In silico and molecular validation of identified putative genes and functional analysis of a NKG2D ligand as a breast cancer biomarkers

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor Philosophiae, in the Department of Biotechnology, University of the Western Cape.

Supervisor: Dr. Ashley Pretorius

Co-supervisor: Prof. Mervin Meyer

September, 2015

ABSTRACT

In silico and molecular validation of identified putative genes and functional analysis of a NKG2D ligand as a breast cancer biomarker

H.A. Bankole

PhD, Thesis, Department of Biotechnology, University of the Western Cape.

The current diagnostic, prognostic, predictive and therapeutic monitoring methods used for breast cancer are limited. Thus, research into more specific, sensitive and effective strategies is required. Breast cancer is the most prevalent form of cancer in women worldwide and accounts for the most common cause of death in women every year. Cancer development is characterized by a wide spread of genetic abnormalities of gene sequences that can be used in detecting and monitoring treatment of the disease as a result of altered gene expression patterns which leave a trail of biomarkers. Seven candidate genes (Gene 1-7) were identified from a previous *in silico* study and their gene products (BRG 1-7) were annotated to be good candidate breast cancer biomarkers. Differential gene expression analysis using quantitative real-time PCR (qRT-PCR) validated the over-expression of Gene 3, Gene 4 and Gene 7 in a breast cancer cell line (MCF7), of which Gene 7, annotated as a Natural killer group 2, member D (NKG2D) ligand, was observed to be the most over-expressed gene.

The innate immune system is the first line of the body's physiological defense against diseases and the natural killer (NK) cells, are central to mediating this type of immunity. NK cells are activated when a specific surface receptor such as the NKG2D receptor binds its ligands expressed by tumor cells. To evade being detected by the immune system, cancer cells are reported to shed off the NKG2D ligands and are

expected to be present in the bodily fluids of cancer patients. Also, chemotherapeutics have been reported to suppress the natural anti-tumour immune response, thus should be taken into account when designing optimal therapy for cancer patients. The aim of this research was to validate these candidate genes as effective breast cancer biomarkers using several *in silico* methods as well as molecular techniques and study the effect of Gene 7 on modulating the effect of several pro-apoptotic compounds.

The *in silico* part of the study investigated the functional, protein interaction, pathways, and tissue expression specificity of the candidate biomarkers using computational software such as DAVID, STRING, KEGG, Genecards and GEA. Also an *in silico* validation of the prognostic/predictive values of the genes was analysed using SurvExpress, KMplot, and GOBO. Protein expression of selected genes was analysed by Western blot, and immunofluorescence analysis. BRG 7 gene was cloned into pcDNA3.1 vector using recombinant DNA technology while commercial shRNA construct was used to "knock-down" Gene 7 expression. The two constructs were used to transfect MCF-7 and MCF-12A cells. Over-expression and "knock down" of BRG 7 in transfected cells was confirmed using western blot analysis. Stably transfected cells were then treated with three pro-apoptotic compounds (Camptothecin, Doxorubicin and DMSO) for 24 hours. The apoptotic cells were stained with 3, 4, 5, 6-tetrachloro-2', 4', 5', 7'-tetraiodofluorescein (TCTF) and then analysed using flow cytometry.

Functional analysis linked Gene 1, Gene 2, Gene 4, Gene 6 and Gene 7 to different cancer related processes. The pathway analysis showed Gene 1, Gene 2, Gene 4 and Gene 7 were involved in pathways that can be linked to cancer modulation. The protein-protein interaction analysis showed only BRG 2 was directly linked to two major hallmarks of cancer (Apoptosis and Autophagy). Breast cancer associated

Transcription factors were shown to regulate these genes. Gene 1 and Gene 5 as well as the three genes observed to be highly expressed in the qRT-PCR study were validated to differentially express in breast cancer. An additional protein (BRG 8) was identified and postulated to be a good biomarker candidate for breast cancer based on its direct interaction with BRG 7 and estrogen receptor protein (ESR).

The prognostic value of the candidate genes were monitored in two datasets (DATA1 and DATA2) in SurvExpress. DATA1 showed that Gene 6 and Gene 8 while DATA2 showed that Gene 3, Gene 6 and Gene 7 were valuable candidate genes in breast cancer prognosis. The survival curves from the two datasets showed the combined genes could predict the outcome of breast cancer patients undergoing treatments. A plot box output from SurvExpress showed most of the genes were differentially expressed comparing two risk groups. The Kaplan Meier plotter confirmed, Gene 1, Gene 3, Gene 4 and Gene 7 have a significant P-value in predicting the survival outcome based on gene differential expression value. GOBO analysis showed the genes may accurately predict the survival outcome of estrogen positive subtype, ERBB2 subtype of estrogen receptor negative and lymph node negative subtype of ER- tumours, but not all subtype of ER- tumours.

Western blot analysis showed BRG 7 may be highly expressed in MCF-7 as compared to MCF-12A, BRG 8 was found to be expressed in all cancer cell types analyzed except for MCF-7 and HT29. BRG 2 was found to be expressed in all cancer types analyzed. Immunofluorescence analysis showed BRG 3, BRG 4 and BRG 7 are differentially expressed in breast cancer cell line and are more localized on the cell membrane when compared to the breast non-cancer cell line.

Over-expression and gene knock down in cells were successfully confirmed with Western blot analysis. Stably transfected MCF-12A cell for over-expression of BRG7 protein, resulted in cell senescent and the cell stopped growing while stably transfected MCF-7 over-expressing BRG7 did not show any morphological changes. Apoptosis was enhanced in cells treated with camptothecin, doxorubicin and DMSO overexpressing BRG7. Apoptosis was reduced in camptothecin and DMSO treated gene “knock-down” cells but not doxorubicin treated. BRG7 gene “knock-down” in transfected cells showed varying response to all three pro-apoptotic compounds.

From this study Gene 3, 5, 7 and 8 and their protein levels were confirmed to be differentially expressed in breast cancer cells and could serve as putative biomarkers for breast cancer. However the variance in the effectiveness of individual genes suggests that the set of genes would perform better than individual gene. The modulating role of BRG7 in drug induced apoptosis, suggest it could probably play an important role in personalised medicine and could serve as a biomarker to monitor the prognosis and/or therapeutic outcome of pro-apoptotic drugs in breast cancer patients. These findings will be further investigated in human breast tissues to validate these data.

Keywords:

Apoptosis

Biomarkers

Breast cancer

Differential expression

In silico

Knock down

Natural Killer cells

NKG2D ligand

Validation

Over-expression

DECLARATION

I declare that "*In silico* and molecular validation of identified putative genes and functional analysis of a NKG2D ligand as a breast cancer biomarker" is my own work that has not been submitted for any degree or examination in any other university and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Habeeb Adebodun Bankole September 2015

Signed: ……..

Table of Contents

List of figures

List of tables

ABBREVIATIONS

1D-SDS	I-dimensional Sodium Dodecyl Sulphate
2DG	2-deoxy-d-glucose
ACS	American Cancer Society
AML	Acute myeloid leukemia
ANOVA	Analysis of variance
AP	Activator Protein
APAF-1	Apoptotic Protease-activating Factor 1
APS	Ammonium persulphate
ASCO	American Society of Clinical Oncology
ATCC	American Type Culture Collection
ATF	Activation transcription factor
ATM	Ataxia telangiectasia mutated
ATP	Adenosine Triphosphate
BACH1	BTB and CNC homology 1
BAD	Bcl-2 homology domain 3
BAK	Bcl-2-associated agonist of cell death
BAX	Bcl-2-associated protein X
Bcl-2	B-cell lymphoma 2
Bcl-2	B-cell lymphoma 2
BECN1	Beclin-1
BH3	Bcl-2 antagonist/killer
BID	BH3-interacting domain
BIM	Bcl-2-interacting mediator
BLAST	Basic Local Alignment Search Tool

BLASTn	Basic Local Alignment Search Tool (Nucleotide database)
BLASTp	Basic Local Alignment Search Tool (Protein database)
BMP-5	Bone Morphogenetic Protein 5
BRCA1	Breast Cancer 1
BRCA2	Breast Cancer 2
BRIP	BRCA1-interacting protein 1
BSA	Bovine Serum Albumin
CA 15-3	Cancer antigen 15-3
CA 27-29	Cancer antigen 27-29
CA-125	Cancer antigen 125
CAI	Codon Adaptation Indices
CBB	Coomassie Brilliant Blue
CBE	Clinical Breast Examination
cDNA	Complementary DNA
CEA	Carcinoembryonic antigen
CG	Co-expressed gene
CGAP	Cancer Genome Anatomy Project
CHEK2	Checkpoint kinase 2
CoA	Coenzyme A
CPT	Camptothecin
CREB	cAMP response element-binding
CTL	CD8+ cytotoxic T cells
CTL	Cytotoxic T lymphocytes
CTL	Cytotoxic T-lymphocyte
CUTL1	Cut-Like Homeobox 1
DAPI	4', 6-diamidino-2-phenylindole

DAVID	Database for Annotation, Visualization, and Integrated Discovery
DCIS	Ductal carcinoma *in-situ*
DMEM	Dulbecco's Modified Eagle's Medium
DMFS	Distant Metastasis-free Survival
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DOX	Doxorubicin
DTT	Dithiothreitol
ECM	Extracellular Matrix
ECM	Extracellular matrix
EDTA	Ethylene Diamine Tetra-acetic acid
EGF	Epidermal growth factor
ELISA	Enzyme-linked Immunosorbent Assay
ENSEMBL	European Bioinformatics Institute and the Wellcome Trust Sanger Institute
ER	Estrogen Receptor
ERp5	Endoplasmic reticulum protein-5
ESR	Estrogen receptor
EST	Expressed Sequence Tag
FasL	Fas Ligand
FBS	Fetal Bovine Serum
FFPE	Formalin-fixed, Paraffin-embedded
FGF	Fibroblast Growth Factor
FISH	Fluorescence *in situ* hybridization
G418	Geneticin
GEA	Gene Expression Atlas
GEO	Gene Expression Omnibus

GF	Growth factor
GO	Gene Ontology
GOBO	Gene Expression-Based Outcome for Breast Cancer Online
GOC	Gene Ontology Consortium
GS	Growth signal
GSA	Gene Set Analysis
HCG	Human Chorionic Gonadotropin
HER2	Human Epidermal growth factor Receptor 2
HER2	Human epidermal growth factor receptor 2
HGCN	HUGO Gene Nomenclature committee
HRG	Histidine-rich glycoprotein
HTLV-1	Human T-lymphotropic virus type 1
HTP	High throughput
HUGO	Human Genome Organization
IBC	Inflammatory breast cancer
IC_{50}	Half maximal inhibitory concentration
ICAM-1	Intercellular Adhesion Molecule 1
IDC	Invasive Ductal Carcinoma
IGS	Invasiveness gene signature
IL-2	Interleukin 2
ILC	Invasive Lobular Carcinoma
ITTACA	Integrated Tumour Transcriptome Array and Clinical data Analysis
KDM	Lysine (K)-specific demethylase
KEGG	Kyoto Encyclopaedia of Genes and Genomes
KMPlot	Kaplan Meier-plotter
LB	Luria broth

MCL-1	Myeloid Cell Leukemia 1
MICA	MHC class I polypeptide-related sequence A
MICB	MHC class I polypeptide-related sequence B
miRNA	micro-RNA
MMP9	Matrix metallopeptidase 9
MOMP	Mitochondrial Outer Membrane Permeabilization
MRI	Magnet Resonance Imaging
mRNA	Messenger RNA
MS	Mass spectrometry
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
MUC-1	Mucin 1
NaCl	Sodium Chloride
NaOH	Sodium Hydroxide
NBS1	Nijmegen breakage syndrome 1
NCBI	National Centre for Biotechnology Information
NK	Natural Killer
NKG2D	Natural-killer group 2, member D
NKG2DL	Natural-killer group 2, member D Ligand
NKT	Natural Killer T-cell
PAGE	Polyacrylamide Gel Electrophoresis
PALB2	Partner and localizer of BRCA2
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDGF	Platelet-derived growth factor
PR	Progesterone receptor
PSA	Prostate Cancer Antigen

PTEN	Phosphatase and Tensin homolog
qRT-PCR	Quantitative Real-time polymerase chain reaction
RAD50	RAD50 Homolog
RAE-1	Retinoic acid early inducible-1
RB	Retinoblastoma
rDNase	recombinant DNase
RNA	Ribonucleic Acid
RNAi	RNA interference
RNA-Seq	RNA Sequencing
ROS	Reaction Oxygen Species
RSV	Respiratory Syncytial Virus
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SBE	Self-Breast Examination
SDS	Sodium Dodecyl Sulphate
SHR	Steroid Hormone Receptor
shRNA	Short hairpin RNA
SIBLING	Small Integrin Binding Ligand *N*-linked Glycoprotein
siRNA	Small interfering RNA
SMAC	Second mitochondria-derived activator of caspases
SOD1	Superoxide dismutase 1
SOM	Self-organized maps
SP	Sample prediction
STRING	Search Tool for the Retrieval of Interacting Genes
TAM	Tumour-associated macrophage
TBE	Tris-borate-EDTA
TBS-Tween	Tris-Buffered Saline and Tween-20

TCTF	3, 4, 5, 6-tetrachloro-2', 4', 5', 7'-tetraiodofluorescein
TE	Tris-EDTA
TEMED	N, N, N', N'-Tetramethylethylenediamine
TF	Transcription Factor
TGFα	Tumour growth factor α
TGFβ	Transforming growth factor beta
TNF	Tumour Necrosis Factor
TNM	Tumour, Nodes, and Metastasis
topo I	Topoisomerase I
topo II	Topoisomerase II
TP53	Tumour protein p53
TSP1	Thrombospondin-1
ULBP1-6	UL16-binding proteins 1–6
UniProt	Universal Protein resource
UniProtKB	UniProt Knowledgebase
UV	Ultraviolet
VEGF	Vascular Endothelial Growth Factor
VEGFA	Vascular Endothelial Growth Factor-A
VEGFR	Vascular Endothelial Growth Factor Receptor
WHO	World Health Organization
XIAP	X-linked inhibitor of apoptosis

ACKNOWLEDGEMENT

I will like to thank the DST/Mintek Nanotechnology Innovation Center and National Research Foundation (NRF) for funding this project and scholarship. Dr. Ashley Pretorius and Prof. Mervin Meyer, for their intellectual inputs and guidance. To all members of the Bioinformatics Research Group in particular and all members of the NIC family in general, I say thank you. Many thanks to families, friends and colleagues both in Nigeria and South Africa for their prayers and several words of encouragements.

I will also like to thank my inlaws, parents, sisters and brother for their support and understanding, I apologise for all the inconveniences my going away caused. I will like to specially acknowledge you mum, for the prayers and words of encouragements. Special thank you to Rashydah (my wife) for her understanding, perseverance and support. I can not thank you all enough, but pray Allah reward you abundantly.

All praises and honour is due to Allah, the Lord of the universe who took me through this path, blessed me with caring and supportive family, help me overcome all challenges and make this dream come through.

DEDICATION

I dedicate this work to you my wife and kids, Rashydah, Haleemah and Aneesah, for your sacrifice

Chapter 1: Literature Review

1.1 Cancer

Cancer is a disease that occurs as a result of a series of molecular events that fundamentally alter the normal properties of cells. It develops due to extensive genetic abnormalities and aberrations in gene expression patterns (Allinen *et al.*, 2004). In cancer, there is a loss of balance between cell division and cell death and cells that should have died, did not receive the signals to do so (Wong, 2011). Cancer has become the leading cause of death worldwide, according to the World Health Organization (WHO), it is projected that death resulting from cancer will rise above 13.1 million by 2030 (Mathers and Bonita, 2009). According to the Cancer Association of South Africa's estimate, there are one in six male and one in seven female cancer patients in South Africa (Jemal *et al.*, 2011).

Two major factors that contribute to the development or progression of cancers are either environmental (external) or hereditary (internal) (Czene *et al.*, 2002). External factors includes tobacco, chemicals, radiation, infectious organisms etc., while internal factors that play a role include hormones, inherited mutations, immune conditions and mutations that result from metabolism (Amir *et al*, 2010). The combination of both these factors gives rise to genetic abnormalities, which can occur throughout the person's lifetime (Amir *et al*, 2010). The abnormalities in cancer cells usually result from mutations in protein-encoding genes that regulate cell division, and over time more genes could become mutated.

1.2 Genetics of cancer

The etiology of cancer is multifactorial, with both genetic and environmental factors interacting to produce a given malignancy. However, knowledge of cancer genetics is

rapidly improving the understanding of cancer biology, helping to identify individuals at risk, improving the ability to characterize malignancies, establishing treatment tailored to the molecular fingerprint of the disease, helping to identify suitable biomarkers for easy diagnosis, prognosis and also leading to the development of new therapeutic modalities. This expanding knowledge base thus has implications for all aspects of cancer management, including prevention, screening, and treatment (Lindor *et al.*, 2008).

Cancer is generally recognized as a multistage disease involving accumulation of a critical number of mutations within a stem cell (Loeb and Loeb, 2000). Alterations in a gene can be associated with cancer in different ways. These malfunctioning genes can be broadly classified into three groups. The first group, called proto-oncogenes; produces protein products that normally enhance cell division or inhibit normal cell death. The mutated forms of these genes are called oncogenes. The second group, called tumour suppressors, makes proteins that normally prevents cell division or cause cell death. The third group is genes involve in DNA repair, which help prevent mutations that lead to cancer. The discovery of oncogenes, tumour suppressor, DNA repair and cancer susceptibility genes has led to the concept that carcinogenesis is a pure endogenous genetic process (Hahn and Weinberg, 2002; Hoyer *et al.*, 2002), but it's been shown that most times these genes are modulated by various environmental factors (Mucci *et al.*, 2001).

Human cancers occur as a consequence of genetic alterations that promote uncontrolled cell proliferation. Mutations in oncogenes and tumour suppressor genes influence various cellular processes, such as signal transduction and gene expression. In normal cells, proto-oncogenes code for the proteins that send a signal to the nucleus to stimulate cell division. These signaling proteins act in a series of steps called signal

transduction cascades or pathways. An oncogene is a modified proto-oncogene that promotes unregulated proliferation of cells, increasing the chance that a normal cell develops into a tumour cell, possibly resulting in cancer (Rigel *et al.*, 2011).

As the pro-oncogene stimulates cell division, the tumour suppressor genes are meant to keep the proliferation of cells in check to bring about homeostasis. Tumour suppressor genes are normal genes that slow down cell division, repair DNA mistakes, or instruct cells when to die (a process known as Apoptosis or Programmed Cell Death). When tumour suppressor genes functions are inadequate, cells can grow out of control, which can lead to the onset and progression of cancer (López-Otín, and Matrisian, 2007).

Tumour suppressor proteins retinoblastoma (RB protein) and TP53 are two well-known tumour suppressor gene products that play a central role in governing the decision for cells to proliferate or alternatively, activate senescence (Hanahan and Weinberg, 2011). Defects in these genes enable cells to proliferate uncontrollably rendering the cells insensitive to growth signals and resistant to apoptosis. Familial cancers are believed to be due to the inheritance of a mutated tumour suppressor gene and that a second mutation event, involving the wild type allele of that gene, leads to malignancy (Knudson, 1985). In addition to these genetic alterations as a result of mutation, changes in DNA methylation, an epigenetic modification that occur in mammalian cells, is another hallmark of human cancer (Herman, 1999; Sharma *et al.*, 2010).

1.3 Cell signaling and cancer

Cells are known to respond to specific chemicals that signal them to divide. While there are countless pathways utilized by cells, they all display the same fundamental

characteristics. All "signal pathways" involve a chemical messenger, a protein receptor, and a cellular response (Hunter, 2000).

Cellular signaling pathways are interconnected to form complex signaling networks. Cells receive information from many different growth factor receptors and from cell-matrix and cell-cell contacts (Figure 1.1). They then integrate this information to regulate diverse processes, such as protein synthesis, cell growth, motility, cell architecture and polarity, differentiation, and programmed cell death (Martin, 2003). Cell proliferation, motility, and survival are regulated by multiple pathways, and the changes that occur in cancer cells are the result of multiple alterations in the cellular signaling machinery (Greenhough *et al.*, 2009).

Cancer cells are genetically unstable, undergo multiple genetic and epigenetic changes, and continuously evolve in response to selective pressures. They even evade inhibition of mutational activated pathways by activating alternate pathways to block the action of the inhibitor, though early stage malignancy may respond to a single inhibitor (Martin, 2003). The complex architecture of signaling networks and the consequences of this complexity are now been looked at for cancer diagnostics and therapeutics.

