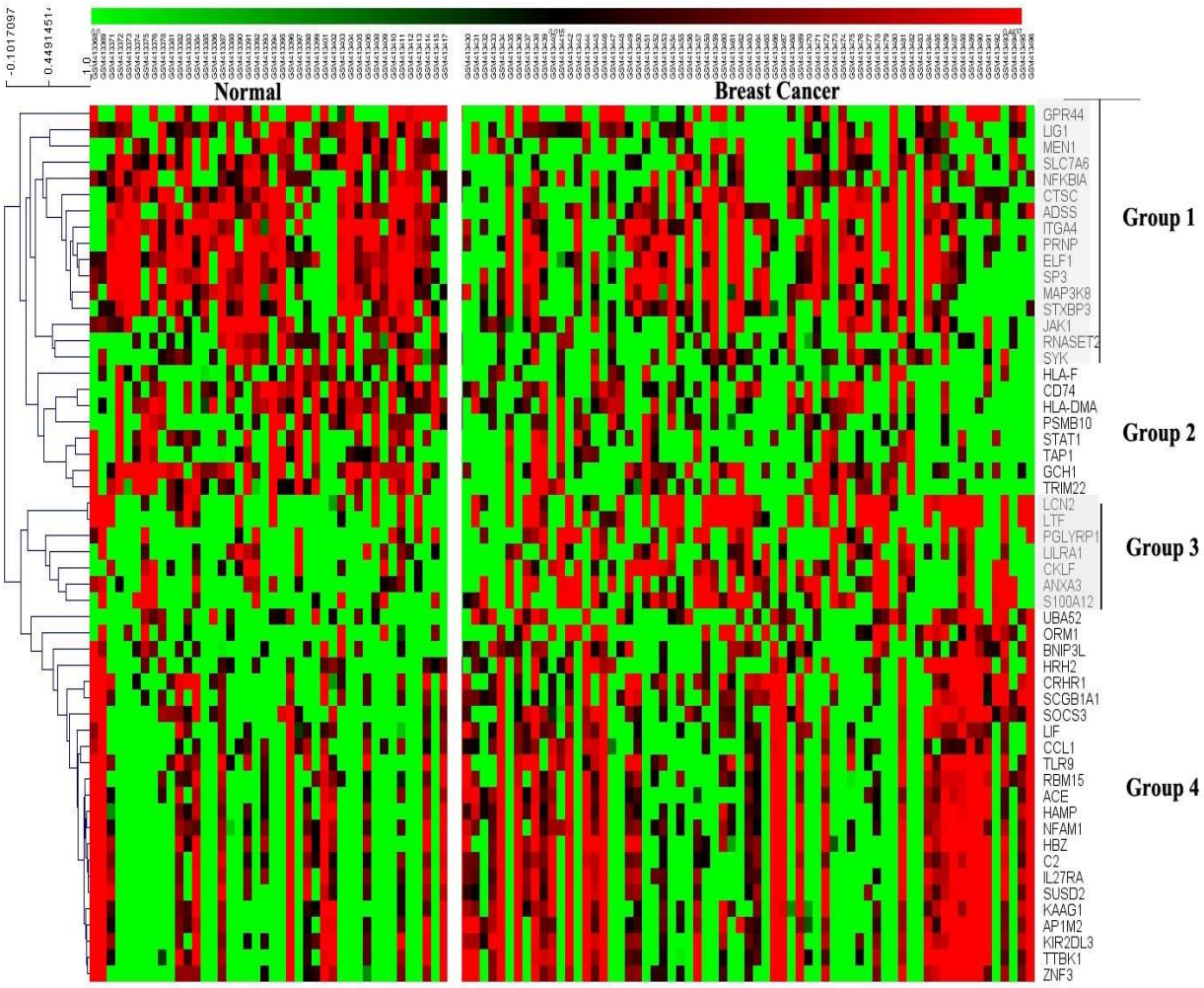


Figure 4.4 Hierarchical clustering of significant genes related to the immune system obtained through the comparison of expression profiles between the WBCs of Healthy Donors and Breast Cancer Patients: Hierarchical Clustering was done using GOFFA. Red indicates that the gene is relatively overexpressed in breast cancer patient samples and green the vice versa. Four distinct clusters were observed among the samples.

Table 4.2 List of differentially expressed genes involved in the immune response obtained from the comparison between breast cancer patients and healthy donors: Significant genes were identified by analysing GSE16443 dataset and gene ontology was done to identify immune system related genes using GOFFA. Red colour indicates overexpressed genes and green colour indicates under expressed genes in the breast cancer patient cohort.

Gene Symbol	Gene Name	Up/Down regulated	Gene Symbol	Gene Name	Up/Down regulated
ACE	angiotensin I converting enzyme (peptidyl-dipeptidase A) 1		*LILRA1*	leukocyte immunoglobulin-like receptor	
ADSS	adenylosuccinate synthase		*LTF*	lactotransferrin	
ANXA3	annexin A3		*MAP3K8*	mitogen-activated protein kinase 8	
AP1M2	adaptor-related protein complex 1, mu 2 subunit		*MEN1*	multiple endocrine neoplasia I	
BNIP3L	BCL2/adenovirus E1B 19kDa interacting protein 3-like		*NFAM1*	NFAT activating protein with ITAM motif 1	
C2	complement component 2		*NFKBIA*	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha	
CCL1	chemokine (C-C motif) ligand 1		*ORM1*	orosomucoid 1	
CD74	CD74 antigen		*PGLYRP1*	peptidoglycan recognition protein 1	
CKLF	chemokine-like factor		*PRNP*	prion protein (p27-30)	
CRHR1	corticotropin releasing hormone receptor 1		*PSMB10*	proteasome subunit, beta type, 10	
CTSC	cathepsin C		*RBM15*	RNA binding motif protein 15	
ELF1	E74-like factor 1		*RNASET2*	ribonuclease T2	
GCH1	GTP cyclohydrolase 1 (dopa-responsive dystonia)		*S100A12*	S100 calcium binding protein A12	

GPR44	G protein-coupled receptor 44		*SCGB1A1*	secretoglobin, family 1A, member 1	
HAMP	hepcidin antimicrobial peptide		*SLC7A6*	solute carrier family 7 member 6	
HBZ	hemoglobin, zeta		*SOCS3*	suppressor of cytokine signaling 3	
HLA-DMA	major histocompatibility complex, class II, DM alpha		*SP3*	Sp3 transcription factor	
HLA-F	major histocompatibility complex, class I, F		*STAT1*	signal transducer and activator of transcription 1	
HRH2	histamine receptor H2		*STXBP3*	syntaxin binding protein 3	
IL27RA	interleukin 27 receptor, alpha		*SUSD2*	sushi domain containing 2	
ITGA4	integrin, alpha 4		*SYK*	spleen tyrosine kinase	
JAK1	Janus kinase 1		*TAP1*	transporter 1, ATP-binding cassette, sub-family B (MDR/TAP)	
KAAG1	kidney associated antigen 1		*TLR9*	toll-like receptor 9	
KIR2DL3	killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 3		*TRIM22*	tripartite motif-containing 22	
LCN2	lipocalin 2		*TTBK1*	tau tubulin kinase 1	
LIF	leukemia inhibitory factor		*UBA52*	ubiquitin A-52 residue ribosomal protein fusion product 1	
LIG1	ligase I, DNA, ATP-dependent		*ZNF3*	zinc finger protein 3 (A8-51)	

Table 4.3 Ontology of the differentially expressed genes involved in immune system processes: The expression profiles of WBCs of breast cancer patients and healthy donors were compared and the differentially expressed genes were obtained. Gene ontology was performed using GOFFA. The genes involved in various immune system processes (GO: 0002376) were identified which are given below.

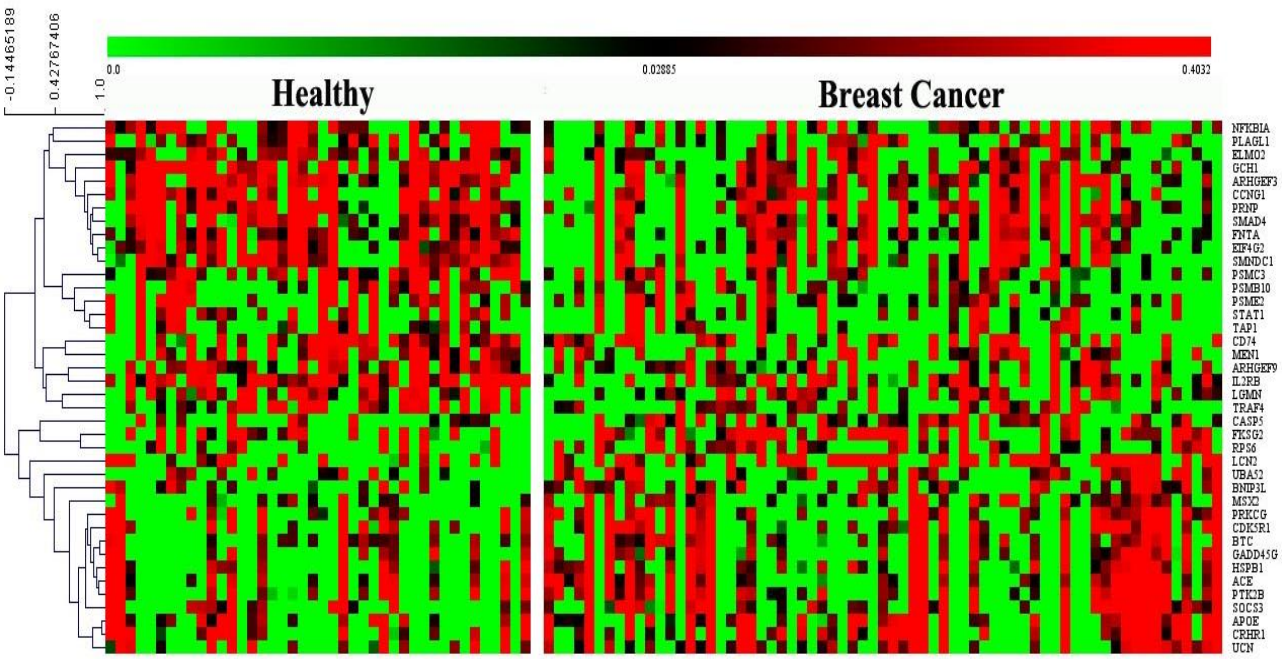
Description	Genes involved	P value
Innate Immune response	*C2,GCH1,HLA-DMA,HLA-F,JAK1,LCN2,NFKBIA, PGLYRP1,S100A12,SOCS3,STAT1,TAP1,TLR9,UBA52*	0.010
Lymphocyte Activation	*CD74,HLA-DMA,ITGA4,MAP3K8,NFAM1, PRNP,SCGB1A1,SP3,SYK*	0.019
Activation of Immune response	*C2,ELF1,HLA-DMA,NFAM1,NFKBIA, PRNP,SYK,TLR9,UBA52*	0.034
Antigen processing and presentation	*CD74,HLA-DMA,HLA-F,TAP1*	0.005
Immune effector process	*ACE,ANXA3,AP1M2,BNIP3L,C2,CD74, HLA-DMA,STXBP3,SYK,TAP1*	0.022
Immune system development	*CD74,HLA-DMA,ITGA4,MAP3K8, NFAM1,PRNP,SCGB1A1,SP3,SYK*	0.0005
Leukocyte activation	*ANXA3,CD74,HLA-DMA,ITGA4,MAP3K8,NFAM1, PRNP,SCGB1A1,SP3,STXBP3,SYK,ZNF3*	0.006
Leukocyte Migration	*CD74,CKLF,ITGA4,SLC7A6,SYK*	0.0125
Regulation of Immune system processes	*AP1M2,C2,CD74,ELF1,HLA-DMA,HLA-F,ITGA4, JAK1,KIR2DL3,LIF,LILRA1,MAP3K8,NFAM1,NFKBIA, ORM1,PRNP,RBM15,SCGB1A1,SOCS3,STAT1,SYK, TAP1,TLR9,UBA52*	0.0012
Adaptive immune response	*C2,CD74,HLA-DMA, TAP1*	0.048
Positive regulation of immune system	*C2,CD74,ELF1,HLA-DMA,MAP3K8, NFAM1,NFKBIA,PRNP,SYK,TAP1,TLR9,UBA52*	0.008
Negative of immune system	*CD74,ELF1,PRNP,SCGB1A1,TAP1,TLR9*	0.067
Inflammatory response	*APOL3,ACE,LIPA,IL20,UCN,NFAM1,APOE, SCGB1A1,TLP9,S100A8,S100A12,ORM1*	0.045

4.3.2.3 Analysis of genes associated with apoptosis

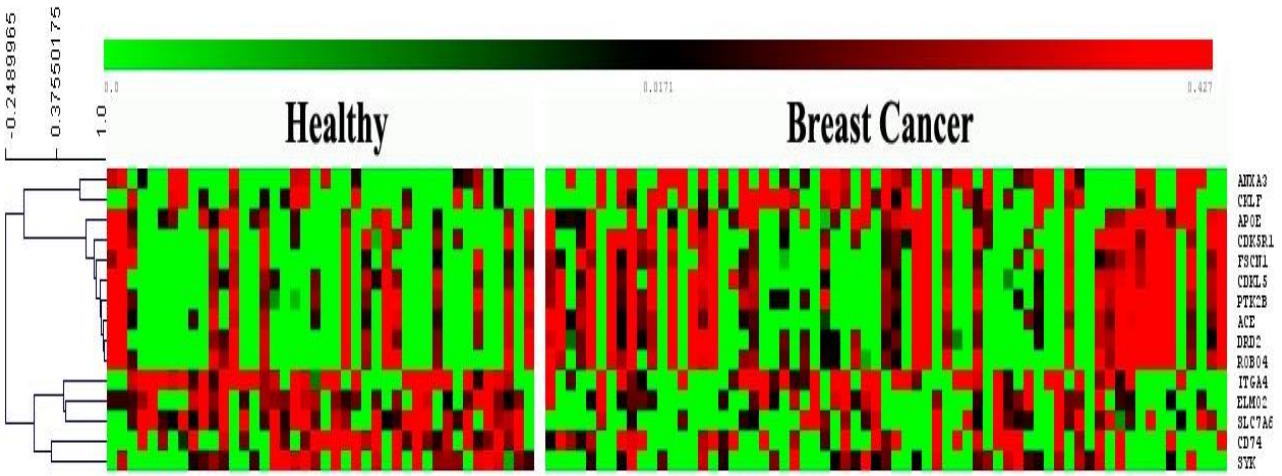
Gene ontology was done on the 506 significant genes using GOFFA. Genes involved in cell death were identified. Figure 4.3 shows that only 8% (n=40) of the 506 genes that changed in the WBCs of breast cancer are involved in cell death. Hierarchical clustering was done on these genes which shown in Figure 4.5A.It was observed that 22 genes involved in the cell death were down regulated in breast cancer patients and 19 genes were up regulated. Thus a conclusion on the effect of tumour on the WBCs cell death could not be achieved. So the genes involved in induction of apoptosis were identified and the genes involved were *GCH1, STAT1,TAP1, SMNDC1, ARHGEF9, ARHGEF3, PLAGL1, BNIP3L, UBA52* and *APOE* which has been shown in Table 4.3. It was observed except for *BNIP3L, UBA52* and *APOE* all the other genes involved in the induction of apoptosis were down regulated which suggests that the WBCs were not driven towards apoptosis. The genes which were involved in cell growth are also given Table 4.3. It was observed that all the genes involved in positive regulation of cell growth were overexpressed and genes involved in negative regulation of cell growth were under-expressed (except for *ADRB1* and *CAMP*). This supports the observation that the WBCs are not driven towards cell death but towards cell proliferation.

4.3.2.4 Analysis of genes involved with cell migration

Cell adhesion and cell migration are very important for WBC function to detect antigens in different tissues and organs. WBC movement to the site of infection is very important for initiation and regulation of both the innate and adaptive immune response. WBC cell migration has been discussed in Chapter 1 (see Section 1.4).Gene ontology analysis was performed on the significant genes differentially expressed in the breast cancer patient cohort and genes involved in cell migration were obtained using GOFFA. Figure 4.5 B shows the hierarchical clustering of genes involved in cell migration and there were two groups of genes: Upregulated (*ACE,ANXA3,APOE,CDK5R1,CDKL5,CKLF,DRD2,FSCN1,PTK2B,ROBO4*) and downregulated genes (*CD74, ITGA4, ELMO2,SLCA6, SYK*) in breast cancer category. Table 4.5shows the functions of the genes involved in cell migration and also the effect on cell migration. All the genes which have negative effect on cell migration were overexpressed. The genes involved in leukocyte migration specifically are *ITGA4*, *CKLF*, *CD74*, *SLCA6* and *SYK* (Table 4.3). Except for *CKLF* all the genes are downregulated. This suggests that cell migration is negatively affected in the WBCs of breast cancer.



A



B

Figure 4.5 Hierarchical clustering of significant genes involved in cell death and cell migration: The genes which were significantly changed in the WBCs of breast cancer were identified through Student's T-Test. Gene ontology was done and genes involved in Cell Death (A) and Cell migration (B) were identified. Hierarchical Clustering was done using GOFFA. Red indicates that the gene is relatively overexpressed and green the vice versa in the breast cancer cohort.

Table 4.4 Ontology of the genes involved in cell death and cell growth obtained by comparison expression profiles in the WBCs of breast cancer patients and healthy donors: The genes which changed significantly in WBCs of breast cancer patients were identified through T-Test and gene ontology was done using GOFFA. Genes involved in Cell death (GO: 0008219) and growth (GO: 0016049) were identified.

Biological Process	Genes involved	P value
Cell Death	GCH1,IL2RB,STAT1,PRNP,PSMC3,MEN1,CCNG1, SMAD4,TAP1,SMNDC1,LGMN,FNTA,CD74,ARHGEF9, EIF4G2,PSME2,TRAF4,ELMO2,ARHGEF3,PLAGL1,PSMB10, NFKBIA,BTC,RPS6,MSX2,HSPB1,CASP5,ACE,GADD45G, UCN,SOCS3,BNIP3L,UBA52,CDK5R1,CRHR1,APOE, PRKCG,FKSG2,PTK2B,LCN2	0.006
Positive regulation of apoptosis	ACE,APOE,ARHGEF3,ARHGEF9,BNIP3L,CDK5R1,GCH1, MEN1,MSX2,PLAGL1,RPS6,SMNDC1,STAT1,TAP1,UBA52	0.002
Negative regulation of apoptosis	APOE,BNIP3L,BTC,CCNG1,CD74,HSPB1,IL2RB,LGMN, MSX2,NFKBIA,PRNP,PTK2B,SOCS3,UBA52	0.022
Induction of apoptosis	GCH1,STAT1,TAP1,SMNDC1,ARHGEF9,ARHGEF3, PLAGL1,BNIP3L,UBA52,APOE	0.0056
Cell growth	ADRB1,APOE,CAMP,CDKL5,DRD2,IGFBP7,LGMN, MEN1,NELL2,PRLH,PTK2B,SMAD4,SMARCA2,SOCS3, TFCP2L1,UCN,ULK2	0.044
Negative regulation of cell growth	ADRB1,CAMP,LGMN,MEN1,SMAD4,SMARCA2,ULK2	0.032
Positive regulation of cell growth	CDKL5,DRD2,PTK2B,TFCP2L1,UCN	0.047

Table 4.5 Ontology of significant genes in WBCs of breast cancer patients relating to Cell migration: Gene ontology was done using GOFFA on differentially regulated genes in the WBCs of breast cancer patients obtained by comparison with healthy donors. Genes involved in Cell migration process (GO: 0016477) were identified**.**

Gene Symbol	Gene name	Function	Fold Change	Effect on migration
ACE	angiotensin I converting enzyme	Involved in blood vessel constriction	+	Decreases
ANXA3	annexin A3	Inhibits phospholipase A	+	Decreases
APOE	apolipoprotein E	Transports vitamins and lipids into the lymph system	+	Decreases
DRD2	dopamine receptor D2	receptor for dopamine	+	Decreases
ROBO4	roundabout homolog 4, magic roundabout (Drosophila)	Receptor for Slit proteins involved in angiogenesis.	+	Decreases
CD74	CD74 antigen	Involved in antigen recognition and interacts with Macrophage inhibitory factor	-	Increases
CKLF	chemokine-like factor	Acts as chemoattractant for neutrophils and lymphocytes	+	Increases
ELMO2	engulfment and cell motility 2	Involved in actin cytoskeletal rerrangement	-	Increases
FSCN1	fascin homolog 1, actin-bundling protein	Increases Actin organisation in filipodia	+	Increases
ITGA4	integrin, alpha 4	Involved in adhesion and transendothelial migration	-	Increases
PTK2B	PTK2B protein tyrosine kinase 2 beta	Regulates reorganization of the actin cytoskeleton, cell migration and adhesion	+	Increases

4.3.2.5 Pathway enrichment analysis of genes differentially expressed in the WBCs of breast cancer patients

Pathway analysis is another approach to interpret the biological meaning of a microarray dataset. This approach can be used to observe subtle and consistent changes in the pathways using functional annotations. Pathway analysis was performed on the 506 significant genes using GenMAPP software as described in Section 2.4.3.3. The pathway database used was KEGG (Kyoto Encyclopedia of Genes and Genomes). The top ten pathways obtained by using GenMAPP are shown in Table 4.6 along with the altered genes. The pathways relevant to immune system processes in the table are antigen processing, Jak-STAT signaling pathways and leukocyte transendothelial migration which are shown in Figure 4.6, Figure 4.7 and Figure 4.8 respectively.

It was observed that all the genes (*HLA-DMA,LGMN,PSME2,HLA-F,CD74,TAP1*) which changed in the antigen processing and presentation pathway was downregulated except for *ARHGDIG* (Brp57) and Killer Cell Immunoglobulin-Like Receptor (*KIR2DL3*).KIR2DL3 is receptor found on natural killer cells which inhibits their activity. Thus the data suggests that antigen presentation is negatively affected in the WBCs of breast cancer patients. All the genes in the Jak-STAT signaling pathway were also downregulated except for Suppressor of cytokine signaling 3 (*SOCS3*), Interleukin 20 (*IL20*) and Leukemia inhibitory factor (*LIF*). SOCS3 binds to tyrosine kinase receptors such as LIF, IL12, GCSF etc instead of STAT4 to inhibit cytokine based signaling and it also inhibits JAK2 kinase (Sasaki et al., 1999; Yamamotoa et al., 2003). But interleukin 20 (IL20) and leukemia inhibitory factor (LIF) belonging to the IL6 family are Upregulated. The receptor for both IL20 and LIF which is Interleukin receptor Interleukin-2 receptor (IL2RB) subunit beta is downregulated. Thus it can be concluded that the Jak-STAT pathway is also downregulated in the WBCs of breast cancer

patients. Antigen presentation is regulated through the Jak-STAT pathway by activation of IL2 and IFN-γ receptors (Schroder et al., 2004). The leukocyte transendothelial migration pathway had five genes changing: Integrin alpha 4 (*ITGA4*), Tyrosine protein kinase (*TXK*) were downregulated and Protein Kinase C Gamma (*PRKCG*), Protein Tyrosine Kinase 2 Beta (*PTK2B*), Claudin 11 (*CLDN11*) were upregulated. All the genes changed in the pathway favours cell migration. It can be observed in Figure 4.8 that the effect of cancer on the leukocyte transendothelial migration in WBCs is inconclusive. But the expression data can be combined with protein data to obtain a better understanding of this pathway.

Table 4.6 Top 10 Pathways obtained by comparison of gene expression profiles of healthy donors and breast cancer patients: List of pathways and genes involved in immune response to tumour obtained by high throughput proteomics by comparison of WBCs from healthy donor and breast cancer patients.

Kegg Pathway	No of Genes	Genes involved	P value
Metabolic pathways	37	*MAT1A,FUT5,AK2,UROC1,GNE,B3GAT3,DPAGT1, ISYNA1,PLCB3,GBGT1,IDI1,AMY2B,GOT2,ALPL,CS, NDUFV3,PCYT2,ALDH18A1,ACAD8,PIGV,SMS, TK1,LIPF,UROS,UGT2A1,ITPA,UCKL1,IDH3A,ADSS, SDS,PRPS2,GCH1,AANAT,HMBS,PTDSS1,PSPH,HDC*	2.49e10
Neuroactive ligand-receptor interaction	10	*ADRA2C,CRHR1,DRD2,MC1R,LHB,ADRB1, ADRA2B,HRH2,HTR1D,GPR35*	0.0014
Pathways in cancer	10	*PRKCG,JAK1,TPM3,MYC,NFKBIA,STAT1, PIAS2,TFG,TRAF4,SMAD4*	0.005
Jak-STAT signaling pathway	9	*LIF,JAK1,MYC,STAT1,PIAS2,IL20,IL28A,I L2RB,SOCS3*	0.0002
Lysosome	8	*CD68,AP1M2,CTSC,CTSO,AP3M2,SUMF1, LGMN,LIPA*	0.0002
Antigen processing and presentation	8	*HLA-DMA,LGMN,PSME2,HLA-F,KIR2DL3, CD74,TAP1,ARHGDIG*	0.0002
Hepatitis C	7	*JAK1,NFKBIA,LDLR,STAT1,SOCS3,PIAS2,CLDN11*	0.0014
Osteoclast differentiation	6	*JAK1,SYK,NFKBIA,LILRA1,STAT1,SOCS3*	0.007
Leishmaniasis	5	*HLA-DMA,JAK1,ITGA4,NFKBIA,STAT1*	0.0044
Leukocyte transendothelial migration	5	*PTK2B,PRKCG,ITGA4,TXK,CLDN11*	0.019

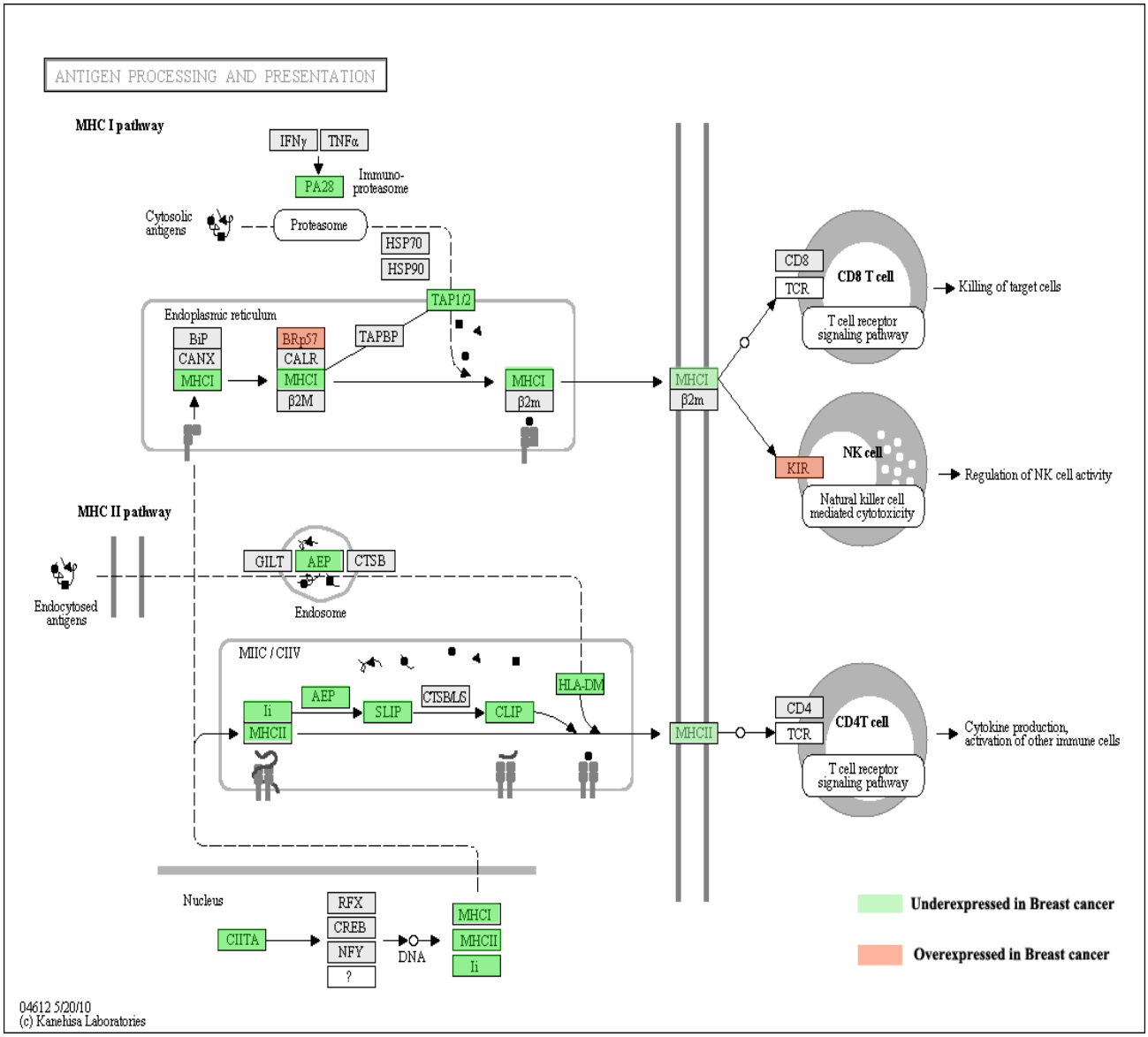


Figure 4.6 Effect of breast tumour on the Antigen presentation pathway in WBCs: Pathway analysis was done using the differentially expressed genes in the WBCs of breast cancer patients. The antigen processing and presentation pathway was significantly changed. GenMAPP was used to draw maps and the pathway database used was KEGG.