Normal cells require mitogenic growth signals (GS) before they can move from a quiescent state into an active proliferative state. These signals are transmitted into the cell by transmembrane receptors that bind distinctive signaling molecules: diffusible growth factors, extracellular matrix components, and cell-to-cell to extracellular superstructures adhesion/interaction molecules. Many of the oncogenes in the cancer catalog, act by mimicking normal growth signaling in one way or another. Three common molecular strategies tumour cells use to achieve GS autonomy involves

alteration of extracellular growth signals, alteration of trans-cellular transducers of those signals, or the alteration of intracellular circuits that translate those signals into action (Hanahan and Weinberg, 2000).

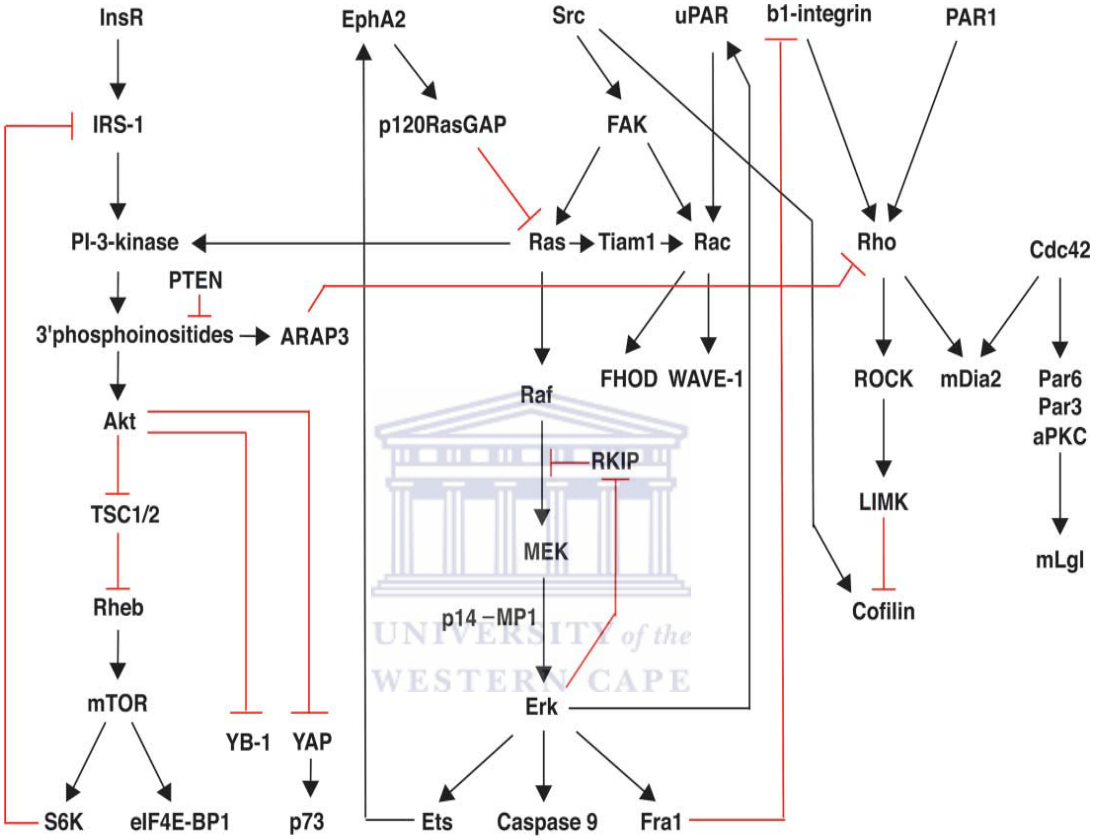


Figure 1.1: Cellular signaling pathways involved in cancer (Pattabiraman and Weinberg, 2014)

While most soluble mitogenic growth factors (GFs) are made by one cell type in order to stimulate proliferation of another, a process known as heterotypic signaling, many cancer cells acquire the ability to synthesize GFs to which they are responsive, creating a positive feedback signaling loop often termed autocrine stimulation (Fedi *et al.*, 1997; Avraham *et al.*, 2011). Clearly, the manufacture of a GF by a cancer cell obviates dependence on GFs from other cells within the tissue. The production of

PDGF (platelet-derived growth factor) and tumour growth factor α (TGFα) by glioblastomas and sarcomas, respectively, are two illustrative examples (Fedi *et al.*, 1997; Hanahan and Weinberg, 2000).

1.4 Hallmarks of cancer

Research over the years has revealed common molecular, biochemical and cellular traits shared by most and perhaps all cancer types (Hanahan and Weinberg, 2000). Tumour cells are more than clustered masses of proliferated cells; they are complex tissues interacting with one another (Hanahan and Weinberg, 2011). The progression of normal cells to a neoplastic state is due to certain capabilities that are unique to cancer cells as compared to normal cells.

Advances in cancer research have generated a rich and complex body of knowledge, suggesting that cancer is a disease involving dynamic changes in the genome. Different cancer phenotypes have been shown to be as a result of mutation(s), which generates oncogenes with dominant gain of function and tumour suppressor genes with recessive loss of function (Bishop and Weinberg, 1996). Hanahan and Weinberg (2011) suggested that there are six hallmarks linked to all cancer types. These are: (1) cancer cells being self-sufficient, (2) insensitivity to anti-growth signals, (3) sustained angiogenesis, (4) tissue invasion and metastasis, (5) limitless replicative capacity, and (6) evading apoptosis (Figure 1.2). Understanding these hallmarks has brought perspective into how this unified theory can be used to identify cancer biomarkers and provide possible ways by which the disease can be managed (Abbott *et al*, 2006).

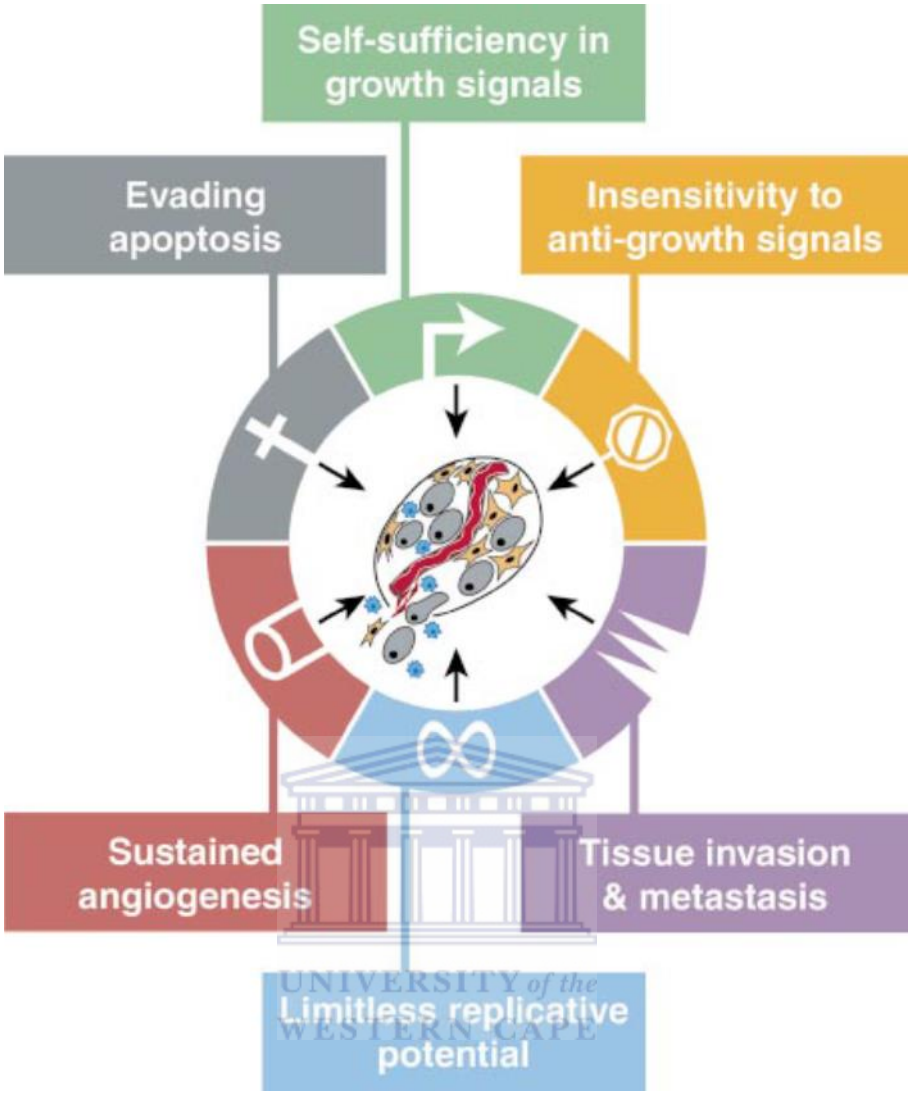


Figure 1.2: Six hallmarks associated with all cancer types (Hanahan and Weinberg, 2011)

1.4.1 Self-sufficiency in growth signalling

Normal tissues are known to carefully manage and control the release of growth signals through the cell cycle. During this cycle, cells are programmed to undergo self-death (apoptosis) at a later stage of the cycle. Also if there are defects in the genome of the cell, a signal will be sent for apoptosis to occur. Likewise, signals are send for cell repair or growth, through signalling pathways in the cell cycle. However,

cancer cells facilitate their growth by either producing growth factor ligands themselves to which they can respond to by the expression of cognate receptors, resulting in autocrine proliferative stimulation or by sending signals to manipulate normal cells into supplying them with growth signals (Cheng *et al.*, 2008; Bhowmick *et al.*, 2004). Also they can deregulate receptor signaling by elevating the levels of receptor proteins displayed on the cancer cell surface, making the cell hyper-responsive to otherwise limited growth factor signals (Hanahan and Weinberg 2011).

1.4.2 Evading growth suppressors

Cancer cells in addition to sustaining growth by activating positive growth signals, also circumvent programs that negatively regulate cell proliferation which most times depend on the actions of tumour suppressor genes. The activities of these tumour suppressor genes in limiting cell growth and proliferation, was discovered through their inactivation in one or another form of cancer by loss or gain of function. Two typical genes of the tumour suppressor genes encode the retinoblastoma (RB) and the TP53 proteins. These proteins operate as central control nodes within two key complementary cellular regulatory circuits that govern the decisions of cells to proliferate or, alternatively, activate senescence and apoptotic programs (Hanahan and Weinberg, 2011).

The RB protein receives different signals from both extracellular and intracellular sources, integrates them and decides if a cell should proceed through the growth cycle (Burkhart and Sage, 2008; Deshpande *et al.*, 2005; Sherr and McCormick, 2002). Cells with defects in the RB pathway, misses the services of a critical gatekeeper of cell-cycle progression and whose absence permits persistent cell proliferation. The RB

protein transduces growth-inhibitory signals that originate largely outside of the cell (Hanahan and Weinberg, 2011).

The tumour suppressor gene *TP53* on the other hand encodes a transcription factor (protein) which is activated in response to several forms of cellular stress and exerts multiple, anti-proliferative functions (Vogelstein *et al.*, 2000). Somatic *TP53* gene alterations are frequent in most human cancers and germline *TP53* mutations are linked to a wide spectrum of early-onset cancers (Hainaut and Hollstein, 2000). Although these two suppressor genes have prominence in regulating cell proliferation, various lines of evidence have shown that they operate in a network that may bring about functional redundancy. For example, clusters of RB null cells in rats, expected to cause neoplasms was found to participate in relatively normal tissue morphogenesis throughout the body (Lipinski and Jacks, 1999). Likewise, TP53 null mice that developed leukemia and sarcomas later in life, largely showed proper cell and tissue homeostasis, and developed normally for the major part of their lives (Ghebranious and Donehower, 1998; Vuong *et al.*, 2012).

1.4.3 Sustained angiogenesis

In order to sustain growth, tumour cells require means of getting nutrients like oxygen as well as possess the ability to take out metabolic waste and carbon dioxide. Normal cells get oxygen and nutrients supplies by vasculature, which necessitates virtually all cells in a tissue to reside within 100 μm of a capillary blood vessel (Hanahan and Weinberg, 2000). Malignant cells form functional micro-vascular networks with red blood cell perfusions (Neovasculature) through the process of angiogenesis to address their nutrient supply and metabolic waste excretion needs. Development of vasculature normally occurs during embryogenesis (Hanahan and Folkman, 1996).

Embryogenesis involves the birth of new endothelial cells and their assembly into tubes (vasculogenesis) as well as sprouting (angiogenesis) of new vessels from existing ones. Following this, the normal vasculature becomes dormant until it is required in adulthood for normal physiologic processes such as wound healing and female reproductive cycling, where angiogenesis is transiently turn on. In contrast, during tumour progression, an angiogenic process is always activated and remains on, allowing normal vasculature to continually sprout new vessels that help sustain expanding neoplastic growths (Hanahan and Weinberg, 2011).

This angiogenic switch is control by regulators, some of which are signaling proteins that bind to stimulatory or inhibitory cell surface receptors displayed by vascular endothelial cells. Two examples of these angiogenesis inducers and inhibitors are vascular endothelial growth factor-A (VEGFA) and thrombospondin-1 (TSP1), respectively (Hanahan and Weinberg, 2011).

The VEGFA gene encodes ligands that stimulate growth of new blood vessels during embryogenic and postnatal development, for maintaining homeostasis for the survival of endothelial cells, as well as being required in maintaining normal physiological and pathological situation in adults (Ferrara, 2010). The VEGFA gene expression is regulated at multiple levels through three tyrosine kinase receptors (VEGFR-1-3), but the up-regulation of the VEGF can be as a result of both hypoxia and oncogene signaling (Ferrara, 2010; Mac Gabhann and Popel, 2008; Carmeliet, 2005). VEGF ligands can be released into the extracellular matrix in an inactive form and later activated by extracellular matrix-degrading proteases such as the matrix metalloproteinase-9 (MMP9) (Kessenbrock *et al.*, 2010). Some pro-angiogenic compounds such as the fibroblast growth factor (FGF) family members have also been shown to be up-regulated in the sustenance of tumour angiogenesis (Baeriswyl and

Christofori, 2009). The anti-angiogenic signal TSP1, in maintaining homeostasis of the angiogenic switch, also binds trans-membrane receptors on the endothelial cells, to evoke suppressive signals that counteract pro-angiogenic stimuli (Kazerounian *et al.*, 2008).

1.4.4 Tissue invasion and metastasis

As most human cancer types develops, primary tumour tissue produce cells that have the ability to invade adjacent tissues and hence may travel to a distant site where they may succeed in forming new colonies. This spread is referred to as metastasis and the new cancer formed is called metastases, which are reported to be the cause of about 90 % of all human cancer deaths (Sporn, 1996; Chaffer and Weinberg, 2011). The ability of cancer cells to leave the primary tumour tissue and invade other tissues to form metastases, provides these cells at least an initial unlimited amount of nutrients and space (Chaffer and Weinberg, 2011).

Cancer cells associated with carcinomas arising from epithelial tissues are altered in shape as well as in their attachment to other cells and the extracellular matrix (ECM) as the carcinomas progresses to higher pathological grades of malignancy (Thompson and Newgreen, 2005).

A well-known alteration is the loss of a key cell-to-cell adhesion protein known as E-cadherin. This protein normally helps to assemble epithelial cell sheets by forming adherence junctions with adjacent epithelial cells and keeps the cells in an inactive state within the sheet (Park *et al.*, 2008). Alteration in the expression of E-cadherin is well established as a modulator of cancer metastasis. The increase expression served as an antagonist of invasion and metastasis. Thus, the down regulation of the gene or its inactivation as a result of mutation, as found in human carcinomas, is a strong

indicator that E-cadherin serve as a key suppressor of metastasis (Berx and van Roy, 2009; Cavallaro and Christofori, 2004).

Alterations in the expression of genes encoding other cell-to-cell or cell-to-ECM molecules have also been shown to be involved in highly aggressive carcinomas (Cavallaro and Christofori, 2004). The process of invasion and metastasis begins with local invasion, then intravasation by the cancer cells into nearby blood and lymphatic vessels. The cancer cells then transit through the lymphatic and blood systems, followed by the release of cancer cells from the lumina of the vessels into the parenchyma of distant tissues (extravasation). The cancer cells then form small nodules (micrometastases), which then finally grows to macroscopic tumours (Talmadge and Fidler, 2010; Fidler, 2003).

1.4.5 Limitless replicative potential

All mammalian cells appear to possess an intrinsic autonomous program independent of the cell-to-cell signaling pathways that limit their multiplication. Unlike normal mammalian cells, which are known to have finite growth-and-division cycles, cancer cells acquire unlimited replicative ability to form macroscopic tumours (Hanahan and Weinberg, 2000). As normal cells replicates, it either gets to a stage of senescence, which is an irreversible, but non-proliferative state or experiences cell death as a result of some “crisis” (Hanahan and Weinberg, 2011; Hanahan and Weinberg, 2000).

The telomeres, which are regions of repetitive nucleotide sequences of a chromosome that protect the chromosome from deterioration, are highly implicated in infinite growth of tumour cells (Blasco, 2005; Shay and Wright, 2000). Thus, the length of the telomeric DNA will determine the number of successive cell generations. The telomerase; a DNA polymerase that adds telomere to the end of telomeric DNA, is

reported to be highly expressed in the majority of immortalized cells, including cancer cells but found to be almost absent in non-immortalized cells (Hanahan and Weinberg, 2011).

The telomerase helps to counter progressive erosion of telomeric DNA by extending telomeric DNA and its active presence (either in spontaneous immortalized cells or cells engineered to express the protein), is correlated with resistance to induction of both senescence and apoptosis. Whereas the suppression of telomerase activity leads to the shortening of the telomere and eventually the activation of either senescence or cell death (Shay *et al.*, 2001).

1.4.6 Evading apoptosis

The programmed cell death (apoptosis) mechanism is triggered in response to various physiologic stresses, most of which cancer cells experiences from the effect of the body's immune defenses, as a result of anticancer therapy or during the course of tumourigenesis itself (Adams and Cory, 2007; Lowe *et al.*, 2004: Evan and Littlewood, 1998). There is profound evidence from both animal studies and cultured cell lines as well as from descriptive analysis of different stages of human carcinogenesis, which showed that cancer cells acquired various means of resisting apoptosis (Fernald and Kurokawa, 2013). For example, genetic inactivation of the Bcl-2 antagonist/killer (BH3) protein or a caspase in mouse genetic models does not only inhibit pro-apoptotic stimuli but also stimulate tumour formation in mice (Hubner *et al.*, 2008; Bean *et al.*, 2013; Parsons *et al.*, 2013). Likewise, other studies showed that disruption of the apoptotic pathways were present in tumours that succeeded in processing high-grade malignancy and those that resist therapy (Adams and Cory, 2007; Lowe *et al.*, 2004).

1.5 Apoptosis and cancer

Organisms have naturally developed a cellular suicidal program to eliminate unhealthy or unnecessary cells from the body in their course of development or following cellular stress. Apoptosis involves a series of cellular events that culminate in the activation of caspases (cysteine proteases) that are generally grouped as 'initiator' caspases (caspase-2, 8, 9, and 10) and the 'executioner' caspases (caspase-3 and 7) (Denault, and Salvesen, 2002).

The initiator caspases are activated by the apoptotic stimuli and in turn they activate the zymogenic 'executioner' caspases, which subsequently cleaves some specific cellular substrates, and the effect leads to cell death (Fernald and Kurokawa, 2013). Thus the apoptotic machinery is said to be composed of both the upstream regulators and the downstream effector components (Adams and Cory, 2007).

The regulators are further divided into those receiving and processing extracellular death signals known as the extrinsic apoptotic program. Their actions culminate in the activation of caspase-8 and 10 through the formation of the death-inducing signal complex (DISC) when extracellular ligands such as Fas, tumour necrosis factor (TNF) bind their cell surface receptors (Fulda and Debatin, 2006). The other regulator is responsible for sensing and integrating intracellular death signals and eventually activates the intrinsic apoptotic program through the mitochondrial outer membrane permeabilization (MOMP) mechanism (Hanahan and Weinberg, 2011).

MOMP triggers the release of pro-apoptotic proteins, including Cytochrome c and second mitochondria-derived activator of caspases (SMAC), into the cytoplasm from the mitochondrial inter-membrane space. Cytochrome c binds to the apoptotic protease-activating factor 1 (APAF-1) to form an apoptosome; a caspase-9-activating

complex, which then initiates a cascade of proteolysis by activating the effector caspases responsible for the execution phase of apoptosis by disassembling the cancer, cells which are then consumed by phagocytic cells (Tait and Green, 2010). SMAC augments Cytochrome c-induced caspase activation by binding and neutralizing X-linked inhibitor of apoptosis protein (XIAP), an inhibitor of caspase-3, -7, and -9 (Hanahan and Weinberg, 2011; Fernald and Kurokawa, 2013).