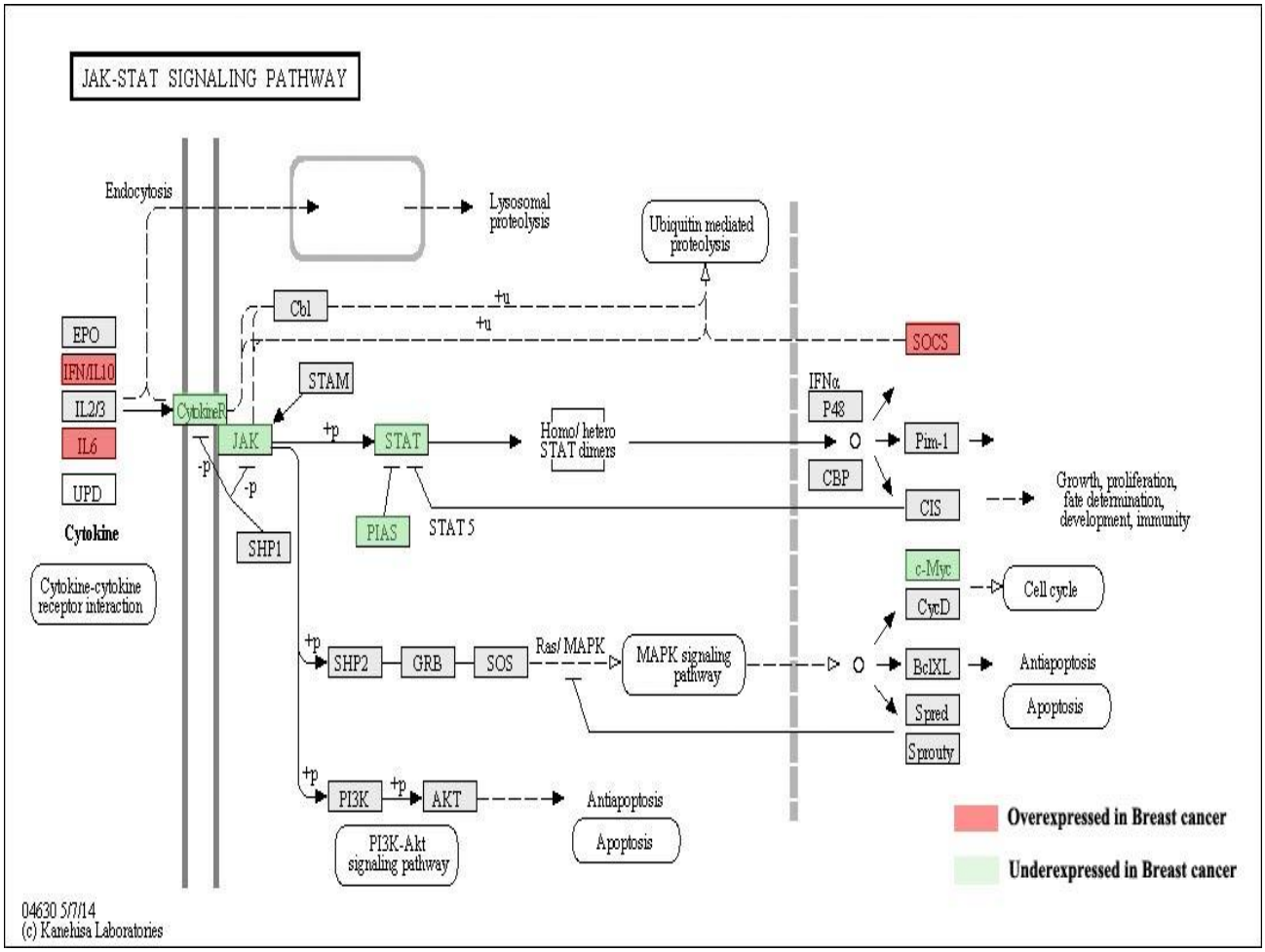


Figure 4.7 Effect of breast tumour on the JAK-STAT pathway in WBCs: Pathway analysis was done using the differentially expressed genes in the WBCs of breast cancer patients. The JAK-STAT pathway was significantly changed. GenMAPP was used to draw maps and the pathway database used was KEGG.

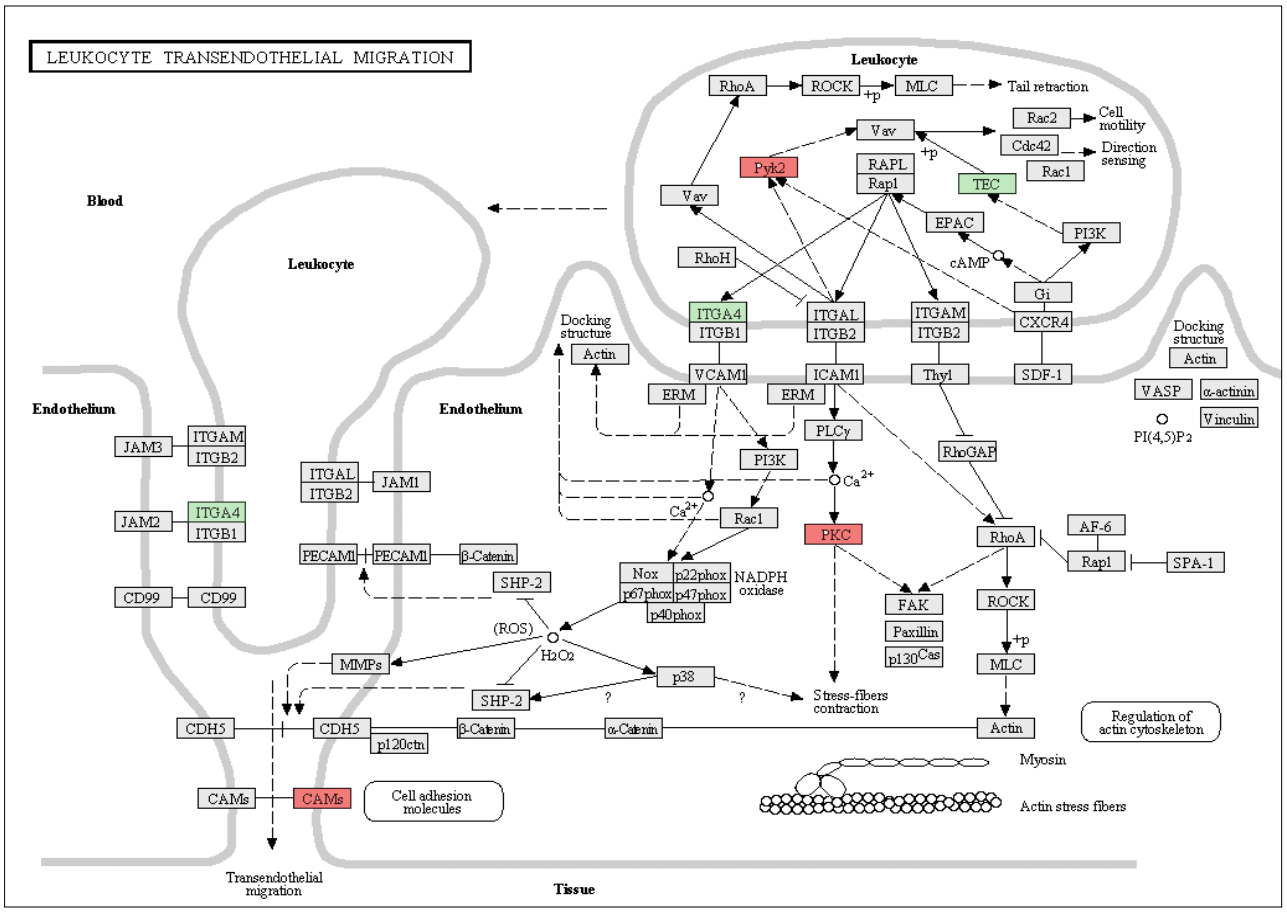


Figure 4.8 Effect of breast tumour on the Leukocyte transendothelial migration in WBCs: Pathway analysis was done using the differentially expressed genes in the WBCs of breast cancer patients. Leukocyte transendothelial migration pathway was significantly changed. GenMAPP was used to draw maps and the pathway database used was KEGG. Genes colored in red were upregulated and green were downregulated in the WBCs of breast cancer patients.

4.3.2.6 Construction of gene regulatory network models using the differentially expressed genes in the WBCs of breast cancer patients

Gene interaction network construction was done using Osprey software. Osprey is a tool for graphical visualisation of complex biological interaction networks using the gene ontology annotated dataset maintained by Biological General Repository for Interaction Datasets (BIOGRID). This has been explained further in Section2.4.3.4. The list of significant genes differentially expressed in the WBCs (Table 4.1) was used to generate the interaction network (Figure 4.9). Hub genes are those genes which are highly connected to other genes and when altered leads to drastic phenotypic changes in the cell which could be lethal. The hub genes which were identified in Figure 4.9 were *MCM7, SMAD4, SYK, STX1A, PTK2B, JAK1, RBBP4, LUC7L2, STAT1, NFKIBA* and *NCOR2*. All the genes except for *MCM7, PTK2B* and *STX1A* were downregulated in the breast cancer patient cohort. A gene regulatory network was also created using the genes involved in immune response (Figure 4.10) using the list in Table 4.2. The hub gene nodes in the immune system gene network were *JAK1, STAT1, SOCS3, MAP3K8 and NFKBIA*. The ontology terms enriched among these hub genes were cell proliferation (*MCM7, NFKBIA, PTK2B, RBBP4, SMAD4, STAT* and *SYK*), regulation of immune system process (*JAK1, MAP3K8, NFKBIA, SOCS3, STAT1*and*SYK)* and regulation of localisation (*NFKBIA, PTK2B, SMAD4, STX1A* and *SYK*). The genes involved in major pathways are *JAK1, STAT1* (Jak-STAT pathway), and *NFKIBA* (NFκB pathway). Jak-STAT pathway is involved in various immune responses such as antigen presentation, interferon-γ signalling and interleukin signalling (IL-2, IL-4, IL-7, IL-9, IL-15 and IL-21) (Shuai and Liu, 2003). NFκB pathway also is very important in activating several immune system pathways such as chemokine signalling, inflammatory processes etc (Richmond, 2002).

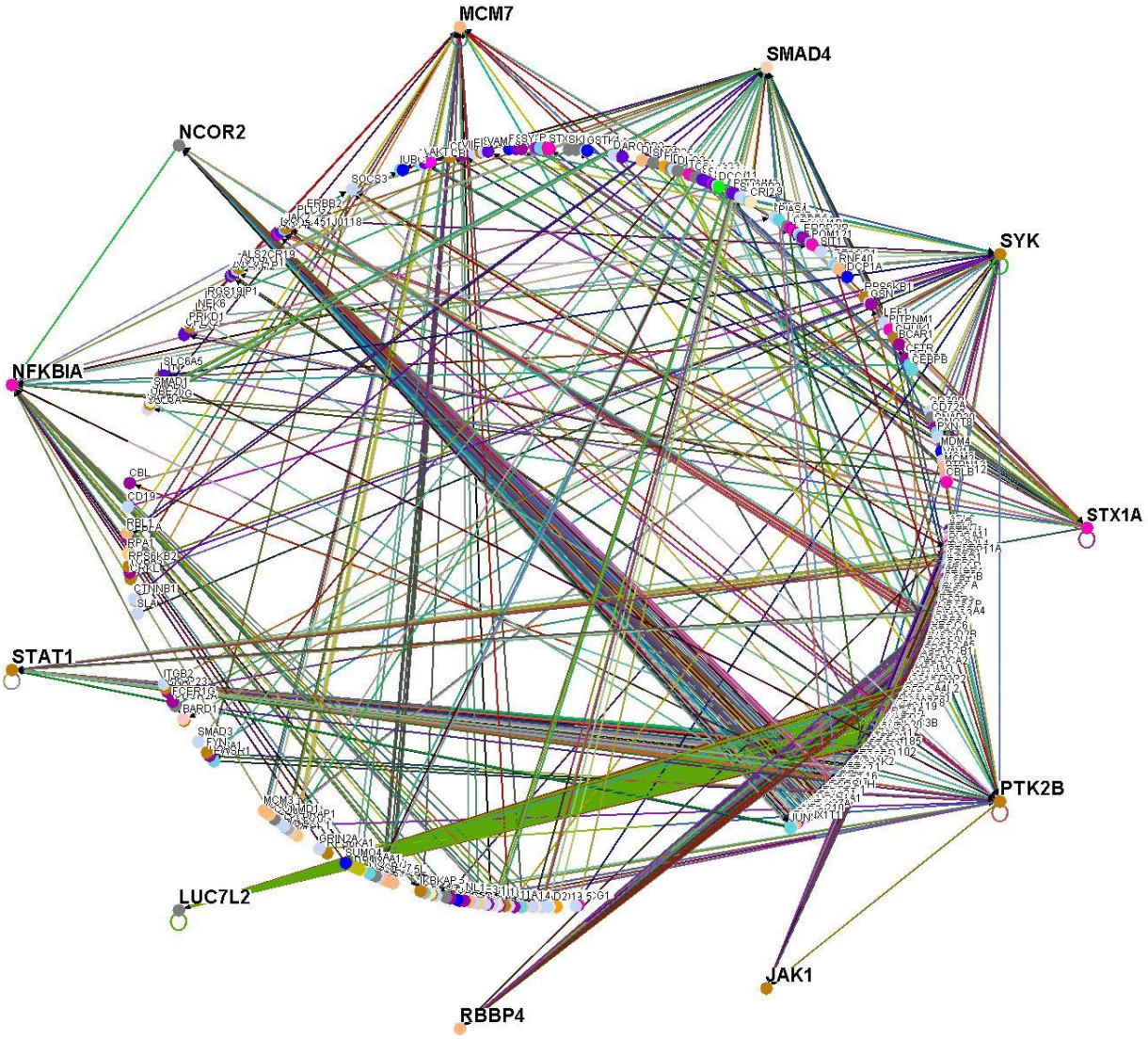


Figure 4.9 Identification Hub gene nodes changed in the WBCs of breast cancer patient cohort: Differentially expressed genes in the WBCs of breast cancer patients were used to construct gene regulatory networks. Network was constructed using Osprey (http://biodata.mshri.on.ca/osprey/servlet/Index). The lines between gene nodes represent interactions and the genes with most interactions (hub genes) are arranged on the outer layer of the network.

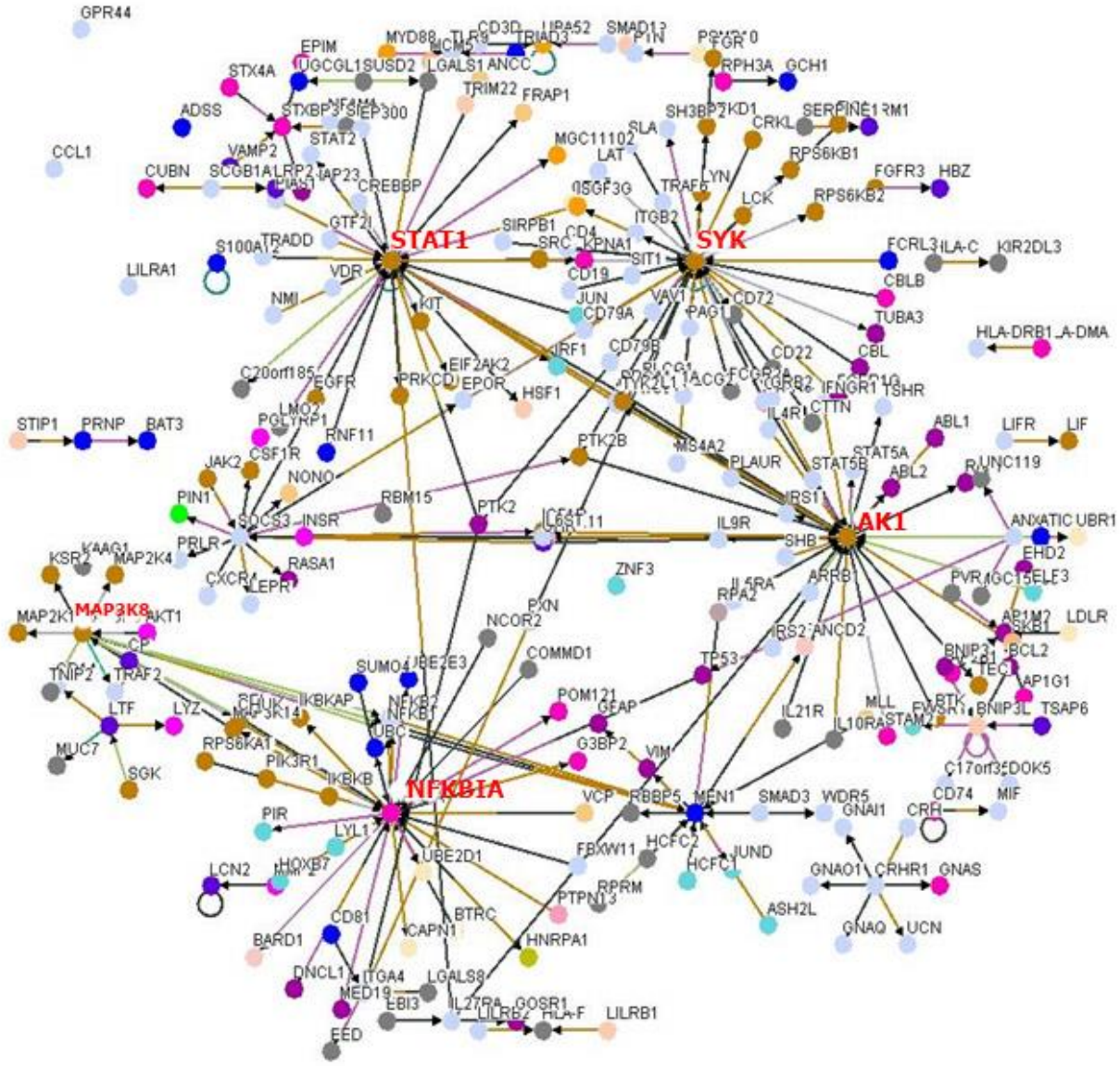


Figure 4.10 Identification Hub gene nodes related to immune response changed in the WBCs of breast cancer patient cohort: Differentially expressed genes involved in the immune response in the WBCs of breast cancer patients were used to construct gene regulatory networks. Network was constructed using Osprey (http://biodata.mshri.on.ca/osprey/servlet/Index). The lines between gene nodes represent interactions and the hub genes with their own network are arranged independent of each other.

4.4 Discussion

The microarray and more recent RNA-Seq technologies, used to generate gene expression profiles in different tissues and biological fluids have been very useful to understand the molecular mechanisms of diseases allowing identification of potential biomarkers for cancer and other conditions, although such studies are also very challenging. The WBCs represent a valuable resource for such studies. Indeed, the potential WBC biomarkers have been obtained by using these technologies for several types of cancers including pancreatic (Baine et al., 2011), colorectal (Xu et al., 2013b), renal (Twine et al., 2003) and lung cancers (Rotunno et al., 2011). These studies have also shown that the expression profiles in the WBC in cancer patients is altered compared with healthy donors. However, it has been generally accepted by the scientific community that in order to further evaluate the data obtained in these experiments, additional analyses of the RNA (e.g. by RT-qPCR) and proteins (e.g. by immunostaining, western blot, ELISA)are required.

The challenge of analysing the dataset of the breast cancer WBC (GSE16443) has been that WBCs are a mixture of different types of cells. Also each sample was taken from different individuals at varying stages of disease progression. This may be the cause for the variability of individual gene levels in the same cohort of healthy donors or breast cancer patients. Such variations were also reported in the peripheral blood mononuclear cells of patients with advanced renal cell carcinoma (Twine et al. 2003). It will be interesting to identify whether this high level of variability could be correlated with any clinical categories of diagnosis or response to treatment.

In this chapter the analysis was done on the gene expression data of the WBCs of breast cancer patients. The gene expression profiles of breast cancer patients were compared with healthy donors using Students T-Test to give 506 significantly expressed genes with a “p”- value

cutoff of <0.05 and fold change of ±1.2 (Table4.1). Compared to the analysis done by Aarøe et al., 2010 there were 49 genes which overlapped with the analysis done in this chapter. Four genes which had a fold change of more than +2 in breast cancer patients. They were Lipocalin2 (*LCN2*), Lactotransferrin (*LTF*), Orosomucoid (*ORM1*) and murine retrovirus integration site 1 homolog (*MRVI1*). LCN2, also known as neutrophil gelatinase-associated lipocalin, is involved in innate immunity and its function is to sequester iron that limits bacterial growth. LCN2 is used as a biomarker for kidney injury (Devarajan, 2010) and is also present at elevated levels in urine and tissue samples of breast cancer patients and increases tumour cell migration (Yang et al., 2009). Lactotransferrin (LTF) is present in secreted fluids such as milk, tears, saliva and also present in neutrophil secondary granules. LTF also has iron binding properties LCN2 and binds to bacterial wall lipopolysaccharides. Both LTF and LCN2 have similar antimicrobial properties and also have anti-inflammatory properties (Adlerova et al., 2008; Shashidharamurthy et al., 2013).

Gene ontology analysis revealed that genes involved in inflammatory response were overexpressed in the WBCs of breast cancer patients. The altered genes which were common with lung cancer patient WBC study (Rotunno et al., 2011) were *CD9,CPA3,DOCK10,HEBP1,NKTR,S100A12* and colorectal cancer WBC study (Xu et al., 2013a) were *DOCK10,FKBP5,EPB41L3,F5,CLEC12A*. Functional analysis was done using gene ontology, pathway enrichment analysis and gene regulatory network construction. Gene ontology analysis was directed at the immune system processes, cell death and cell migration. The number of genes which were involved in immune system processes was 52. Further gene ontology analysis revealed that genes positively regulating WBC activation, antigen recognition, processing and presentation were under-expressed in the breast cancer cohort. Some of the altered genes involved in the antigen presentation are also responsible for lymphocyte activation were down-regulated. But the genes involved in the inflammatory response were up-regulated.

These ontology results correlated with pathway analysis where the antigen processing and presentation pathway was down-regulated as was the JAK-STAT signaling pathway which positively regulates the antigen presentation pathway. The immune response ontology analysis and pathway analysis suggests that antigen presentation in the innate immune cells, leukocyte activation and lymphocyte activation are affected in the WBCs of breast cancer patients. Although the genes involved in inflammation were up-regulated. Chronic inflammation caused by diseases such as Crohn's disease, schistosomiasis, Hepatitis C infection and *Heliobacter pylori* infection can cause cancer in the bowel, liver and stomach respectively (Coussens and Werb, 2002). It has also been reported that immunosupression after organ transplant increases the probability of developing a tumour between 5-6% (Penn and Starzl, 1973).

Notably, the WBCs in breast cancer patients had properties associated with cell proliferation. Thus, the genes involved in induction of apoptosis were down-regulated and genes in positive regulation of cell growth were up-regulated. It was previously reported that Fas ligand attracts inflammatory cells into the tumour microenvironment to induce apoptosis (O'connell et al., 1996; Whiteside, 2002), however the data in this study only represent peripheral blood WBCs not the WBCs infiltrating the tumour.

It has been shown in several studies that leukocyte migration is inhibited in various carcinomas (Brandes and Goldenberg, 1976; Kadish et al., 1976; Lee et al., 1977). Analysis of differentially expressed genes involved in cell migration revealed that all the genes negatively affecting cell migration were overexpressed in the breast cancer patient cohort. The gene network analysis also demonstrated that the hub genes which changed were involved in cell proliferation and regulation of cell localisation or migration. The hub genes which were involved in major cell processes were *JAK1, STAT1* and *NFKBIA*. *NFKBIA* codes for Nuclear Factor of Kappa Light Polypeptide Gene Enhancer in B-Cellsa member of the NFκB inhibitor family. NFκB has been implicated to play a central role in inflammation and cancer, and is required for

B-cell maturation and survival (Karin, 2006). TNFα, IL61, VEGF and IL6 have been known to be secreted in tumour microenvironments which also activate NFκB(De Visser et al., 2006). The gene expression analysis provided potential biomarker candidates which would be compared in tandem with the proteomics data from the previous chapter to provide a shortlisted list of biomarker candidates.

Chapter 5: Characterisation and Validation of novel Biomarkers for Breast Cancer from White Blood Cells

5.1 Introduction

The understanding of molecular signatures of different types of cells has greatly increased over the past decade due to rapid technological advances. The clinical need in finding cancer biomarkers led to huge investments into this area. Since 1986 there has been over 5000 projects funded by the US National institute of Health for biomarker discovery. Between 1990 and 2008 the number of publications relating to biomarkers alone have exceeded 4000 (Ptolemy and Rifai, 2010). Approaches using technologies such as microarrays, highthroughput DNA sequencing, highthroughput proteomics and other modern methods have been useful in identifying potential biomarkers for different disease conditions. However, despite huge research efforts the number of biomarkers used in the clinic is very small. One of main problems is the failure of the biomarkers to pass the validation or clinical stages when the study is expanded to include larger sample numbers. This has been explained and discussed in detail in Chapter 1 (Section 1.10).Another challenge in biomarker discovery is data reduction to shortlist candidates for the next step in the biomarker discovery pipeline. Different research groups have combined both the genomics and proteomics approaches to identify biomarkers with the results of the combined data better than just analysing the data from one approach (Chen et al., 2002; Nishizuka et al., 2003; Orntoft et al., 2001).Using both genomic and proteomic approaches will reduce the overwhelming volume of data to be considered for preliminary biomarker selection. However the integration of gene expression and protein data needs to be approached with caution as the experimental and data processing methods vary which might lead to increase in the probability of obtaining false positives. In this chapter the results from the microarray dataset (Chapter 4) and the proteomic analysis (Chapter3) have been combined to shortlist a panel of

biomarkers. Additionally data from another study conducted by Dr. Dawn Farrar from our laboratory was also included in the shortlisting process. The latter study is described in the following Section.

5.2 Advanced breast cancer pilot study

The aim of this investigation (acronymed Abs) was to explore the changes in the WBCs in patients with advanced breast cancer (also referred to as metastatic breast cancer, or MBC) in response to chemotherapy and endocrine therapy, using the 2D-gel electrophoresis method. This was a collaborative study between the Department of Oncology at the Colchester Hospital University NHS Foundation trust and the University of Essex. Experiments for this study were conducted by Dr. Dawn Farrar (Senior Research Fellow) from Prof. Elena Klenova's laboratory

5.2.1 Background of the study

The advanced (recurrent) breast cancer is often associated with distant spread or metastasis of tumour cells from the primary tumour. The incidence of advanced breast cancer is between 5%-10% of total number of breast cancers diagnosed newly and the mortality rate in this group is 80% (Cardoso et al., 2012). The pathology of advanced metastatic breast cancer is discussed in detail in Chapter 1.2. Currently the assessment of response to therapy relies on both subjective and objective measures. Subjective assessments can be confounded by the side effects of the therapy and objective measurements are not usually indicative of response until a number of cycles of chemotherapy or several months of hormonal therapy have been given. Some patients do not have easily measurable disease and the response to treatment can only be assessed on purely subjective measures. The currently available circulating markers e.g. CEA and CA15-3 are not ideal in metastatic breast cancer as they are not routinely elevated and do not reliably show an early change with therapy.

The clinical response is categorised subjectively based on the concept above as complete response (CR), partial response (PR), stable disease (SD, progressive disease (PD) and non-evaluable disease (NE). This type of classification of treatment response is based purely on disease progression which changes rapidly in patients because metastatic breast disease is very aggressive and breast cancer at molecular level is diverse. The discovery of a suitably sensitive blood biomarker could potentially replace and/or complement the assessment currently performed using imaging such as CT, Isotope bone scans and MRI. The focus of the investigation was on the identification of WBC–based biomarkers for monitoring of anti-cancer treatment based on simpler biochemical tests. Such biomarkers are required to be sensitive, specific and more reliable for prognosis of metastatic breast cancer to therapeutic treatment which can also complement the existing methods of investigation. The focus of the investigation was on the identification of WBC–based biomarkers for monitoring of anti-cancer treatment based on simpler biochemical tests. The study was conducted in collaboration between the Colchester General Hospital and the University of Essex.

5.2.2 Patient recruitment and blood sample collection

Patients with MBC were recruited for the study before the treatment commenced, with the written consent obtained from the participants. Two types of treatment were as follows: Chemotherapy (6 cycles) and Endocrine therapy (3 cycles). Patients in the final stages of the disease with less than 3 months of life expectancy did not participate in the study. The blood (10 ml) was collected from the patients before the beginning of the treatment and before each treatment cycle. Patients were assessed by physicians before each cycle using radiological investigation techniques. Samples were collected at 6 weeks and 12 weeks after the beginning of treatment for endocrine therapy patients. The WBCs were fractionated from whole blood samples in the lab according to the protocol in Chapter2 (Section2.1) and the WBCs were split into different parts for various experiments. 2D-gel electrophoresis was performed on pre- and

post-treatment WBC samples according the protocol in Chapter 2 (Section 2.2.3). Protein spots of interest were identified through image analysis using Samespots and the protein identification was done by Mass spectrometry.

5.2.3 Results for the Advanced breast cancer pilot study

Thirty one patients were recruited for this study. Twenty patients had chemotherapy, two patients had both therapies and nine had endocrine therapy. The patients responses at the end of the treatment were varied; CR=1, PR=8, SD=4, PD=14 and NE=1. Three patients had died during or immediately after the course of treatment. The 2D gel electrophoresis was performed on the WBC samples and 64 protein spots were identified to be potential biomarker candidates for prognosis which showed correlation with treatment response and treatment type. Figure 5.1 shows the 2D- gel image of the WBCs from patient Abs013 where the64 significant protein spots are shown. The protein spots were identified using mass spectrometry and 75 protein identifications were obtained. Table 5.1 gives the list of these proteins.