1.5.1 Inhibition of Apoptosis in cancer

Inhibition of apoptosis is strongly believed to be critical to cancer cells survival as well as tumour development. The B cell lymphoma 2 (Bcl-2) family proteins plays an important role in regulating apoptosis, serving as a pro- or anti-apoptotic signals, depending on the member of the protein family that is controlling MOMP (Tait, and Green, 2010).

The BH3-only proteins e.g., Bcl-2-interacting mediator of cell death (BIM), BH3-interacting domain death agonist (BID), and Bcl-2-associated agonist of cell death (BAD) and multi-BH domain proteins e.g., Bcl-2-associated protein X (BAX) and Bcl-2 antagonist/killer (BAK) promotes MOMP as shown in figure 1.3. Thus the Bcl-2 protein family are mostly pro-apoptotic (Youle, and Strasser, 2008). The BH3-only proteins initiate oligomerization of BAX (or BAK) following cytotoxic or genotoxic stress and this results in MOMP (Figure 1.3). However the pro-survival members of the protein family for example, the Bcl-2, B cell lymphoma extra-large (Bcl-xL), and induced myeloid leukemia cell differentiation protein (MCL-1), inhibits the activities of pro-apoptotic substances (Martinou and Youle, 2011; Stresser *et al.*, 2011).

As a result of several extrinsic factors including oncogenic factors, genomic instability, and cellular hypoxia, cancer cells are constantly under physiological stress,

which stimulate apoptosis, and in response to such stimuli, the intrinsic pathway of apoptosis would normally be activated. However, cancer cells will often avoid this cellular response by disabling the apoptotic pathways (Fernald and Kurokawa, 2013). Though forced expression of anti-apoptotic Bcl-2 family proteins can significantly aid tumour development induced by an oncogene, the over-expression of the anti-apoptotic protein alone may not result in tumour formation similar to that seen with an oncogene alone (Lowe and Lin 2000; Fernald and Kurokawa, 2013).

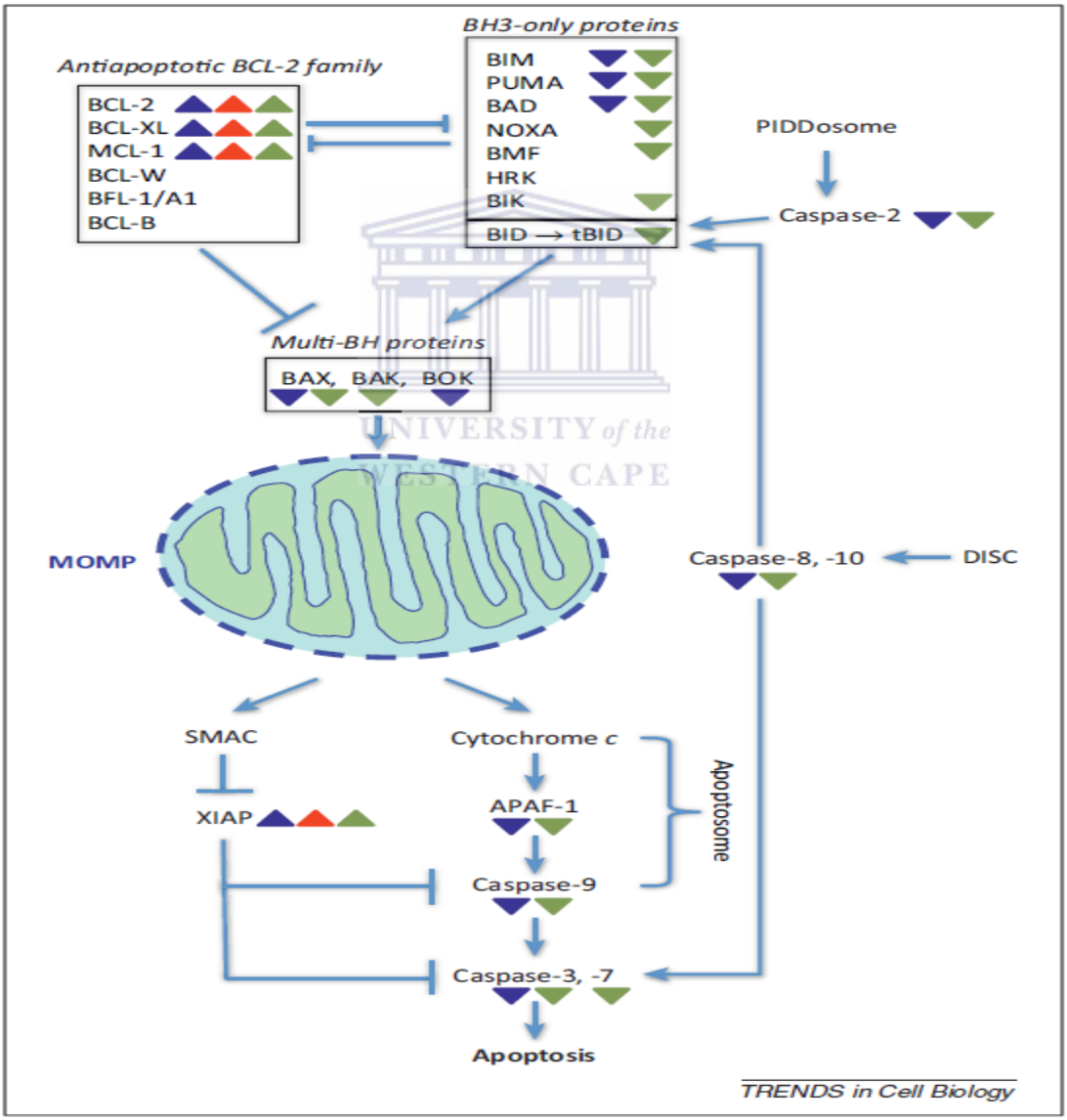


Figure 1.3: Apoptosis as it is been triggered in response to internal or external stimuli (Fernald and Kurokawa, 2013).

1.6 Cancer and the immune system

Another important mechanism by which cancer progresses, is by evading the immune surveillance machinery. An immune-surveillance hypothesis of an immunological response to tumours was first proposed as early as the 1950s with no evidence until the last two decades when tumour specific antigens and immune modulation leading to tumour regression were identified (Waldner *et al.*, 2006).

The long-standing theory of immune surveillance proposes that cells and tissues are constantly monitored by the immune system, and that such immune surveillance is responsible for recognizing and eliminating the vast majority of incipient cancer cells and thus nascent tumours (Hanahan and Weinberg, 2011). The activation of the host immune system by tumour cells is a complex cascade that involves both the innate and adaptive immune systems. Several research evidence have shown that both cell-mediated innate and adaptive immunity are essential in preventing primary tumour outgrowth and rejecting transplanted tissues (Dunn *et al.*, 2004; Pardoll, 2003).

Tumour-associated macrophages (TAMs) and T cells are the most commonly found cells within the tumour microenvironment. High levels of TAM have been correlated with poor cancer prognosis (Murdoch *et al.*, 2008), and it is known to promote angiogenesis, invasion and metastasis (Condeelis and Pollard, 2006). Mature T cells are majorly of two types based on the receptors they expressed. They are either gamma delta (gd/γδ) or alpha beta (ab/αβ) T cells, which are further classified based on their effector functions as CD8+ cytotoxic T cells (CTLs), and CD4+ helper T cells (Th) (which include Th1, Th2, Th17 and T regulatory (Treg) cells, as well as natural killer T (NK) cells). Based on their effector function, T cells can exert both tumour-suppressive as well as tumour-promoting effects (DeNardo *et al.*, 2009; Langowski *et al.*, 2007; Smyth *et al.*, 2006).

Actions of tumour promoting T cells are mediated by cytokines, whereas both cytokines and cytotoxic processes mediate the anti-tumourigenic effects of T lymphocytes. However, the NK cells are the only immune surveillance cells that lacks a pro-tumourigenic effect (Grivennikov *et al.*, 2010).

The CTL and NK cells recognize and kill virus-infected or transformed cells through three main pathways. CTLs use ligands of the tumour necrosis factor super family on their cell surface to bind and eliminate target cells expressing the corresponding receptors (Cullen *et al.*, 2010). The CTL/NK cells can also target cells using the granule exocytosis pathway by delivering cytotoxic granules to the surface of target cells. A third process is the perforin/Granzyme B pathway by which apoptosis is initiated in cells as shown in figure 1.4 (Fadeel and Orrenius, 2005; Lawen, 2003).

Membrane-disrupting toxins in the cytoplasm known as perforin as well as a family of proteins, which are structurally similar to the serine proteases, (the granzymes) with substrate specificity, are secreted by exocytosis to induce apoptosis on the target cells (Trapani and Smyth, 2002). Perforin, promotes granzyme delivery into the cytosol of target cells and these proteases (granzymes) cleave their cohort of substrates to promote rapid and efficient cell death (Cullen *et al.*, 2010). The major constituents of cytotoxic granules are perforin and granzymes. The granzymes are contained in the granules preventing them from damaging the host cells. When CTL/NK cell recognizes a tumour cell, cytotoxic granules move along microtubules and polarize at the plasma membrane, adjacent to the target cell, where they are then secreted into the immunological synapse between the two cells (Cullen and Martin, 2008).

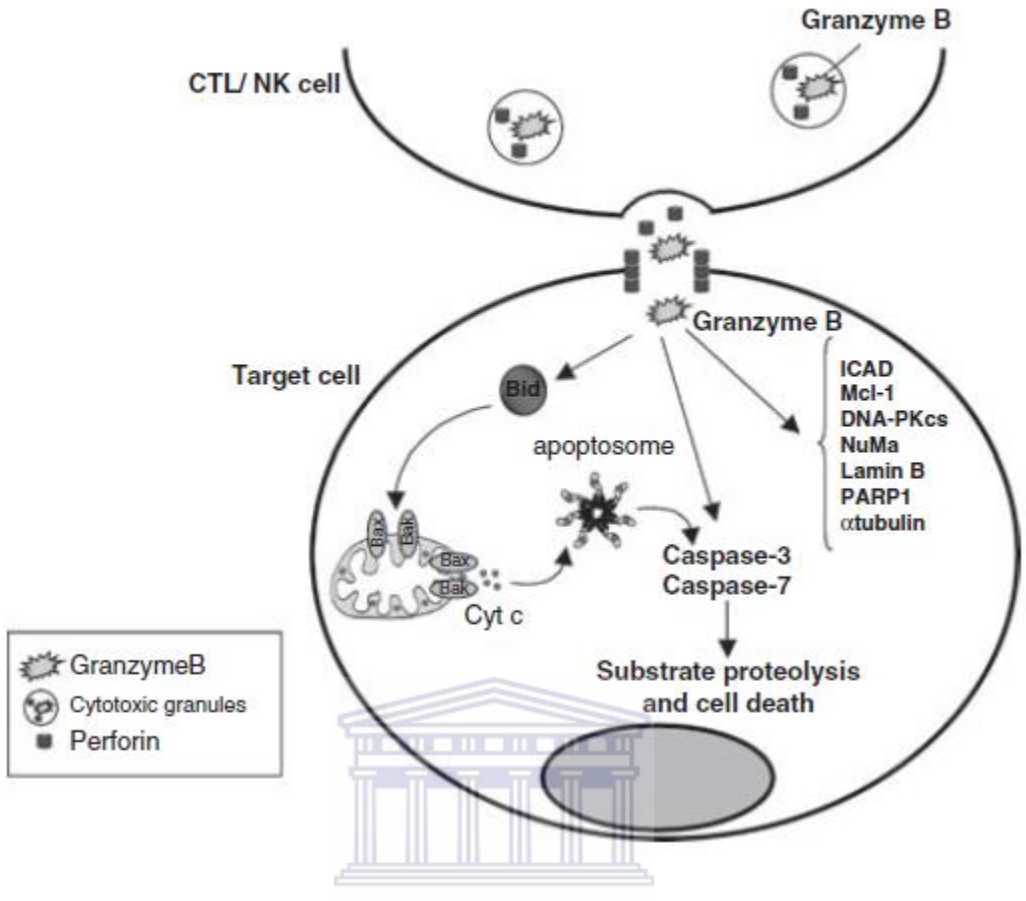


Figure 1.4: Pathways of granzyme B-mediated cell death (Cullen *et al.*, 2010).

Cancers are made up of different cell types, which include the fibroblasts and epithelial cells, as well as the innate and adaptive immune cells, and specialized mesenchymal cell types that are unique to each tissue microenvironment. Based on the site of cancer origin, there are different types and subtypes of cancer, named by the tissue they primarily originate from (Visvader, 2011). The main function of the mammalian immune system is to maintain tissue homeostasis, prevent invading or infectious pathogens and eliminate damaged cells (de Visser, *et al.*, 2006). The innate and adaptive immune systems are involved in preventing cancer originating from any tissue of the body in humans (Vesely, *et al.*, 2011). Expression patterns of lymphocytes (T cells, T regulatory cells, NK cells) and their cytokine have been implicated in preventing breast cancer and its recurrence (Standish, *et al.*, 2008)

1.7 Breast cancer

Breast cancer refers to a malignant tumour that has developed from cells in the breast. Usually the malignancy either begins in the cells of the milk-producing glands known as the lobules, or the ducts, which is the passages that drain milk from the lobules to the nipple (Kanagathara and kavitha, 2014). Breast cancer can also start in the stromal tissues, which include the fatty and fibrous connective tissues of the breast, although this is rare (Devadoss *et al.*, 2013).

The cancer cells can invade nearby breast tissue with time and make their way into the small organs that filter out foreign substances in the body known as the lymph nodes. The lymph nodes then serve as a pathway for these cancer cells into other parts of the body (Devadoss *et al.*, 2013). The metastasized cancer cells may attach to other tissues and grow to form new tumours of those tissues. For breast cancer, the most common sites for metastatic tumours are the bones, liver, lungs, and the brain (DuPrie, 2010). Breast cancer is the most frequent cancer among women with a global estimation of about 23 % of newly diagnosed cases in 2008 of all cancers, and also the most common cause of cancer deaths in women in both developing and developed countries (Kartal *et al.*, 2014).

Epidemiologic studies have shown a marked geographic variation in the breast cancer burden worldwide. Northern Europe, Australia, and New Zealand average 95–100 cases for 100,000 persons. Surprisingly, the incidence rate in Western Africa and Eastern Asia is approximately 20 per 100,000. Also there is a substantial variation in breast cancer incidence and mortality rates among regions within the continent of Africa. An estimated age-standardized incidence rate per 100,000 women are 20.2 in Eastern Africa, 13.5 in Middle Africa, 24.8 in Western Africa, and 31.8 in Southern Africa (Fregene and Newman, 2005; Ferlay *et al.*, 2010).

Significant problems exists for accurate estimation of breast cancer incidences in developing countries (Ferlay *et al.*, 2010). Challenges such as, countries not having cancer registries that can provide incidence rates covering intervals of 15 years and above or fail in obtaining accurate incidence rates (Fregene and Newman, 2005). Few countries have cancer registries that can provide accurate information whilst several of the most mature registries cover relatively small populations. Also, difficulty in estimating population denominators further creates an additional barrier to the calculation of incidence and mortality rates (Fregene and Newman, 2005).

South Africa is one of the few African countries that collates population-based cancer data, through the National Cancer Registry. However, it is suspected that the registry may underestimate actual cancer burden because it relies on information submitted by pathology laboratories and cases of a 1arge population group do not always get to the pathologist (Vorobiof *et al.*, 2001). A comparison between populations subsets in South Africa reveal progressively decreasing lifetime risk of breast cancer for white South Africans (1 in 13) versus mixed ethnicity South Africans (1 in 63) versus black South Africans (1in 81). However, it is believed these figures are not accurate (Vorobiof *et al.*, 2001).

1.7.1 Types of breast cancer

Normally the mammary glands are located at the level of the III to VII ribs separated by a deepening called the cavity (sinus mammarum) and surrounded by fatty tissue except for the nipples and the areola (Reece *et al.*, 2009). In average it is 10-12 cm in diameter when it is not lactating, 2-3 cm in thickness and weighs between 150-200 g (300-900 g during lactation) (Sahoo *et al.* 2015). Each segment of the mammary gland is divided into the lobules (lobuli mammae), separated from each other by connective

tissue and consisting of alveoli and the adipose tissue. The adipose tissue carries all lobular lymph nodes in space. The lymph nodes are small, bean-shaped collections of immune system cells that are connected by lymphatic vessels (Sahoo *et al.* 2015).

Based on the location of the malignancy, breast cancer is classified majorly into two forms, either ductal or lobular. Furthermore, the ductal or lobular carcinoma can further be classified based on pathological distinction as *in-situ* or invasive, indicating if the cancer remain within the ducts or lobules (localised) or if it has spread to surrounding tissues (metastasize) (Hammer *et al*, 2008). Ductal carcinoma *in-situ* of the breast is sometimes referred to as non-invasive or pre-invasive breast cancer because it might develop into an invasive breast cancer if left untreated. When cancer cells are confined to the lobules it is called lobular carcinoma *in-situ* (Allred *et al*., 2001).

The Invasive Ductal Carcinoma (IDC), which is the most common type of breast cancer, starts in the milk ducts of the breast, breaks through the wall of the duct, and grows into the fatty tissue of the breast. While the Invasive Lobular Carcinoma (ILC), starts from the milk producing glands (lobules). Like IDC, it can spread (metastasize) to other parts of the body too (Sahoo *et al.,* 2015).

DNA microarray expression profile analysis indicates five distinct and reproducible groups of invasive breast cancinomas which are: luminal A, luminal B, HER2-enriched, +/estrogen receptor (ER), basal-like, and normal breast-like (Rakha *et al*., 2008). The basal-like subtype has been shown to have the highest proliferation rates and poor clinical outcomes (Rakha *et al*., 2008). They are typically ER and HER2 receptor negative and are described in association with BRCA1 modulated carcinomas showing some characteristics of breast myoepithelial cells, but little is known with

regard to their development and prevention (Livasy *et al.*, 2006). Triple-negative breast cancer is used to describe breast cancers (usually invasive ductal carcinomas) whose cells lack estrogen receptors and progesterone receptors, and do not have an excess of the HER2 protein on their surfaces (Sahoo *et al.,* 2015).

The luminal (glandular) cells, basal/myoepithelial cells, and stem cells are the three types of epithelial cell component of a normal breast duct. The cytokeratin expression pattern of the epithelial cells distinguishes the myoepithelial cells from the luminal epithelia cells. Myoepithelial cells are known to express cytokeratin 5/6 and cytokeratin 17, while luminal epithelial cells expresses cytokeratins 8 and 18 (Böcker *et al.*, 2002). Immunohistochemical staining for these cytokeratins are helpful in subtyping invasive breast carcinomas into distinct biological subtypes, although the staining patterns for cytokeratins 5/6 and 17 is not absolute (Livasy *et al.*, 2006).

An adenocarcinoma is a type of cancer that starts from tissue that makes and secretes substances, like glandular tissue for example. The ducts and lobules of the breast are glandular tissues because they produce and secrete the breast milk, so cancers starting in these areas are often referred to as adenocarcinomas (Sahoo *et al.* 2015).

Inflammatory breast cancer (IBC) is an uncommon type of invasive breast cancer that is found in about 1 to 6 % of all breast cancers (Yang *et al.*, 2008). The affected breast may become larger or firmer, tender, or itchy. Usually this type of cancer does not present any lump or tumour and it might not be detected by a mammogram (Sahoo *et al.* 2015). This cancer makes the skin on the breast look red and feels warm. It may also make the breast skin appear thick and pitted in appearance (the breast looks a lot like an orange peel). These changes in appearance are as a result of cancer cells blocking lymph vessels of the skin. Inflammatory breast cancer is often mistaken for

an infection of the breast (called mastitis) in its early stages and treated with antibiotics, but in most cases biopsy will reveal cancer when there is no improvement in treatment. This type of breast cancer tends to have a higher chance of spreading and a worse outlook (prognosis) than typical invasive ductal or lobular cancer (Sahoo *et al.* 2015; Yang *et al.*, 2008; Bertucci *et al.*, 2005).

Paget disease of the nipple and phyllodes are two rare breast cancer types that occur in about 1 % of breast cancer cases (Sahoo *et al.* 2015). Paget disease of the nipple normally starts in the breast ducts and then spreads to the skin of the nipple, the areola and then finally the dark circle around the nipple (Rosen 2001). The skin of the nipple and areola often appears crusted, scaly, and red, with areas of bleeding or oozing. The woman may notice burning or itching. It is most times always associated with either ductal carcinoma *in-situ* (DCIS) or infiltrating ductal carcinoma (Fu *et al.*, 2001). The prognosis of this disease is excellent if it is associated with DCIS, but the prognosis is not too good if it leads to invasive cancer and the cancer will need to be staged and treated like any other invasive cancer (Sahoo *et al.*, 2015).

The phyllodes tumour on the other hand is a very rare breast tumour that develops in the stroma (connective tissue) of the breast, in contrast to carcinomas, which develop in the ducts or lobules (Lee, 2008). Other names for these tumours include phyllodes tumour and cystosarcoma phyllodes. They are usually benign but on rare occasions may be malignant. The benign phyllodes tumours are usually removed along with a margin of normal breast tissue, while a malignant type is treated by removing it along with a wider margin of normal tissue, or by mastectomy (Hunt *et al.*, 2012; Sahoo *et al.* 2015).

Another rare type of breast cancer is Angiosarcoma. This form of cancer rarely occurs in the breasts and when it does it usually starts in cells that line blood vessels or lymph vessels resulting from complication of previous radiation used as treatment (Hunt *et al.*, 2012). It is an extremely rare complication of breast radiation therapy that can develop about 5 to 10 years after radiation. It is also found in the arms of women who develop lymphedema as a result of lymph node surgery or radiation therapy to treat breast cancer (Sahoo *et al.,* 2015).

1.7.2 Breast cancer risk factors and diagnosis

The knowledge of breast cancer risks and the risks and benefits of the intervention, greatly assist in choosing an appropriate strategy in managing the disease. It is central to making accurate decisions regarding the level of prevention for individual risk assessment. Women who are at high risk of breast cancer can be offered more intensive surveillance or prophylactic measures, such as surgery or chemoprevention. Generally, the population quoted for breast cancer diagnosis, is estimated to be a lifetime risk of one in eight to one in twelve (Amir *et al.*, 2010). Genetic and familial factors can substantially increase the lifetime risk of developing breast cancer and are associated with the development of cancer at a young age. Other risk factors, such as endocrine factors, and host factors including breast density and history of benign proliferative breast disorders, can substantially modify the risk of developing breast cancer (Amir *et al.*, 2010).

Risk assessments are grouped into two categories namely the risk of carrying a mutation in a high-penetrance gene such as BRCA1 or BRCA2 and the risk of developing breast cancer with or without such a mutation (Amir *et al.*, 2010). Other germline mutations that are associated with familial breast cancer have been identified

in other genes and they include TP53, PTEN, ATM, CHEK2, NBS1, RAD50, BRIP, and PALB2 (Walsh and King, 2007). It is important to assess as many factors as possible accurately over time before concluding if there is a risk of breast cancer development or not. A number of risk factors for breast cancer have been identified and quantified. Some of these risk factors are:

1.7.2.1 Family history of breast cancer

Apart from age, the most important risk factor associated with the development of breast cancer is the presence of a substantial family history of breast cancer. Consequently, a family history evaluation is necessary to assess the likelihood of predisposing genes for breast cancer in a family (Amir *et al.*, 2010).

Normally two-pronged approaches are used to assess familial breast cancer risk in an individual. Firstly, those patients who are at risk of carrying a germline mutation are identified and offered formal genetic testing. Secondly, for those who do not meet the criteria for genetic testing or who test negative for germline mutations, there is a need to quantify the risk of developing cancer over a specified length of time. With the resulting information, surveillance, lifestyle, pharmacological, or surgical interventions can be instituted to improve patients risk (Amir *et al.*, 2010).

The age of the patient at onset of breast cancer, family history of whether there are multiple cases or not on both the maternal or paternal side, cases of unilateral versus bilateral disease, the degree of relationship (first or greater), cases of other related tumours with early-onset (e.g., ovary, sarcoma), and the number of members of a family that are unaffected are some of the parameters considered (Kolonel *et al.*, 2004). Women with one affected first-degree relative have twice the risk of breast cancer, those with two first-degree relatives have thrice the risk, and those with three

or more first-degree relatives have quadruple the risk when compared with women with no affected relatives. Breast cancer diagnosed at a younger age in a family member is associated with an increased risk of breast cancer in a first-degree relative (Kolonel, *et al.*, 2004; Amir *et al.*, 2010).

1.7.2.2 Hormonal and reproductive risk factors

Hormonal and reproductive factors have long been recognized to be important in the development of breast cancer. Certain established factors, such as postmenopausal obesity, age at menarche, and use of exogenous hormones, have been shown to contribute to breast cancer risk by increasing systemic exposure to hormones (Schneider *et al.*, 2008). Early age at menarche is associated with a 4 % per-year increase in the relative risk of breast cancer, whereas late menopause is associated with a 3 % per-year increase (Amir *et al.*, 2010).

There is a risk of diagnosing breast cancer in women taking contraceptives, but no significant increased risk of having breast cancer ten or more years following the cessation of the oral contraceptive agent. Cancers diagnosed in women taking the oral contraceptive may not likely be in a clinically advanced state when compared to those diagnosed in women who have never used oral contraceptives (Sestak *et al.*, 2012). A meta-analysis based study showed a 1-2 % per year increase risk of breast cancer in women taking hormone replacement therapy but disappeared within 5 years of stopping treatment. There is a higher risk associated with estrogen and progestin combined hormone replacement therapy as compared to estrogen only whose risk may be negligible (Lacey, *et al.*, 2006). Another meta-analysis study suggested an increased risk of about 24 % of breast cancer diagnoses during current use of the combined oral contraceptive and 10 years after discontinuation (Amir *et al.*, 2010).

1.7.2.3 Proliferative benign breast disease

Some types of benign breast disease present a higher risk of breast cancer compared to others. For example, lobular carcinoma *in-situ* is associated with a 10-fold increase in breast cancer risk when compared to ductal carcinoma *in-situ* (Amir *et al.*, 2010). There is a four to five times higher risk of developing breast cancer in women with severe atypical epithelial hyperplasia as compared to women who do not have any proliferative changes in their breast and the risk increase nine fold in women with a family history of breast cancer (first degree relative) (Mcpherson *et al.*, 2000).

1.7.2.4 Mammographic density

The breast density can rapidly be determined using a mammogram, thus, the mammographic data can present reliable risk prediction data (Amir *et al.*, 2010). Mammographic density is generally looked at as the proportion of the breast tissue that appears dense on the mammogram. It is the most assessable important risk factor and its presentation has a substantial heritable component. The presentation has some geographical distribution link; approximately 5 % of the white female population worldwide has mammographic density covering more than 75 % of the breast (Boyd *et al.*, 2007).

The increase in relative risk for women with 50–75 % mammographic breast density is approximately twice that of women with mammographic breast density of less than 10 %, and these women comprise approximately 14 % of white women. This white female population have an increased risk of about fivefold when compared with women with mammographic breast density of less than 10 %. (Boyd *et al.*, 2007).

1.7.2.5 Other risk factors

The risk of breast cancer observed in teenage girls exposed to radiation during the Second World War doubled (Mcpherson *et al.*, 2006). It was also observed that ionisation radiation during breast formation increase the risk later in life. Obesity in post-menopausal women showed an increase risk of breast cancer whereas a reduced incidence is associated with premenopausal women (Mcpherson *et al.*, 2006). There is an inconsistent report of alcohol and smoking as risk factors of breast cancer. Some studies show a link between alcohol consumption and an increase in incidence rate, but the relationship was inconsistent and it was concluded that the association might be influence by other dietary factors other than alcohol (Mcpherson *et al.*, 2000).

1.8 Breast cancer screening methods

Breast cancer screening is reported to be an effective tool for reducing breast cancer mortality in most recent publications (Kartal *et al.*, 2014). Breast cancer is a progressive disease, and small tumours when detected early could be treated successfully, and present a much better prognosis. So the main benefit of breast cancer screening is that it allows for the detection of breast cancers before it become palpable (Smith *et al.*, 2003).

Although it is true that breast screening is associated with reduced morbidity and mortality, most women who participate in screening do not develop breast cancer in their lifetime. Also, screening do not benefit all women diagnosed as having breast cancer, since it can harm women who undergo biopsy for abnormalities that are not breast cancer as well as women that are over-treated for lumps like ductal carcinoma *in-situ* (DCIS) that may not progress to breast cancer (Smith *et al.*, 2003).

There is a smaller benefit associated with screening in younger women probably as a result of lower incidence, more rapidly growing tumours, and greater radiographic density of breast tissue in women less than 50 years of age. However, screening is most times offered to women with a genetic predisposition to breast cancer that are under the age of 50 years (Pissano *et al.*, 2005).

Various screening methods, such as mammography (sensitivity 71–96 %, specificity 94–97 %), clinical breast examination (CBE; sensitivity 40–69 %, specificity 86–99 %), and self-breast examination (SBE; sensitivity 12–14 %), have a big role in detecting breast cancer (Ma *et al.*, 2012). Self-breast examination and clinical breast examination as diagnostic techniques for breast cancer remain controversial with recommendations ranging from those advocating for SBE/CBE to others stating there is insufficient evidence for their efficacy or a lack of utility in performing these examinations (Ma *et al.*, 2012).

The American Cancer Society (ACS) guidelines for breast cancer screening recommended that women should begin annual screening at age 40 noting that there was no chronological age at which screening should stop, and further stated that a woman in good health would most likely benefit from breast cancer screening (Saslow *et al.*, 2007). Also annual CBE was recommended for women 40 and above scheduled close to the time of, and before, their annual mammograms (Smith *et al.*, 2003).

Mammographic screening of women who are between 50 and 70 years of age have been demonstrated to reduce mortality from breast cancer, however, there is no consensus about the value of mammographic screening among women who are 40 to 49 years old. One of the reasons for the lack of agreement is the difficulty in detecting tumours by mammographic screening in younger women, who have denser breasts

than postmenopausal women (Kriege *et al.*, 2004). There is a reduction in sensitivity in mammography as compared to clinical breast examination in detecting tumours in women less than 50 years, especially in carriers of a BRCA mutation. The reason could be typical changes seen on screening mammograms and specific histopathological characteristics seen in carriers of BRCA mutations, as compared with non-carriers of the same age (Kriege *et al.*, 2004).

Although CBE and SBE could be effective in screening for palpable breast cancer, these methods could be very subjective and inadequate. The magnetic resonance imaging (MRI) is a sensitive method in a diagnostic setting generally and breast imaging in particular (Saslow *et al.*, 2007). Comparing MRI screening in women with a familial or genetic predisposition to breast cancer revealed that MRI appears to be more sensitive than mammography in detecting tumours in women with an inherited susceptibility to breast cancer, however it is highly influenced by breast density making the specificity variable and also the technologies cost are high (Kriege *et al.*, 2004). Despite the availability and recommended use of mammography, MRI, SBE and CBE as routine screening methods for screening breast cancer in women, their effectiveness in reducing overall population mortality from breast cancer is still a big issue. Currently, tumour markers are being investigated for use in breast cancer detection (Li *et al.*, 2002).

1.9 Biomarkers

The National Institutes of Health define biomarkers as a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to therapeutic intervention (Coccia, 2012). Generally, cancer biomarkers are defined as any measurable cellular,

subcellular, or humoral factors that indicate the presence or absence of malignance, its potential, or predict tumour behaviour, prognosis, or response to treatment. These indicators could be nucleic acids, proteins, carbohydrates, lipids, small metabolites, whole cells or biophysical characteristics of tissue (Bahr *et al.*, 2013).

Breast cancer lack early symptoms in most cases and breast cancer detected at an early stage is treatable, but presently there is no Food and Drug Administration-approved tests for early detection of the disease, by the time the cancer is finally diagnosed, it is often in an advance stage of progression and untreatable. Thus, there is a need for reliable breast cancer biomarkers, which can indicate the potential, or presence of the disease (Kirmiz *et al.*, 2007).

Cancer cells display a large range of genetic alterations that include gene rearrangements, point mutations, and gene amplifications, which may lead to change in molecular pathways regulating cell growth, survival, as well as leading to metastasis. Such changes in a specific type of tumour in the majority of patients can serve as biomarkers for diagnosis, prognosis and predicting response to treatment (Karley *et al.*, 2011). Identification of biomarkers for cancer processes required pathway related studies, which include those responsible for proper regulation of various cell functions, whilst paying attention to specific component of the pathway. Study of a single component might not be an easy task because there are lots of components involved in the regulation, thus most times a comparative study of two or more components in the process of carcinogenesis is employed to identify biomarkers which can track the event at an early stage as depicted in the figure below (Karley *et al.*, 2011).

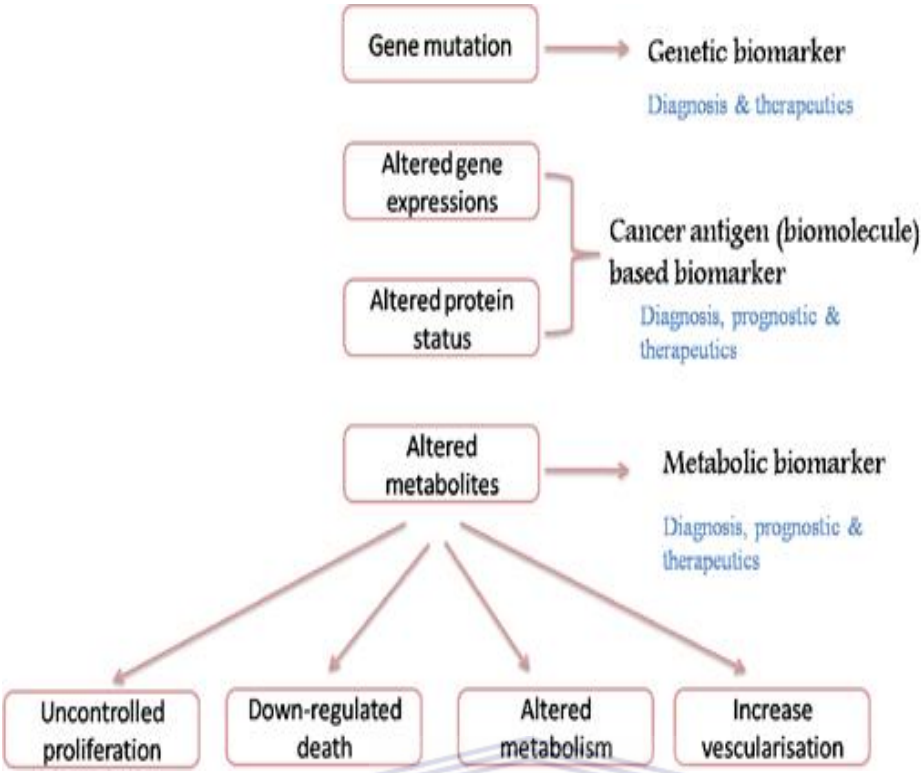


Figure 1.5: The process of carcinogenesis, showing opportunities of identifying biomarker (Karley *et al.*, 2011).

1.9.1 Biomarkers of disease

Every cell type has a unique molecular signature, which are distinguishable characteristics such as, levels of activities of some genes, proteins or other molecular features, and are referred to as biomarkers. Biomarkers are subject to dynamic modulation, and are expected to serve as an indicator of drug metabolism, drug action, efficacy, and safety. They can also facilitate molecular definition of diseases, provide information about the course of disease and predict response to therapies (Bhatt *et al.*, 2010).

The advent of various technologies such as the mRNA (cDNA) microarrays in recent years has helped in generating significant amount of data pertaining to the diagnosis of human diseases. These new techniques have helped in identifying new disease

subtypes that would not have been possible using conventional techniques, which had led to new molecular based classifications of some types of cancers (Kulasingam, and Diamandis, 2008). In most diseases, proteins are subjected to various changes including proteolytic cleavage and/or post-translation modification. Also there is differential expression of some proteins, which result in over-expression or under-expression in disease conditions (Karley *et al.*, 2011).

In cancers, there are alterations in the expression of proteins involved in the process of signal transduction. There are many instances where the protein products of proto-oncogenes modulate the signal transduction process and alteration in these genes may lead to uncontrolled cellular signalling (Alaiya *et al.*, 2000). Messenger-RNA (mRNA) expression profiling normally carried out with the aid of microarray analysis present a first line of investigation in the search for disease biomarker, however, they fall short of presenting absolute biomarkers for disease diagnosis. This is because the microarray that provides information on differential expression of genes will not provide information on post-translational modification of the protein products. For example, alternative splicing of the mRNA can produce different forms of proteins, also in some cases there is no correlation between the level of mRNA and protein concentration (Karley *et al.*, 2011).

1.9.2 Tumour and cancer biomarkers

Biomarkers that will guide physicians at every stage of cancer onset, progression and its management are ideal for the future of cancer management. The cancer biomarker will be useful for accurate evaluation of a disease stage and can be used for predicting several outcomes during the course of disease including: early detection, outcome prediction and detection of disease recurrence (Chatterjee and Zetter, 2005).

Effective tumour markers have the potential to reduce cancer mortality rates by either facilitating early cancer diagnosis or helping to individualize treatments, thus they are in great demand. They are endogenous molecules (nucleic acids, proteins, metabolites, etc.) whose amounts or modifications can serve as indicators of the presence or absence of a tumour, the state of the tumour, progression, characteristics, or response to therapies, which could be present in the tumour tissues or body fluids (Karley *et al.*, 2011).

Serum biomarkers are molecules or antigens produced by body organs or tumours on cell surfaces detected in blood and could be suggestive of tumour activities. Most can be produced by normal organs and thus may be non-specific for cancer. An example is the prostate cancer antigen (PSA) produced by normal prostate cells but whose increase in production is in high positive correlation with prostate cancer. Nevertheless, it is probably the only serum biomarker currently used consistently in primary care (Adam *et al.*, 2002). However, cancer antigen 125 (CA-125) should be a biomarker for risk of ovarian cancer (OC) or an indicator of the malignancy, but it's also found to be present in high concentration in pancreatitis, kidney or liver diseased patients and so cannot be a specific biomarker for OC (Cho, 2007). CA-125 can however be used to monitor progress of treatment, or predict treatment failure if the levels rises after treatment.

Some biomarkers are affected by environmental factors, for example, carcinoembryonic antigen (CEA) is a biomarker elevated in colorectal, breast, lung, or pancreatic cancer patients, but which smoking can elevate. This biomarker could be used to determine the adequacy of post-operative therapy in colon cancer patients (Karley *et al.*, 2011). However, a combination of several tumour markers is used for

diagnosis, prognosis, or risk predictors for someone whose family history for the disease is quite high (Cho, 2007).

1.9.3 Classes of cancer biomarkers

Tumour markers can be used for specific purposes such as screening the general population, for differential diagnosis in symptomatic patients, and for clinical staging of cancer. They can also be used to estimate tumour volume, to evaluate response to treatment, to assess disease recurrence through monitoring, or as prognostic indicators of disease progression (Kulasingam and Diamandis, 2008). Based on these cancer biomarkers are generally classified as:

1.9.3.1 Diagnostic biomarkers

A diagnostic marker is used to detect and identify a given type of cancer in an individual. This class of biomarkers play an important role in early detection of cancer disease so they are expected to be specific and sensitive (Kulasingam and Diamandis, 2008). Current clinical and pathological markers poorly predict early disease development and response to treatment. Standard diagnostic methods, including tissue histopathology are now shifting rapidly toward molecular diagnosis due to the rapid progress in facilities for proteomic studies. This powerful technology can identify all proteins and their posttranslational modifications in disease conditions, and hence will greatly accelerate progress toward novel diagnostic and predictive tools to track early disease and tailor made treatments to specific patients (Simpson *et al.*, 2008).

The advent of various new technologies in biomarker research should bring about development of novel markers that will detect cancer early enough, as well as predict the risk of tumours by screening for invasive cancer. However, the major challenge in utilizing circulating molecular markers to detect cancer is that very small tumours,

which are to be detected and removed before they metastasized to other organs, may not produce enough of the marker that can be easily detected in bodily fluids. To overcome the challenge, it is necessary to first develop new ultrasensitive methods for detecting very low circulating levels of the analytes (Chatterjee and Zetter, 2005).

1.9.3.2 Prognostics biomarkers

These are markers used once the disease status has been established. These biomarkers are used to predict the recurrence and probable course of the disease before it is treated, and are therefore important in determining the aggressiveness of therapy. For example, in testicular teratoma, Human Chorionic Gonadotropin (HCG) and alpha-fetoprotein levels can discriminate two groups with different survival rates (Simpson *et al.*, 2008). Tumour classification, staging and sometimes grade are normally used to assess prognosis. However staging could incorporate other parameters that improve prognosis bringing about additional cost. Biomarker expression often replaces or complements tumour classification, stage and grade when biologically targeted therapeutics is under consideration, but addition of markers could similarly fragment the staging process, thereby limiting it's utility (Ludwig and Weinstein, 2005).

1.9.3.3 Predictive (Stratification) biomarkers

These are biomarkers used to predict the response of a patient to a drug before starting treatment. It classifies individuals as likely responders or non-responders to a particular treatment. These biomarkers mainly arise from array-type experiments that make is possible to predict clinical outcome from the molecular characteristics of the patient's tumour (Kulasingam and Diamandis, 2008).