5.3 Aims of the Chapter

In this Chapter genomic data from the GSE16443 dataset analysis (Chapter 4), proteomic data (both high throughput proteomics and 2D-PAGE) (Chapter 3) and data from the advanced breast cancer study conducted by Dr. D Farrar have been combined to identify a panel of potential WBC biomarkers to be validated using the reverse transcriptase quantitative PCR (RT-qPCR) method first. Based on the results from RT-qPCR promising candidates were chosen for further validation by western Blotting (WB).

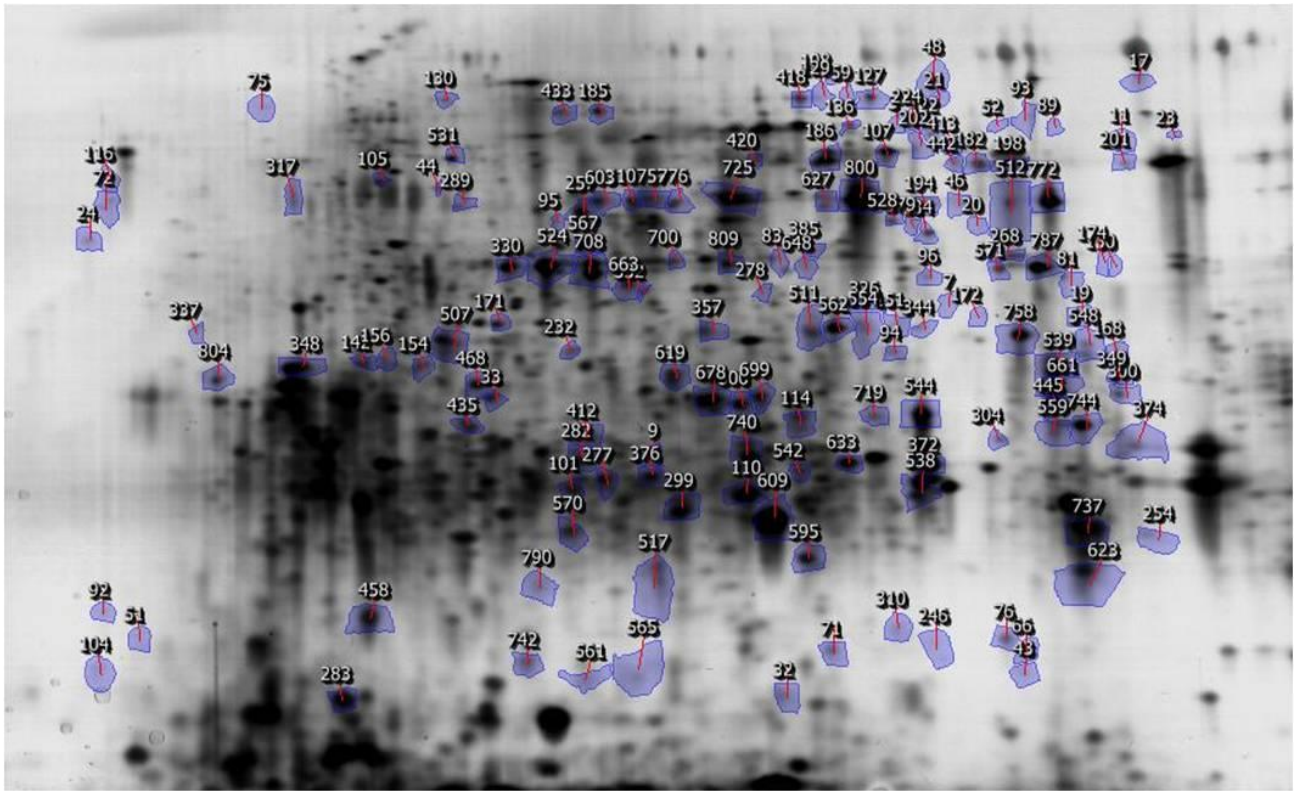


Figure 5.1 Identification of novel Metastatic Breast Cancer biomarkers from WBCs: Blood samples were collected from MBC patients before treatment and over the course of treatment. 2D- gel electrophoresis was carried out on each sample and gels were run in triplicates. Protein spots (n=64) which changed over the course of treatment and with types of metastasis were identified using Samespots software. Fold change in protein concentration was taken into consideration. Protein ids were obtained through mass spectrometry. (Experiments were performed by Dr D. Farrar).

Table 5.1 List of protein spot identifications obtained by comparing WBC protein expression profiles over the course of chemotherapy and endocrine therapy from Metastatic breast cancer (MBC) patients: Blood samples were collected from MBC patients before treatment and over the course of treatment. 2D- gel electrophoresis was done on each sample and gels were run in triplicates. Protein spots (n=64) which changed over the course of treatment and with types of metastasis were identified using Samespots software. Fold change in protein concentration was taken into consideration. Protein ids were obtained through mass spectrometry.

Gene symbol	Gene name	Gene symbol	Gene name
ACTN1	Actinin, alpha 1	LDHB	Lactate dehydrogenase B
ADH5	Alcohol dehydrogenase 5	LTF	Lactotransferrin
ALDOA	Aldolase A, fructose-bisphosphate	MMP8	Matrix metallopeptidase 8
ALDOC	Aldolase C, fructose-bisphosphate	MMP9	Matrix metallopeptidase 9
ANXA11	Annexin A11	MPO	Myeloperoxidase
ANXA3	Annexin A3	MSN	Moesin
ANXA6	Annexin A6	MYH9	Myosin, heavy chain 9, non-muscle
ANXA7	Annexin A7	NONO	Non-POU domain-containing octamer-binding protein
ARHGDIB	Rho GDP dissociation inhibitor (GDI) beta	NCF4	Neutrophil cytosolic factor 4, 40kDa
ARPC2	Actin related protein 2/3 complex, subunit 2, 34kDa	P4HB	Procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), beta polypeptide
BASP1	Brain abundant, membrane attached signal protein 1	PCNA	Proliferating cell nuclear antigen
CALR	Calreticulin	PDIA4	Protein disulfide isomerase family A, member 4
CAPG	Capping protein (actin filament), gelsolin-like	PGAM1	Phosphoglyceratemutase 1 (brain)
CAPZA1	Capping protein (actin filament) muscle Z-line, alpha	PGK1	Phosphoglycerate kinase 1

	1		
CAT	Catalase	PGLYRP1	Peptidoglycan recognition protein 1
CHIT1	Chitinase 1 (chitotriosidase)	PKM2	Pyruvate kinase, muscle
CLC	Charcot-Leyden crystal protein	PNP	Purine nucleoside phosphorylase
CORO1A	Coronin, actin binding protein, 1A	PPP1CA	Protein phosphatase 1, catalytic subunit, alpha isoform
CPNE3	Copine III	PPP2CA	Protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform
CTSG	Cathepsin G	PRKCSH	Protein kinase C substrate 80K-H
DEF6	Differentially expressed in FDCP 6 homolog (mouse)	PRTN3	Proteinase 3
DSP	Desmoplakin	PTPRC	Protein tyrosine phosphatase, receptor type, C
FLNA	Filamin A, alpha (actin binding protein 280)	PYGL	Phosphorylase, glycogen
G6PD	Glucose-6-phosphate dehydrogenase	RAN	RAN, member RAS oncogene family
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	RPIA	Ribose 5-phosphate isomerase A (ribose 5-phosphate epimerase)
GNAI3	Guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 3	RPS3A	Ribosomal protein S3A
GNB2L1	Guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1	S100A8	S100 calcium binding protein A8
HBB	Hemoglobin, beta	SERPINA1	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1
HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1	SERPINA3	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3
HNRNPA2B1	Heterogeneous nuclear ribonucleoprotein A2/B1	SERPINB1	Serpin peptidase inhibitor, clade B (ovalbumin), member 1
HSP90AA1	Heat shock protein 90kDa alpha (cytosolic), class A member 1	SERPINB10	Serpin peptidase inhibitor, clade B (ovalbumin), member 10
HSPA1A	Heat shock 70kDa protein 1A	SERPINB6	Serpin peptidase inhibitor, clade B (ovalbumin), member 6

HSPA8	Heat shock 70kDa protein 8	TALDO1	Transaldolase 1
ITGAM	Integrin, alpha M	TKT	Transketolase
ITGB2	Integrin, beta 2	TPM3	Tropomyosin 3
LCN2	Lipocalin 2 (oncogene 24p3)	VCP	Valosin-containing protein
LDHA	Lactase dehydrogenase A	WDR1	WD repeat domain 1
XRCC6	X-ray repair complementing defective repair in Chinese hamster cells 6	YWHAE	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, epsilon polypeptide

5.4 Results

5.4.1 Selection process of new WBC candidate biomarkers for further characterisation and validation

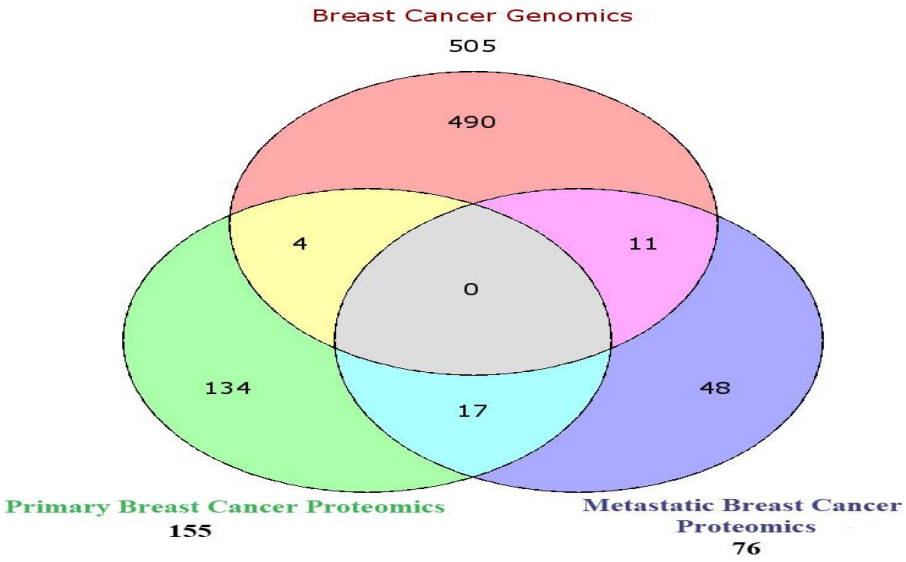
5.4.1.1 Comparing genomic and proteomic data to identify common overlaps

The short listing of biomarkers involved comparing the data from both genomics and proteomics approaches. Significant genes or proteins had been obtained by comparing expression profiles of the WBCs of breast cancer patients and healthy donors, using appropriate fold change and p-value cut-offs. The next step was to compare the list of these significant genes to find common overlaps. The data from the metastatic breast cancer study was also was also used to identify common overlaps. Figure 5.2 shows the comparison of the three datasets and the number of overlapping molecules between each of the datasets. The list from the highthroughput protein analysis and 2D-gel electrophoresis analysis of primary breast cancer patients were combined to give the primary breast cancer proteomic dataset. It was surprising to observe that there were only four overlaps between the microarray dataset and proteomics dataset of primary breast cancer patients. Seventeen proteins overlapped between the proteomic datasets of primary breast cancer and metastatic breast cancer. Eleven overlaps were found between primary breast cancer microarray dataset and the metastatic breast cancer dataset. There were no common proteins between all the three datasets. The table in Figure 5.2 gives the list of those genes which have overlapped in more than one dataset.

5.4.1.2 Short listing of candidates for validation using RT-qPCR

The common overlaps were identified by comparing the list of genes or proteins between different datasets. The next step in short listing involved several criteria such as: normal presence in WBC, correlation with clinical response/data, relative levels to healthy donor, function in immune response and association with cancer through literature search. Genes or proteins which

were involved in purely metabolic processes were excluded. Genes with immune system functions were given priority while shortlisting. Fifteen genes which appeared in more than one dataset were chosen for validation on these criteria for validation by RT-qPCR. For example, ALOX5, OSTF1 and FGL2 were chosen based on their WBC expression profiles of breast cancer patients from the highthroughput proteomics dataset. FGL2 was not present in the healthy donor cohort but present in the breast cancer cohort. ALOX5 and OSTF1 were chosen because of their fold change of +1.6 in the breast cancer cohort. The list of the proteins is given in Table 5.2 along with the details of their presence in different types of datasets.The presence in metastatic breast cancer cohort is denoted by 'X' and the fold change is not included since the initial analysis did not include comparison with healthy donors.



A

Category	Overlapping candidate biomarkers
Between Primary Breast cancer Genomics and Proteomics data (4 candidates)	CAMP,CSTA,ITGA4,RHOA
Between Primary Breast cancer and Metastatic Breast cancer proteomics data (17 candidates)	ACTN1,ARPC2,CAPG,CAPZA1,CHIT1,CORO1A,GNAI3,HNRNPA1,ITGAM,MSN,NCF4,PGAM1,PYGL,SERPINB1,SERPINB6,TKT,YWHAE
Between Primary Breast cancer Genomics data and Metastatic Breast cancer proteomics data (11 candidates)	ANXA3,BASP1,CALR,CPNE3,LCN2,LTF,NONO,PGLYRP1,S100A8,TPM3,WDR1

B

Figure 5.2 Overlapping data between genomic and proteomic approaches presented as Venn diagram. Data was obtained by comparing the WBC profiles of breast cancer patientsand healthy donors using microarray expression dataset (GSE16443) and proteomic dataset (highthroughput protein profiling and 2D-gel electrophoresis. This was compared with the 2D-gel electrophoresis analysis results for metastatic breast cancer patient WBC to find overlapping proteins to be shortlisted for characterisation and validation.

Table 5.2 Short listed candidates chosen for validation by RT-qPCR: Genes common from different datasets comparing healthy donors, primary breast cancer patients and metastatic breast cancer patients were obtained. The overlapping genes were further shortlisted based on several criteria; normal presence in WBC, correlation with clinical response/data, relative levels to healthy donor, function in immune response and association with cancer through literature search to give a 15 panel gene list to be validated using RT-qPCR. Red coloured cells indicate overexpression; green indicates underexpression; pink overexpression only in low grade breast cancer.The presence in metastatic breast cancer cohort is denoted by 'X' and the fold change is not included since the initial analysis did not include comparison with healthy donors

Candidates	Primary Breast cancer-2D-gel electrophoresis	Primary Breast cancer-High throughput proteomics	Primary Breast cancer -Microarray (GSE16443)	Metastatic Breast cancer-2D-gel electrophoresis	No of categories of the overlap
SERPINB1	+1.3	+1.46		X	3
CALR			+1.4	X	2
YWHAE		+1.46		X	2
CPNE3			+1.2	X	2
ANXA3			+1.97	X	2
ANXA1	+1.3	+1.49			2
LTF			+2.37	X	2
NONO			+1.21	X	2
ITGA4		-3.4	-1.28		2
WDR1			-1.23	X	2
RHOA		+1.4	+1.5		2
LCN2			+2.16	X	2
ALOX5		+1.6			1
FGL2		+			1
OSTF1		+1.23			1

5.4.1.3 Description of genes selected for validation using RT-qPCR

5.4.1.3.1 SerpinB1 (SERPINB1)

SerpinB1 is also known as neutrophil elastase inhibitor. The protein encoded belongs to a class of protease inhibitors. SerpinB1 inhibits neutrophil serine protease, elastase, cathepsin G and proteinase-3 (Farley et al., 2012). It also is involved in neutrophil extravasation and formation neutrophil extracellular traps SerpinB1 is associated with cancer as both tumour suppressor and promoter (Chou et al., 2012; Cui et al., 2014; Tseng et al., 2009).

5.4.1.3.2 Calreticulin (CALR)

Calreticulin is a chaperone protein which binds to calcium and is involved in several functions of the immune system such as Major histocompatibility complex folding and assembly, phagocytosis (Raghavan et al., 2013). Calreticulin is also implicated in breast cancer aggressiveness and metastasis (Lee et al., 2012; Lwin et al., 2010). It is also proposed that Calreticulin is serum biomarker for lung cancer and bladder cancer (Kageyama et al., 2004; Liu et al., 2012a).

5.4.1.3.3 Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, epsilon (YWHAE)

YWHAE belongs to the family of 14-3-3 proteins which are involved in various biological processes such as cell signalling, cell cycle, apoptosis and regulation of transcription (Tzivion et al., 2001). Reduced YWHAE expression is shown to contribute to the epithelial to mesenchymal transition of hepatocellular and gastric cancer cells (Leal et al., 2012; Liu et al., 2013). Increased expression has also been reported in breast cancer and lung cancer (Li et al., 2006; Qi et al., 2005).

5.4.1.3.4 Copine 3 (CPNE3)

Copine3 is a ubiquitously expressed protein but its function is unknown. It belongs to a family of phospholipid binding protein. Copine 3 is upregulated in tumour cells and interacts with ErbB2 to promote tumour cell migration (Cowland and Carter, 2003; Heinrich et al., 2010).

5.4.1.3.5 Annexin A1 and Annexin A3

Annexins are also called lipocortins and belong to family of calcium dependent phospholipid binding proteins located on the cytosolic face of the plasma membrane. ANXA1 and ANXA3 both inhibit phospholipase A2 which causes inflammation (Walther et al., 2000).Reduced ANXA1 expression is linked to invasiveness of breast cancer, gastric cancer and neck cancer (Pedrero et al., 2004; Yom et al., 2011; Yu et al., 2008a). ANXA3 also correlates to tumour progression in thyroid cancer and hepatocellular carcinoma (Jung et al., 2010; Pan et al., 2013).

5.4.1.3.6 Lactoferrin (LTF)

Lactoferrin is secreted protein that is found in various bodily fluids such as milk, saliva, tears and nasal secretions. LTF is an iron-binding protein antibacterial, anti-carcinogenic and antioxidant effects. LTF is expressed in three different isoforms and first was isolated form neutrophils (Levay and Viljoen, 1995). LTF has been shown to inhibit cancer growth and metastasis (Bezault et al., 1994; Damiens and Yazidi, 1999).

5.4.1.3.7 Non-POU Domain Containing, Octamer-Binding (NONO)

NONO is a RNA-binding protein which plays various roles in cellular processes including transcriptional regulation and RNA splicing (Amelio et al., 2007). NONO is implicated in progression of malignant melanoma (Schiffner et al., 2011).

5.4.1.3.8 Integrin Alpha 4 (ITGA4)

ITGA4 is also known as CD49d belongs to the family integrin alpha chain proteins which are involved in cell adhesion and migration.Integrin receptors are critical for cell attachment to the extracellular matrix (ECM) and this is mediated through integrin-fibronectin interaction (Barczyk et al., 2010). Integrins are also involved in transendothelial migration of leukocytes which is very important for the functioning of the innate immune system (Williams and Solomkin, 1999).

5.4.1.3.9 WD Repeat domain 1 (WDR1)

WDR1 protein contains nine WD domains that are involved on protein-protein interactions. The function of WDR1 is unclear but it may help actin cytoskeletal rearrangement and cell migration. This protein has been reported to be overexpressed in ovarian cancer (Haslene-Hox et al., 2013).

5.4.1.3.10 The Ras Homolog Family Member A (RHOA)

RHOA is a small GTPase protein which regulates the formation of actin fibers linking which connects the receptors on the plasma membrane. RHOA overexpression has been linked to tumour progression/invasion in breast cancer, prostate cancer and testicular cancer (Hodge et al., 2003; Kamai et al., 2004; Pillé et al., 2005).

5.4.1.3.11 Lipocalin 2 (LCN2)

LCN2 protein is associated with multiple cellular processes such as apoptosis, innate immunity etc. Its molecular function is sequestering iron siderophores which limits bacterial growth. LCN2 overexpression was found to be associated with aggressive breast tumors (Yang et al., 2009) and linked to breast cancer patient response to neoadjuvant chemotherapy (Wenners et al., 2012).

5.4.1.3.12 Arachidonate 5 Lipoxygenase (ALOX5)

ALOX5 plays a role in synthesis of leukotrienes from arachidonic acid. Leukotrienes are important mediators of inflammation. ALOX5 has been linked to promoting prostate cancer and breast cancer cell survival (Ghosh, 2003; Hu et al., 2011).

5.4.1.3.13 Fibrinogen-like 2 (FGL2)

The function of FGL2 is unknown but it may play a role physiologic lymphocyte function at mucosal sites.FGL2 has been implicated playing a role in conversion of prothrombin to thrombin. It has shown to be overexpressed in colon, breast, lung, gastric, oesophageal and cervicalcancers (Liu et al., 2012b)**.**

5.4.1.3.14 Osteoclast Stimulating Factor (OSTF1)

OSTF1 is produced by osteoclasts and indirectly induces osteoclast formation and bone resorption. Breast cancers commonly cause bone metastasis and this process is dependent on osteoclast-mediated bone resorption (Thomas et al., 1999).

5.4.2 Optimisation of RNA extraction and quality control to study gene expression from WBCs

5.4.2.1 Background

Relative quantification of RNA is a popular method to study the gene expression changes to cells in response to biological stimuli. To obtain high quality results for this analysis it is important to (i) obtain high quality RNA and (ii) identify a suitable reference gene to normalise the expression of the target gene. Purification of RNA from white blood cells (fresh or cryopreserved) can be very challenging. The RNA yield is often very low when extracting due to the presence of residual red blood cells in the WBC mixture, the RNA quality is poor because of the salts carried over from RNAlater and purity is not satisfactory due to contaminating genomic

DNA. Furthermore, RNA degrades rapidly by exogenous and endogenous RNA nucleases, storage in cryostate over the time and also exposure to heat. These changes in RNA integrity can lead to falsely altered gene expression patterns.

5.4.2.2 Optimisation of RNA extraction from WBCs

In order to obtain RNA samples of sufficient quality from WBCs to be used for RT-qPCR the procedure of RNA purification needed to be optimised as the initial experiments resulted in the WBCs RNA of poor quality and low amounts. The conventional phenol extraction protocol was modified to suit WBCs to extract RNA of satisfactory quality and purity (see Section 2.3.1). The total RNA yield was low when $5x10^6$ cells were originally used and stored in RNAlater. The WBCs numbers were therefore increased up to $1x10^7$ cells for preservation in RNAlater at -80^0C. The purity of RNA which was extracted was assessed by the use of Nanodrop UV/vis spectrophotometer (see Section 2.3.3). A ratio of absorbance at 260, 230and 280 nm (A260:A280 and A260:A230) greater than 1.8 is usually considered as an acceptable indicator of RNA purity (Glasel, 1995). It was essential that the chloroform step was to be repeated twice for the 260/280 values to be higher than 1.8. The RNA had to be washed at least twice with 70% ethanol for the 260/230 values to be higher than 1.8.

Treatment of RNA with DNase is recommended to eliminate DNA contamination after RNA extraction in tissues and cell lines (Becker et al., 2010).In our experiments, RNA was treated with DNase (see Section 2.3.2) to eliminate DNA contamination. After DNase treatment, RNA was evaluated for their integrity using Agilent 2100 bioanalyzer and the results are presented in Figure 5.3. The bioanalyzer analysis does not show the presence of DNA contaminations. The electropherogram peaks in Figure 5.3 indicate the RNA marker, 5sRNA, 18sRNA and 28sRNA in the order of appearance from left to right respectively. A value of above1.5 for the ratio between 28sRNA and 18sRNA is generally viewed as acceptable for RNA

integrity. The gel image also shows if there is degradation in RNA samples. Examples of both good quality RNA and poor quality RNA are shown in Figure 5.3. It can be observed from the gel image and the electropherogram that RNA obtained from sample Brca1501 had degraded while RNA from sample HD3 is not. The final product was then converted to cDNA using RNA reverse transcriptase using the protocol described in Section 2.3.5. The absence of contamination with genomic DNA was confirmed in RNA preparations treated with DNAse and also in cDNA samples by PCR using β-actin primers (Forward primer- CTGGGACGACATGGAGAAA and reverse primer- GGGATAGCACAGCCTGGATA). The actin primers were designed across different exons so that they can both amplify cDNA and genomic DNA. After quality control the cDNA was used to gene expression studies to validate WBC biomarkers using RT-qPCR.

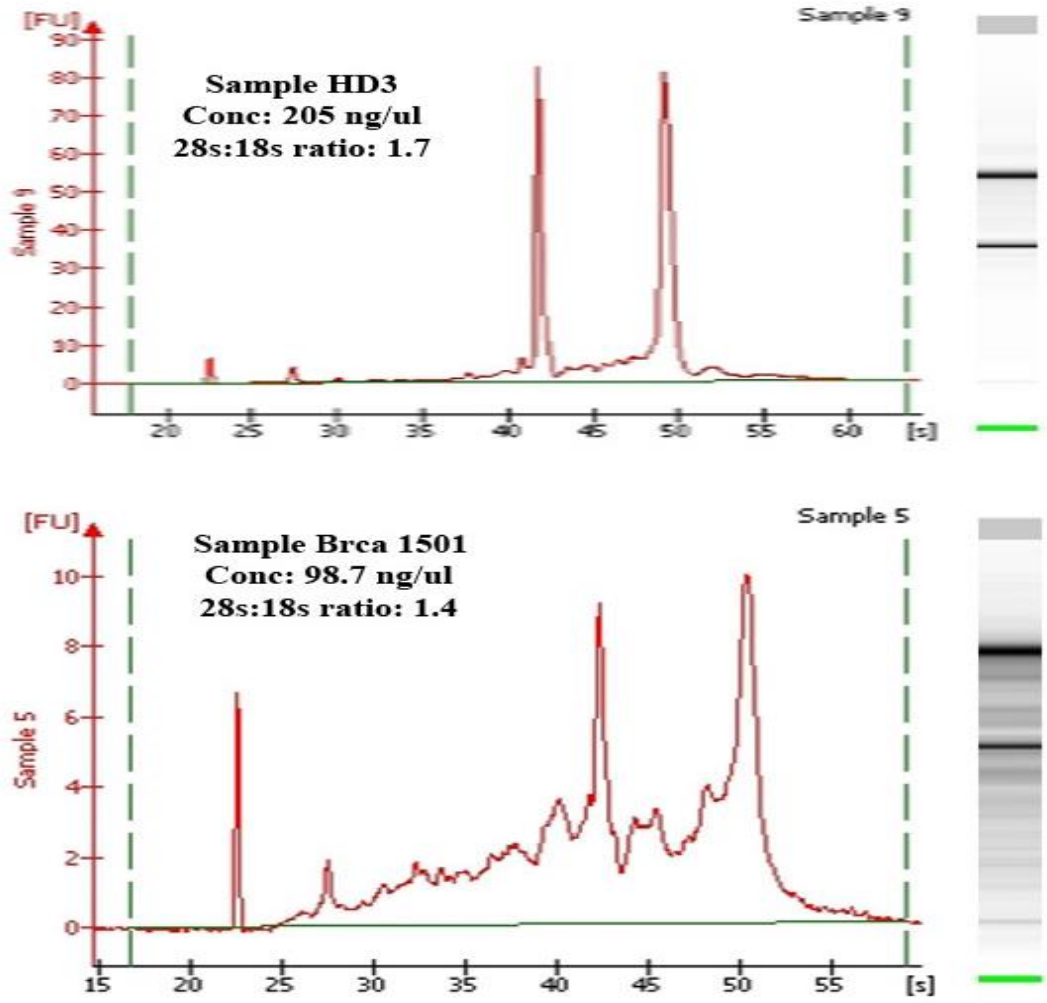


Figure 5.3 Quality control of Total RNA extracted from WBCs using Agilent 2100 Bioanalyzer: Total RNA was extracted by a modified phenol-chloroform protocol as described in Section 2.3.1. Electropherograms for two samples (HD3 and Brca1501) have been shown with the gel images on the right. The samples were treated for DNA contamination with DNase and Agilent 2100 bioanalyzer was used to assess RNA integrity. Samples with 28s:18s ratio lower than 1.5 were not selected for further analysis.

5.4.3 Selection of suitable reference gene for RT-qPCR reactions from cDNAs

The selection of reference gene to normalise the expression of a target gene across different samples from different conditions is important to obtain and interpret the biological meaning of the data. In this study we are analysing for the difference in gene expression of candidate biomarkers in the WBCs of breast cancer patients. The samples are obtained from different participants at different stages of the disease (Healthy donor and breast cancer) and breast cancer is a heterogeneous disease at the molecular and cellular level. Also, the WBCs are a mixture of different types of cells at varying proportions among participants. Therefore, it is important to identify a reference gene which can be applied for specific conditions and the varying nature of samples. Although several reference genes have previously been suggested for the whole blood (Dheda et al., 2004) and neutrophils (Ledderose et al., 2011), it was important to test a panel of the candidate reference genes in our experimental settings.

The reference gene should meet several criteria: have a constant level of expression across different samples, have stable expression across different disease conditions and have a low Ct range across the samples. To achieve this aim equal amount (500ng) of RNA was taken and converted to cDNA. Expression of six reference genes (HuPO- human acidic ribosomal protein; CycloB- cyclophylinB; B2M- β2-microglobulin; HPRT- hypoxanthine phosphoribosyltransferase; 18sRNA- 18S ribosomal RNA; RPL32- ribosomal protein 32; TBP- Tata box Binding Protein) were evaluated to find the reference gene with the lowest variance among the samples. The primers used for each reference gene are given in Table 2.2 in Chapter 2. The cDNA from six WBC samples (Breast cancer-4, Healthy Donor-2) were used for the analysis. Quantitative PCR was performed using the cDNAs and the mean Cq values were obtained and the Box plots were constructed using SPSS (Figure 5.4). It was observed that RPL32 was the best, with high consistency and low variance among the samples.TBP showed low variation but

the Cq values were high. B2M and HuP0 had the highest variability. RPL32 was chosen as the housekeeping gene to normalise the expression of the target genes to be validated.