1.10 Breast Cancer Biomarkers; Uses and Limitation

Breast cancer markers currently in use includes the cancer antigen 15-3 (CA 15-3) and carcinoembryonic antigen (CEA). The BR27.29 marker (also known as CA27.29) detects the same antigen i.e. MUC-1 protein just like the CA 15-3 in the serum, and they provide similar clinical information, however, the CA 15-3 has been more investigated (Harris *et al.*, 2007). The serum levels of CA 15-3 and CEA are related to tumour size and nodal involvement and have been recommended by the American Society of Clinical Oncology (ASCO) to be used in conjunction with diagnostic imaging, history and physical examination in monitoring patients with metastatic disease during active therapy (Harris *et al.*, 2007). These markers are not sensitive enough for screening and early diagnosis of primary breast cancer, but serum cancer levels reflect tumour burden (Kulasingam and Diamandis, 2008).

Humoral response against many tumour associated proteins such as c-erbB-2/HER2/neu, RS/DJ-1, and mucin-related antigens have been detected in breast cancer (López-Árias *et al.*, 2012). The proto-oncogene c-erbB-2/HER2/neu, which encodes a growth factor receptor, was found to be over-expressed in about 30 % of patient with breast cancer. Also the autoantibodies to this oncogene have been observed in 11 % of breast cancer cases and this figure correlates with the amount of protein over-expressed in tumour tissues. In addition, elevated amount of c-erbB-2/HER2/neu protein was found in the serum of 29 % of patients with breast carcinoma associated with poor prognosis (Yi *et al.*, 2009).

The non-serological markers like the hormone receptors estrogen and progesterone are used in the clinic as breast cancer tissue-based markers. The estrogen receptors is regarded as a weak prognostic marker, it is also used for predicting response to hormone therapy in conjunction with progesterone receptor in both early and

advanced breast cancer treatment (Reis-Filho and Pusztai, 2011). A meta-analysis study, showed that ER-positive patients were 7-times less likely to develop recurrent disease than ER-negative patients after 5 years of adjuvant treatment with an estrogen receptor antagonist, tamoxifen. Her-2/neu is another tissue-based marker used in the clinic for breast cancer prognosis and it is most useful in node-positive patients (Duffy, 2001). For the node-negative patients, there are conflicting reports on prognosis markers (McShane *et al.*, 2005). The levels of Her-2 in tissue are also used for determining patients with either early or metastatic breast cancer for treatment with Trastuzumab (Herceptin) (Kulasingam and Diamandis, 2008). BRCA1 and BRCA2 are two genetic markers used to identify individuals who are at high risk of developing breast or ovarian cancer in high-risk families (Kulasingam and Diamandis, 2008).

Although many candidate biomarkers for breast cancer disease have been identified and evaluated, efforts are still very intensive for the discovery of novel biomarkers. Some major problems in biomarker research are that biomarker detection assays may not be sufficiently sensitive or specific enough to be reliable for early diagnosis, or prognosis. Presently, great efforts are being made toward the discovery of better biomarkers for the early diagnosis and prognosis of disease (Yi *et al.*, 2009). For this purpose, bioinformatics is one of the new powerful techniques that could be exploited to identify new biomarkers from various data generated from different biological samples and powerful technological platforms.

1.11 Bioinformatics and biomarkers

The availability of the complete human genome has paved the way for the systematic understanding of human diseases. Recent technological advances in functional genomics and proteomics using microarray and mass spectrometry, coupled with

improved bioinformatics and statistical tools, have fuelled the interest in identifying biomarkers for complex diseases such as the neurodegenerative diseases and cancer (Karley *et al.*, 2011). Functional genomics involves the use of large-scale data produced by high throughput (HTP) technologies to analyse the expression pattern of large number of genes in diseased and normal tissue, and to understand the function of such genes and other parts of the genome. Also, much effort is directed towards the development of strategies for quantitative protein profiling using different biological samples (Karley *et al.*, 2011).

The increasing size and complexity of the data generated by HTP methods and the microarray technology provide challenges for researchers to extract the biologically relevant information. Bioinformatics serve to link scientific data with clinical information. This is particularly important in relation to the recent development of powerful bioinformatics algorithms, which can interpret multiple parameters much more efficiently than most traditional approaches (Kulasingam and Diamandis, 2008).

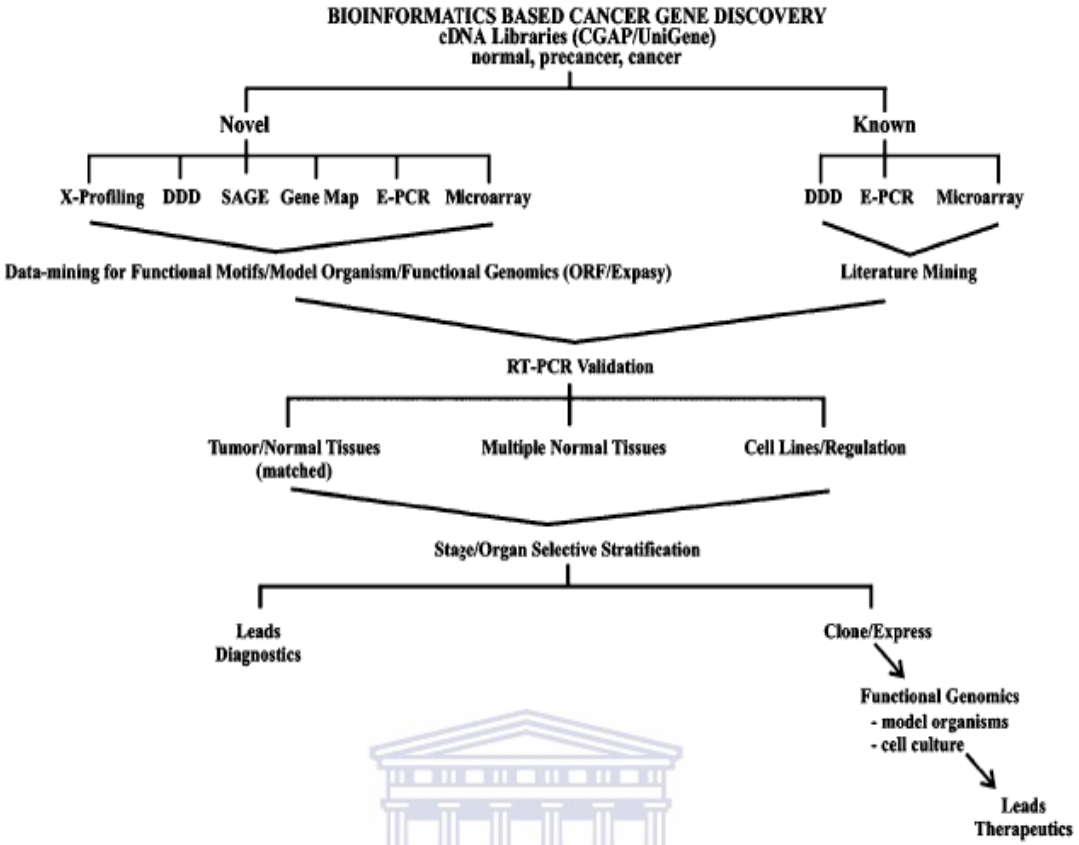


Figure 1.6: A proposed approach to cancer gene discovery from the CGAP database. Both novel and known ESTs are identified using multiple data-mining tools from this database (Narayanan, 2007).

Genomics provides the blueprint for the discovery of possible gene products that can serve as biomarkers. Although there are about 30,000 genes in the human genome (http://www.ncbi.nlm.nih.gov/genome/guide/human/), the protein complement of these genes (the proteome) of a cell or tissue is larger and more dynamic in nature. This as a result of different modification processes like alternative splicing of transcripts, which leads to different isoforms of a protein, glycosylation, phosphorylation, myristylation etc. that leads to different active proteins per gene (Packer *et al.*, 2007).

The genomics and transcriptomics approach to biomarker discovery provide powerful tools for identifying candidate markers, but an important shortcoming of these approaches is that there are no correlation between the changes in expression of mRNA and level of protein. Thus many technologies are being developed and exploited to collect information about proteins (Patterson and Aebersold 2003). Most of the proteomics studies begin with isolation of total proteins from test samples and another from a control group. Analysis of these proteins is then carried out by mass spectrometry (MS) or 2-D gel electrophoresis. The data generated is then subjected to data mining (Bioinformatics) to recognize a complex pattern and identify proteins involved, which could serve as biomarkers (Patterson and Aebersold 2003).

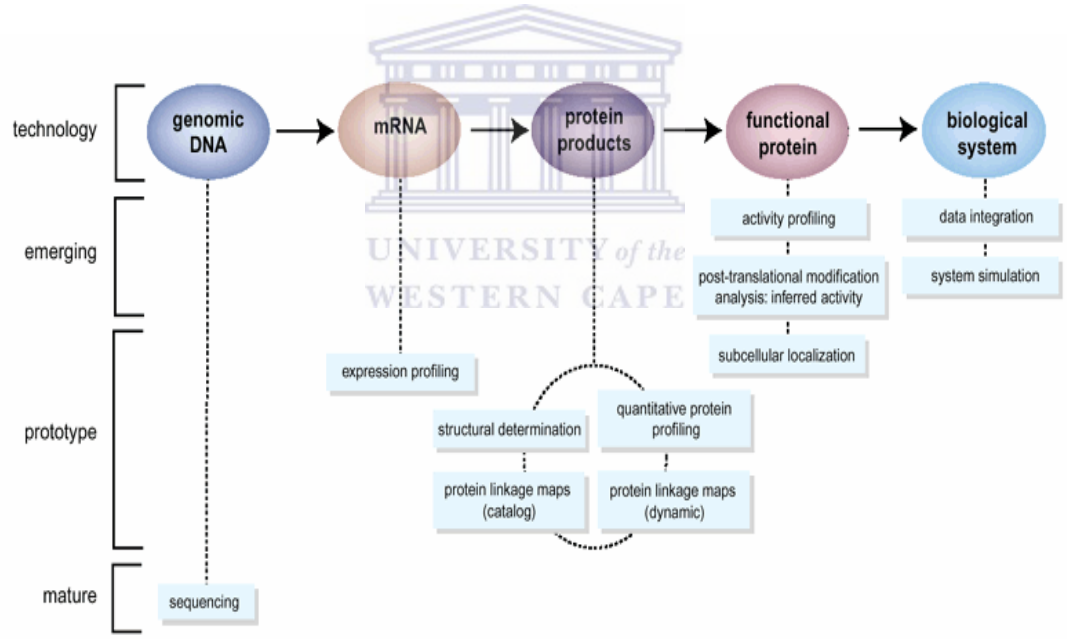


Figure 1.7: Current status of proteomic technologies in biomarker discovery (Patterson and Aebersold 2003).

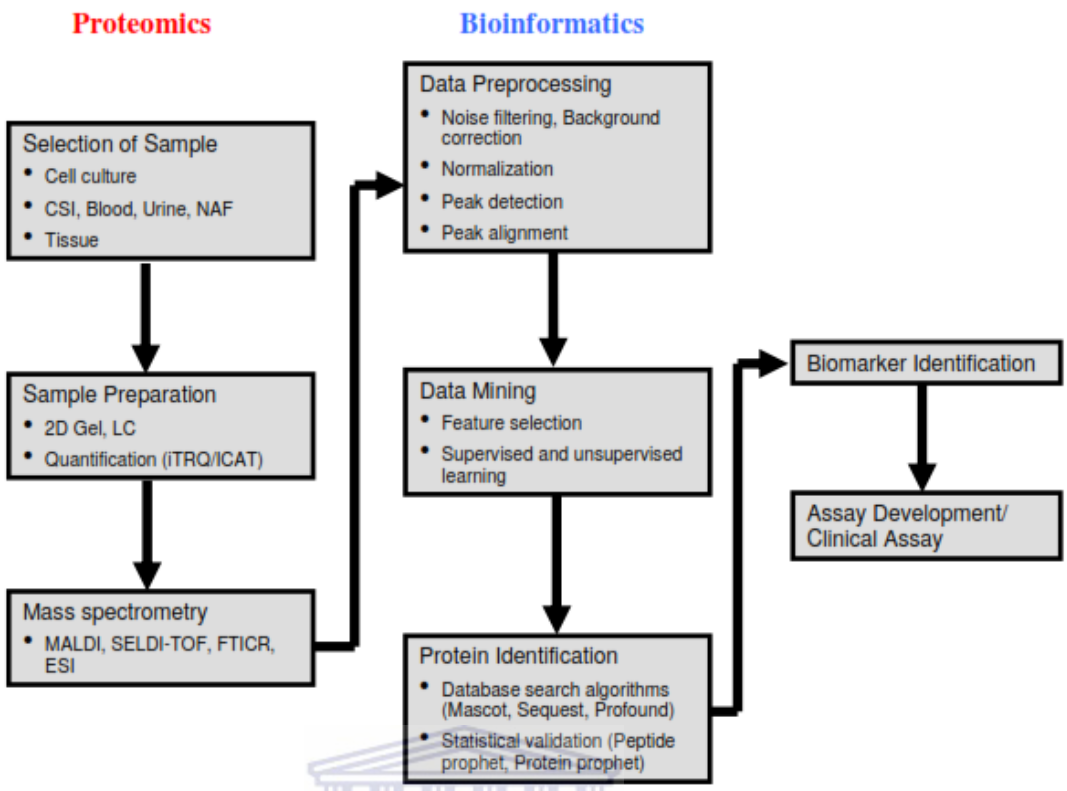


Figure 1.8: Schematic representation of the proteomics process in biomarker discovery (Verma, 2007).

With increasing availability of public data, there is a need for rigorous comparisons of data, pre-processing and data mining in the course of searching for biomarkers. Bioinformatics approaches appear to be very critical for effective analyses and mining of data in providing insights into disease biology that provides good candidate markers (Moore, 2010). Many algorithms exist to mine large datasets, but no specific approach is ideal or applicable to all study designs, different approaches are normally used in parallel before arriving at a final algorithm. Finally, it is anticipated that existing and emerging computational data mining approaches along with rigorous and systematic evaluation, will help to unleash the full biological potential of proteomic profiling (Sugimoto *et al.*, 2012).

1.13 Aim and Objectives

1.13.1 Aim

The aims of this research was to validate several candidate genes identified from a previous study as effective breast cancer biomarkers using several *in silico* methods as well as molecular techniques and to study the modulating effect on pro-apoptotic compounds of one of these genes in particular (Gene 7).

1.13.2 Objectives

- Functional characterization of the identified candidate biomarkers using a bioinformatics (*in silico*) approach.
- Prognostic and predictive validation of the candidate biomarkers using a bioinformatics approach
- Protein expression analysis of the candidate biomarkers in cancer cell lines
- Construction of "knock-down" and over-expression vectors of Gene 7 in breast cell lines
- Effect of Gene 7 "knock-down" and over-expression in a breast cancer cell line on Apoptosis using different apoptotic inducers.

Chapter 2; Functional characterisation of identified genes using *in silico* approach

2.1.1 Introduction

The field of bioinformatics has grown in recent years with a concurrent increase in genomic and proteomic data (Li *et al.*, 2010). Data intensive research has spurred the development of computational methods to ease analysis of databases, especially those of a cross-disciplinary nature, which serve as resources for data collection and integration (Ramakrishnan *et al.*, 2005). Databases come in all forms, and some examples are protein interaction databases and pathway databases to mention only a few. These are valuable tools that allow for genes and proteins to be organised within their larger biological context, making it possible for inferences to be made regarding their role in disease pathways (Draghici *et al.*, 2007).

The widespread use of DNA microarray technology to perform experiments on thousands of gene fragments at the same time has generated lots of expression data. In handling these huge datasets, whose interrelationships is well understood, exploratory and visualization techniques are necessary (Gibbons and Roths, 2002). Different bioinformatics techniques such as placing members of a set of genes to one of a much smaller number of classes (Clustering) are exploited in inferring gene function(s) (Aittokallio and Schwikowski, 2006). Gene clustering encompasses many techniques, such as hierarchical clustering (which encompassing single-, complete-, and average linkage variants), *k*-means clustering, and self-organized maps (SOM), used to assign genes into similar and relevant biological groups (Gibbons and Roths, 2002).

The identification of candidate genes or proteins that may be involved in cancer disease is important in expanding our knowledge of the disease process which may

help in early diagnosis, prognosis and therapeutics. Investigating the role of identified genes and other proteins they interact with could provide clues of their molecular function and regulation within a disease process such as cancer (Baylin and Jones, 2011). Also identifying associated cellular pathways could reveal the functional environment of a candidate gene (Hartman *et al.*, 2001). Various databases made up of experimentally verified genes from published literature as well as data of predicted genes, based on trends from the genome wide sequence, could provide a framework for further experiments and studies using bioinformatics platforms.

The advances in technology and the advent of high-throughput genomic, proteomic combined with bioinformatics techniques, allows investigators to simultaneously examine the changes, expression level or pattern and regulation of genome-wide genes under certain biological conditions, unlike the traditional biological approaches that typically study one gene at a time (Huang *et al.*, 2009). The output of these high-throughput techniques are large 'interesting' gene lists that needs to be further analysed with the aim of achieving a smaller sized gene list that can be validated using traditional biological techniques. However interpreting the biological functions and relevance of the genes in such a list is still a challenging and daunting task. This is because of the size of the generated gene list, but with the aid of bioinformatics methods, the biological knowledge discovered or predicted and stored in public databases could help to sieve out and assemble a summary of the most enriched and pertinent candidate gene(s) (Huang *et al.*, 2009).

2.1.2 Previous study

An *in silico* approach was used to identify cell surface genes implicated in the development or progression of breast cancer through mining of biological databases

such as ONCOMINE and Gene Expression Atlas (GEA). ONCOMINE (www.oncomine.org) is a popular cancer micro-array database, while the GEA (www.ebi.ac.uk) is a general micro-array database. A set criterion that searched for upregulated, cell surface genes in breast cancer was used, which resulted in extraction of several gene lists from these databases. The resulting lists were harmonized, refined and the final output was ranked based on the intersection of the data types. Candidate genes were selected based on their ranking and subjected to Gene Ontology (GO) analysis for functional characterization to determine their cellular localization. Twenty-three genes whose protein products are categorized as cell surface proteins were selected based on the GO analysis. Cell surface proteins were of particular interest as it is anticipated that their protein products will be shed into human bodily fluids and could easily be assessed for early diagnosis of breast cancer. A literature search on these candidate genes was done and fourteen were found to be experimentally linked to breast cancer, one was in the process of clinical validation as a biomarker, one had FDA approval as breast cancer biomarker whilst seven were not inferred or proven as breast cancer biomarkers.

The expression of these 7 genes were subsequently analyzed in breast cancer cell lines as well as other types of cancers, using semi-quantitative PCR and qRT-PCR. Gene expression analyses performed on a breast cancer cell line and a non-cancerous breast cell line, revealed differential expression of some of these genes between the cancer and the non-cancerous cells. In addition, a differential expression analysis comparing the expression level of these genes in breast cancer cell line with several other cancer types revealed that some of these putative genes were also over-expressed in the other cancer types, whilst three of the genes were found to be differentially expressed in the breast cancer cell line (MFC-7) (Ngcoza, 2012).

2.1.3 Aim

The aim of this chapter was to functionally characterize the seven identified genes as putative breast cancer biomarkers using several *in silico* approaches.

2.1.4 Objectives

- Gene clustering analysis of the candidate genes
- Gene pathway analysis of the candidate genes
- Protein interaction analysis of the candidate genes
- Transcription factor binding analysis of the candidate genes
- Tissue expression anlysis of the candidate genes

2.2.0 Gene clustering (Gene Ontology)

The Gene Ontology (GO) is an annotation tool set to present consistent descriptions of gene products across databases. The project was founded in 1998 by the GO consortium (GOC), incorporating many databases, including several major repositories for plant, animal and microbial genomes (Lee and Sonnhammer, 2003).

The Gene Ontology Consortium (GOC) provides a structured language that is applied to the distribution of differentially expressed genes and proteins in all organisms into functional categories (Rhee *et al.*, 2008). GO is structured into three main categories namely: Biological Process, Molecular Function, and Cellular Component. Biological Process includes broad biological processes, which are accomplished by ordered assemblies of molecular functions for example, the purine metabolism process. Molecular Function annotate genes based on the tasks performed by their individual gene products; examples are transcription factor activity and DNA Helicase activity. The Cellular Component annotations include classification of genes into sub-cellular structures, locations, and macromolecular complexes; examples include nucleus, telomere, and origin recognition complex (Coulibaly and Page, 2008).