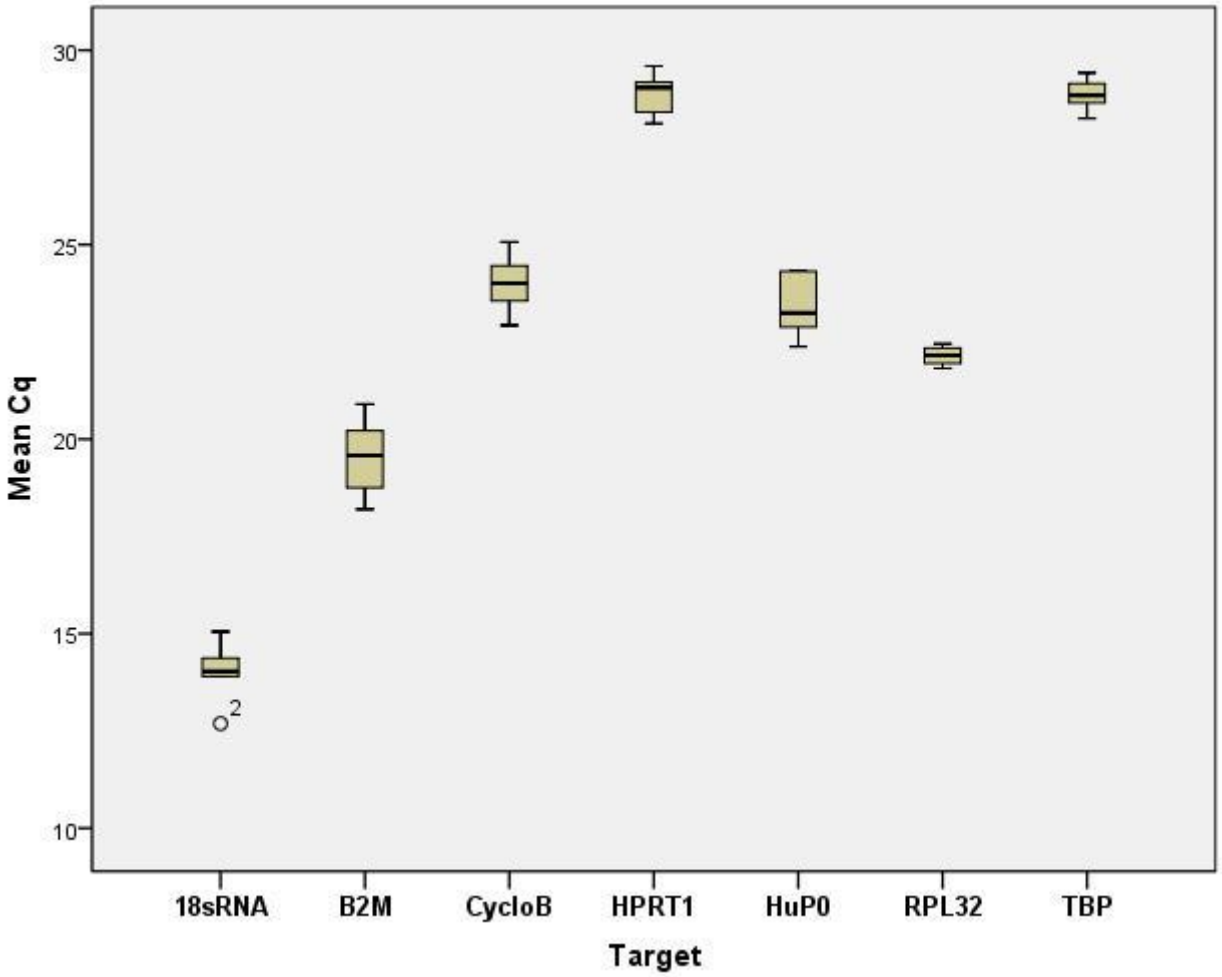


Figure 5.4 The RT-qPCR cycle threshold values in RNA fromWBC samples: Expression levels of genes are shown as medians (lines), 25th percentile to the 75th percentile (boxes), and ranges (whiskers) for 6 human WBC samples (4 breast cancer patients and 2 healthy donors). HuPO, human acidic ribosomal protein; CycloB, cyclophylinB; B2M, β2-microglobulin; HPRT, hypoxanthine phosphoribosyltransferase; 18sRNA, 18S ribosomal RNA; RPL32, ribosomal protein 32; TBP, Tata box Binding Protein.

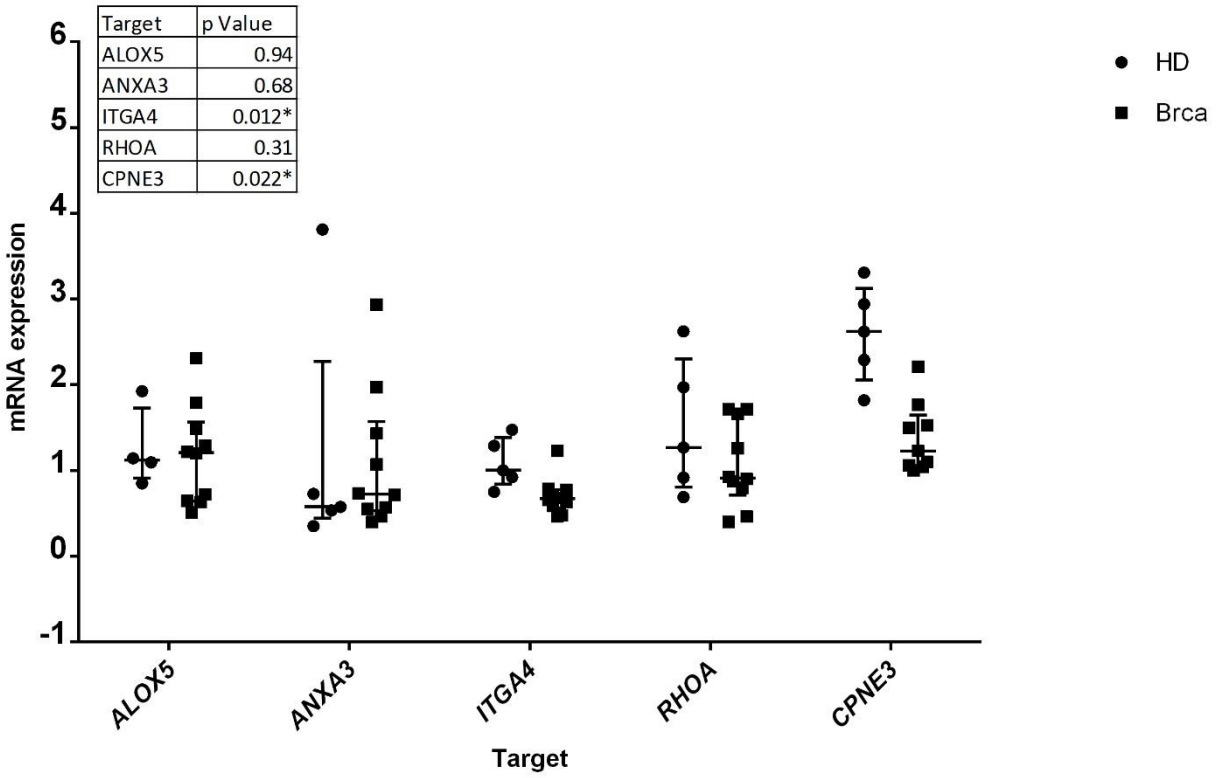
5.4.4 Validation of a panel of 15 biomarkers using RT-qPCR in the WBCs of breast cancer patients

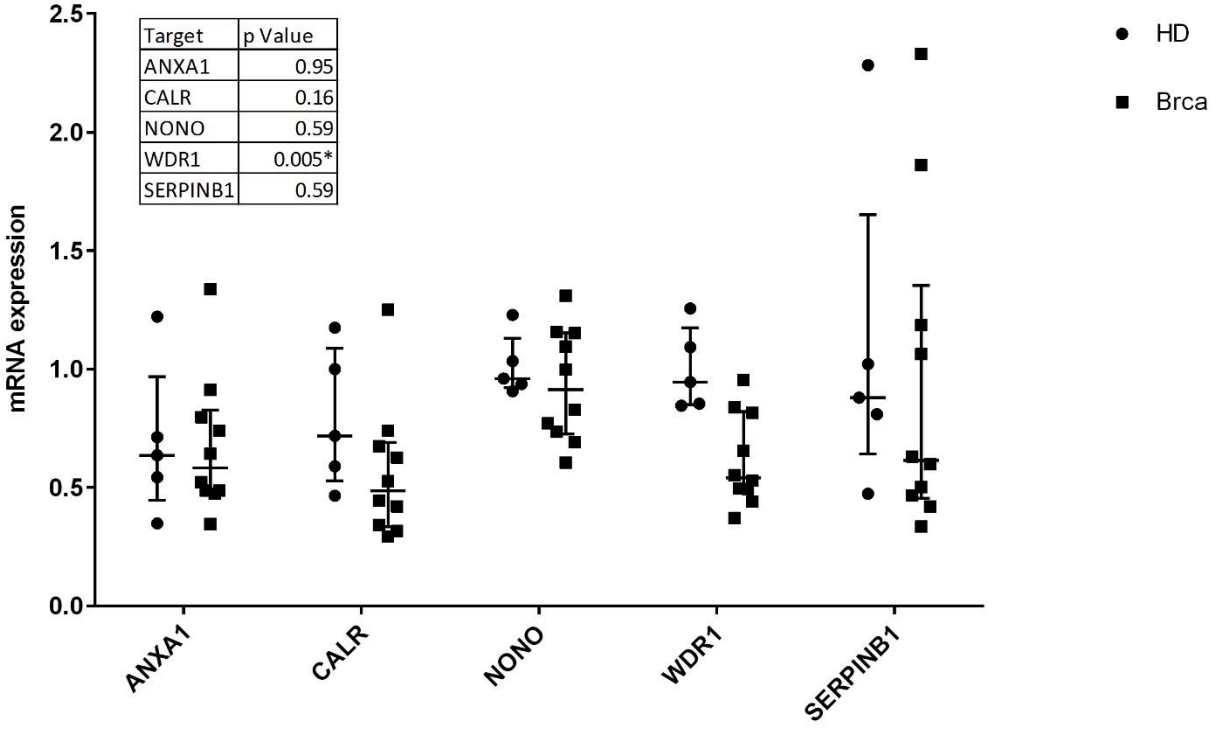
For experiments described in this section RNA was extracted by a modified protocol as described in Section 5.4.2. DNase treatment was done and the integrity of RNA was assessed using the Agilent 2100 bioanalyzer. The RNA was then converted to cDNA (refer to Section 2.3.5). This cDNA was used to validate the biomarker candidates obtained by the selection process described in Section 5.4.1. The primers for the candidate genes were designed to span different exons. The details about the primers and the cycling conditions uses are given in Table 2.3. The RT-qPCR was conducted as explained in Section 2.3.7 using *RPL32* as the reference gene. Three technical repeats were performed for the experiment. In the initial series of experiments, the WBC samples from ten breast cancer patients (Panel 1) and five healthy donors were used (the patient details are given in Table 5.3).The qPCR was performed and the data was analysed using the Biorad CFX manager software. The software calculated the Cq values and normalised the expression values with respect to *RPL32*. These expression values for each sample were taken and box plots were constructed using SPSS. The Mann-Whitney U test was performed between the healthy donors and breast cancer patients with a p value cut-off of 0.05 to find significant differences in gene expression. RT-qPCR of the 15 different genes revealed that the differences in expression of three genes, *ITGA4*, *CPNE3* and *LCN2* were significant (P value<0.05) between the WBCs of breast cancer patients and healthy donors. The respective fold changes were -1.03,-1.6 and -3.28 in breast cancer patients. The box plots of the expression values are given in Figure 5.5 along with the P values obtained from the T-test.*WDR1* had a p value of 0.053 which is very close to the cut-off and had a fold change of -1.4. All the other genes tested had higher p values. The genes which passed the threshold were tested again with an extended panel of primary breast cancer patients and metastatic breast cancer patients. This is explained in the following Section.

Table 5.3 Clinical data for breast cancer patient samples used for validation by RT-qPCR: T size- Tumour size, T Grade- Tumour grade, T stage- Tumour stage, DCIS- Ductal Carcinoma in situ, IDC- Invasive Ductal Carcinoma, LCIS- Lobular carcinoma in situ and ILC- Invasive Lobular Carcinoma. Nodes indicate if the tumour had spread to the lymph nodes.

Study No	AGE	T SIZE	T STAGE	T GRADE	NODES	DIAGNOSIS
Panel 1						
1535	67	20	II	1	NEG	IDC
1477	46	8	I	1	NEG	IDC,DCIS(LOW GRADE)
1560	82	20	II	3	NEG	IDC
1532	73	10	II	3	NEG	IDC+DCIS
1492	47	>50	III	1	POS	LCIS+ILC
1556	46	25	NA	High	NA	DCIS
1512	74	23	II	2	POS	ILC+LCIS
1561	41	25	II	2	POS	ILC+DCIS
1438	53	9	NA	Intermediate	ND	DCIS
1490	81	40	II	2	NA	IDC+ILC+DCIS+LCIS
Panel 2						
1276	64	10	II	1	POS	Tubulo-lobular+DCIS
1499	67	15	I	1	NEG	IDC+DCIS+Tubulo-lobular
1432	63	20	II	2	POS	IDC+DCIS
1500	65	23	II	2	POS	IDC+DCIS
1553	36	20	II	2	POS	IDC+DCIS

1558	85	15	II	2	POS	IDC
1516	46	14	III	3	NEG	IDC+DCIS
1528	84	20	II	3	NEG	IDC+DCIS
1494	66	25	II	3	NEG	IDC+DCIS
1554	65	30&3	II	3	NEG	IDC+DCIS


Target | p Value
ALOX5 | 0.94
ANXA3 | 0.68
ITGA4 | 0.012*
RHOA | 0.31
CPNE3 | 0.022*
mRNA expression
6
5
4
3
2
1
0
-1
ALOX5
ANXA3
ITGA4
RHOA
CPNE3
Target
HD
Brca


Target | p Value
ANXA1 | 0.95
CALR | 0.16
NONO | 0.59
WDR1 | 0.005*
SERPINB1 | 0.59
mRNA expression
2.5
2.0
1.5
1.0
0.5
0.0
ANXA1
CALR
NONO
WDR1
SERPINB1
HD
Brca

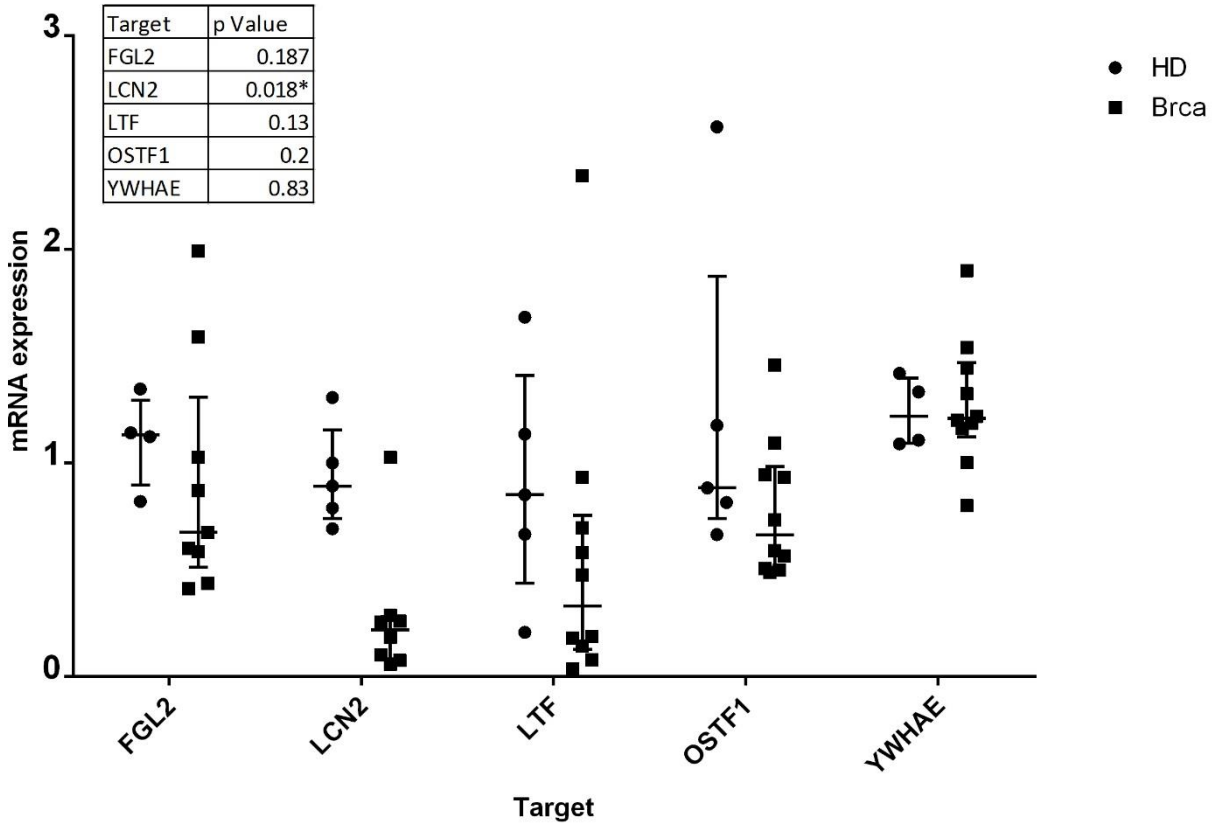


Figure 5.5 Validation of the 15-gene signature in WBCs of breast cancer patients using RT-qPCR: RT-qPCR was performed on the cDNA obtained from the WBCs of breast cancer patients (n=10) and healthy donors (n=5). *RPL32* was used as the housekeeping gene to normalise the target genes and the expression values were used to construct a box plot using SPSS. Mann-Whitney U test was performed to identify the significance of the expression values between both groups to identify gene targets which have p value<0.05. Expression levels of genes are shown with medians and their quartile ranges. HD-Healthy Donor, Brca-Primary breast cancer. *pvalue<0.05

5.4.4.1 Validation of ITGA4, CPNE3 and LCN2 using an extended panel of primary and metastatic breast cancer patients

The WBC expression profiles of *ITGA4, CPNE3* and *LCN2* showed a significant difference between the healthy donors and primary breast cancer patients when tested using RT-qPCR. These three genes were therefore chosen for validation using an extended panel which included ten more participants (Panel 2, Table 5.3). The RNA samples from 7 pre-treatment metastatic breast cancer patients were also included in this panel. The rationale behind including metastatic samples was that it may provide biomarkers to diagnose or predict distant organ tumour metastasis. After the samples were analysed by RT-qPCR, the data were compiled together by normalising different runs using the expression values of a common sample used in both RT-qPCR runs and scatter plots were generated in SPSS. The different groups were then compared for significance by performing a Kruskal-Wallis test and a confidence interval of 95% (p value<0.05) was considered.

Figures 5.6, Figure 5.7 and Figure 5.8 show the box plots of the normalised expression data of *ITGA4*, *CPNE3* and *LCN2* respectively (see Figure legends for details). It was observed that *ITGA4* expression was significantly less in the WBCs of primary breast cancer (average fold change= -1.54) and metastatic breast cancer patients (average fold change= -2.21) when compared to the healthy donor cohort. *CPNE3* expression was significantly higher (fold change= +2.37) in the metastatic breast cancer cohort when compared to both healthy donor and primary breast cancer cohorts. *LCN2* expression was less in the primary breast cancer patients (fold change= -1.35) and significantly higher (fold change= +3.13) in the metastatic breast cancer patients.

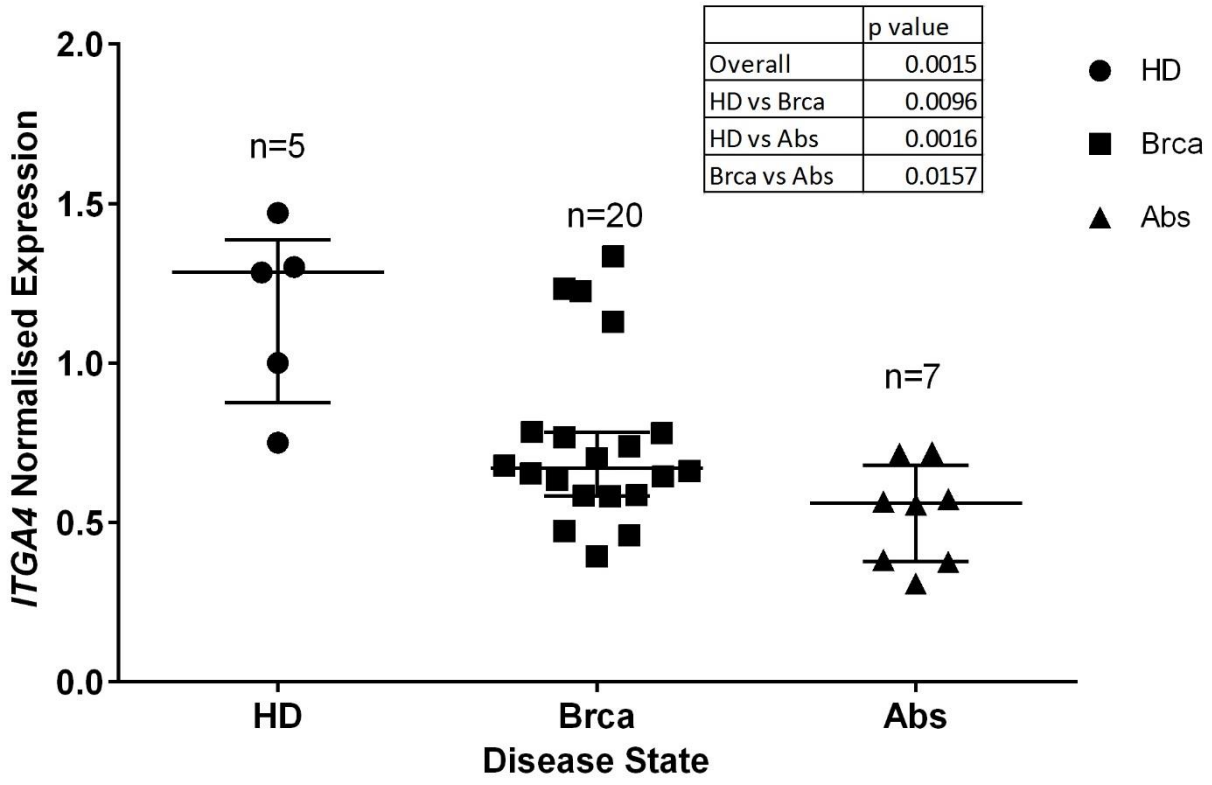


Figure 5.6 RNA expression profiles of Integrin Alpha 4 (ITGA4) in the WBCs of Healthy donors and Breast cancer patients: RT-qPCR was performed using RNA obtained from the WBCs of healthy donors, primary breast cancer patients and metastatic breast cancer patients to detect the expression of *ITGA4*. *RPL32* was used as the reference gene in all experiments. The data was normalised using the expression value of sample HD3 which was used in all experiments. Kruskal-Wallis test was performed to test for significance between the different sample cohorts used. Expression levels of genes are shown with medians and their quartile ranges. HD-Healthy Donor, Brca-Primary breast cancer, ABS- Advanced breast cancer..

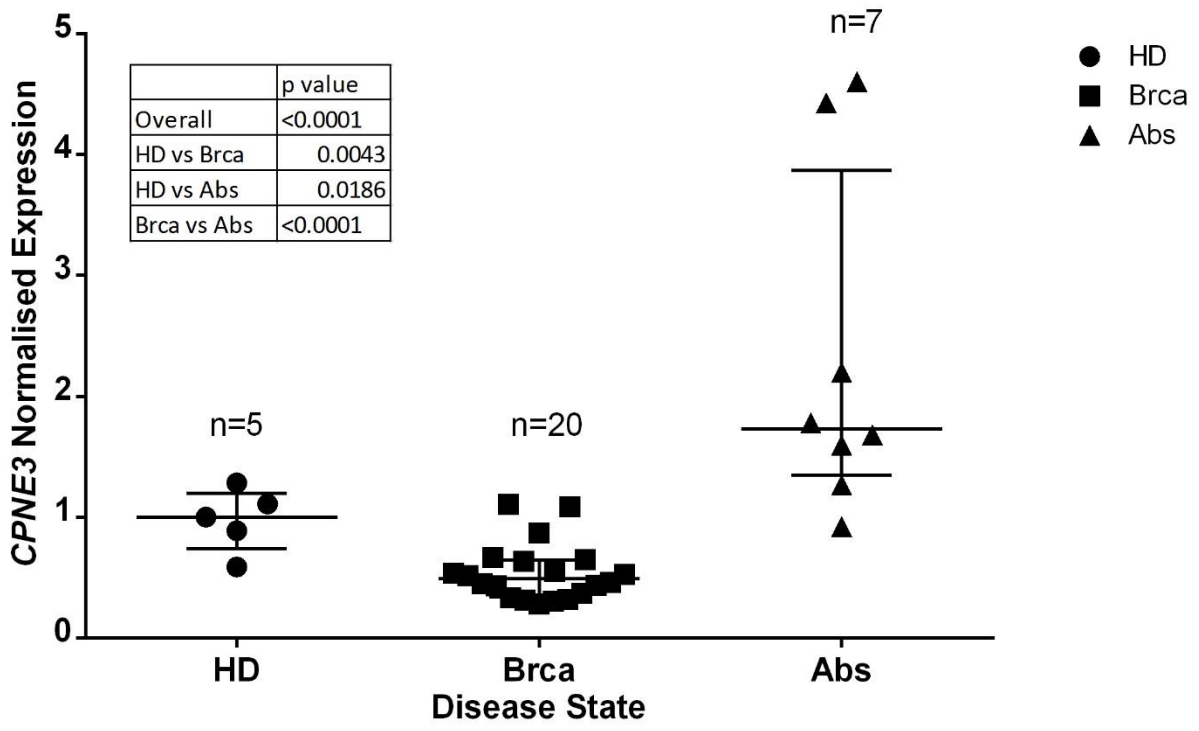


Figure 5.7 RNA expression profiles of Copine3 (CPNE3) in the WBCs of Healthy donors and Breast cancer patients: RT-qPCR was performed using RNA obtained from the WBCs of healthy donors, primary breast cancer patients and metastatic breast cancer patients to detect the expression of *CPNE3*.*RPL32* was used as the reference gene in all experiments. The data was normalised using the expression value of sample HD3 which was used in all experiments. Kruskal-Wallis test was performed to test for significance between the different sample cohorts used. Expression levels of genes are shown with medians and their quartile ranges. HD-Healthy Donor, Brca-Primary breast cancer, ABS- Advanced breast cancer.

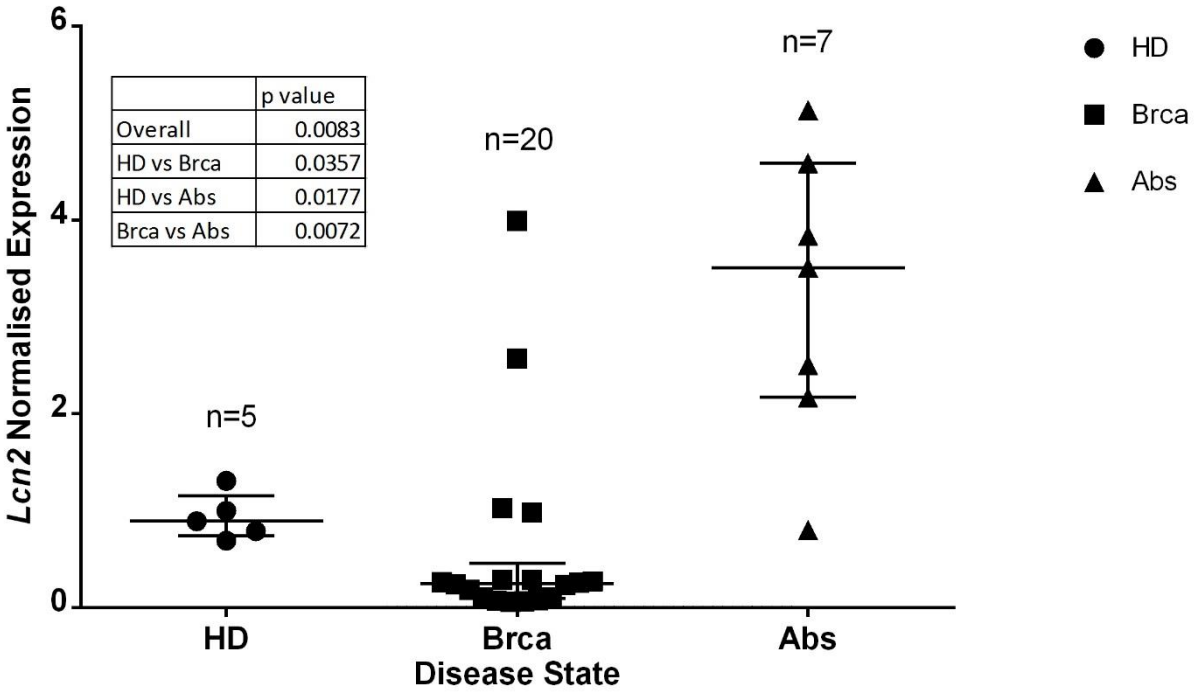


Figure 5.8 RNA expression profiles of Lipocalin2 (LCN2) in the WBCs of Healthy donors and Breast cancer patients: RT-qPCR was performed using RNA obtained from the WBCs of healthy donors, primary breast cancer patients and metastatic breast cancer patients to detect the expression of *LCN2*. *RPL32* was used as the reference gene in all experiments. The data was normalised using the expression value of sample HD3 which was used in all experiments. Kruskal-Wallis test was performed to test for significance between the different sample cohorts used. Expression levels of genes are shown with medians and their quartile ranges. HD-Healthy Donor, Brca-Primary breast cancer, ABS- Advanced breast cancer.

5.4.4.2 Validation of *SERPINB1*using RT-qPCR in an extended panelof primary and metastatic breast cancer patients

Leukocyte elastase inhibitor (SERPINB1) was found to be upregulated in the WBCs of breast cancer patients in this study by both highthroughput proteomics (Fold change= +1.46) and 2D gel electrophoresis (Fold change= +1.46). SERPINB1 was also found to be positively correlated with mortality and disease progression in the experiments using the WBCs of metastatic breast cancer patients. These combined results are presented in Figure 5.9.

Although the RNA levels of SERPINB1 were measured using RT-qPCR in the WBCs of primary breast cancer patients and results were not significant when compared to the levels in the healthy donors (Section 5.4.4), it was possible that SERPINB1 may be linked to metastatic breast cancer and not primary breast cancer. We therefore assessed *SERPINB1* mRNA levels using the extended panel with metastatic patients. The results presented in Figure 5.10 demonstrate significantly higher levels of *SERPINB1* mRNA (Fold Change= +3.4) in the WBCs of metastatic breast cancer patients when compared to both healthy donors and primary breast cancer patients. The difference was not significant between primary breast cancer cohort and the healthy donors.

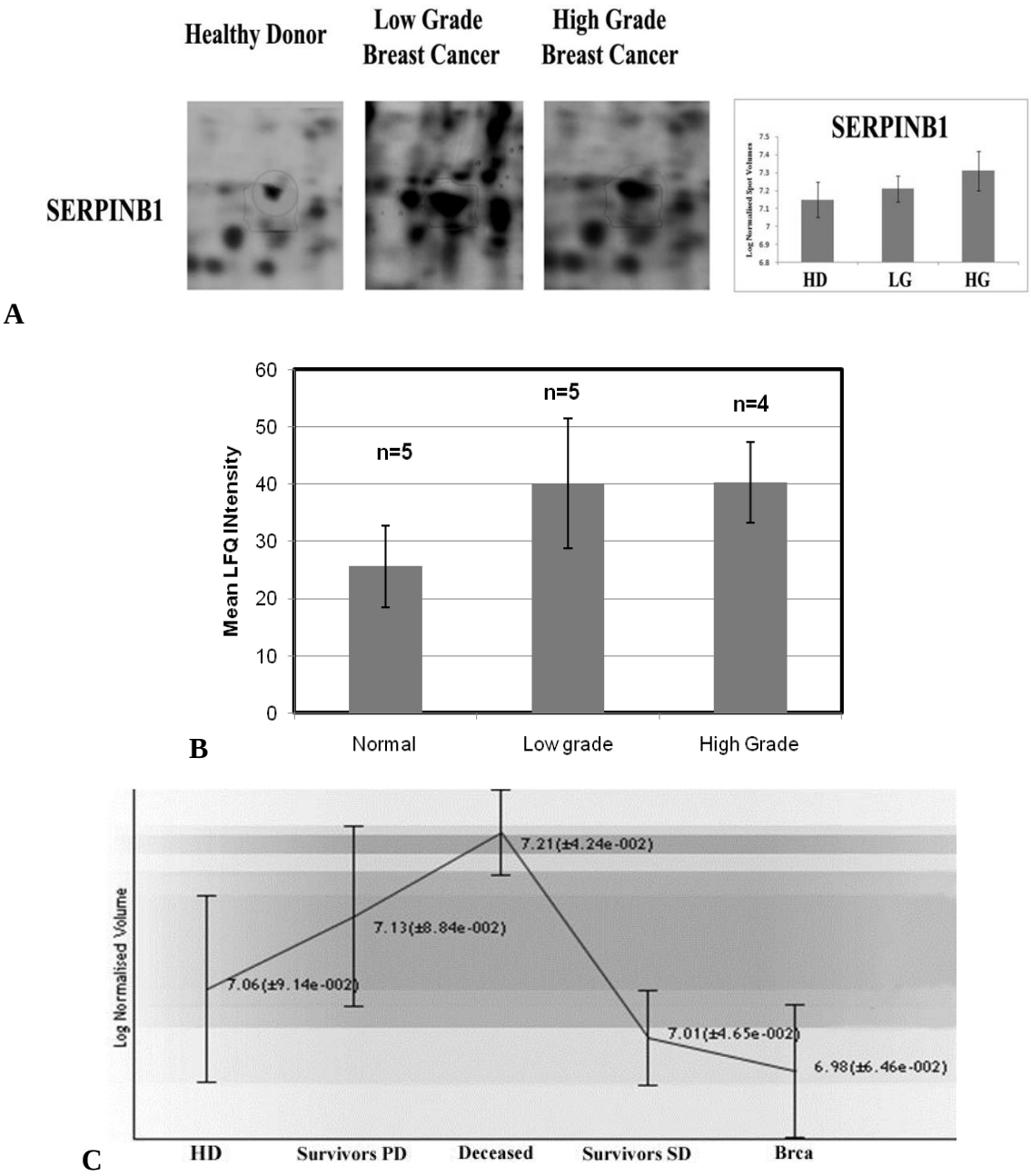


Figure 5.9 Levels of SERPINB1 in WBCs of Breast cancer patients: protein spots identified as SERPINB1 in primary breast cancer patients by (A) 2D-gel electrophoresis,(B) Highthroughput proteomics and in (C) metastatic breast cancer patients by 2D-gel electrophoresis. PD- Progressive Disease, SD- Stable Disease, HD-Healthy donor, Brca- Breast cancer. The error bars indicate the standard deviation.

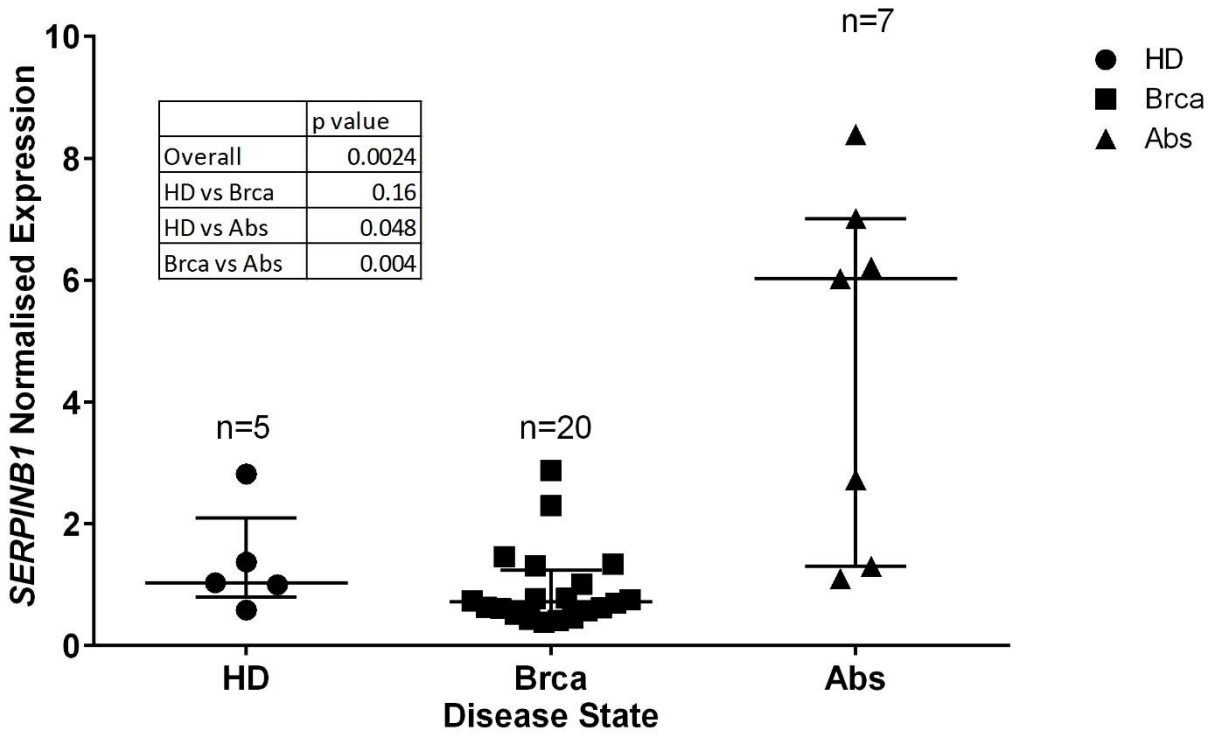


Figure 5.10 RNA expression profiles of SerpinB1 in the WBCs of Healthy donors and Breast cancer patients: RT-qPCR was performed using RNA obtained from the WBCs of healthy donors, primary breast cancer patients and metastatic breast cancer patients to detect the expression of *SERPINB1*. *RPL32* was used as the reference gene in all experiments. The data was normalised using the expression value of sample HD3 which was used in all experiments. Kruskal-Wallis test was performed to test for significance between the different sample cohorts used. Expression levels of genes are shown with medians and their quartile ranges. HD-Healthy Donor, Brca-Primary breast cancer, ABS- Advanced breast cancer.

5.4.5 Validation of biomarkers at protein level using western blotting

The fifteen shortlisted candidates obtained by combining the data from different approaches were validated using RT-qPCR. Four genes *ITGA4*, *CPNE3*, *LCN2* and *SERPINB1*, passed the significance test and showed considerable difference between the healthy donor and primary breast cancer and/or metastatic breast cancer cohorts. Out of these candidates *CPNE3*, *LCN2* and *SERPINB1* were chosen for further validation at protein level using western blotting. *ITGA4* was not chosen because the difference in Cq values and the normalised expression ratios between healthy donors and breast cancer patients were low. Antibodies for Copine3 (cat no: ab97919), SerpinB1 (cat no: ab47731), and Lipocalin2 (cat no: ab63929) were purchased from Abcam. All antibodies were rabbit polyclonal antibodies and, according to Abcam specifications, were suitable for western blotting. The secondary antibody used was anti-rabbit horseradish peroxidase (HRP) conjugated from Abcam (cat no: ab99697). The antibody for loading control (β-Actin) was the anti β-Actin antibody from Sigma (cat no: A1978). Western blotting was carried out as described in Chapter 2 (Section 2.2.1) and the antibody signal was visualized. The samples were normalised with each other by normalising the target signal with the actin signal.

5.4.5.1 Evaluation of Copine 3 protein expression in breast cancer patients

Breast cancer patients' samples (Primary and Metastatic) were compared with healthy donors using western blotting. Samples from nine healthy donors, 22 primary breast cancer and twelve metastatic breast cancer patients were used for this analysis. The membrane was probed with the anti-Copine 3 (CPNE3) antibody; the results of western blotting are shown in Figure 5.11. The positive control was the lysate from 293T cell and the negative control was the lysate from NB4 cell. A band around 60kDa was observed in the WBC samples and the positive control. The intensity values were obtained using Image J as described in Section 2.2.1.4; the CPNE3 values were normalised to β-Actin for each sample. The values between different western blotting analyses were normalised with each other using the expression values of

common samples in different westerns assays. The combined results from different westerns are represented as a box plot in Figure 5.12 and Kruskal-Wallis was performed using SPSS to confirm significance in difference between different sample cohorts. It was found that CPNE3 expression in WBCs was highly variable and there were no significant differences between the healthy donor and primary breast cancer and metastatic cancer cohorts. The expression data did not show any correlation between any clinical data and we therefore concluded that Copine 3 was not likely to be a suitable biomarker for diagnosis and prognosis for breast cancer.

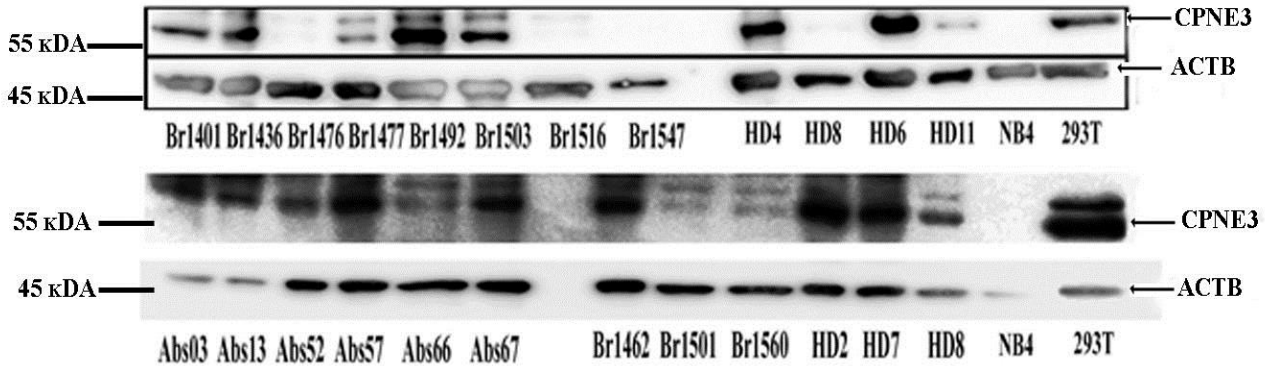


Figure 5.11 Protein expression of CPNE3 in WBCs of breast cancer patients using western blotting: western blot analysis showing expression of CPNE3 in the WBCs of healthy donors, primary breast cancer patients and metastatic breast cancer patients. Lysates from breast cancer cell line 293T was used as a positive control and actin as loading control. The proteins were resolved by SDS - PAGE, blotted and probed with CPNE3 IgG (rabbit) antibody and visualized.(Br - Breast cancer, HD - Healthy donor, Abs- advanced breast cancer study, ACTB- beta-actin).

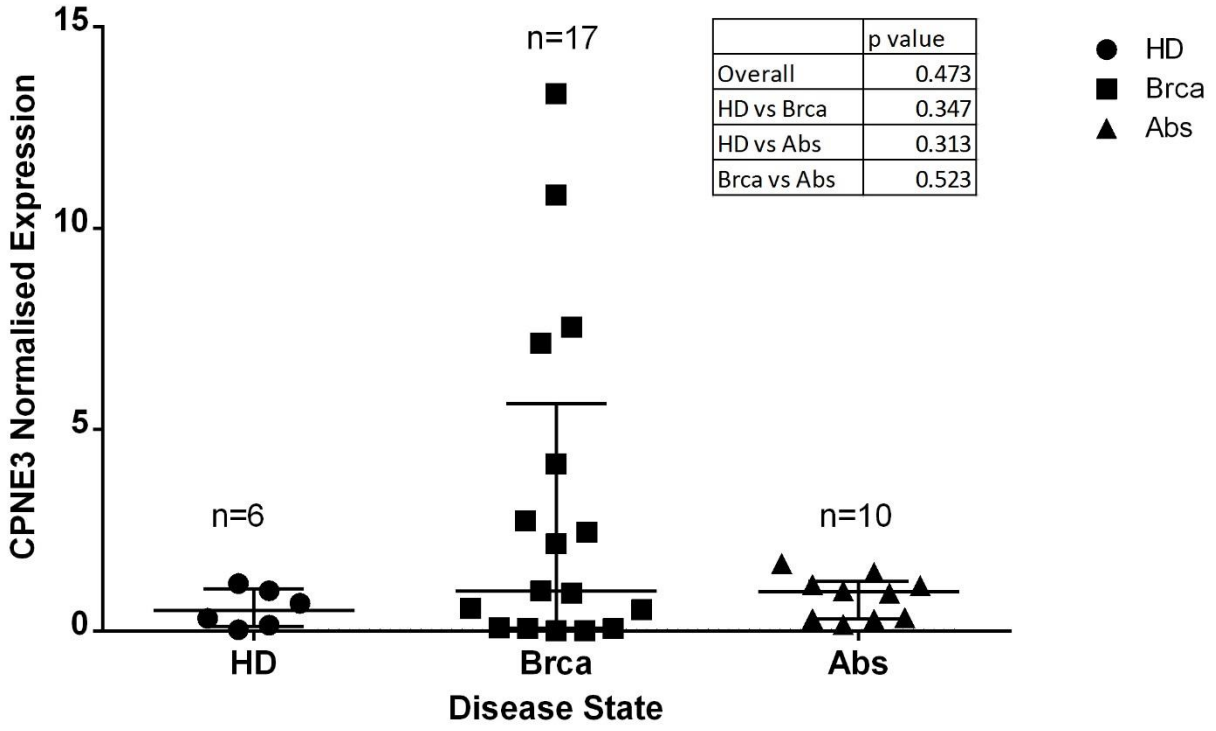


Figure 5.12 Scatter plot of Protein expression of Copine3 (CPNE3) in WBCs of breast cancer patients: western blotting was performed using the lysates obtained from the WBCs of healthy donors, primary breast cancer patients and metastatic breast cancer patients to detect the expression of CPNE3. B-Actin was used as the loading control. Kruskal-Wallis test was performed to test for significance between the different sample cohorts used. Expression levels of genes are shown with medians and their quartile ranges. HD-Healthy Donor, Brca-Primary breast cancer, ABS- Advanced breast cancer.

5.4.5.2 Analysis of SERPINB1protein levels in breast cancer patients

SERPINB1 protein expression was evaluated by western blotting in the breast cancer cohorts and compared with healthy donors. A band of approximately 43kDa was observed. The SERPINB1 signal was normalised using the β-Actin using Image J. Expression values between westerns were compared by normalising the expression values from the common sample used in each experiment (usually a sample from a healthy donor). Figure 5.13 represents a typical western showing expression of SERPINB1. It can be clearly observed that expression in metastatic patients is much higher than in healthy donors and primary breast cancer patients. Figure 5.14 shows the combined data in scatter plots. The fold change in primary breast cancer was +3.9 and metastatic breast cancer was +7 when compared to healthy donor cohort. Around 47% of primary breast cancer patients had higher levels of SERPINB1 when compared to median of healthy donors. Also 90% of metastatic breast cancer patients had higher SERPINB1 levels when compared to primary breast cancer patients. The p values were low (<0.05) between the different cohorts. Since SERPINB1 levels were higher in metastatic breast cancer patients, SERPINB1 expression was correlated in primary breast cancer patients with the node status. Figure 5.15 shows the expression of SERPINB1 between the node positive and node negative breast cancer cohorts. The p value was not significant but the median of node positive samples was higher than node negative samples.

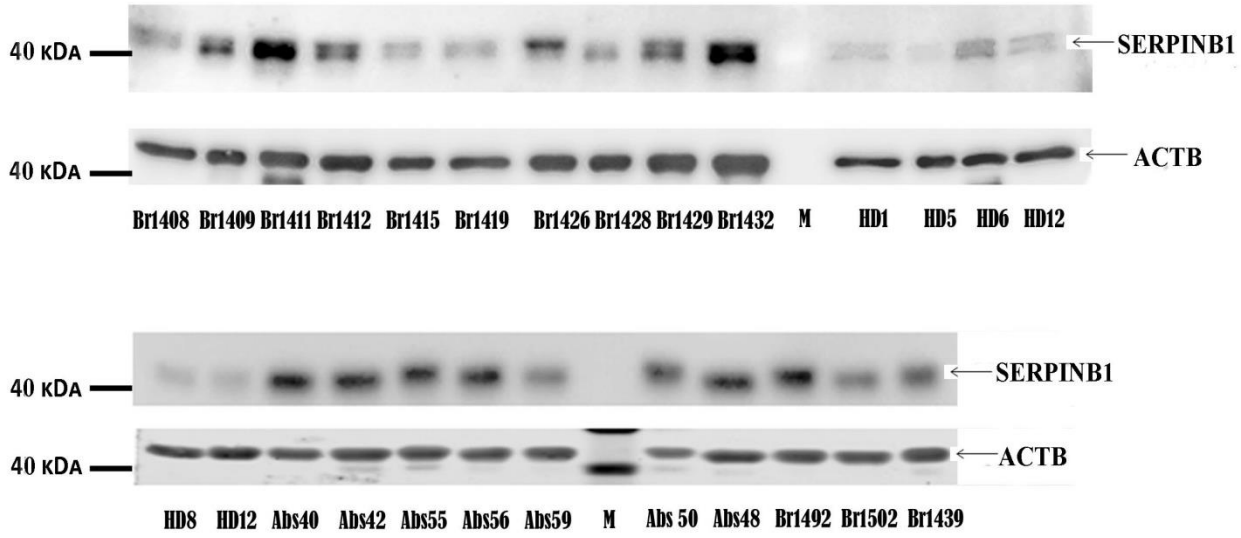


Figure 5.13 Protein expression of SerpinB1 in WBCs of breast cancer patients using western blotting: western Blot analysis showing expression of SERPINB1 in the WBCs of healthy donors, primary breast cancer patients and metastatic breast cancer patients. β-Actin was used as loading control. The proteins were resolved by SDS - PAGE, blotted and probed with SERPINB1 IgG (rabbit) antibody and visualized. (Br - Breast cancer, HD - Healthy donor, Abs - advanced breast cancer study, ACTB- β Actin, M-protein marker).

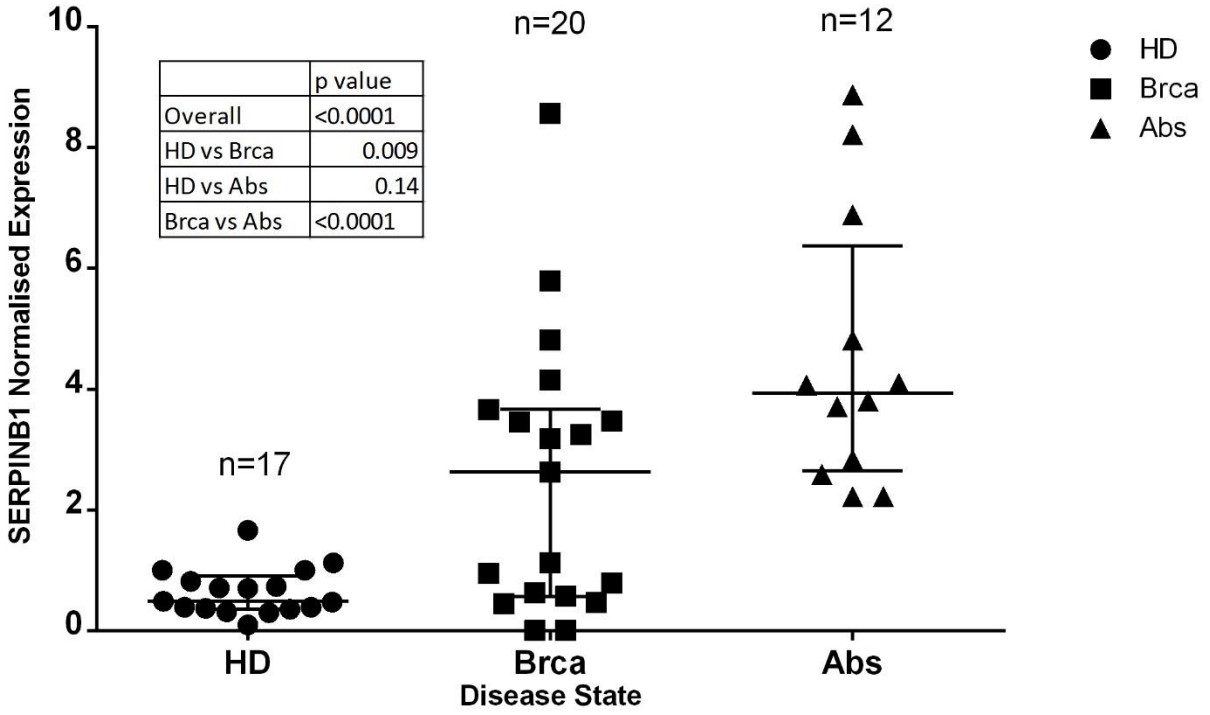


Figure 5.14 Scatter plot of Protein expression of SerpinB1 in WBCs of breast cancer patients: western blotting was performed using the lysates obtained from the WBCs of healthy donors, primary breast cancer patients and metastatic breast cancer patients to detect the expression of SERPINB1. B-Actin was used as the loading control. Kruskal-Wallis test was performed to test for significance between the different sample cohorts used. Expression levels of genes are shown with medians and their quartile ranges. HD-Healthy Donor, Brca-Primary breast cancer, ABS- Advanced breast cancer.

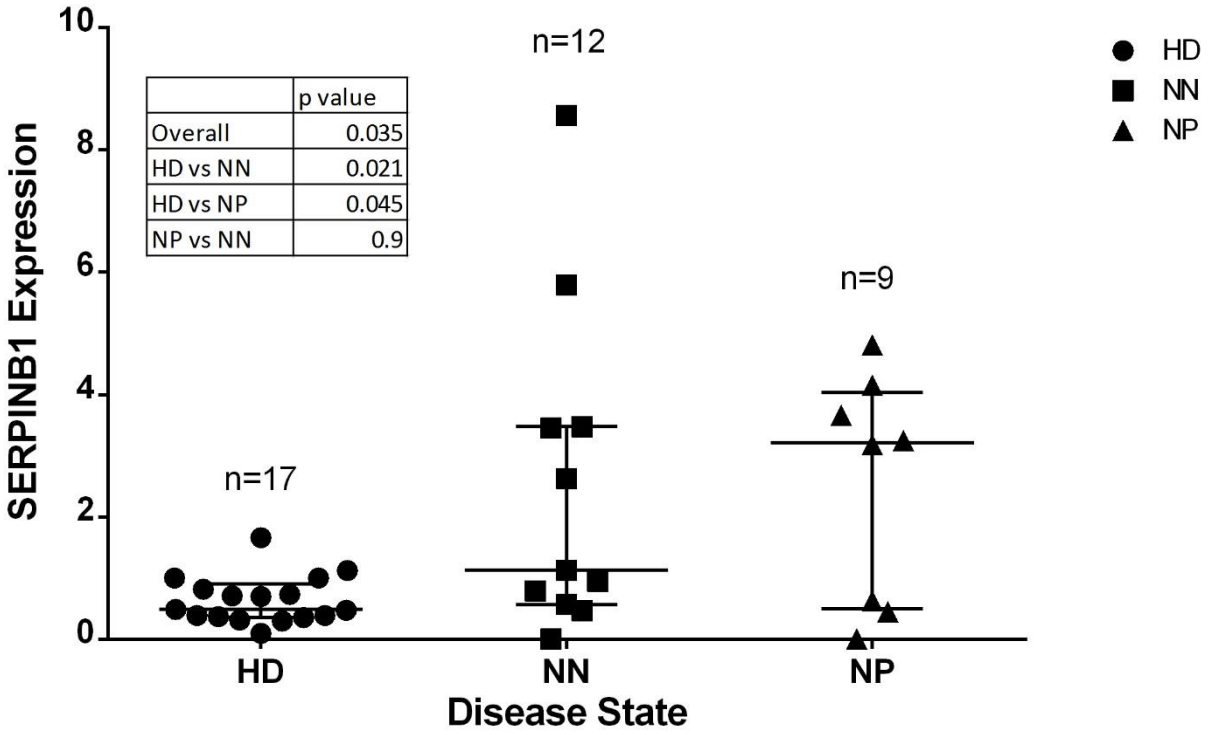


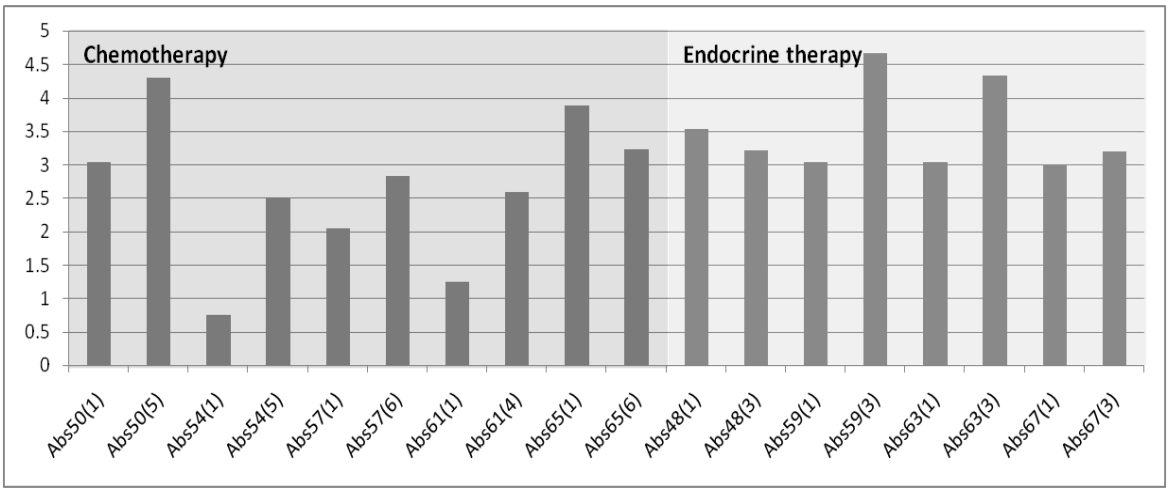
Figure 5.15 Scatter plot of Protein expression of SerpinB1 in WBCs of breast cancer patients according to node status: western blotting was performed using the lysates obtained from the WBCs of healthy donors and primary breast cancer patients to detect the expression of SERPINB1. B-Actin was used as the loading control. Kruskal-Wallis test was performed to test for significance between the different sample cohorts used. Expression levels of genes are shown with medians and their quartile ranges. HD-Healthy Donor, NN-Node Negative primary breast cancer, NP-Node Positive primary breast cancer.