2.2.1 Database for annotation, visualization, and integrated discovery (DAVID)

Various tools have been developed to assess the statistical significance of association of a list of genes with GO annotations terms over the years. Most of these tools provide exceptional depth and coverage of the functional data available for a given gene, but are not designed to effectively explore the biological knowledge associated with hundreds or thousands of genes in parallel (Dennis *et al.*, 2003).

The Database for Annotation, Visualization, and Integrated Discovery (DAVID) provides a set of data-mining tools that describe the functions of genes in a set of data

and systematically combine the functionally descriptive data with graphical displays. DAVID provides exploratory visualization tools that promote discovery through functional classification, biochemical pathway maps, and conserved protein domain architectures, while simultaneously remaining linked to rich sources of biological annotation (Huang *et al.*, 2008).

The DAVID Gene Functional Classification Tool (http://david.ncifcrf.gov) annotate and combine a gene list of similar properties of associated biological terms into groups based on reported or predicted functions. These classes of related genes are referred to as biological modules and are accomplished by mining the biological co-occurrences found in different sources of functional annotation (Huang *et al.*, 2008).

2.2.2 Pathway analysis

The advents of high-throughput techniques for gene sequences and gene/protein profiling have transformed biological research by providing an effective means of monitoring a biological system. These high-throughput techniques usually present a list of differentially expressed genes after analysis, indicating genes that may play a role in a given phenomenon or phenotype. However, these lists most times often do not provide insight into the underlying biological condition being studied (Khatri *et al.*, 2012).

A major application of identified regulatory networks or the identification of clusters of functionally related genes is to associate the genes and their interconnections with known metabolic pathways (Barabási *et al.*, 2011). Biochemists summarized the sequence of enzymes that catalyzed metabolic reactions between biomolecules into network of interactions that results from the conversion of one organic substance (substrate) to another (product). These biochemical networks represent the potential

link that exists between genes and gene products, involved in specific biological processes (Coulibaly and Page, 2008).

There are lots of specialized databases available with large amount of information on metabolic reactions and one of such is the Kyoto Encyclopaedia of Genes and Genomes (KEGG), used for this study.

2.2.3 Kyoto encyclopaedia of genes and genomes (KEGG)

The Kyoto Encyclopaedia of Genes and Genomes (KEGG) is a database resource of the Japanese GenomeNet service (http://www.genome.ad.jp/) use in understanding the functions of cells or the organisms' base on their genomic information (Rashidi and Buehler, 1999).

KEGG consists of a database of pathways (PATHWAY database) for analysing molecular interaction and networks, the database of genes (GENES database) for accessing information about genes and their protein products which are generated by genome sequencing projects, and the database with information about chemical compounds and chemical reactions that are relevant to cellular processes (LIGAND database) (Kanehisa *et al.*, 2002). The KEGG pathway maps are widely used for biological interpretation of genome sequences and other high-throughput data. The database is a reference database for pathways developed by capturing and organizing experimental results from published literature, which initially focused on metabolic pathways but later incorporated other cellular processes (Kanehisa *et al.*, 2014).

2.2.4 Search tool for the retrieval of interacting genes/proteins (STRING)

STRING is a database and web resource dedicated to protein–protein interactions, including both physical and functional interactions (Jensen *et al.*, 2009). Proteins interact with one another not only by physical binding, but also indirectly by sharing

an intermediary substrate in a metabolic pathway, by regulating the transcription of each other either through feedback or co-expression, or by participating in larger multi-protein assemblies (Von Mering *et al.*, 2005).

These various interactions which may be transient and as well functional pathways, are context dependent and requires various experimental techniques which are not similar, and thus, not easily elucidated and understood (Ben-Hur and Noble, 2005). Nevertheless, protein–protein interaction networks or functional association networks are crucial in understanding the cellular machineries in organisms (Barabasi and Oltvai, 2004). Furthermore, protein networks can help fulfill other practical purposes such as filtering and assessing high-throughput functional genomics data, and providing intuitive visual scaffolds for annotating the structural, functional and evolutionary properties of proteins (Jensen *et al.*, 2009). STRING present an aggregate of available information on protein–protein associations, scores and weights them, and also augments the list with predicted interactions, as well as results from automatic literature-mining searches (Jensen *et al.*, 2009).

2.2.5 GeneCards® Version 3: the human gene integrator

GeneCards (www.genecards.org) is a comprehensive, authoritative compendium of list of annotated information of human genes. Its content are automatically mined and collated from over 80 data sources of genome, transcriptome, proteome and disease data. These databases includes large public databases like ENSEMBL, UniprotKB, National Centre for Biotechnology Information (NCBI) and HUGO Gene Nomenclature committee (HGCN) among others (Safran *et al.*, 2010).

The database was established by the Weizmann Institute of Science in collaboration with the Life Map Sciences Incorporated with the aim to integrate various data

generated by the human genome project and different specialized databases derived from this project and other expression analysis resources (Crusio and Gerlai, 1999). GeneCards version 3 contains more than 73 000 entries, which includes cards for known and predicted protein coding genes as well as pseudo-genes and RNA genes (Safran *et al.*, 2010).

The GeneCards® database also provide multifaceted tools for gene characterisation from a nucleotide sequence starting point, to the analysis of gene expression and protein modification, including protein involvement in cellular and disease pathways, listing transcription factors associated with protein of interest and providing information on molecular tools to aid research (Safran *et al.*, 2010).

2.2.6 TfactS®: (Transcription Factor activity database)

TfactS® is a database designed to predict regulation, inhibition or the activation of transcription factors (TFs) in a biological system based on gene lists of upregulated and down regulated genes generated in microarray experiments (Essaghir and Demoulin, 2012). List of up- and/or down-regulated genes are submitted for TF query to the TfactS® input dialogue boxes, the list is compared with a catalogue of annotated target genes, and returns three lists of transcription factors whose annotated target genes show a significant overlap with the query genes. One of the list shows the regulated Transcription Factors (TF), another list consist of activated TF while the third shows a list of repressed TF. Both the activated and repressed lists are produced using the Sign-Sensitive catalogue whilst the Sign-Less catalogue does not provide this information (i.e., up-/down-regulated TF) (Essaghir *et al.*, 2010).

2.2.7 Gene Expression Atlas

Gene Expression Atlas (GEA) is a general micro-array database that makes use of information obtained from experiments carried out using both tissues and cell line samples in various disease states from humans (Lein, *et al.*, 2007). GEA integrates data from 5.372 human samples representing 369 different cell and tissue types. It makes use of over 9000 raw data files generated on the human gene expression array Affymetrix U133A from public databases, Gene Expression Omnibus and Array Express. GEA thus integrates all micro-array studies pertaining to a disease and allows the user to search for a gene of interest as well as to define conditions, and provides an overall output extracted from all the information collected in the database regarding gene expression in that disease condition. (Kapushesky *et al.*, 2010).

2.3 Materials and methods

2.3.1 Functional annotation

The seven genes identified as putative breast cancer biomarkers (Ngcoza, 2012) were classified based on annotated roles in cancer disease using the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 (Niaid, 2006). The functional annotation tool was selected from the home page of the software (http://david.abcc.ncifcrf.gov/). The HGCN gene symbols of the 7 genes were typed into "paste list" box, "official gene symbol" was selected as the unique identifier, and "gene list" was selected as list type after which the "submit list" button was selected. Functional annotation clustering was selected from the output and different processes implicated in cancer were selected from the different clusters. These biochemical processes were tabulated against the annotated genes.

2.3.2 Pathway analysis.

Each of the seven genes was submitted to the KEGG pathways database (http://www.genome.jp/kegg/pathway.html) after selecting *Homo sapiens* as the organism and the output were tabulated against annotated pathways. The default setting was used for the pathway analysis.

2.3.3 Protein-protein interaction analysis

The interactions of the protein products of the putative gene, with proteins known to be involved in apoptosis and those involve in autophagy where monitored using the STRING protein database version 9.1 (http://string-db.org/). The protein symbols for each gene was pasted in the "list of protein" box after selecting multiple names as input option for the database, *Homo sapiens* was selected as organism and "go" button was clicked for interaction search. Also Gene 7, observed to be the most differentially

expressed from the qRT-PCR analysis (Ngcoza, 2012), was also separately analyzed by inputting the gene symbol into the search box after selecting the "search by name" option. The default confidence score of 0.400 was changed to 0.700 in order to restrict the result to interactions with a high confidence level.

2.3.4 Validating expression of candidate genes in breast cancer

The Gene Expression Atlas database was accessed at http://www.ebi.ac.uk/gxa and the seven genes were queried by inserting the gene names (HGCN) into the gene box, the "up/down expression" option in *Homo sapiens* was selected under organism and "cancer, cancerous" were selected under condition. The result was viewed in a heat map format showing the genes, the number of studies done, and their expression level in different cancer types. The number of times a gene was reported to be over-, under-, or not reportedly (not determined) expressed in breast, colon, lung and prostate cancer types was recorded by assigning 2, 1 and 0 to over-, under- or not reported expression respectively. The experimental outcome in different cancer cell lines and tissue for each of the cancer types was then plotted against the genes using the Graph pad prism version 5.00 statistical tool GraphPad Software (San Diego, California USA).

2.4 Result and Discussion

2.4.1 Functional annotation

Identifying specific gene expression signatures of biological relevance from the genome-wide expression or microarray generated data would further our understanding of central tumour forming biological processes. Their controlling factors and might also help to delineate therapeutic, diagnostics or prognostic considerations for cancer care, as well as play an important role in the development of novel targeted therapies (Fredlund *et al.*, 2012).

The aim of this chapter was to validate seven genes identified in an earlier study (Ngcoza, 2012) as potential breast cancer biomarkers using additional *in silico* approaches to functionally characterize these candidate genes. The involvement of these genes and their protein products were analysed to identify their functional relationship and links with other genes and protein products already implicated in cancer in general and breast cancer in particular.

Figure 2.1 shows the genes were clustered into several functional groups using the DAVID database. The result showed that several of the identified genes are involved in processes that have been implicated in cancer pathways (Figure 2.1). Three of the seven genes were present in the list of genes annotated as secreted proteins. These genes included Gene 7, which showed high differential expression between the non-cancerous breast cell line, (MCF-12A) and the breast cancer cell line (MCF-7), as well as Gene 1 and Gene 2, that did not show differential expressession amongst the cell lines analysed based on the qRT-PCR analysis (Ngcoza, 2012).

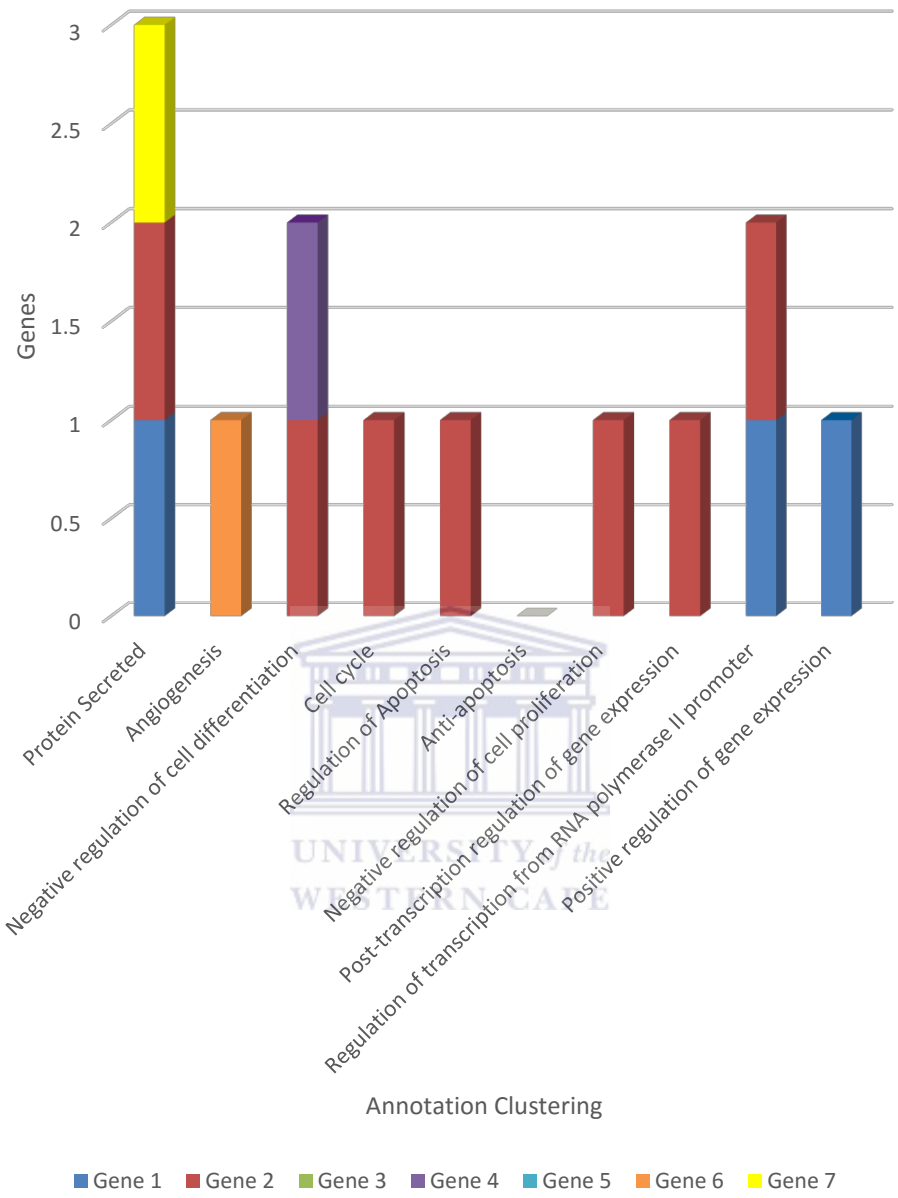


Figure 2.1: Functional annotation clustering of the seven putative breast cancer genes using GO analysis.

A number of gene families of secreted and trans-membrane proteins related by homology have been identified and classified into groups known to have key roles in important biological processes such as morphogenesis, cellular differentiation, angiogenesis, apoptosis, and modulation of the immune response, as well as disease processes such as cancer progression. These genes are members of diverse gene families, which include tumour necrosis factors, growth factors, cytokines, chemokines, interferons, and angiopoietin-related protein families, as well as the protein families of their receptors and other receptors such as the Toll-like receptors, integrins, and disintegrins (Clark *et al.*, 2003). This result suggests that the protein products of these putative genes shown to be secreted, might be present in bodily fluids of breast cancer patients, and may easily be accessible if they are to be used as a biological indicator of breast cancer in a point of care device.

Apart from serving as markers for cancer discovery and monitoring, secreted protein could play a key role in cancer therapeutics and prognosis. They also play an integral role at a specific state of a disease or the biological pathway leading to disease, and their identification and characterization may lead to understanding the disease paradigm. Based on their location on the cell they possess good properties of therapeutic agents or targets, since they are accessible to various drug delivery mechanisms (Rangiah *et al.*, 2009). Rituxan and Herceptin are two therapeutic antibodies that target B lymphocyte-specific CD20 protein and HER2 receptor protein for the treatment of non-Hodgkin's lymphoma and breast cancer respectively and are considered as markers for these diseases (Clark *et al.*, 2003).

Some of these genes were also found to be involved in processes that are hallmarks of cancer like angiogenesis (Gene 6), apoptosis, cell cycle and regulation of cell differentiation (Gene 2). Angiogenesis is the development of new vessels from

existing vasculature, a process necessary in tumour growth, survival, and progression. Angiogenesis is regulated at the molecular level by pro and anti-angiogenic cytokines. The cytokines that majorly control angiogenesis is the relatively small family of vascular endothelial growth factors (VEGFs) (Wehland *et al.*, 2013). VEGF was among the twenty-three genes identified in the earlier *in silico* approach to identifying breast cancer biomarkers (Ngcoza, 2012).

Several angiogenic particles have been studied as prospective predictive biomarkers for different breast cancer types, but not all breast cancer patient have a good prognostic outcome for anti-angiogenic therapy. Thus, so far no pattern has been established to select biomarkers that will serve this purpose (Wehland *et al.*, 2013). Although none of the three highly differentially expressed genes based on qRT-PCR analysis (Ngoza, 2012) were shown to be involve in angiogenesis, the functional annotaion analysis showed one of the seven genes (Gene 6) to be involved in this biochemical process (Figure 2.1). This suggests that the expression pattern of the protein product of Gene 6, could potentially serve a prognostic biomarker role in anti-angiogenic therapy in breast cancer following extensive molecular validation.

Figure 2.1 also showed one of the seven candidate genes (Gene 2) to be involved in regulating apoptosis. Deregulation of apoptosis is a known characteristic of human carcinogenesis and molecules involved in this process are expected to be good candidate markers. Several apoptotic markers like Bcl-2, caspase-3, P53, Survivin etc. have been suggested as prognostic markers for different cancer types (Karam *et al.*, 2007). Also Gene 2 is the only gene out of the seven candidate genes associated with other cancer causing processes like; cell cycle, negative regulation of cell proliferation, and post-transcriptional regulation of gene expression (Figure 2.1).

Interestingly, Gene 2 was not shown to be differentially expressed in the qRT-PCR analysis (Ngcoza, 2012), but could be considered a good candidate biomarker based on its recurring presence in most cancer related processes based on the *in silico* functional annotation analysis thus far. The expression level of this gene could potentially be repress under breast cancer conditions. Gene repression can also serve as a good indicator of disease onset or progression (Frank and Hargreaves, 2003), which could be an additional basis for Gene 2 being a potential biomarker for breast cancer. Post transcriptional regulation of gene expression play a key role in human disease states. MicroRNAs (miRNAs), a class of small non-coding RNAs is known to modulate post-transcriptional outcomes of gene expression by targeting mRNAs and triggering either translational repression or RNA degradation (Iorio *et al.*, 2005). Also, Gene 2 was the only gene out of the seven candidate genes linked with cell cycle, regulation of apoptosis and negative regulation of cell proliferation and not classified as an anti-apoptotic molecule from the annotation analyses (Figure 2.1), which suggest other possible reasons why it could be repressed by an oncogene to enhanced cancer proliferation.

2.4.2 Pathway analysis

Grouping and/or assigning genes, proteins, and/or other biological molecules to the pathways they are involved in, helps to minimise the number of pathways that needs to be dealt with when characterizing candidate biomarkers on a molecular level. Also identifying distinct active pathways between candidate genes under different conditions can have more validating power than a simple list of differentially expressed genes or protein. Thus, pathway analysis has become the first choice for gaining insight into the underlying biology of differentially expressed genes and protein (Khatri *et al.*, 2012).

From the pathway analysis result using KEGG (table 2.1), the seven genes showed no known direct and specific link to breast cancer. However, some of the genes are shown to be involved in pathways that may enhanced, inhibit or regulate macromolecule metabolic pathways, cell membrane structure, or the body's immune system, thus processes that have been implicated in cancer and its proliferation.

From table 2.1, Gene 4 was specified to be involved in nucleic acid metabolism, carbohydrate metabolism as well as the biosynthesis of vitamins that may signal the stimulation of cell proliferation as well as partake in reorganization of metabolic activities. The onset of cell proliferation through the cell cycle has been reported to require doubling of total biomass (proteins, lipids and nucleic acids) a phenomenon that poses a great metabolic challenge to cells and which must be overcome for cells to respond to proliferative stimuli (DeBerardinis *et al.*, 2008). Thus deregulation of these pathways should enhance tumour growth, hence the probable role of this gene and/ or is protein product in cancer onset or progression.

Metabolism in cancer cells differ from the normal dormant cells in high rates of glycolysis leading to an increase in lactate production, increase production of lipids and other macromolecules. Proliferating cells often take up nutrients in excess of energy flow required and link metabolites into pathways that support a platform for biosynthesis for achieving the three basic needs of dividing cells which are rapid ATP generation to maintain energy status, increased biosynthesis of macromolecules and tightened maintenance of the appropriate cellular redox status (DeBerardinis *et al.*, 2008).

Table 2.1: Pathways linked to the different identified putative biomarker genes using KEGG

GENES	ASSOCIATED PATHWAYS
Gene 1	Extracellular matrix (ECM)-receptor interaction
Gene 2	Phagosome formation, Antigen processing and presentation, Protein processing in endoplasmic reticulum, Chagas disease (American trypanosomiasis), Human T-lymphotropic virus type 1 (HTLV-1) infection
Gene 3	No pathway associated in KEGG
Gene 4	Nicotinate and nicotinamide metabolism, Purine metabolism, Starch and sucrose metabolism, Riboflavin metabolism, Pantothenate and CoA biosynthesis, Metabolic pathways
Gene 5	No pathway associated in KEGG
Gene 6	Leukocyte transendothelial migration
Gene 7	Natural killer cell mediated cytotoxicity

Gene 4 was also linked to carbohydrate metabolism (table 2.1), which suggests a role for it in sustaining cancer cells. The best studied metabolic process observed in tumour cells is the Warburg effect. This is a process of ATP generation through glycolysis by converting glucose to lactate even in the presence of oxygen, a shift from the normal ATP generation through oxidative phosphorylation (Cairns *et al.*, 2011).