5.4.5.3 Evaluation of SerpinB1 protein expression in metastatic breast cancer patients as a marker of prognosis

Since SERPINB1 was present in the metastatic breast cancer dataset its levels were tested in WBCs of metastatic breast cancer patients before and after treatment to assess whether it may be a marker of treatment efficacy and prognosis. Figure 5.16A shows the levels of SERPINB1 in such paired metastatic breast cancer WBC samples; five paired samples from patients who underwent chemotherapy and four samples from patients who underwent endocrine therapy patients were used. Out the nine patient samples tested, in two samples SERPINB1 levels decreased after treatment (Abs065 and Abs048) and in all other samples increased. The patient data for the samples used in these experiments are presented in Figure 5.16,including the information regarding the type of metastasis and survival in months. The overall tendency was for increased levels of SERPINB1 to correlate with diminished response to therapy and poorer survival. For example, in the case of participant Abs048, complete response for the treatment was observed, the participant was alive after 18 months and the SERPINB1 levels decreased after treatment. In participants Abs054 and Abs061 the amount of SERPINB1 was about two-fold higher after treatment; these patients survived for only two and four months, respectively. However, the pattern of changes in SERPINB1 levels in participant Abs059 cannot be explained by this hypothesis. Unfortunately, due to time restraints, it was not possible to perform these experiments using a large panel of samples, which in addition to the first and the last samples would include intermediate time points. These experiments are currently being carried out in our laboratory.

5.4.5.4 Evaluation of Lipocalin 2 protein expression in breast cancer patients

The Lipocalin 2 (LCN2) protein was another promising WBC biomarker and it was selected for further validation by western blotting (Figure 5.17 shows these results). The predicted size for LCN2 is 23 kDa; there was no signal from the anti-LCN2 antibody in this range. The LCN2 protein was previously reported to be present in MCF7 cells (Yang et al., 2009). Therefore the MCF7 cell lysate was included in this experiment as a positive control; however no band was detected in this sample. The presence of sufficient amount of protein in all specimens was confirmed be re-probing the membrane with the anti β-Actin antibody. We concluded that the anti-LCN2 antibody tested was not suitable for western blotting experiments; these assays will need to be repeated with another antibody.



A

Sample ID	SERPINB1 Fold Change after Treatment	Treatment Type	Metastasis type	Prognosis	Survival in months (as on 31-05-2012)
Abs048	-1.10078	Endocrine	Lung	CR	Alive
Abs050	1.421607	Chemotherapy	Liver, Lung, Bone	SD	10
Abs054	3.328652	Chemotherapy	Liver, Bone	SD	4
Abs057	1.384492	Chemotherapy	Liver, Lung, Bone, Brain	PD	8
Abs059	1.565282	Endocrine	Lung	SD	Alive
Abs061	2.062586	Chemotherapy	Lymph nodes	PR	2
Abs063	1.427723	Endocrine	Bone	PR	7
Abs065	-1.2011	Chemotherapy	Liver, Lung, Bone, Brain	PR	6
Abs067	1.068564	Endocrine	Liver, Bone, Skin	SD	10

B

Figure 5.16 Expression of SERPINB1 in metastatic breast cancer patient samples from before treatment and after treatment: (A) The WBC samples from metastatic breast cancer patients who underwent two types of treatment (Chemotherapy and endocrine therapy) were used. SERPINB1 levels were assessed by western blotting. Samples were loaded analysed in pairs (before and after treatment). The images obtained by western blotting was analysed by Image J and SERPINB1 levels were normalised against β-Actin values. (B) Patient information and levels of SERPINB1 for samples used in the experiments shown in panel A. The fold change of SERPINB1 after treatment is also given along with disease survival. SD= Stable Disease, PD = Progressive Disease, CR = Complete Response, PR = Partial Response.

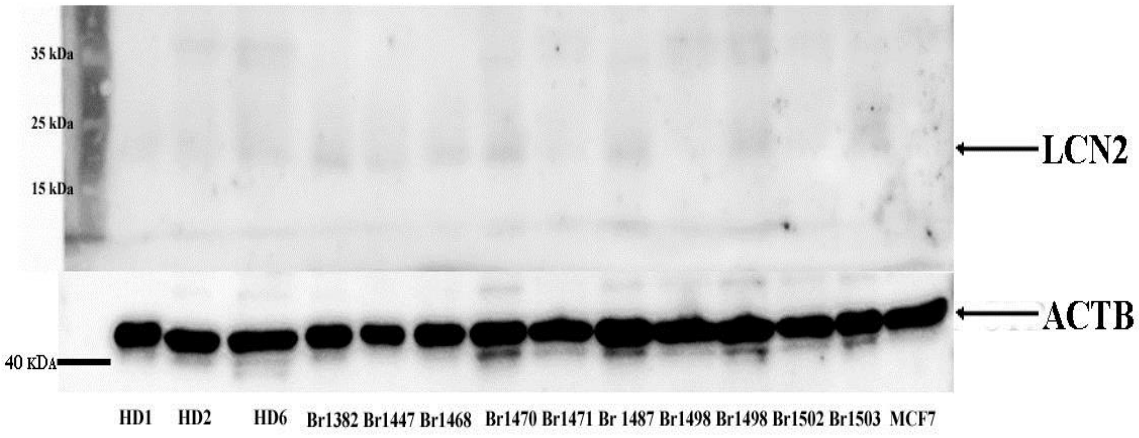


Figure 5.17 Protein expression of Lipocalin 2 in WBCs of breast cancer patients using western blotting: Expression of LCN2 in the WBCs of healthy donors, primary breast cancer patients and metastatic breast cancer patients has been analysed; β-Actin was used as loading control. The proteins were resolved by SDS - PAGE, blotted and probed with the anti-LCN2 IgG (rabbit) antibody and visualized. Br - Breast cancer, HD - Healthy donor and ACTB- β Actin.

5.5 Discussion

The aim of the work described in this chapter was to identify potential biomarkers using three different experimental datasets and then validate the selected candidates using RT-qPCR and western blotting.

In order to carry out RT-qPCR, RNA had to be extracted from WBCs and converted into cDNA. The WBCs were stored in RNAlater solution at -80^{0}C so that RNA could be extracted at convenience. Extracting RNA from WBCs proved to be difficult as the yield and the quality of RNA (purity and integrity) were initially very poor. Indeed this has been noted by several groups using RNA form peripheral blood (Debey et al., 2004; Feezor et al., 2004; Wang et al., 2004). According to Minimum Information of Quantitative real-time PCR experiment (MIQE) guidelines the assessment of nucleic acid purity and integrity is essential (Bustin et al., 2009). According to the modified protocol for RNA extraction described in this chapter the second chloroform step and an additional ethanol wash was required to obtain pure RNA without protein or salt contamination. The yield of RNA from WBCs which were stored in RNAlater was very low (10 million cells were used to obtain only around ~200ng of RNA. Alternatively the RNA yield was observed to be higher when WBCs were freshly lysed after separation from whole blood using Trisure (300ng/10 million cells). This lysate prepared by using Trisure could only be stored at -80^{0}C but only for a maximum of four weeks according to the manufacturer's instruction. Due to large numbers of sample collection and delayed availability of clinical diagnosis data RNAlater was used and RNA extraction protocol was optimised to it.

DNA contaminations in RNA preparations can cause significant problems in gene expression analysis. The conventional phenol-chloroform method for RNA extraction does not completely eliminate DNA. There are no RNA extraction procedures available that completely avoid DNA contamination and also there are no techniques available to detect DNA in RNA

preparations other than PCR. In our study the RNA preparations had DNA contamination which was detected in cDNA preparations. RT-qPCR analysis and cDNA preparations are expensive, therefore DNase treatment is essential and ensured that RNA was DNA-free.

The selection of a suitable reference gene in RT-qPCR reactions is important as it determines the significance of the results. In this study it was observed that reference genes like β-actin, human acidic ribosomal protein and β2-microglobulinproduced false results when the target genes were normalised with the reference genes, especially if the change in target gene levels were low. *RPL32*showed minimum transcription range across the tested samples and was chosen as a reference gene. A recent study (Xu et al. 2013) used several reference genes to validate candidate genes from the WBCs of colorectal cancer patients. Although this approach it is very labor- and time-consuming, it ultimately improves the outcomes of the RT-qPCR reactions.

The shortlisting of candidates from the three different datasets proved to be very challenging as the datasets were very large and involved significant amount of computing. The main criteria for the shortlisting process were (1) presence of the same candidate in more than one dataset, (2) biological function in WBC (in particular, in immune response) and (3) association with cancer. An initial panel of 15 biomarkers was chosen using the above criteria to be validated by RT-qPCR in the first instance. Table 5.4 shows the list of candidates chosen to be validated by RT-qPCR and the results in the form of fold change. Only *YWHAE, ANXA3* and *ALOX5* were overexpressed in the breast cancer cohort. All the other genes tested were overexpressed in the primary breast cancer cohort. It was observed that out of the 12 genes which were overexpressed in the high throughput data (both genomic and proteomic) only *CPNE3, ANXA3* and *ALOX5* followed the same trend in RT-qPCR analysis. *ITGA4* and *WDR1* showed consistent negative change in breast cancer patients in RT-qPCR analysis and the microarray data. This result is consistent with other groups where only 25-30% of genes

correlate at both RNA and protein levels (Lundberg et al., 2010; Tian et al., 2004; Vogel et al., 2010).*ITGA4, CPNE3* and *LCN2* showed significant difference (T-test p value<0.05) in the primary breast cancer cohort. Although *SERPINB1* did not show significance at RNA levels in primary breast cancer patients it was chosen for further validation based on the information obtained from proteomics and also the RNA expression in metastatic breast cancer patients.

All four genes (*ITGA4, CPNE3, LCN2* and *SERPINB1*) showed significant difference at the RNA expression levels in the metastatic breast patient cohort (Section 5.4.4.1 and 5.4.4.2). Only *ITGA4* had negative expression in both breast cancer and metastatic breast cancer cohort. Others showed significantly higher expression levels in metastatic breast cancer patients.

Of note, it has been reported in the literature that protein levels do not always correlate with the levels of corresponding mRNA (Lundberg et al., 2010; Tian et al., 2004; Vogel et al., 2010). Therefore, it is likely that the number of potential candidate biomarkers discovered in the course of this study is much higher. The integration of genomics and proteomics provided stricter pre-requisitions and helped us to shortlist a manageable panel of candidates for validation.

Table 5.4 Candidate WBC biomarkers chosen for validation by RT-qPCR and the summary of RT-qPCR results: Genes common from different datasets comparing healthy donors, primary breast cancer patients and metastatic breast cancer patients were obtained. The overlapping genes were further shortlisted to give a 15 panel gene list to be validated using RT-qPCR. Fold change refers to fold change in primary breast cancer cohort. Red coloured cells indicate overexpression; green indicates underexpression; pink overexpression only in low grade breast cancer. *T-test p value<0.05. The presence in metastatic breast cancer cohort is denoted by 'X' and the fold change is not included since the initial analysis did not include comparison with healthy donors.

Candidates	Primary Breast cancer-2D-gel electrophoresis	Primary Breast cancer-High throughput proteomics	Primary Breast cancer -Microarray (GSE16443)	Metastatic Breast cancer-2D-gel electrophoresis	Fold Change in RT-qPCR
SERPINB1	+1.3	+1.46		X	-1.16
CALR			+1.4	X	-1.4
YWHAE		+1.46		X	+1.03
CPNE3			+1.2	X	-1.47*
ANXA3			+1.97	X	+1.98
ANXA1	+1.3	+1.49			-1.02
LTF			+2.37	X	-1.60
NONO			+1.21	X	-1.08
ITGA4		-3.4	-1.28		-1.4*
WDR1			-1.23	X	-1.46
RHOA		+1.4	+1.5		-1.39
LCN2			+2.16	X	-1.35*
ALOX5		+1.6			+1.28
FGL2		+			-1.23
OSTF1		+1.23			-1.56

The results of RT-qPCR identified four candidates: ITGA4, LCN2 and CPNE3. The LCN2 and CPNE3 proteins were then validated at a protein level using western blotting. Additionally, SERPINB1 was added to this panel to be validated by western blotting. Although CPNE3 expression at RNA levels in both primary breast cancer and metastatic breast cancer cohorts were significantly different from the healthy donor cohort, western blotting revealed that CPNE3 expression was not significantly different in both breast cancer cohorts, with CPNE3 expression hugely varying in the primary breast cancer cohort. If the panel was extended CPNE3 could be correlated with the clinical data to observe if it could be used as prognosis marker.

SERPINB1 levels were significantly higher (Fc=+7) in the metastatic breast cancer cohort and also in the primary breast cancer cohort (Fc=+3.5). From these data it can be concluded that SERPINB1 represents a very attractive candidate to be used as a biomarker of metastatic breast cancer. There was also positive correlation of SERPINB1 protein levels to lymph node metastasis state in the primary breast cancer cohort, whereby participants who had lymph node metastasis had higher levels of SERPINB1.Therefore, higher SERPINB1 may be an indicator/predictor of metastasis and hence used as a prognostic biomarker. Further retrospective studies will be required to assess how levels of SERPINB1 correlate with the recurrence and survival. This was not possible with the samples collected for this study as the information was not available.

The levels of SERPINB1 were also measured to observe if it could be used as a prognostic marker for treatment response and survival. The initial 2D gel electrophoresis study revealed that higher SERPINB1 levels could indicate poor prognosis. The analysis carried out in this study was not conclusive as the sample numbers were low. Regarding ITGA4 and LCN2 the RT-qPCR results look very promising and antibodies could be used to validate them at protein levels in the future.

There is still lack of standardised protocol for protein biomarker development pipeline while the validation protocols are well developed for the pharmaceutical industry for drug development (Buyse et al., 2010).Western blotting is an effective method to validate biomarker candidates initially but it is limited to the sample number which could be used at the same time. Although many protein analytical techniques are available for validation ELISA still dominates in a clinical setting. ELISA is a very effective technique to validate potential biomarkers in a clinical setting quantitatively and also a large number of samples could be used. Further validation of SERPINB1 by ELISA would be helpful to gain further data to continue on to the clinical phase in the biomarker development pipeline.

Chapter 6: General Discussion

6.1 The challenging field of breast cancer biomarkers

Although the survival for female breast cancer has greatly improved in the recent years, breast cancer remains a very serious condition. Breast cancer is challenging to screen or diagnose, it is fast growing once it reaches critical mass and possesses high ability to metastasise to multiple sites in the body. It has been suggested that breast cancer metastasis occurs 5 years even before the primary tumour is clinically detected (Engel et al., 2003). The current major clinical diagnostic technique for breast cancer is the use of mammograms which detects abnormal lumps of tissue in the breast. Mammography has a high rate of false negatives (10-15%) where the tumour is present and is not detected (Heywang-Köbrunner et al., 2011); this may endanger patients' lives. On the other hand, over-diagnosis may also take place in these screenings because the number of false positives in mammography is also high. In addition, mammograms do not provide information about the abnormal lump i.e. whether it is benign or malignant. As a result, for final diagnosis tissue biopsies may be required which puts patients under unnecessary anxiety and stress.

This justifies the need for biomarkers for screening, diagnosis and prognosis of breast cancer. Currently there are six FDA approved biomarkers: three are tissue based (estrogen receptor, progesterone receptor and HER2/neu) and three plasma based (circulating tumour cells, CA 15-3 and CA 27-29), to monitor breast cancer disease progression and response to therapy (Füzéry et al., 2013). These biomarkers are not accurate in predicting response to treatment and also cannot be used for diagnosis or screening. With the rapid development of the "omics" technologies several groups have tried to identify breast cancer biomarkers in tissue and plasma (refer to Section 1.7, 1.8 and 1.9); however still no biomarkers have been discovered to screen or diagnose breast cancer. In this study we have hypothesized that breast tumour would interact with the immune system causing a systemic change in immune expression profile. This could be

exploited to find novel biomarkers for breast cancer and in this study we have explored biomarkers in a novel source, the white blood cells 9WBCs). We also applied these findings to elucidate the effects of the breast tumour on the immune system.

6.2 Protein profiling expression analysis of WBCs from breast cancer patients

In this study we used several approaches for biomarker discovery. The High throughput proteomics using mass spectrometry is more efficient than genomics technologies because almost all of the clinical diagnostics techniques are antibody based assays targeting proteins. Moreover RNA levels do not always correspond to the translated protein levels in the cell. Drabovich et.al 2014 has proposed an integrated protein biomarker discovery pipeline. This study has almost followed the steps in the proposed pipeline even though the group published the work just recently (Figure 1.3). However it has to be noted that none of the existing breast cancer biomarkers have been discovered using mass spectrometry based approaches. This may be due to the lengthy period of validation, clinical trials and the approval process following the protein biomarker discovery. It is therefore possible that the candidate biomarkers first discovered in the 2000s may obtain the FDA approval very soon if they satisfy all the criteria. Regardless of the status with the biomarkers currently in the FDA pipeline, the search for breast cancer biomarkers continues.

The proteomic technologies used in this study were the high throughput profiling by OrbitrapVelos mass spectrometer and conventional 2D-PAGE. The high throughput proteomics approach resulted in the detection of around 2577 proteins in the membrane fraction of WBCs. This comparison illustrates the sensitivity of the mass spectrometry over the 2D-PAGE. The comparison between the protein profiles of healthy donors and breast cancer patients using high throughput proteomics gave 136 significant proteins while 2D-PAGE gave only 18 significant proteins. Performing 2D-PAGE was time consuming, costly and prone to experimental errors such as inter-gel variation, differences in staining intensities, low sensitivity to low abundant proteins,

poor inter-gel normalisation techniques and reduced resolution compared to mass spectrometry. The results from the both analyses showed that the genes which changed had less fold change range when compared to the data obtained from tumour tissues (from literature). The heterogeneity (variability) in the significant protein levels was also high due to the multiple cell type composition of WBCs and also the dynamic nature of the immune system. The heterogeneity of breast cancer subtypes might also be the cause for this. Three proteins appeared in both analyses; SERPINB1, JUP and ANXA1. SERPINB1 became an important candidate later in our biomarker pipeline. This will be discussed in the following sections. The data from the soluble fraction of the high throughput proteomics is still to be analysed. This will lead to discovery of more biomarkers.

6.3 Gene expression analysis of WBCs in breast cancer patients

There is a vast amount of information in the microarray data produced and deposited so far. This includes the gene expression data for peripheral blood samples in breast cancer patients, although such data have been limited to only a few studies. In our investigation, we used these existing data to compare the expression profiles of blood samples from breast cancer patients and healthy donors. These data deposited in the GEO (id-GSE16443) (Aarøe et al., 2010) were from the gene expression analysis of RNA samples extracted from peripheral blood cells of 67 primary breast cancer patients and 54 healthy donors. The microarray method is fast and high throughput enabling to compare many genes simultaneously but it has several limitations. Experimental method drawbacks include frequent problems with quality/integrity of RNA, artefacts in the image after hybridisation, batch effects of hybridisation, difficulty in data reduction and lack of standard procedures for data collection/analysis. The high throughput RNA-sequencing (RNA-Seq) has outdated the microarray technology and also overcame several limitations of microarray technology such as high background noise, poor detection of unknown transcripts and relative

quantification. Any future experiments related to WBC biomarker discovery are now likely to be performed using the RNA-Seq to avoid the limitations of microarrays.

The facts that for the gene expression studies the RNA samples were extracted from the whole blood (all blood cells plus serum) and also the heterogeneous nature of blood cells were important factors to consider in the data analysis performed in Chapter 4. This was reflected, for example, in the variation in expression levels of individual genes obtained after comparison between the individuals in breast cancer patient cohort and the individuals in the healthy donor cohort. Another notable phenomenon was that the majority of the genes obtained after comparison between the two cohorts was with fold change were less than ±2. The systemic changes in the blood would have been minimal because of the influence of tumour on immune cells, which is reason the fold change cut-off for genes was set at ±1.2. The comparison resulted in the identification of 506 differentially expressed genes in the breast cancer patient cohort. These data were combined with the proteomics data (Chapter 3) to shortlist biomarkers for further validation by RT-qPCR and Western blotting. Furthermore microRNAs are another promising source of biomarkers in blood. MicroRNAs like (miR-10b, miR-21, miR-125b, miR-145, miR-155 miR-191 and miR-382) (George and Mittal, 2010; Mar-Aguilar et al., 2013) have shown promise and are currently in the development pipeline.

6.4 Integrating data to understand the effect of breast tumours on the immune system

The relationship between tumours and the immune system is very complex and not fully understood. Tumours interact with the surrounding tissues and the immune system, and the changes in the cells of the immune system reflect such communications. The tumour, for example, manipulates the immune system to escape antigen recognition and also utilises the immune response for survival leading to metastasis (describe in Section 1.10. in more detail). In this study,

together with discovery of novel cancer biomarkers from WBCs, we used the data to investigate differences related to the immune system functions.

Functional analysis of both datasets (genomics and proteomics) showed overall consistent functional changes in the WBCs of breast cancer patients. The GO analysis of the gene expression data revealed that genes involved in WBC activation, antigen recognition, processing and presentation were downregulated in breast cancer patients. These observations were complemented by the outcomes of the pathway analysis performed using KEGG database, showing that the JAK-STAT pathway downregulated. This analysis also revealed that some genes involved in the transendothelial migration, inflammatory response and cell proliferation were upregulated, but others down-regulated in the breast cancer cohort. The protein profiling data demonstrated that inflammatory response was upregulated in the breast cancer patients. The GO analysis also revealed that proteins involved in cell migration and cell survival were upregulated. The pathway analysis showed that leukocyte transendothelial migration pathway was upregulated.

The data obtained in these analyses suggest that systemic inflammatory response is increased in breast cancer patients. Breast cancer cells secrete various pro-inflammatory factors such as TNF-α, IL-8, IL-10, VEGF etc along with several chemokines which attract the immune cells to sites of inflammation. Interestingly, systemic inflammation markers such as neutrophil-lymphocyte ratios, levels of platelets and platelet-lymphocyte ratios have been used as prognostic indicators of cancer outcome (Fox et al., 2013). In addition, elevated neutrophil counts, elevated platelet counts and low lymphocyte counts were predictors of low overall survival in renal carcinoma patients. This observation could play an important role in biomarker discovery process as several proteins are expressed at different levels in different cell types of the immune system.

The GO analysis specifically indicated that 30% of proteins which changed were involved in cellular localisation. In particular, proteins involved in cell migration were positively regulated

in the WBCs of breast cancer patients. This complements the fact that highly invasive tumours have increased immune cell (Dendritic cell, T-cell) infiltrations in breast tumour which correlates with poor prognosis (DeNardo et al., 2011; Leek et al., 1996; Ruffell et al., 2012). Some of the genes involved in WBC migration are also involved in tumour cell metastasis where a process similar to WBC migration occurs. The cell migration factors secreted by WBCs could be used by tumour cells to enhance cell adherence and invasion. It has been shown in several studies that leukocyte migration is inhibited in various carcinomas (Brandes and Goldenberg, 1976; Kadish et al., 1976; Lee et al., 1977).

The observations regarding changes in the functions of the WBC could be validated using primary tumour tissue or cells and the WBC samples. Extracts from tumour tissues or cell culture media supernatants from breast cancer cell lines can be applied to WBC to induce changes in their functions related to cancer. These functions could be measured by appropriate assays such as migration, apoptosis, survival, reactive oxygen species measurement and other assays. Components of breast tumour tissue supernatants could be profiled using mass spectrometry to identify factors inducing WBC functions. This work is already progressing in our laboratory (data not shown) with the aims to identify factors inducing tumour specific molecules in WBCs. This study shows the importance of the effect of breast cancer on systemic WBC migration.

6.5 Shortlisting and validation of breast cancer biomarker candidates

The shortlisting of candidates for validation had been very time consuming as the volumes of data generated by both genomic and proteomic approaches were very high. The results from the third independent study performed using the WBCs of metastatic breast cancer patients were also used in the study (see Section 5.2) to complement the shortlisting process. It was surprising to find that only 15 candidates overlapped between proteomics and genomics considering the

large number of significant genes obtained from gene expression analysis, while 21 candidates were common in both primary breast cancer and metastatic breast cancer proteomics datasets.

The first step of validation was to assess the RNA levels of 15 genes obtained from the shortlisting described above, in both primary breast cancer cohort and healthy donor cohort. RT-qPCR revealed that out of the nine genes chosen for validation from the genomics dataset only three genes *ANXA3, ITGA4* and *WDR1* had the same trends in RNA expression as in the microarray data. Also only three genes *CPNE3, LCN2* and *ITGA4* passed the significance test in the RT-qPCR analyses. These genes were tested in an extended panel of primary breast cancer patients and metastatic breast cancer patients. *ITGA4* levels were downregulated in both primary and metastatic breast cancer patients. Both *CPNE3* and *LCN2* showed similar RNA expression patterns, namely, downregulated in the WBCs of primary breast cancer patients and upregulated in metastatic breast cancer patients. *SERPINB1*did not pass the significance test in the primary breast cancer panel at mRNA levels, but it was significantly (by more than 3-fold) higher in the WBCs of metastatic patients. Two candidates, SERPINB1 and CPNE3, were validated at the protein levels to identify their potential use as breast cancer biomarkers. The levels of CPNE3 were not significantly changed in the WBCs of primary breast cancer patients or metastatic breast cancer patients; the variance in the levels of CPNE3 was very high suggesting very low specificity. This disqualifies CPNE3 as a potential biomarker to be taken to the clinical stage. On the other hand, promising results were observed for SERPINB1.

6.6 SERPINB1 as a potential biomarker for breast cancer progression and metastasis

6.6.1 Biological functions of SERPINB1

SERPINB1 is the neutrophil elastase inhibitor, it was first characterised in neutrophils but it is present in most of the subsets of WBCs. SERPINB1 inhibits serine proteases elastase, cathepsin G and proteinase-3 (Farley et al., 2012). Neutrophil elastase is an inflammatory protein

which is involved in anti-microbial action and is released at sites of inflammation. Neutrophil elastase dissolves cellular junctions and aids in transmigration of WBCs across epithelial cells (Ginzberg et al., 2001). SERPINB1 also inhibits Granzyme H (Wang et al., 2013) which is used by natural killer cells to induce cell death in the antigen presenting cells and tumour cells (Trapani and Smyth, 2002). SERPINB1 is involved in neutrophil extravasation and formation neutrophil extracellular traps to contain inflammation (Farley et al., 2012).The association of SERPINB1 with cancer is complex, thus BERPINB1 can be a *tumour suppressor and promoter*, at the same time, depending on the specific biology of the disease, microenvironment and other factors (Cui et al., 2014; Tseng et al., 2009)**.**

6.6.2 SERPINB1 as a prospective biomarker

SERPINB1 showed promising results when the protein levels in WBCs were measured by Western blotting. Thus, SERPINB1 levels were 3.5 times higher in 50% of the primary breast patient WBCs and 7 times higher in more than 90% of the metastatic breast cancer patient WBCs. Furthermore, SERPINB1 levels were higher in the WBCs in participants with node positive primary breast cancer than in node negative participants. If follow up clinical data could be collected for primary breast cancer patients to observe for recurrence and metastasis, SERPINB1 could be correlated to those parameters to assess further its prognostic significance. Collectively the data suggests the possibility that SERPINB1 could be a marker of lymph node metastasis for primary breast cancer patients and a general metastatic breast cancer biomarker of prognosis. Metastatic breast cancer patients who initially had higher levels of SERPINB1 had poorer prognosis. More samples from metastatic breast cancer patients should be evaluated to identify changes in levels of SERPINB1 to validate it as a prognostic biomarker for disease progression and treatment response. These findings demonstrate that SERPINB1 could be taken to the next phase of biomarker development. It would be also important to design an ELISA based test for SERPINB1 because Western blotting will not be practical for clinical settings. The function of

SERPINB1 in the WBCs of breast cancer patients could point towards other biomarkers by searching for the factor that induces elevated SERPINB1 expression in WBCs of breast cancer patients. The expression of SERPINB1 could also be evaluated in the WBCs obtained from patients who have other types of cancer.

6.7 Future work

Developing a clinical biomarker for diagnosis and screening is a time consuming process which is typically much longer than the period of a postgraduate research study. In this study we developed a pipeline for biomarker discovery and identified one prospective biomarker, SERPINB1, for prognosis of the disease and response to treatment. This investigation is continuing and Lyudmyla Pavlova, a current PhD student, who has taken over this project, confirmed the main findings described in this report using a larger panel of WBCs whereby the levels of Serpin-B1 were higher in the WBCs of patients with breast cancer than in healthy donors, the WBCs from lymph node positive participants had more SERPINB1 than from node negative and the highest levels of Serpin-B1 were in the WBCs from the metastatic breast cancer cohort (Figure 6.1). Another possible application for SERPINB1 is the biomarker of response to treatment. Recent experiments performed by Lyudmyla Pavlova in WBCs from the metastatic breast cancer cohort during chemotherapy and endocrine therapy treatment revealed the potential of SERPINB1 to predict the treatment efficacy and outcome (data not shown). SERPINB1 will also need to be evaluated in the WBCs of other cancer patients (e.g. prostate cancer and colorectal cancer patients- these materials are available in our laboratory and will be tested in due course). Development of the methods based on ELISA will be important; they will facilitate screening of larger numbers of samples and precisely establish the levels of SERPINB1 in WBCs, which in turn can be more accurately correlated with clinical parameters. Proteins ITGA4 and LCN2 are still to be validated at the protein levels as they showed promising results in RT-qPCR experiments.

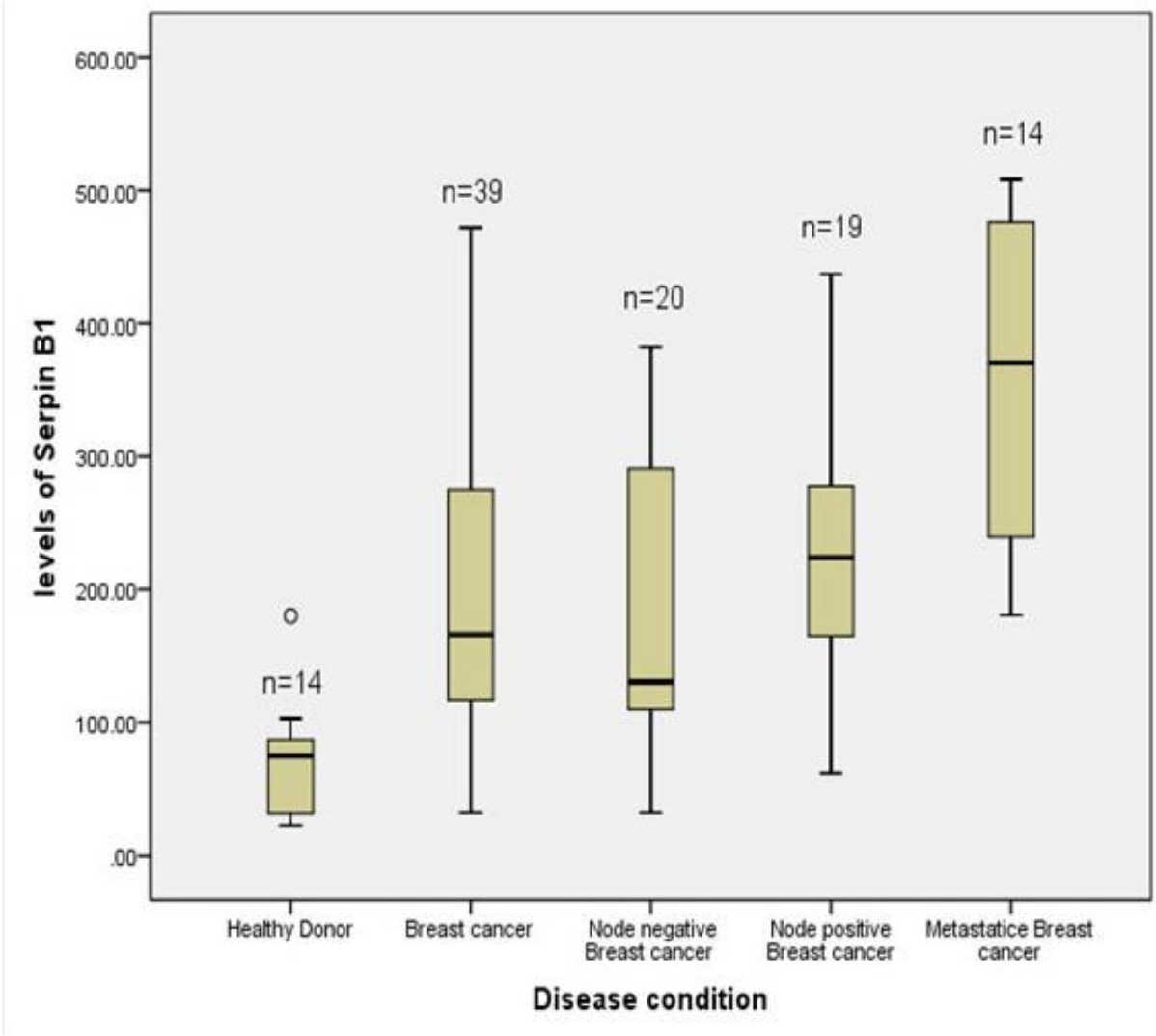


Figure 6.1 Expression of SERPINB1 in the WBCs of breast cancer patients: SERPINB1 protein levels were assessed in the WBCs by Western blotting and normalised to β-Actin levels. Data were produced by Lyudmyla Pavlova.

It has to be noted that only 15 candidates were chosen in the validation phase after a very stringent shortlisting process. More candidates for validation can still be obtained from our datasets. For example S100 family of proteins have been associated with the progression of breast cancer (Emberley et al., 2004). S100A8 and S100A12 appear in the microarray data analysed. S100A8 also appears in the metastatic breast cancer patient cohort. SERPINB6 was observed in the primary breast cancer and metastatic breast cancer proteomic datasets. The validation pipeline used in this study can be applied for validation using initially a small number of samples and then expanded with larger panel.

The value of this investigation will not only be relevant to the biomarker field but also for basic research, i.e. to elucidate the mechanisms of biological processes involved in the interaction between the tumour and immune system. The data generated for this study has given insight of how the immune system is altered in the breast cancer patients. The effect of breast tumour supernatant on immune cells have been previously studied (Eichbaum et al., 2011; Kusmartsev and Gabrilovich, 2006; Luboshits et al., 1999). Peripheral WBCs or cell lines can be treated with conditioned media after exposure with fresh tissues (normal and breast tumour) or selected regulatory molecules (chemokines, cytokines, growth factors). The effects of these factors on WBCs biology in general, or expression of particular genes, including the candidate biomarkers identified in this study can be evaluated. Alternatively, immune cell functions such as antigen presentation (evaluation of innate and adaptive immune receptors), migration (cell migration assays), reactive oxygen species production etc can be tested to identify the systemic effect of tumour on the immune system. Purification of individual cell types from the whole WBCs population will help understand the function of these cells with the tumours.

6.8 General conclusions

The proposition that the tumour affects the WBCs in breast cancer patients resulting in the changes in the RNA and protein profiles provided the rationale for this study to discover novel breast cancer biomarkers. This aim was achieved and a panel of potential biomarkers has been generated. One biomarker in particular (SERPINB1) has shown promising results, but more samples will have to be analysed before it could be taken to the clinical phase of validation. Other shortlisted biomarkers found in this study (ITGA4 and LCN2) show promise but they need to be tested further. More additional candidate biomarkers can be selected from the initial list of biomarkers using this pipeline. The functional analysis of the data has also provided the insight into the immune response to cancer which now needs to be validated by laboratory based functional assays.

References

Aarøe, J., Lindahl, T., Dumeaux, V., Saebø, S., Tobin, D., Hagen, N., Skaane, P., Lönneborg, A., Sharma, P., and Børresen-Dale, A.-L. (2010). Gene expression profiling of peripheral blood cells for early detection of breast cancer. Breast Cancer Res. *12*, R7.

Abba, M.C., Lacunza, E., Butti, M., and Aldaz, C.M. (2010). Breast cancer biomarker discovery in the functional genomic age: a systematic review of 42 gene expression signatures. Biomark. Insights *5*, 103–118.

Adlerova, L., Bartoskova, A., and Faldyna, M. (2008). Lactoferrin: a review. Vet. Med. (Praha). *53*, 457–468.

Aktary, Z., and Pasdar, M. (2012). Plakoglobin: role in tumorigenesis and metastasis. Int. J. Cell Biol. *2012*, 189521.

Alkam, Y., Mitomi, H., Nakai, K., Himuro, T., Saito, T., Takahashi, M., Arakawa, A., Yao, T., and Saito, M. (2013). Protein expression and methylation of DNA repair genes hMLH1, hMSH2, MGMT and BRCA1 and their correlation with clinicopathological parameters and prognosis in basal-like breast cancer. Histopathology *63*, 713–725.

Alldridge, L., Metodieva, G., and Greenwood, C. (2008). Proteome profiling of breast tumors by gel electrophoresis and nanoscale electrospray ionization mass spectrometry. J. Proteome Res. *7*, 1458–1469.

Amelio, A.L., Miraglia, L.J., Conkright, J.J., Mercer, B.A., Batalov, S., Cavett, V., Orth, A.P., Busby, J., Hogenesch, J.B., and Conkright, M.D. (2007). A coactivator trap identifies NONO (p54nrb) as a component of the cAMP-signaling pathway. Proc. Natl. Acad. Sci. *104*, 20314–20319.

Ashburner, M., Ball, C., Blake, J., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., S., S.D., T., J.E., et al. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29.

Augenlicht, L., Wahrman, M., Halsey, H., Anderson, L., Taylor, J., and Lipkin, M. (1987). Expression of cloned sequences in biopsies of human colonic tissue and in colonic carcinoma cells induced to differentiate in vitro. Cancer Res. *47*, 6017–6021.

Baine, M.J., Chakraborty, S., Smith, L.M., Mallya, K., Sasson, A.R., Brand, R.E., and Batra, S.K. (2011). Transcriptional profiling of peripheral blood mononuclear cells in pancreatic cancer patients identifies novel genes with potential diagnostic utility. PLoS One *6*, e17014.

Balkwill, F. (2004). Cancer and the chemokine network. Nat. Rev. Cancer *4*, 540–550.

Barczyk, M., Carracedo, S., and Gullberg, D. (2010). Integrins. Cell Tissue Res. *339*, 269–280.

Baylin, S.B., and Herman, J.G. (2000). DNA hypermethylation in tumorigenesis: epigenetics joins genetics. Trends Genet. *16*, 168–174.

Becker, C., Hammerle-Fickinger, A., Riedmaier, I., and Pfaffl, M. (2010). mRNA and microRNA qulaity control for RT-qPCR analysis. Methods *50*, 237–243.

Ben-Baruch, A. (2003). Host Microenvironment in breast cancer development: Inflammatory cells, cytokines and chemokines in breast cancer progression-reciprocal tumor–microenvironment interactions. Breast Cancer Res. *5*, 31–36.

Bernstein, J.L., Godbold, J.H., Raptis, G., Watson, M. a, Levinson, B., Aaronson, S. a, and Fleming, T.P. (2005). Identification of mammaglobin as a novel serum marker for breast cancer. Clin. Cancer Res. *11*, 6528–6535.

Bezault, J., Bhimani, R., Wiprovnick, J., and Furmanski, P. (1994). Human lactoferrin inhibits growth of solid tumors and development of experimental metastases in mice. Cancer Res. *54*, 2310–2312.

Boja, E., Hiltke, T., Rivers, R., Kinsinger, C., Rahbar, A., Mesri, M., and Rodriguez, H. (2011). Evolution of clinical proteomics and its role in medicine. J. Proteome Res. *10*, 66–84.

Brandes, L.J., and Goldenberg, G.J. (1976). Peripheral Leukocyte Migration Inhibition Reactivity to Breast Cancer Antigens in Patients with Breast Cancer and in Normal Controls Peripheral Leukocyte Migration Inhibition Reactivity to Breast Cancer Antigens in Patients with Breast Cancer and in Norma. 3707–3710.

Brennan, K., Garcia-Closas, M., Orr, N., Fletcher, O., Jones, M., Ashworth, A., Swerdlow, A., Thorne, H., Riboli, E., Vineis, P., et al. (2012). Intragenic ATM methylation in peripheral blood DNA as a biomarker of breast cancer risk. Cancer Res. *72*, 2304–2313.

Bustin, S., and Nolan, T. (2004). Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. J. Biomol. Tech. *15*, 155–166.

Bustin, S. a, Benes, V., Garson, J. a, Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. *55*, 611–622.

Van Buul, J.D., and Hordijk, P.L. (2004). Signaling in leukocyte transendothelial migration. Arterioscler. Thromb. Vasc. Biol. *24*, 824–833.

Buyse, M., Sargent, D.J., Grothey, A., Matheson, A., and de Gramont, A. (2010). Biomarkers and surrogate end points-the challenge of statistical validation. Nat. Rev. Clin. Oncol. *7*, 309–317.

Cardoso, F., Harbeck, N., Fallowfield, L., Kyriakides, S., and Senkus, E. (2012). Locally recurrent or metastatic breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann. Oncol. *23*, 11–19.

Carlsson, A., Wingrena, C., Kristenssona, M., Roseb, C., Fernöc, M., Olssonc, H., Jernströmc, H., Eka, S., Gustavssona, E., Ingvard, C., et al. (2011). Molecular serum portraits in patients with primary breast cancer predict the development of distant metastases. Proc. Natl. Acad. Sci. *108*, 14252–14257.

Chen, G., Gharib, T.G., Huang, C.-C., Taylor, J.M.G., Misek, D.E., Kardia, S.L.R., Giordano, T.J., Iannattoni, M.D., Orringer, M.B., Hanash, S.M., et al. (2002). Discordant Protein and mRNA Expression in Lung Adenocarcinomas. Mol. Cell. Proteomics *1*, 304–313.

Cheng, P., Corzo, C. a, Luetteke, N., Yu, B., Nagaraj, S., Bui, M.M., Ortiz, M., Nacken, W., Sorg, C., Vogl, T., et al. (2008). Inhibition of dendritic cell differentiation and accumulation of myeloid-derived suppressor cells in cancer is regulated by S100A9 protein. J. Exp. Med. *205*, 2235–2249.

Chou, R.-H., Wen, H.-C., Liang, W.-G., Lin, S.-C., Yuan, H.-W., Wu, C.-W., and Chang, W.-S.W. (2012). Suppression of the invasion and migration of cancer cells by SERPINB family genes and their derived peptides. Oncol. Rep. *27*, 238–245.

Chouaib, S., Asselin-Paturel, C., Mami-Chouaib, F., Caignard, A., and Blay, J.Y. (1997). The host-tumor immune conflict: from immunosuppression to resistance and destruction. Immunol. Today *18*, 493–497.

Chrambach, A., and Rodbard, D. (1971). Polyacrylamide gel electrophoresis. Science (80-.). *172*, 440–451.

Chung, L., Moore, K., Phillips, L., Boyle, F.M., Marsh, D.J., and Baxter, R.C. (2014). Novel serum protein biomarker panel revealed by mass spectrometry and its prognostic value in breast cancer. Breast Cancer Res. *16*, R63.

Cianfrocca, M., and Gradishar, W. (2009). New molecular classifications of breast cancer. CA. Cancer J. Clin. *59*, 303–313.

Conklin, D. (1995). 14-3-3 proteins associate with cdc25 phosphatases. Proc. Natl. Acad. Sci. *92*, 7892–7896.

Cousineau, I., Abaji, C., and Belmaaza, A. (2005). BRCA1 regulates RAD51 function in response to DNA damage and suppresses spontaneous sister chromatid replication slippage: implications for sister chromatid cohesion, genome stability, and carcinogenesis. Cancer Res. *65*, 11384–11391.

Coussens, L.M., and Werb, Z. (2002). Inflammation and cancer. Nature *420*, 860–867.

Cowland, J., and Carter, D. (2003). Tissue expression of copines and isolation of copines I and III from the cytosol of human neutrophils. J. Leukoc. Biol. *74*, 379–388.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367 – 1372.

Crowther, M., Brown, N.J., Bishop, E.T., and Lewis, C.E. (2001). Microenvironmental influence on macrophage regulation of angiogenesis in wounds and malignant tumors. J. Leukoc. Biol. *70*, 478–490.

Cui, X., Liu, Y., Wan, C., Lu, C., Cai, J., He, S., Ni, T., Zhu, J., Wei, L., Zhang, Y., et al. (2014). Decreased expression of SERPINB1 correlates with tumor invasion and poor prognosis in hepatocellular carcinoma. J. Mol. Histol. *45*, 59–68.

Curtis, C., Shah, S.P., Chin, S.-F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352.

D'Arcy, V., Abdullaev, Z.K., Pore, N., Docquier, F., Torrano, V., Chernukhin, I., Smart, M., Farrar, D., Metodiev, M., Fernandez, N., et al. (2006). The potential of BORIS detected in the leukocytes of breast cancer patients as an early marker of tumorigenesis. Clin. Cancer Res. *12*, 5978–5986.

Damiens, E., and Yazidi, I. El (1999). Lactoferrin inhibits G1 cyclin- dependent kinases during growth arrest of human breast carcinoma cells. J. Cell. Biochem. *74*, 486–498.

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

Debey, S., Schoenbeck, U., Hellmich, M., Gathof, B.S., Pillai, R., Zander, T., and Schultze, J.L. (2004). Comparison of different isolation techniques prior gene expression profiling of blood derived cells: impact on physiological responses, on overall expression and the role of different cell types. Pharmacogenomics J. *4*, 193–207.

DeNardo, D.G., Brennan, D.J., Rexhepaj, E., Ruffell, B., Shiao, S.L., Madden, S.F., Gallagher, W.M., Wadhwani, N., Keil, S.D., Junaid, S. a, et al. (2011). Leukocyte complexity predicts breast cancer survival and functionally regulates response to chemotherapy. Cancer Discov. *1*, 54–67.

Devarajan, P. (2010). Review: Neutrophil gelatinase-associated lipocalin: A troponin-like biomarker for human acute kidney injury. Nephrology *15*, 419–428.

Dheda, K., Huggett, J., and Bustin, S. (2004). Validation of housekeeping genes for normalizing RNA expression in real-time PCR. Biotechniques *37*, 112–119.

Drabovich, A.P., Martínez-Morillo, E., and Diamandis, E.P. (2014). Toward an integrated pipeline for protein biomarker development. Biochim. Biophys. Acta *S1570-9639*.

Duffy, M.J. (2006). Serum tumor markers in breast cancer: are they of clinical value? Clin. Chem. *52*, 345–351.

Dunn, G., Bruce, A., and Ikeda, H. (2002). Cancer immunoediting: from immunosurveillance to tumor escape. Nat. Immunol. *3*, 991–998.

Ebeling, F., Schmitt, U., Untch, M., Nagel, D., Fateh-Moghadam A, Stieber, P., and Seidel, D. (1999). Tumour markers CEA and CA 15-3 as Prognostic factors in breast cancer--univariate and multivariate analysis. Anticancer Res. *19*, 2545–2550.

Edgar, R., Domrachev, M., and Lash, A. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*, 207–210.

Eichbaum, C., Meyer, A.-S., Wang, N., Bischofs, E., Steinborn, A., Bruckner, T., Brodt, P., Sohn, C., and Eichbaum, M.H.R. (2011). Breast cancer cell-derived cytokines, macrophages and cell adhesion: implications for metastasis. Anticancer Res. *31*, 3219–3227.

Emberley, E.D., Murphy, L.C., and Watson, P.H. (2004). S100A7 and the progression of breast cancer. Breast Cancer Res. *6*, 153–159.

Engel, J., Eckel, R., Kerr, J., Schmidt, M., Fürstenberger, G., Richter, R., Sauer, H., Senn, H.-J., and Hölzel, D. (2003). The process of metastasisation for breast cancer. Eur. J. Cancer *39*, 1794–1806.

Farley, K., Stolley, J.M., Zhao, P., Cooley, J., and Remold-O'Donnell, E. (2012). A serpinB1 regulatory mechanism is essential for restricting neutrophil extracellular trap generation. J. Immunol. *189*, 4574–4581.

Feezor, R.J., Baker, H. V, Mindrinos, M., Hayden, D., Tannahill, C.L., Brownstein, B.H., Fay, A., MacMillan, S., Laramie, J., Xiao, W., et al. (2004). Whole blood and leukocyte RNA isolation for gene expression analyses. Physiol. Genomics *19*, 247–254.

Fentiman, I., and D'Arrigo, C. (2004). Pathogenesis of breast carcinoma. Int. J. Clin. Pract. *58*, 35–40.

Fidler, I. (2003). The pathogenesis of cancer metastasis: the'seed and soil'hypothesis revisited. Nat. Rev. Cancer *3*, 453–458.

Forster, T., Roy, D., and Ghazal, P. (2003). Experiments using microarray technology: limitations and standard operating procedures. J. Endocrinol. *178*, 195–204.

Fox, P., Hudson, M., Brown, C., Lord, S., Gebski, V., De Souza, P., and Lee, C.K. (2013). Markers of systemic inflammation predict survival in patients with advanced renal cell cancer. Br. J. Cancer *109*, 147–153.

Fridlender, Z., Sun, J., Kim, S., and Kapoor, V. (2009). Polarization of tumor-associated neutrophil phenotype by TGF-β:"N1" versus "N2" TAN. Cancer Cell *16*, 183–194.

Friedrichs, K., Gluba, S., Eidtmann, H., and Jonat, W. (1993). Overexpression of p53 and prognosis in breast cancer. Cancer *72*, 3641–3647.

Füzéry, A.K., Levin, J., Chan, M.M., and Chan, D.W. (2013). Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. Clin. Proteomics *10*, 13.

Gabrilovich, D.I., Ostrand-Rosenberg, S., and Bronte, V. (2012). Coordinated regulation of myeloid cells by tumours. Nat. Rev. Immunol. *12*, 253–268.

George, G., and Mittal, R. (2010). MicroRNAs: Potential biomarkers in cancer. Indian J. Clin. Biochem. *25*, 4–14.

Gharbi, S. (2001). Evaluation of Two-dimensional Differential Gel Electrophoresis for Proteomic Expression Analysis of a Model Breast Cancer Cell System. Mol. Cell. Proteomics *1*, 91–98.

Ghosh, J. (2003). Inhibition of arachidonate 5-lipoxygenase triggers prostate cancer cell death through rapid activation of c-Jun N-terminal kinase. Biochem. Biophys. Res. Commun. *307*, 2342–2349.

Ginzberg, H.H., Cherapanov, V., Dong, Q., Cantin, A., McCulloch, C.A.G., Shannon, P.T., and Downey, G.P. (2001). Neutrophil-mediated epithelial injury during transmigration: role of elastase. Am. J. Physiol. *8*, 705–717.

Glasel, J. (1995). Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. Biotechniques *18*, 62–63.

Glinsky, G., Higashiyama, T., and Glinskii, A. (2004). Classification of human breast cancer using gene expression profiling as a component of the survival predictor algorithm. Clin. Cancer Res. *10*, 2272–2283.

Gregory, A.D., and Houghton, a M. (2011). Tumor-associated neutrophils: new targets for cancer therapy. Cancer Res. *71*, 2411–2416.

Grivennikov, S.I., Greten, F.R., and Karin, M. (2010). Immunity, inflammation, and cancer. Cell *140*, 883–899.

Halliday, W., Koppi, T., Khan, J., and Davis, N. (1980). Leukocyte adherence inhibition: tumor specificity of cellular and serum-blocking reactions in human melanoma, breast cancer, and colorectal cancer. J. Natl. Cancer Inst. *65*, 372–332.

Harris, S.L., and Levine, A.J. (2005). The p53 pathway: positive and negative feedback loops. Oncogene *24*, 2899–2908.

Haslene-Hox, H., Oveland, E., Woie, K., Salvesen, H., Wiig, H., and Tenstad, O. (2013). Increased WD-repeat containing protein 1 in interstitial fluid from ovarian carcinomas shown by comparative proteomic analysis of malignant and healthy gynecological tissue. Biochim. Biophys. Acta *1834*, 2347–2359.

Haupt, S., Berger, M., Goldberg, Z., and Haupt, Y. (2003). Apoptosis - the p53 network. J. Cell Sci. *116*, 4077–4085.

Heinrich, C., Keller, C., Boulay, A., Vecchi, M., Bianchi, M., Sack, R., Lienhard, S., Duss, S., Hofsteenge, J., and Hynes, N.E. (2010). Copine-III interacts with ErbB2 and promotes tumor cell migration. Oncogene *29*, 1598–1610.

Herbig, U., and Sedivy, J.M. (2006). Regulation of growth arrest in senescence: telomere damage is not the end of the story. Mech. Ageing Dev. *127*, 16–24.

Heywang-Köbrunner, S.H., Hacker, A., and Sedlacek, S. (2011). Advantages and Disadvantages of Mammography Screening. Breast Care *6*, 199–207.

Hiratsuka, S., Watanabe, A., Aburatani, H., and Maru, Y. (2006). Tumour-mediated upregulation of chemoattractants and recruitment of myeloid cells predetermines lung metastasis. Nat. Cell Biol. *8*, 1369–1375.

Hodge, J., Bub, J., Kaul, S., Kajdacsy-Balla, A., and Lindholm, P.F. (2003). Requirement of RhoA activity for increased nuclear factor κB activity and PC-3 human prostate cancer cell invasion. Cancer Res. *63*, 1359–1364.

Hofmann, a, Raguénès-Nicol, C., Favier-Perron, B., Mesonero, J., Huber, R., Russo-Marie, F., and Lewit-Bentley, A. (2000). The annexin A3-membrane interaction is modulated by an N-terminal tryptophan. Biochemistry *39*, 7712–7721.

Hondermarck, H., Vercoutter-Edouart, A.-S., Révillion, F., Lemoine, J., El-Yazidi-Belkoura, I., Nurcombe, V., and Peyrat, J.-P. (2001). Proteomics of breast cancer for marker discovery and signal pathway profiling. Proteomics *1*, 1216–1232.

Hu, N., Li, Y., Zhao, Y., Wang, Q., You, J., Zhang, X., and Ye, L. (2011). A novel positive feedback loop involving FASN/p-ERK1/2/5-LOX/LTB4/FASN sustains high growth of breast cancer cells. Acta Pharmacol. Sin. *32*, 921–929.

Hu, Q., Noll, R.J., Li, H., Makarov, A., Hardman, M., and Graham Cooks, R. (2005). The Orbitrap: a new mass spectrometer. J. Mass Spectrom. *40*, 430–443.

Itahana, K., Dimri, G., and Campisi, J. (2001). Regulation of cellular senescence by p53. Eur. J. Biochem. *268*, 2784–2791.

Izbicka, E., Streeper, R.T., Michalek, J.E., Louden, C.L., Diaz, A., and Campos, D.R. (2012). Plasma biomarkers distinguish non-small cell lung cancer from asthma and differ in men and women. Cancer Genomics Proteomics *9*, 27–35.

Jackson, J., Seed, M., Kircher, C., Willoughby, D.A., and Winkler, J.D. (1997). The codependence of angiogenesis and chronic inflammation. FASEB J. *11*, 457–465.

Jandus, C., Boligan, K., Chijioke, O., Liu, H., Dahlhaus, M., Demoulins, T., Schneider, C., Wehrli, M., Hunger, R.E., Baerlocher, G.M., et al. (2014). Interactions between Siglec-7/9 receptors and ligands influence NK cell–dependent tumor immunosurveillance. J. Clin. Invest. *124*, 1810–1820.

Jöhrer, K., Janke, K., Krugmann, J., Fiegl, M., and Greil, R. (2004). Transendothelial migration of myeloma cells is increased by tumor necrosis factor (TNF)-α via TNF receptor 2 and autocrine up-regulation of MCP-1. Clin. Cancer Res. *10*, 1901–1910.

De Jong, J., van Diest, P., van der Valk, P., and Baak, J. (1998). Expression of growth factors, growth inhibiting factors, and their receptors in invasive breast cancer. I: An inventory in search of autocrine and paracrine loops. J. Pathol. *184*, 44–52.

Jung, E.-J., Moon, H.-G., Park, S.-T., Cho, B.-I., Lee, S.-M., Jeong, C.-Y., Ju, Y.-T., Jeong, S.-H., Lee, Y.-J., Choi, S.-K., et al. (2010). Decreased annexin A3 expression correlates with tumor progression in papillary thyroid cancer. Proteomics. Clin. Appl. *4*, 528–537.

Kadish, A.S., Marcus, D.M., and Bloom, B.R. (1976). Inhibition of leukocyte migration by human breast-cancer-associated antigens. Int. J. Cancer *18*, 581–586.

Kageyama, S., Isono, T., Iwaki, H., Wakabayashi, Y., Okada, Y., Kontani, K., Yoshimura, K., Terai, A., Arai, Y., and Yoshiki, T. (2004). Identification by proteomic analysis of calreticulin as a marker for bladder cancer and evaluation of the diagnostic accuracy of its detection in urine. Clin. Chem. *50*, 857–866.

Kamai, T., Yamanishi, T., Shirataki, H., Takagi, K., Asami, H., Ito, Y., and Yoshida, K.-I. (2004). Overexpression of RhoA, Rac1, and Cdc42 GTPases is associated with progression in testicular cancer. Clin. Cancer Res. *10*, 4799–4805.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. *40*, D109–14.

Karin, M. (2006). Nuclear factor-kappaB in cancer development and progression. Nature *441*, 431–436.

Keshaviah, A., Dellapasqua, S., Rotmensz, N., Lindtner, J., Crivellari, D., Collins, J., Colleoni, M., Thürlimann, B., Mendiola, C., Aebi, S., et al. (2007). CA15-3 and alkaline phosphatase as predictors for breast cancer recurrence: a combined analysis of seven International Breast Cancer Study Group trials. Ann. Oncol. *18*, 701–708.

Khadir, A., Ali, T., and Tiss, A. (2013). Proteomics Approaches towards Early Detection and Diagnosis of Cancer. J. Carcinog. Mutagen. *S14*.

Kim, R., Emi, M., and Tanabe, K. (2007). Cancer immunoediting from immune surveillance to immune escape. Immunology *121*, 1–14.

Klimov, E., Selivanova, N., Razumnova, G., Rudko, O., and Golovatenko-Abramov, P. (2013). Short Communication RHOA, SEMA3B, and CKAP2 Expression in Leukaemia of Different Types: the Results of a Pilot Experiment. Folia Biol. (Praha). *206*, 204–206.