The seven candidate genes also included Gene 1 a member of the integrin family that are involved in extracellular matrix receptor interaction activity (table 2.1). The integrin family of adhesion receptors transduces signals from the extracellular matrix that regulate growth, gene expression and differentiation, cell shape, motility and cytoskeletal architecture. It is reported that integrin in conjunction with other classes of adhesion receptors or growth factor receptors help to transduce signals (Damsky and Werb, 1993). Fisher *et al.* (2004), reported that, different members of the small integrin binding ligand *N*-linked glycoprotein (SIBLING) gene family, have been reported to modulate the activities of the cell surface proteins known as the matrix metallo-proteinases (MMPs). MMPs was shown to be differentially expressed in different cancer types including breast cancer. The study concluded that SIBLINGs are potential markers of early disease progression in a number of different cancer types, some of which lack vigorous clinical markers (Fisher *et al.*, 2004). This suggest that Gene 1, being in the same class as the SIBLINGs could also serve as a biomarker for breast cancer.

Gene 7 is a member of the protein family that binds to natural killer cells to initiate their activities. The pathway analysis result confirmed this by associating this gene to natural killer (NK) cell mediated cytotoxicity (table 2.1). NK cells are important components of the innate immune system, which is regulated by a complex balance resulting from signals received via inhibitory and activating receptors (Rölle *et al.*,

2003). NKG2D is an activating receptor found on natural killer cells. They serve as immune-receptor of cytotoxic lymphocyte. Tumours can escape immune surveillance by shedding off ligands that binds NKG2D receptors for NK cell stimulation. These soluble ligands thus are present in cancer patients' sera and could serve as biomarkers (Paschen *et al.*, 2009). Ligands of the NK receptors have strongly been suggested as biomarkers for monitoring cancer initiation, progression and therapeutics (Neagu *et al.*, 2010). Since Gene 7 is a NKG2D ligand family member and was annotated as a secreted protein (Figure 2.1), it could thus be shed into bodily fluids of breast cancer patients and could serve as a useful biological marker for the disease.

2.4.3 Protein Interaction Analysis

Protein-protein interactions are very important in understanding biological processes, in that it provide insights into protein function(s). It may also help to uncover the generic organization principles of functional cellular networks, when both spatial and temporal aspects of interactions are considered (Stelzl *et al.*, 2005). The protein composition in a cell is very dynamic and changes during development, disease condition and in response to external stimuli, and these proteins form large interaction networks, in which they regulate and support each other (Von Mering *et al.*, 2002).

Using the *S*earch Tool for the Retrieval of Interacting Genes/Proteins (STRING) database, the interaction among the seven candidate genes was analyzed. For the purpose of this thesis, the protein products of Gene 1-7 will be referred to as BRG 1-7 respectively.

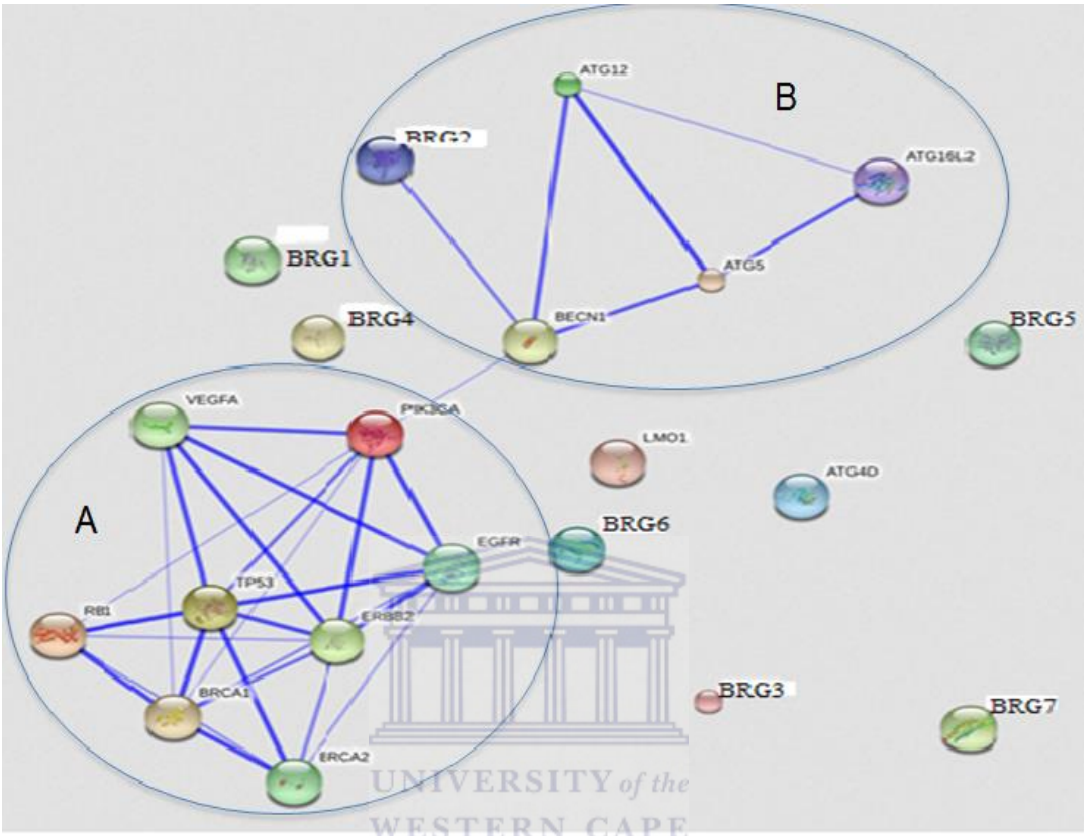


Figure 2.2: Protein interaction analysis of candidate genes using STRING. The figure showed the interaction between the gene products (BRG 1-7) and their link to Apoptosis (Circle A) and Autophagy (Circle B)

The seven proteins showed no direct interaction or linked with each other (Figure 2.2). This implies that these genes might not directly interact with each other, but may present an expression pattern for breast cancer onset or progression. Based on the criteria used by STRING to establish the interactions, the database might also not hold enough information to confirm the link between these proteins. However, it has been reported that sets of genes used as biomarkers for disease does not necessarily directly interact with each other (Barabási *et al.*, 2011).

To determine if these proteins modulate the processes of apoptosis, cell cycle and autophagy (three major hallmarks of cancer), the STRING software was used to query the interaction of the seven candidate genes and major proteins in the above mentioned processes. The result (Figure 2.2) suggest that most of the protein products of the candidate genes might not directly modulate these processes as no direct interaction was shown to well-known proteins linked to these processes. However, this does not imply that they are not involved in these pathways, but may be linked through interaction with intermediary proteins, which has not be reported or predicted to date by the STRING database. Most proteins functional linkages are not necessarily stable physical interaction. They may interact by catalyzing subsequent reactions in a metabolic pathway, regulate each other transcriptionally or post-transcriptionally, or jointly contribute to larger, structural assemblies without ever making direct contact (Franceschini *et al.*, 2013).

Figure 2.2 also showed that BRG 2 is directly linked to a protein (BECN1) involved with the biochemical process of autophagy and this protein directly linked apoptosis to autophagy. An interesting part of this result was that this protein was also indicated to be involved with most of the cancer processes examined in the functional annotation

analysis (Figure 2.1), but was not shown to be differentially expressed in the qRT-PCR analysis (Ngcoza, 2012).

2.4.4 Transcription Factors Associated with genes

Complete knowledge of all proteins that directly and/or indirectly interacts with a protein in a given cell type or condition would present an important milestone towards a comprehensive description of cellular mechanisms and functions (Franceschini *et al.*, 2013). These proteins could also be regulated by the same transcription factors.

The transcription factors (TF) were annotated to the putative genes using the Qiagen and/or regulatory elements and epigenetics data from the Genecards database (Safran *et al.*, 2010). Identifying genes to whose promoters either non-oncogenic or oncogenic transcription factors binds could provide insight into the transcriptional regulation of such genes and thus their role in the control of normal cell growth, differentiation as well as in cancer development.

The transcription factors that were predicted to target the promoter region of the candidate genes have been reported to be involve in cancer initiation, repression or progression (table 2.3). For example, p53 is a tumour suppressor protein that plays a critical role in most cancer types. Though this gene was initially thought to be an oncogene, it was later discovered that the mutation of the gene is involved in many common human malignancy (Hollstein *et al.*, 1991). This TF was annotated to regulate three of the candidate genes, all of which that did not show differential expression in breast cancer versus a normal breast cell line in the qRT-PCR analysis (Ngcoza, 2012).

Table 2.2: Transcription Factors associated with the seven putative breast cancer genes.

Gene	Regulatory Transcription Factor
Gene 1	CREB, p53, deltaCREB, MyoD, STAT3
Gene 2	AP-1, p53, ATF-2, CUTL1, Nkx2-5, c-Jun
Gene 3	AhR, Sox5, Pax-5, RREB-1, GATA-1, Arnt, ZID, NF-Y, CBF(2)
Gene 4	E2F-3a, E2F-4, E2F-5, FOXF2, Pax-2, Pax-2a CUTL1, E2F-2, E2F-1,E2F
Gene 5	Bach1, Pbx1a, GATA-3, AML1a, MEF-2, Nkx2-5, XBP-1, C/EBPalpha, MEF-2A, aMEF-2
Gene 6	c-Fos, p53, AP-1, GR-beta, ATF-2, STAT3, c-Jun
Gene 7	AP-2alpha isoform 3, AP-2alpha isoform 4, AP-2alpha isoform 2, AP-2alpha, AP-2alphaA

CREB has been shown to be over-expressed in leukemic blast cells from patients with acute myeloid leukemia (AML) and its over-expression was associated with poor prognosis in AML patients. However the over-expression of the CREB protein did not correlate closely with CREB mRNA levels, suggesting that post-transcriptional mechanisms may contribute to the elevated expression (Sakamoto and Frank, 2009). This TF showed association with one of the seven putative genes (Gene 1) as shown in table 2.2.

Table 2.2 showed three of the seven genes (Gene 2, Gene 6 and Gene 7) to be associated with a family of genes that encodes the AP transcription factors. The mammalian AP-1 protein is a dimer protein made up of a group of DNA binding protein comprising of the Jun, Fos and ATF subgroups of transcription factors. They have been found to regulate cell proliferation, neoplastic transformation and apoptosis (Shaulian and Karin, 2001). There is evidence that AP-1 proteins, mostly those that belong to the Jun group, control cell life and death through their ability to regulate the expression and function of cell cycle regulators such as Cyclin D1, p53, $p21^{cip1/waf1}$, $p19^{ARF}$ and p16 (Shaulian and Karin, 2001).

The Stat-3 transcription factors has been reported to be an oncogenic signaling pathway component whose expression suppresses tumour expression of pro-inflammatory mediators and whose block in expression increases expression of pro-inflammatory cytokines and chemokines that activate the innate and the adaptive immune systems (Wang *et al.*, 2004). E2A-Pbx1 is a fusion onco-protein found in about 10 % of acute leukemia in humans (Thorsteinsdottir *et al.*, 1999). This TF showed association to Gene 6 (table 2.2).

Table 2.2, also showed a list of transcription factors that have been directly linked to different types of breast cancer. Gene 2 and Gene 4 are two of the seven genes that were associated with CUTL1. CUTL1, is a known transcription factor that acts as a major regulator of cell motility and invasion and an important downstream effector of TGFβ, a key factor modulating cell migration, invasion and tumour progression. It was highly expressed in high-grade carcinomas and its expression is inversely correlated with survival in breast cancer (Michl *et al.*, 2005).

Over-expression of c-Jun (a transcription factor associated with Gene 6) in MCF-7 cells was reported to lead to an increased aggressiveness in cancer cell growth, as a result of an increased in cellular motility, and increased expression of a matrix-degrading enzyme (MMP-9). Breast cancer cell lines that over-expressed c-Jun were also reported to be unresponsive to estrogen and tamoxifen (Smith *et al.*, 1999). There is an indication that the over-expression of c-Jun represents an estrogen-independent phenotype of breast cancer. The observed phenotype for MCF-7 cells with c-Jun over-expression is similar to that observed clinically in advanced breast cancer, which do not respond to hormonal treatment (Smith *et al.*, 1999).

A germline mutation of the tumour suppressor gene BRCA1 is an indicator of an increased lifetime risk of breast and/or ovarian cancer (Cantor *et al.*, 2001). BACH1 is a nuclear protein that helps in DNA repairs by direct interaction with the highly conserved, C-terminal BRCT repeats of BRCA1. A disruption in this interaction as a result of mutation in the BRCA1 or BACH1 genes may lead to defects in DNA repairs, which eventually results to breast and/or ovarian cancers (Cantor *et al.*, 2003). There are reports on some women with early-onset breast cancer with no defects in either BRCA1 or BRCA2 possessing a mutated germline BACH1 coding sequence,

suggesting that a dysfunction in BACH1 contribute to the tumour induction in breast cancer (Cantor *et al.*, 2003). BACH1 was linked to Gene 5 (table 2.2).

The transcription factor GATA-3 was also linked to Gene 5 (table 2.2). This TF is majorly responsible for maintaining luminal epithelial differentiation in adult mammary gland, which raises important implications for the pathogenesis of breast cancer. It was shown that a deletion of GATA-3 in mice lead to severe defects in mammary development as a result of failure in terminal end buds formation during puberty. This resulted in an undifferentiated luminal cell expansion with basement-membrane detachment, which led to caspase mediated cell death (Kouros-Mehr *et al.*, 2006).

Table 2.2 also showed the AP-2 family of transcription factors to be linked to Gene 7. This AP-2 family of transcription factors has also strongly been implicated in different subtypes of breast cancer and modulations of genes associated with this transcription factor, based on their characteristics in these different types of breast cancer, could serve as good biomarkers (Vogel *et al.*, 2003).

Perissi *et al.* (2000), reported that the AP-2 transcription factors are implicated in the estrogenic regulation of the ERBB2 gene expression. The ERBB2 is an oncogene repressed by estrogen in human breast cancer (Perissi *et al.*, 2000). Also, the level of AP-2alpha and AP-2gamma were shown to be highly expressed in the majority of breast cancer cell lines over-expressing c-erbB-2 (Bosher *et al.*, 1996). Gene 7, the candidate gene associated with this transcription factor, was the most highly differentially expressed gene from the qRT-PCR result (Ngcoza, 2012). Taken together the association of these various TFs to the putative breast cancer biomarkers,

suggest that these genes could be good prognostic markers as well as markers for monitoring therapeutic outcomes in various types of breast cancers.

2.4.5 Expression pattern of the candidate biomarkers in different cancer types

The expression pattern of the candidate genes was monitored in different cancer types to determine their specificity using the gene expression atlas database (Kapushesky *et al.*, 2012). Only cancer types that showed expression result for four or more experiments for at least one of the candidate genes was considered. The result obtained showed the expression level of the seven putative genes has been monitored in breast, colon lung and prostate cancer (figure 2.3).

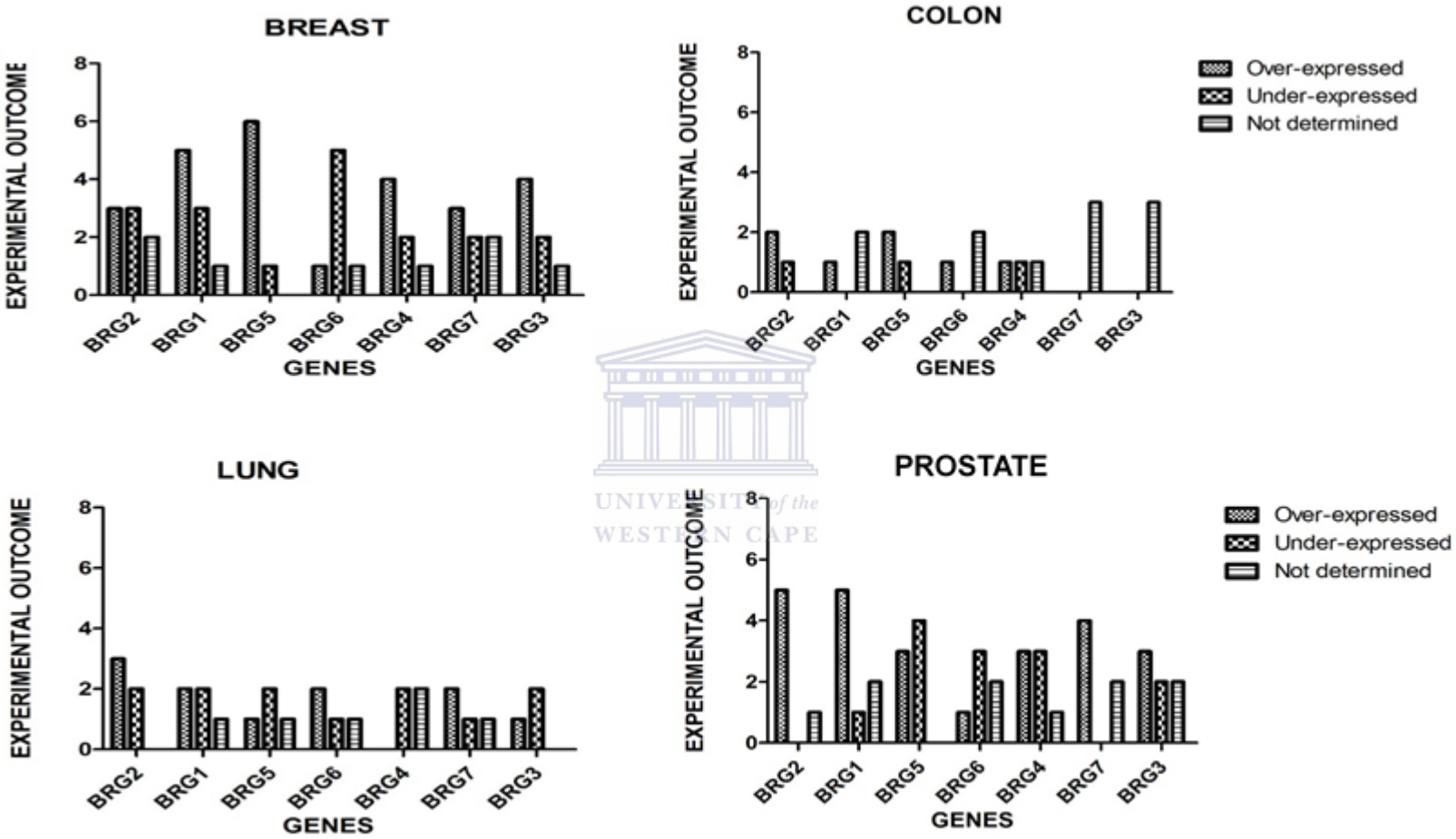


Figure 2.3: Gene expression pattern in breast, colon, lung and prostate cancer types.

Figure 2.3 shows that Gene 1 and Gene 5 in addition to the three genes (Gene 3. 4 and 7) that showed high expression levels from the qRT-PCR analysis (Ngoza, 2012) were highly differentially expressed in breast cancer based on this *in silico* analysis. The result obtained for Gene 1 and Gene 5 that were not shown to be differentially expressed based on qRT-PCR analysis, might be as a result of the methods used to determine their expression levels as reported in this database.

The Gene Expression Atlas is a representation of gene expression across tissues, cell types and cell lines under different biological conditions, including development stages, physiological states, disease state and phenotypes in different organism curated directly from array express Archive of Functional Genomics Experiments and the European Nucleotide Archive (Kapushesky *et al*., 2012). Identifying these genes to be differentially expressed in breast cancer types further validate them as valuable biomarker prospects for breast cancer diagnosis, prognosis or predictive purposes.

The seven putative genes were differentially expressed in other cancer types as well, which bring into concern the specificity of these genes as breast cancer biomarkers. However, the expression pattern of these genes in the different cancer types differs. For example, Gene 7 was differentially expressed in breast, lung and prostate cancer but not expressed in colon cancer from this *in silico* analysis (Figure 2.3). Gene 2 was highly expressed in prostate cancer but was differentially expressed in breast cancer as shown by the GEA output. However, Gene 2 was linked to most of the processes associated with cancer in the earlier sections (Figure 2.1 and 2.2), an indication that the expression of this gene could probably be repressed in breast cancer. These differential expression signatures serve as a suggestion that a combination of this set of genes used in combination will serve the purpose of biomarkers for breast cancer detection or monitoring.