Komatsu, N., Matsueda, S., Tashiro, K., Ioji, T., Shichijo, S., Noguchi, M., Yamada, A., Doi, A., Suekane, S., Moriya, F., et al. (2012). Gene expression profiles in peripheral blood as a biomarker in cancer patients receiving peptide vaccination. Cancer *118*, 3208–3221.

Koyasu, S. (2003). The role of PI3K in immune cells. Nat. Immunol. *4*, 313–320.

Krasteva, M.E., Bozhanov, S.S., Antov, G.G., Gospodinova, Z.I., Angelova, S.G., and Georgieva, E.I. (2012). Breast cancer patients with hypermethylation in the promoter of BRCA1 gene exhibit favorable clinical status. Neoplasma *59*, 85–91.

Krause, D., Fackler, M., CI, C., and WS, M. (1996). CD34: structure, biology, and clinical utility. Blood *87*, 1–13.

Kuhns, D., Long Priel, D., and Gallin, J. (1995). Loss of L-selectin (CD62L) on human neutrophils following exudation in vivo. Cell. Immunol. *164*, 306–310.

Kusmartsev, S., and Gabrilovich, D. (2006). Effect of tumor-derived cytokines and growth factors on differentiation and immune suppressive features of myeloid cells in cancer. Cancer Metastasis Rev. *25*, 323–331.

LaBreche, H.G., Nevins, J.R., and Huang, E. (2011). Integrating factor analysis and a transgenic mouse model to reveal a peripheral blood predictor of breast tumors. BMC Med. Genomics *4*, 61.

Lam, L., Aktary, Z., Bishay, M., Werkman, C., Kuo, C.-Y., Heacock, M., Srivastava, N., Mackey, J.R., and Pasdar, M. (2012). Regulation of subcellular distribution and oncogenic potential of nucleophosmin by plakoglobin. Oncogenesis *1*, e4.

Leal, M.F., Calcagno, D.Q., Demachki, S., Assumpção, P.P., Chammas, R., Burbano, R.R., and Smith, M.D.A.C. (2012). Clinical implication of 14-3-3 epsilon expression in gastric cancer. World J. Gastroenterol. *18*, 1531–1537.

Leber, M., and Efferth, T. (2009). Molecular principles of cancer invasion and metastasis (review). Int. J. Oncol. *34*, 881–895.

Ledderose, C., Heyn, J., Limbeck, E., and Kreth, S. (2011). Selection of reliable reference genes for quantitative real-time PCR in human T cells and neutrophils. BMC Res. Notes *4*, 427.

Lee, C., Chen, S., and Lin, T. (1977). Inhibition of Leukocyte Migration by Tumor-associated Antigen in Soluble Extracts of Human Hepatoma. Cancer Res. *37*, 918–921.

Lee, H.-H., Lim, C.-A., Cheong, Y.-T., Singh, M., and Gam, L.-H. (2012). Comparison of protein expression profiles of different stages of lymph nodes metastasis in breast cancer. Int. J. Biol. Sci. *8*, 353–362.

Leek, R., Lewis, C., Whitehouse, R., Greenall, J., and Harris, A. (1996). Association of macrophage infiltration with angiogenesis and prognosis in invasive breast carcinoma. Cancer Res. *56*, 4625–4629.

Leek, R.D., Landers, R., Fox, S.B., Ng, F., Harris, a L., and Lewis, C.E. (1998). Association of tumour necrosis factor alpha and its receptors with thymidine phosphorylase expression in invasive breast carcinoma. Br. J. Cancer *77*, 2246–2251.

Levay, P.F., and Viljoen, M. (1995). Lactoferrin: a general review. Haematologica *80*, 252–267.

Li, D.-Q., Wang, L., Fei, F., Hou, Y.-F., Luo, J.-M., Zeng, R., Wu, J., Lu, J.-S., Di, G.-H., Ou, Z.-L., et al. (2006). Identification of breast cancer metastasis-associated proteins in an isogenic tumor metastasis model using two-dimensional gel electrophoresis and liquid chromatography-ion trap-mass spectrometry. Proteomics *6*, 3352–3368.

Liu, R., Gong, J., Chen, J., Li, Q., Song, C., Zhang, J., Li, Y., Liu, Z., Dong, Y., Chen, L., et al. (2012a). Calreticulin as a potential diagnostic biomarker for lung cancer. Cancer Immunol. *61*, 855–864.

Liu, T.-A., Jan, Y.-J., Ko, B.-S., Liang, S.-M., Chen, S.-C., Wang, J., Hsu, C., Wu, Y.-M., and Liou, J.-Y. (2013). 14-3-3E Overexpression Contributes To Epithelial-Mesenchymal Transition of Hepatocellular Carcinoma. PLoS One *8*, e57968.

Liu, Y., Xu, L., Zeng, Q., Wang, J., Wang, M., Xi, D., Wang, X., Yang, D., Luo, X., and Ning, Q. (2012b). Downregulation of FGL2/prothrombinase delays HCCLM6 xenograft tumour growth and decreases tumour angiogenesis. Liver Int. *32*, 1585–1595.

Lopez, M. (1999). 2-D electrophoresis using carrier ampholytes in the first dimension. Methods Mol. Biol. *112*, 111–127.

Lord, C.J., and Ashworth, A. (2012). The DNA damage response and cancer therapy. Nature *481*, 287–294.

Luboshits, G., Shina, S., Kaplan, O., Engelberg, S., Nass, D., Lifshitz-mercer, B., Chaitchik, S., Keydar, I., and Ben-baruch, A. (1999). Elevated Expression of the CC Chemokine Regulated on Activation , Normal T Cell Expressed and Secreted (RANTES) in Advanced Breast Carcinoma 1. Cancer Res. *59*, 4681–4687.

Ludwig, J. a, and Weinstein, J.N. (2005). Biomarkers in cancer staging, prognosis and treatment selection. Nat. Rev. Cancer *5*, 845–856.

Lundberg, E., Fagerberg, L., Klevebring, D., Matic, I., Geiger, T., Cox, J., Algenäs, C., Lundeberg, J., Mann, M., and Uhlen, M. (2010). Defining the transcriptome and proteome in three functionally different human cell lines. Mol. Syst. Biol. *6*, 450.

Luster, A.D., Alon, R., and von Andrian, U.H. (2005). Immune cell migration in inflammation: present and future therapeutic targets. Nat. Immunol. *6*, 1182–1190.

Lwin, Z.-M., Guo, C., Salim, A., Yip, G.W.-C., Chew, F.-T., Nan, J., Thike, A.A., Tan, P.-H., and Bay, B.-H. (2010). Clinicopathological significance of calreticulin in breast invasive ductal carcinoma. Mod. Pathol. *23*, 1559–1566.

Mangold, S., Wu, S.K., Norwood, S.J., Collins, B.M., Hamilton, N.A., Thorn, P., and Yap, A.S. (2011). Hepatocyte growth factor acutely perturbs actin filament anchorage at the epithelial zonula adherens. Curr. Biol. *21*, 503–507.

Mar-Aguilar, F., Mendoza-Ramírez, J. a, Malagón-Santiago, I., Espino-Silva, P.K., Santuario-Facio, S.K., Ruiz-Flores, P., Rodríguez-Padilla, C., and Reséndez-Pérez, D. (2013). Serum circulating microRNA profiling for identification of potential breast cancer biomarkers. Dis. Markers *34*, 163–169.

McCourt, M., Wang, J.H., Sookhai, S., and Redmond, H.P. (2001). Activated human neutrophils release hepatocyte growth factor/scatter factor. Eur. J. Surg. Oncol. *27*, 396–403.

Mcpherson, K., Steel, C.M., and Dixon, J.M. (2000). Breast cancer — epidemiology , risk factors , and genetics Risk factors for breast cancer. Br. Med. J. *321*, 624.

Metzeler, K.H., Hummel, M., Bloomfield, C.D., Spiekermann, K., Braess, J., Sauerland, M.-C., Heinecke, A., Radmacher, M., Marcucci, G., Whitman, S.P., et al. (2008). An 86-probe-set gene-

expression signature predicts survival in cytogenetically normal acute myeloid leukemia. Blood *112*, 4193–4201.

Misek, D.E., and Kim, E.H. (2011). Protein biomarkers for the early detection of breast cancer. Int. J. Proteomics *2011*, ID 343582.

Moll, R. (1994). Cytokeratins in the histological diagnosis of malignant tumors. Int. J. Biol. Markers *9*, 63–69.

Moll, U.M., Riou, G., and Levine, a J. (1992). Two distinct mechanisms alter p53 in breast cancer: mutation and nuclear exclusion. Proc. Natl. Acad. Sci. U. S. A. *89*, 7262–7266.

Möröy, T., and Zörnig, M. (1996). Regulators of Life and Death: The Bcl-2 Gene Family. Cell. Physiol. Biochem. *6*, 312–336.

Müller, a, Homey, B., Soto, H., Ge, N., Catron, D., Buchanan, M.E., McClanahan, T., Murphy, E., Yuan, W., Wagner, S.N., et al. (2001). Involvement of chemokine receptors in breast cancer metastasis. Nature *410*, 50–56.

Muller, W. a (2009). Mechanisms of transendothelial migration of leukocytes. Circ. Res. *105*, 223–230.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp. Quant. Biol. *51*, 263–273.

Narod, S. a, and Foulkes, W.D. (2004). BRCA1 and BRCA2: 1994 and beyond. Nat. Rev. Cancer *4*, 665–676.

Naylor, S. (2010). Biomarkers: current perspectives and future prospects. Expert Rev. Mol. Diagn. *3*, 525–529.

Nishizuka, S., Chen, S., Gwadry, F., Alexander, J., Major, S.M., Scherf, U., Reinhold, W.C., Waltham, M., Charboneau, L., Young, L., et al. (2003). Diagnostic Markers That Distinguish Colon and Ovarian Adenocarcinomas Identification by Genomic, Proteomic, and Tissue Array Profiling. Cancer Res. *63*, 5243–5250.

Nolen, B.M., Marks, J.R., Ta'san, S., Rand, A., Luong, T.M., Wang, Y., Blackwell, K., and Lokshin, A.E. (2008). Serum biomarker profiles and response to neoadjuvant chemotherapy for locally advanced breast cancer. Breast Cancer Res. *10*, R45.

Nowarski, R., Gagliani, N., Huber, S., and Flavell, R. a (2013). Innate immune cells in inflammation and cancer. Cancer Immunol. Res. *1*, 77–84.

O'connell, J., O'sullivan, G., Collins, K.J., and Shanagan, F. (1996). The Fas counterattack: Fas-mediated T cell killing by colon cancer cells expressing Fas ligand. J. Exp. Med. *184*, 1075–1082.

O'Farell, P. (1975). High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. *250*, 4007–4021.

Ohtsuka, M., and Arase, H. (2004). NFAM1, an immunoreceptor tyrosine-based activation motif-bearing molecule that regulates B cell development and signaling. Proc. Natl. Acad. Sci. *101*, 8126–8131.

Olsen, J. V, Schwartz, J.C., Griep-Raming, J., Nielsen, M.L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., et al. (2009). A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. Mol. Cell. Proteomics *8*, 2759–2769.

Onrust, S. V, Hartl, P.M., Rosen, S.D., and Hanahan, D. (1996). Modulation of L-selectin ligand expression during an immune response accompanying tumorigenesis in transgenic mice. J. Clin. Invest. *97*, 54–64.

Orntoft, T.F., Thykjaer, T., Waldman, F.M., Wolf, H., and Celis, J.E. (2001). Genome-wide Study of Gene Copy Numbers, Transcripts, and Protein Levels in Pairs of Non-invasive and Invasive Human Transitional Cell Carcinomas. Mol. Cell. Proteomics *1*, 37–45.

Pan, Q.-Z., Pan, K., Weng, D.-S., Zhao, J.-J., Zhang, X.-F., Wang, D.-D., Lv, L., Jiang, S.-S., Zheng, H.-X., and Xia, J.-C. (2013). Annexin A3 promotes tumorigenesis and resistance to chemotherapy in hepatocellular carcinoma. Mol. Carcinog. *4*, 1–10.

Pedrero, J.M.G., Fernandez, M.P., Morgan, R.O., Zapatero, H.A., Gonzalez, M.V., Nieto, C.S., and Rodrigo, J.P. (2004). Annexin A1 Down-Regulation in Head and Neck Cancer Is Associated with Epithelial Differentiation Status. Am. J. Pathol. *164*, 73–79.

Penn, I., and Starzl, T. (1973). Immunosuppression and cancer. Transplant. Proc. *5*, 943–947.

Peri, a, Cordella-Miele, E., Miele, L., and Mukherjee, a B. (1993). Tissue-specific expression of the gene coding for human Clara cell 10-kD protein, a phospholipase A2-inhibitory protein. J. Clin. Invest. *92*, 2099–2109.

Perou, C., Sørlie, T., Eisen, M., van de Rijn, M., Jeffrey, S., Rees, C., Pollack, J., Ross, D., Johnsen, H., Akslen, L., et al. (2000). Molecular portraits of human breast tumours. Nature *533*, 747–752.

Perretti, M., and D'Acquisto, F. (2009). Annexin A1 and glucocorticoids as effectors of the resolution of inflammation. Nat. Rev. Immunol. *713*, 62–70.

Pillé, J.-Y., Denoyelle, C., Varet, J., Bertrand, J.-R., Soria, J., Opolon, P., Lu, H., Pritchard, L.-L., Vannier, J.-P., Malvy, C., et al. (2005). Anti-RhoA and anti-RhoC siRNAs inhibit the proliferation and invasiveness of MDA-MB-231 breast cancer cells in vitro and in vivo. Mol. Ther. *11*, 267–274.

Pitteri, S.J., Faca, V.M., Kelly-spratt, K.S., Kasarda, A.E., Wang, H., Zhang, Q., Newcomb, L., Krasnoselsky, A., Paczesny, S., Choi, G., et al. (2008). Plasma Proteome Profiling of a Mouse Model of Breast Cancer Identifies a Set of Up-Regulated Proteins in Common with Human Breast Cancer Cells research articles. J. Proteome Res. *7*, 1481–1489.

Pitteri, S.J., Kelly-Spratt, K.S., Gurley, K.E., Kennedy, J., Buson, T.B., Chin, A., Wang, H., Zhang, Q., Wong, C.-H., Chodosh, L. a, et al. (2011). Tumor microenvironment-derived proteins

dominate the plasma proteome response during breast cancer induction and progression. Cancer Res. *71*, 5090–5100.

Ptolemy, A.S., and Rifai, N. (2010). What is a biomarker? Research investments and lack of clinical integration necessitate a review of biomarker terminology and validation schema. Scand. J. Clin. Lab. Investig. *242*, 6–14.

Qi, W., Liu, X., Qiao, D., and Martinez, J.D. (2005). Isoform-specific expression of 14-3-3 proteins in human lung cancer tissues. Int. J. Cancer. *113*, 359–363.

Qiu, F.-M., Yu, J.-K., Chen, Y.-D., Jin, Q.-F., Sui, M.-H., and Huang, J. (2009). Mining novel biomarkers for prognosis of gastric cancer with serum proteomics. J. Exp. Clin. Cancer Res. *28*, 126.

Queen, M.M., Ryan, R.E., Holzer, R.G., Keller-Peck, C.R., and Jorcyk, C.L. (2005). Breast cancer cells stimulate neutrophils to produce oncostatin M: potential implications for tumor progression. Cancer Res. *65*, 8896–8904.

Raghavan, M., Wijeyesakere, S.J., Peters, L.R., and Del Cid, N. (2013). Calreticulin in the immune system: ins and outs. Trends Immunol. *34*, 13–21.

Di Renzo, M.F., Olivero, M., Giacomini, A., Porte, H., Chastre, E., Mirossay, L., Nordlinger, B., Bretti, S., Bottardi, S., and Giordano, S. (1995). Overexpression and amplification of the met/HGF receptor gene during the progression of colorectal cancer. Clin. Cancer Res. *1*, 147–154.

Richmond, A. (2002). Nf-kappa B, chemokine gene transcription and tumour growth. Nat. Rev. Immunol. *2*, 664–674.

Robinson, S.C., Scott, K. a, and Balkwill, F.R. (2002). Chemokine stimulation of monocyte matrix metalloproteinase-9 requires endogenous TNF-alpha. Eur. J. Immunol. *32*, 404–412.

Rodier, F., and Campisi, J. (2011). Four faces of cellular senescence. J. Cell Biol. *192*, 547–556.

Rotunno, M., Hu, N., Su, H., Wang, C., Goldstein, A.M., Bergen, A.W., Consonni, D., Pesatori, A.C., Bertazzi, P.A., Wacholder, S., et al. (2011). A gene expression signature from peripheral whole blood for stage I lung adenocarcinoma. Cancer Prev. Res. *4*, 1599–1608.

Ruffell, B., Au, A., Rugo, H.S., Esserman, L.J., Hwang, E.S., and Coussens, L.M. (2012). Leukocyte composition of human breast cancer. Proc. Natl. Acad. Sci. *109*, 2796–2801.

Russo, J., and Hu, Y. (2000). Developmental, cellular, and molecular basis of human breast cancer. J. Natl. Cancer Institutes. Monogr. *27*, 17–37.

Sasaki, A., Yasukawa, H., and Suzuki, A. (1999). Cytokine inducible SH2 protein 3 (CIS3/SOCS3) inhibits Janus tyrosine kinase by binding through the N- terminal kinase inhibitory region as well as SH2 domain. Genes to Cell *4*, 339–351.

Schena, M., Shalon, D., Davis, R., and Brown, P. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science (80-.). *270*, 467–470.

Schiffner, S., Zimara, N., Schmid, R., and Bosserhoff, A.-K. (2011). P54Nrb Is a New Regulator of Progression of Malignant Melanoma. Carcinogenesis *32*, 1176–1182.

Schroder, K., Hertzog, P.P., Timothy, R., and Hume, D.A. (2004). Interferon-γ: an overview of signals, mechanisms and functions. J. Leukoc. Biol. *75*, 163–189.

Scigelova, M., and Makarov, A. (2006). Orbitrap mass analyzer--overview and applications in proteomics. Proteomics *6*, 16–21.

Scully, R., and Livingston, D.M. (2000). In search of the tumour-suppressor functions of BRCA1 and BRCA2. Nature *408*, 429–432.

Sharma, P., Sahni, N.S., Tibshirani, R., Skaane, P., Urdal, P., Berghagen, H., Jensen, M., Kristiansen, L., Moen, C., Sharma, P., et al. (2005). Early detection of breast cancer based on gene-expression patterns in peripheral blood cells. Breast Cancer Res. *7*, 634–644.

Shashidharamurthy, R., Machiah, D., Aitken, J.D., Putty, K., Srinivasan, G., Chassaing, B., Parkos, C. a, Selvaraj, P., and Vijay-Kumar, M. (2013). Differential role of lipocalin 2 during immune complex-mediated acute and chronic inflammation in mice. Arthritis Rheum. *65*, 1064–1073.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. Anal. Chem. *68*, 850–858.

Showe, M.K., Vachani, A., Kossenkov, A. V, Yousef, M., Nikonova, E. V, Chang, C., Kucharczuk, J., Tran, B., Yie, T.A., Speicher, D., et al. (2009). Gene Expression Profiles in Peripheral Blood Mononuclear Cells Can Distinguish Patients with Non-Small-Cell Lung Cancer from Patients with Non-Malignant Lung Disease. Cancer Res. *69*, 9202–9210.

Shuai, K., and Liu, B. (2003). Regulation of JAK-STAT signalling in the immune system. Nat. Rev. Immunol. *3*, 900–911.

Slee, E. a, O'Connor, D.J., and Lu, X. (2004). To die or not to die: how does p53 decide? Oncogene *23*, 2809–2818.

Sodowich, B.I., Fadl, I., and Burns, C. (2007). Method validation of in vitro RNA transcript analysis on the Agilent 2100 Bioanalyzer. Electrophoresis *28*, 2368–2378.

Somiari, R.I., Sullivan, A., Russell, S., Somiari, S., Hu, H., Jordan, R., George, A., Katenhusen, R., Buchowiecka, A., Arciero, C., et al. (2003). High-throughput proteomic analysis of human infiltrating ductal carcinoma of the breast. Proteomics *3*, 1863–1873.

Sørliea, T., Perou, C.M., Tibshiranie, R., Aasf, T., Geislerg, S., Johnsenb, H., Hastiee, T., Eisenh, M.B., Rijni, M. van de, Jeffrey, S.S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. *98*, 10869–10874.

Sreekumar, A., Nyati, M.K., Varambally, S., Proteins, R., Barrette, T.R., Ghosh, D., Lawrence, T.S., and Chinnaiyan, A.M. (2001). Profiling of Cancer Cells Using Protein Microarrays :

Discovery of Novel Radiation-regulated Proteins Profiling of Cancer Cells Using Protein Microarrays : Discovery of Novel. Cancer Res. *61*, 7585–7593.

Srinivas, P.R., Verma, M., Zhao, Y., and Srivastava, S. (2002). Proteomics for cancer biomarker discovery. Clin. Chem. *48*, 1160–1169.

Standish, L.J., Sweet, E.S., Novack, J., Wenner, C.A., Bridge, C., Nelson, A., Martzen, M., and Torkelson, C. (2008). Breast cancer and the immune system. J. Soc. Integr. Oncol. *6*, 158–168.

Sulli, G., Micco, R. Di, and Fagagna, F. di (2012). Crosstalk between chromatin state and DNA damage response in cellular senescence and cancer. Nat. Rev. Cancer *12*, 265–273.

Thomas, R., Guise, T., and Yin, J. (1999). Breast Cancer Cells Interact with Osteoblasts to Support Osteoclast Formation 1. Endocrinology *140*, 4451–4458.

Tian, Q., Stepaniants, S.B., Mao, M., Weng, L., Feetham, M.C., Doyle, M.J., Yi, E.C., Dai, H., Thorsson, V., Eng, J., et al. (2004). Integrated genomic and proteomic analyses of gene expression in Mammalian cells. Mol. Cell. Proteomics *3*, 960–969.

Torácica, A. de la P., and Mama, A. y (2006). Anatomy of the thoracic wall, axilla and breast. Int. J. Morphol. *24*, 691–704.

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. *76*, 4350–4354.

Trapani, J., and Smyth, M. (2002). Functional significance of the perforin/granzyme cell death pathway. Nat. Rev. Immunol. *2*, 735–747.

Tseng, M., Liu, S., Chen, H., Wu, Y., Chiu, C., Chan, P., Chiang, W., Liu, Y., Lu, C., Jou, Y., et al. (2009). Serine protease inhibitor (SERPIN) B1 promotes oral cancer cell motility and is over-expressed in invasive oral squamous cell carcinoma. Oral Oncol. *45*, 771–776.

Twine, N.C., Stover, J.A., Marshall, B., Dukart, G., Hidalgo, M., Stadler, W., Logan, T., Dutcher, J., Hudes, G., Dorner, A.J., et al. (2003). Disease-associated Expression Profiles in Peripheral Blood Mononuclear Cells from Patients with Advanced Renal Cell Carcinoma Disease-associated Expression Profiles in Peripheral Blood Mononuclear Cells from Patients with Advanced Renal Cell Carcinoma. Cancer Res. *63*, 6069–6075.

Tzivion, G., Shen, Y., and Zhu, J. (2001). 14-3-3 proteins; bringing new definitions to scaffolding. Oncogene *4*, 6331–6338.

Van't Veer, L.J., Dai, H., Vijver, M.J. Van De, Kooy, K. Van Der, Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Bernards, Â., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530–536.

Verrills, N. (2006). Clinical proteomics: present and future prospects. Clin. Biochem. Rev. *27*, 99–116.

De Visser, K.E., Eichten, A., and Coussens, L.M. (2006). Paradoxical roles of the immune system during cancer development. Nat. Rev. Cancer *6*, 24–37.

Vogel, C., Abreu, R.D.S., Ko, D., Le, S.-Y., Shapiro, B. a, Burns, S.C., Sandhu, D., Boutz, D.R., Marcotte, E.M., and Penalva, L.O. (2010). Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. Mol. Syst. Biol. *6*, 400.

Wagner, J.G., and Roth, R. a (2000). Neutrophil migration mechanisms, with an emphasis on the pulmonary vasculature. Pharmacol. Rev. *52*, 349–374.

Walther, A., Riehemann, K., and Gerke, V. (2000). A novel ligand of the formyl peptide receptor: annexin I regulates neutrophil extravasation by interacting with the FPR. Mol. Cell *5*, 831–840.

Wang, J., Robinson, J.F., Khan, H.M.R., Carter, D.E., McKinney, J., Miskie, B. a, and Hegele, R. a (2004). Optimizing RNA extraction yield from whole blood for microarray gene expression analysis. Clin. Biochem. *37*, 741–744.

Wang, L., Li, Q., Wu, L., Liu, S., Zhang, Y., Yang, X., Zhu, P., Zhang, H., Zhang, K., Lou, J., et al. (2013). Identification of SERPINB1 as a physiological inhibitor of human granzyme H. J. Immunol. *190*, 1319–1330.

Waugh, D.J.J., and Wilson, C. (2008). The interleukin-8 pathway in cancer. Clin. Cancer Res. *14*, 6735–6741.

Weinberg, R. (2006). The Biology of Cancer (Garland Science).

Wenners, A.S., Mehta, K., Loibl, S., Park, H., Mueller, B., Arnold, N., Hamann, S., Weimer, J., Ataseven, B., Darb-Esfahani, S., et al. (2012). Neutrophil gelatinase-associated lipocalin (NGAL) predicts response to neoadjuvant chemotherapy and clinical outcome in primary human breast cancer. PLoS One *7*, e45826.

Whiteside, T.L. (2002). Tumor-induced death of immune cells: its mechanisms and consequences. Semin. Cancer Biol. *12*, 43–50.

Wilker, E., Lu, J., and Rho, O. (2005). Role of PI3K/Akt signaling in insulin- like growth factor1 (IGF1) skin tumor promotion. Mol. Carcinog. *44*, 137–145.

Williams, M. a, and Solomkin, J.S. (1999). Integrin-mediated signaling in human neutrophil functioning. J. Leukoc. Biol. *65*, 725–736.

Wislez, M., Rabbe, N., Marchal, J., Milleron, B., Crestani, B., Mayaud, C., Antoine, M., Soler, P., and Cadranel, J. (2003). Hepatocyte Growth Factor Production by Neutrophils Infiltrating Bronchioloalveolar Subtype Pulmonary Adenocarcinoma : Role in Tumor Progression and Death Hepatocyte Growth Factor Production by Neutrophils Infiltrating Bronchioloalveolar Subtype Pulmonary. Cancer Res. *63*, 1405–1412.

Xu, Y., Xu, Q., Yang, L., Ye, X., Liu, F., Wu, F., Ni, S., Tan, C., Cai, G., Meng, X., et al. (2013a). Identification and Validation of a Blood-Based 18-Gene Expression Signature in Colorectal Cancer. Clin. Cancer Res. *19*, 3039–3049.

Xu, Y., Xu, Q., Yang, L., Ye, X., Liu, F., Wu, F., Ni, S., Tan, C., Cai, G., Meng, X., et al. (2013b). Identification and Validation of a Blood-Based 18-Gene Expression Signature in Colorectal Cancer. Clin. Cancer Res. *19*, 3039–3049.

Yamamotoa, K., Mitsuko, Y., Nobuyuki, M., and Osamu, M. (2003). SOCS-3 inhibits IL-12-induced STAT4 activation by binding through its SH2 domain to the STAT4 docking site in the IL-12 receptor β2 subunit. Biochem. Biophys. Res. Commun. *310*, 1188–1193.

Yang, J., Bielenberg, D.R., Rodigd, S.J., Doirona, R., Cliftone, M.C., Kung, A.L., Stronge, R.K., Zurakowskic, D., and Mosesa, M.A. (2009). Lipocalin 2 promotes breast cancer progression. Proc. Natl. Acad. Sci. *106*, 3913–3918.

Yom, C.K., Han, W., Kim, S.-W., Kim, H.S., Shin, H.-C., Chang, J.N., Koo, M., Noh, D.-Y., and Moon, B.-I. (2011). Clinical significance of annexin A1 expression in breast cancer. J. Breast Cancer *14*, 262–268.

Yousef, E.M., Laperrière, D., Ramzan-Tahir, M., Mader, S., and Gaboury, L.A. (2013). Deregulated Expression of ANXA1 in Human High-Grade Breast Cancers. J. Mol. Biomark. Diagn. *4*, 1–10.

Yu, G., Wang, J., Chen, Y., Wang, X., Pan, J., Li, Q., and Xie, K. (2008a). Tissue microarray analysis reveals strong clinical evidence for a close association between loss of annexin A1 expression and nodal metastasis in gastric cancer. Clin. Exp. Metastasis *25*, 695–701.

Yu, J., Yu, J., Cordero, K.E., Johnson, M.D., Ghosh, D., Rae, J.M., Chinnaiyan, A.M., and Lippman, M.E. (2008b). A Transcriptional Fingerprint of Estrogen in Human Breast Cancer Predicts Patient Survival. Neoplasia *10*, 79–88.

Yuryev, A., and Wennogle, L. (2003). Novel raf kinase protein-protein interactions found by an exhaustive yeast two-hybrid analysis. Genomics *81*, 112–125.

Zuckerman, N.S., Yu, H., Simons, D.L., Bhattacharya, N., Carcamo-Cavazos, V., Yan, N., Dirbas, F.M., Johnson, D.L., Schwartz, E.J., and Lee, P.P. (2013). Altered local and systemic immune profiles underlie lymph node metastasis in breast cancer patients. Int. J. Cancer *132*, 2537–2547.