Cancer cells are known to display a broad spectrum of genetic alterations that include gene rearrangements, point mutations and gene amplifications, leading to disturbances in molecular pathways. These alterations generates a pattern of secreted varying protein signatures in surrounding cells or at more distal and easily sampled sites of the body and could serve as biomarkers for better management of cancer (Bhatt *et al.*, 2010). The expression signature of these different candidate genes could be more appropriate as a biomarker panel for either breast cancer diagnosis or prognosis.

It has been reported that multiple proteins will be more appropriate as biomarkers for cancer detection and monitoring. Although single biomarker may serve in selected cases, there is growing consensus that multiple (e.g., three to five) markers used individually or as part of an integrated panel will be required for most applications, and the performance advantages of panels have been confirmed in several publications (Rifai *et al.*, 2006; Hanash *et al.*, 2008; Hu *et al.*, 2008).

2.4.6 Protein interaction analysis of BRG 7

To further determine the biochemical role of BRG 7, proteins that physically interact with this protein were assessed using the STRING database (figure 2.4).

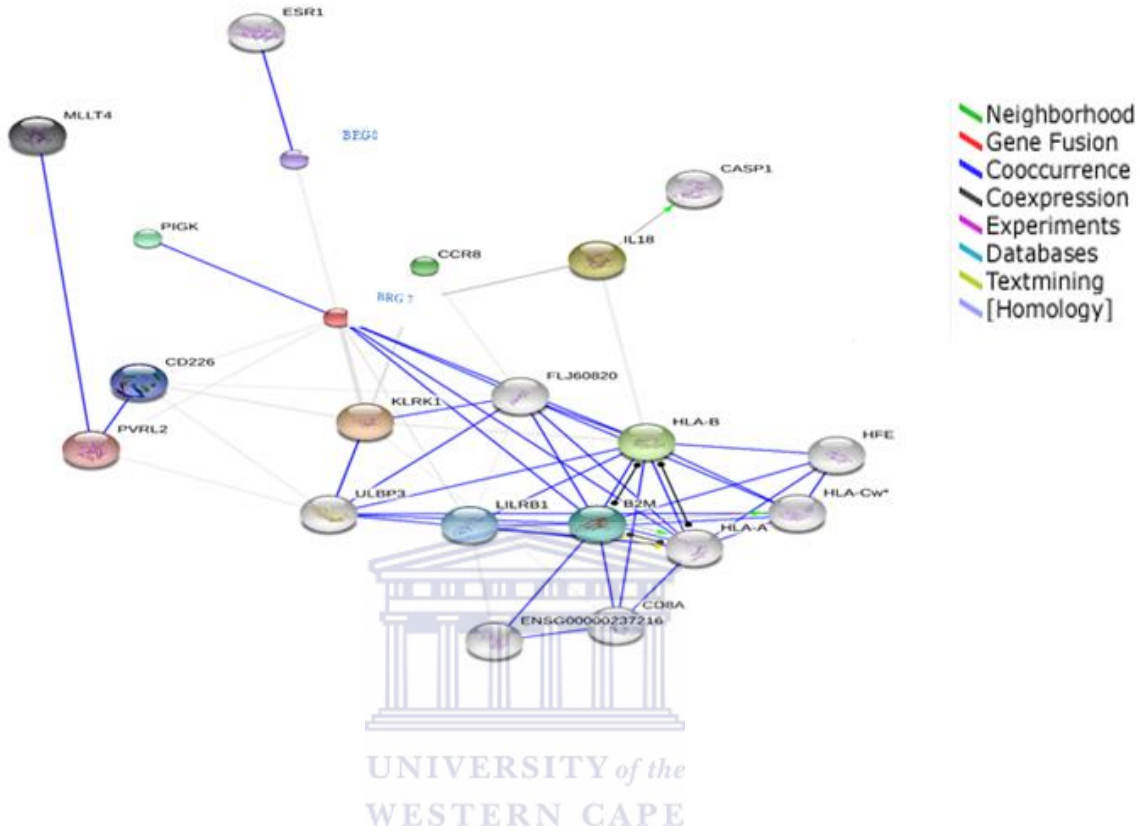


Figure 2.4: Protein interaction analysis of Gene 7 using STRING

Figure 2.4 shows the protein-protein interaction network of the protein product of Gene 7 (BRG 7). BRG 7 is a member of an immune-receptor ligand family expressed on tumour cells and it has been reported that cancer cells could escape from immune-surveillance by shedding off this protein (Paschen *et al.*, 2009). Most proteins that have direct physical interaction with BRG 7 are members of the immune-receptor ligands and their receptor proteins (figure 2.4). Apart from this group of proteins that interacted directly with BRG 7, another membrane bound protein (BRG 8) also interact with this protein. BRG 8 also showed a direct physical interaction with estrogen receptor protein (ESR1). This protein (whose gene and protein product will be referred to as Gene 8 and BRG 8 respectively in the subsequent chapters of this thesis) was also considered a very good putative breast cancer biomarker based on its direct interaction with BRG 7 and ESR1 as shown by STRING analysis (Figure 2.4).

BRG 8 has been reported to act as a mediator of transcriptional repression of an estrogen receptor (ER)-selective co-regulator that potentiates the inhibitory activities of anti-estrogens and represses the activity of estrogens (Manavathi *et al.*, 2014). Estrogen receptor (ER) status of invasive carcinoma has been reported to be a useful prognostic and predictive factor and has become a common practice in the management of breast carcinomas. ER positivity predicts for response to endocrine therapy such as anti-estrogen (tamoxifen) administration or ovarian suppression (Bauer *et al.*, 2007).

Five different morphological subtypes of invasive breast cancer were recently reported from microarray profiling analysis. They include: luminal A, luminal B, normal breast-like, HER2-overexpressing, and basal-like. The basal-like subtype is the most aggressive histologically and is normally characterized by negativity to estrogen receptor, progesterone receptor and HER2 expression. It accounts for about

15 % of breast cancer cases and it is associated with poor prognosis, unresponsiveness to the usual endocrine therapies, shorter survival, and BRCA1-related breast cancer (Bauer *et al.*, 2007). A direct link of this protein using STRING analysis, gives more credence to BRG 7 as a probable biomarker for breast cancer management.

Steroid hormone receptors (SHRs), utilize different tissue-specific co-regulators for their activity in a continuum manner. Co-regulators act as a bridging or helper molecules forming large protein complexes to modulate the activity on target gene chromatin. These co-regulators may positively or negatively influence the receptor's transcriptional activity by modifying the target gene chromatin. The co-regulators that enhance the transcriptional activity of SHRs are called co-activators, and those that decrease its activity are called co-repressors (Manavathi *et al.*, 2014). BRG 8 is a co-repressor of the estrogen receptor and could serve in conjunction with BRG 7, as good biomarker to distinguish different types of breast cancer from each other, however, more *in silico* and molecular analysis is required to validate this.

2.6 Conclusion

Good biomarkers are urgently required to improve diagnosis, prognosis, guide-targeted therapy, and monitor therapeutic responses across the various types of breast cancer. Currently thousands of proteins curated from genomic analyses and gene expression data have been predicted and published either as individual or a group of proteins as candidate cancer biomarkers (Kulasingam and Diamandis, 2008). Interest in proteins as biomarkers is based on their record as biomarkers for other disease states as a result of their function or family relation (Hanash *et al.*, 2008).

Understanding the molecular basis of cancer should provide means for early detection as well as better diagnosis, prognosis and staging of disease and aid in the evaluation of response to therapy, prevention and establishing novel treatment strategies (Seth and Watson, 2005).

The candidate genes from an earlier study (Ngcoza, 2012) were primarily identified from DNA microarray experiments that were published or stored in different databanks. DNA microarrays allows for simultaneous determination of the expression levels of thousands of genes based on certain treatments, diseases, or at the developmental stages of such diseases, and it has been widely applied in cancer research for better diagnosis and prediction of the disease states (Huang *et al.*, 2010). Functional analyses done in this chapter using an *in silico* approach, confirmed these genes as having good prospects as potential biomarkers for breast cancer based on their involvement in different cancer processes.

In addition to the three genes (Gene 3, Gene 4 and Gene 7) identified to be differentially expressed from the qRT-PCR result (Ngcoza, 2012), the protein products of two other genes were also hypothesized to be good candidates as breast

cancer biomarkers based on the *in silico* analysis performed in this chapter. One of the two genes (Gene 2) was part of the seven genes identified in the previous study but was not shown to be differentially expressed in a breast cancer versus normal cell line based on qRT-PCR analysis. This gene was found to be involved in most of the processes linked to breast cancer and cancer in general based on the GO, TF and protein interaction analysis. We postulate that the protein expression pattern of this protein could serve as an additional means for monitoring breast cancer onset, progression as well as monitor therapy.

The specific differential expression signatures of the candidate genes in different cancer types, suggest that a single candidate gene or protein may not be very specific as a breast cancer biomarker, but combining these genes/proteins may be more appropriate.

An additional candidate biomarker (BRG 8) was identified as it showed direct interaction with BRG 7 through STRING analysis. This protein is also reported to modulate the expression of estrogen receptor (ESR), a prognostic factor in a large group of breast cancer types.

Based on the result from this chapter, it was concluded that the protein products of these genes or at least some of them could serve as a panel of biomarkers for diagnosis, prognosis or predictive markers for breast cancer management. To further validate these genes as biomarkers, their prognostic value was analysed using different bioinformatics tools in the subsequent chapter.

Chapter 3: Prognostic and Predictive validation of Genes using an *in silico* approach

3.1: Background

After the introduction of markers such as the estrogen receptor and HER2 for the evaluation of the clinical course of breast cancer, the search for breast cancer biomarker is now shifting towards the identification of gene or protein expression levels capable of predicting the outcome or prognosis of breast cancer (Györffy *et al.*, 2010). Clinician used these signatures when examining early stage cancers, which aid in evaluating the type of therapy best, suited or if adjuvant therapy is required after surgical resection (Khaleel *et al.*, 2014). Identifying cell surface genes whose expression signatures and protein products could serve as biomarkers for breast cancer outcomes, would be a useful addition to current methods used in breast cancer management.

Breast cancer still remains one of the leading causes of cancer deaths despite the increase in the availability of effective adjuvant medical treatment. Although regular mammography screening programs provides a means of detecting early stage (<2cm), node-negative breast cancer and bring about better prognosis, which also contribute to improved outcomes, there is still a need to avoid overtreatment in patients who hardly benefit, whilst suffering from toxic side effects. On the other hand, under treatment or incorrect treatment also have to be avoided (Weigel and Dowset, 2010). Thus, the need for identifying breast cancer biomarkers that could serve as prognostic or predictive indicators of the disease.

Prognostic biomarkers aims to predict objectively and independently patient clinical outcome before treatment. These biomarkers must have the capacity to define

outcome in patient cohorts that did not undergo adjuvant treatment. Whilst the predictive biomarkers tends to foretell the response of a patient to a specific therapeutic intervention and are associated with tumour sensitivity or resistance to a type of therapy (Weigel and Dowset. 2010).

Classical clinical features used for cancer prognosis are mostly pathological practices, which include tumour size, histological subtype and grade, lymph node metastases, and lymphovascular invasion. The TNM (tumour, nodes, and metastasis) system integrates these into tumour staging, which have major prognostic value by careful histological examination of a primary breast cancer sample (Weigel and Dowset. 2010). However, the advent of high-throughput methods, have provided large data on genes and protein candidates that can serve as prognostic and predictive biomarkers (Hayes *et al.*, 2007) for a more effective and less laborious means of evaluating the prognosis of a breast cancer patient as suppose to the pathological methods.

Gene-expression profiling of tumourigenic breast-cancer cells has been used as a prognostic tool in assessing survival in patient with breast cancer (Sotiriou, *et al.*, 2003). Liu *et al.*, (2007) identified 187 genes that are differentially expressed in tumourigenic breast cancer cells compared to normal breast epithelium and reported a distinct gene signature for the disease that differs from an earlier reported gene signature in breast cancer. They concluded that this signature is an indication that a breast tumour would likely metastasize and their study referred to it as “invasiveness” gene signature (IGS). The IGS was also reported to be associated with the risk of death and metastasis not only in breast cancer but also in lung cancer, prostate cancer, and medulloblastoma, suggesting that the IGS could be a general biological feature shared by several different types of tumours (Liu *et al.*, 2007).

With the advent of high throughput technologies, like microarray analysis and qRT-PCR for gene expression profiling, many candidate genes have been identified and suggested to be good biomarkers for monitoring outcomes in breast cancer treatment, however most of these candidate biomarkers have not been validated either in multivariate analyses nor was their discriminative power validated in large clinical cohorts (Győrffy, *et al.*, 2012).

The discovery of biomarkers must be followed up with validation studies. Before drawing conclusion about the expression of candidate genes, independent experimental validation using several techniques such as reverse transcriptase-PCR, Northern blot analysis, or tissue microarrays on human samples that generates large number of datasets, should be carried out (Rhodes, *et al.*, 2002). Uniting several datasets from different studies, employing several statistical methods can enhance and provide more accurate estimates of biomarkers for cancer diagnosis and prognosis (Gyorffy, *et al.*, 2013). With the increasing number of publicly available datasets, a meta-analysis of multiple datasets that address similar hypotheses can be used to validate and statistically assess all the valuable candidate biomarkers simultaneously, potentially yielding significant, differentially expressed genes, void of inadequacies from individual studies (Rhodes, *et al.*, 2002).

For the analysis of candidate biomarkers for their prognostic value, several tools such as ITTACA, KMPlot, RecurrenceOnline, bc- GeneExMiner, GOBO, PrognoScan, and many others have been proposed (Aguirre-Gamboa, *et al.*, 2013). Some of these survival analysis data tools were employed to validate the eight candidate biomarkers in this study.

3.2: Kaplan Meier plot

The relative gene expression level using the microarray technology is regarded as an effective preliminary test tool to evaluate prognostic biomarkers (Fan, *et al.*, 2006). A prudent means of exploiting these biomarkers is by looking into already existing clinical microarray cohorts generated (Györffy *et al.*, 2010). Validating candidate genes as prognostic or predictive biomarkers in well researched breast cancer cohorts is of utmost interest in monitoring treatment and predicting the correct treatment (Weigel and Dowsett, 2010).

The Kaplan–Meier estimator (also known as the product limit estimator) is one of the best statistical options used in estimating the fraction of subjects living for a certain number of years after treatment. The effect of an intervention in clinical trials or community trials is assessed by measuring the number of subjects that survived after an intervention over a period of time (Goel *et al.*, 2010).

It is one of the simplest statistical tool that can be used to compute and estimate the survival of human subjects over time in spite of all the difficulties associated with subjects or situations. For example, it takes into account losses from the cohort, if a patient withdraws from a study, by computing the probabilities of occurrence of the event at the point of exit and multiplying this successive probabilities by any earlier computed probabilities to get the final estimate. The association between a clinical parameter (or biomarker) and survival can be visualized by drawing a Kaplan–Meier plot in which patients are split into groups according to the parameter (Györffy *et al.*, 2010).

3.1.2: Kaplan Meier plotter®

The Kaplan Meier plotter is an online statistical tool used to draw survival plots, to determine the clinical relevance of the expression levels of various genes on the outcome of both treated and untreated cancer patients (Győrffy *et al.*, 2010). The tool can assess the effect of 22,277 genes from 10,188 cancer samples, which include 4,142 breast, 1,648 ovarian, 2,437 lung, and 765 gastric cancer patients. The background database is manually curated and handled by a MySQL server, which simultaneously curate gene expression and clinical data of different cancer patients. The data is analysed using the R statistical tool for calculations (Győrffy *et al.*, 2010). The package calculates the ‘‘survival’’ level against input gene and the Kaplan–Meier survival curves is plotted, indicating the number of patient at risk below the main plot. The hazard ratio with 95 % confidence intervals and logrank P are calculated and displayed on the plot. The feedback is received on the webpage. The Kaplan–Meier plotter is set up using a central server, which can be accessed over the Internet with a primary purpose of meta-analysis based *in silico* biomarker assessment (Győrffy *et al.*, 2013).

3.2: SurvExpress®

The SurvExpress (http://bioinformatica.mty.itesm.mx/SurvExpress) is a cancer-wide gene expression database with clinical outcomes and a web based tool that provides survival analysis and risk assessment of cancer datasets. It is a bioinformatics tool used to look at the performance, validates survival and prognostic biomarkers for cancer outcomes. The tool is a cancer dataset that comprises 20,000 samples and 130 datasets with censors’ clinical information covering over 20 different types of tissues. The web interface can easily be used to perform *in silico* biomarker validation, and compare genes within the database. The Input page accepts gene list based on NCBI

compatible gene identifiers (official symbol, Entrez, Ensembl, HGNC, or others) and dataset of interest can be selected from the list. Users can also choose how to treat genes detected using different experimental methods (Aguirre-Gamboa *et al.*, 2013). Compared to most tools for analysing the validity of biomarkers for cancer outcomes, SurvExpress is the largest and the most versatile free tool for analysing differentially expressed multi-gene biomarkers in human cancers at the same time (Aguirre-Gamboa *et al.*, 2013). The SurvExpress results include a Kaplan-Meier curve for risk groups, clinical information available related to risk group, heat map representation of the gene expression values, a box plot across risk groups, and tables with the summary of the Cox fitting and the prognostic indexes (Aguirre-Gamboa *et al.*, 2013).

3.3: Gene Expression-Based Outcome for Breast Cancer Online (GOBO)

Various databases provide large repertoire of prognostic and/or treatment predictive gene expression studies on breast cancer e.g. Gene Expression Omnibus (Hess *et al.*, 2006). The major aim of the Gene Expression-Based Outcome for Breast Cancer Online (GOBO) an online bioinformatics tool (http://co.bmc.lu.se/gobo) is to provide a means to examine the differential expression level of various genes in an 1881 sample breast tumour dataset and a 51 breast cancer cell lines dataset, generated on Affymetrix U133A microarrays. The 1881-sample breast tumour set comprises 11 public datasets analysed using Affymetrix U133A arrays. GOBO allows the assessment of precompiled datasets to be interrogated for three main classes which are: gene set analysis (GSA) in different subgroups of breast tumours and breast cancer cell lines, identification of co-expressed genes (CG), and sample prediction (SP) which is an expression levels of single genes, sets of genes, or gene signatures in multiple subgroups of the 1881-sample breast cancer data set. GSA analysis using the GOBO web interface is further divided into outcome analysis in breast tumours (GSA-

Tumour) and expression patterns in breast cancer cell lines (GSA-Cell line) (Ringnér *et al.*, 2011).

3.4: Aim of study

The aim of this chapter was to evaluate the prognostic and predictive values of the identified putative gene as biomarkers using different bioinformatics tools.

3.4.1 Objectives

- Prognostic/ predictive validation of candidate genes using two datasets from SurvExpress
- Predictive analysis of individual biomarker candidate genes using the Kaplan Meier plotter
- Prognosis/ Predictive analysis of candidate genes for different breast cancer types using GOBO

3.5: Materials and methods

3.5.1: SurvExpress

The online biomarker validation tool SurvExpress was assessed online at http://bioinformatica.mty.itesm.mx:8080/Biomatec/SurvivaX.jsp. The eight candidate gene symbols were used as input in the space provided for gene list and "breast" option was selected as tissue type and the default setting was used for the option "duplicated genes". The genes were then analyzed twice using two different broad datasets; (1) a dataset comprising 9 datasets from the same platform compiled by seven authors using 1574 samples (for this study termed DATA 1) and (2) metabases constructed by performing a quantile normalization compiled by SurvExpress with 1901 samples (termed DATA 2). Both databases are recurrence, meta-analysis clinical data (Aguirre-Gamboa *et al.*, 2013). The analysed button was clicked and on the next page, "Recurrence. Free. Survival. Years" was selected and the output was downloaded in pdf. Format.

3.5.2: Kaplan-Meier Plotter

Kaplan Meier plotter an online cancer survival analysis tool was used to assessed the effect of each candidate gene on breast cancer outcome at http://kmplot.com/analysis/index.php?p=service&cancer=breast. "Breast" was chosen as cancer type and each gene symbol for the candidate genes were used as input into the gene space provided. All other default setting was used after which the "draw Kaplan-Meier plot" option was selected. The output results were downloaded in pdf. Format.

3.5.3: GOBO databases gene analysis

GOBO (Gene Expression-Based Outcome for Breast Cancer Online) database (http://co.bmc.lu.se/gobo) was used to validate the prognostic value of the candidate genes. The gene sets were queried in the Gene Set Analysis-outcome in the breast tumours (GSA-Tumours) application. The gene symbols was used as input and the option, "different tumour types" were selected and analyzed one at a time, using the 3 groups (quantiles) and "full years censoring" on the default settings (Ringnér, *et al.*, 2011).

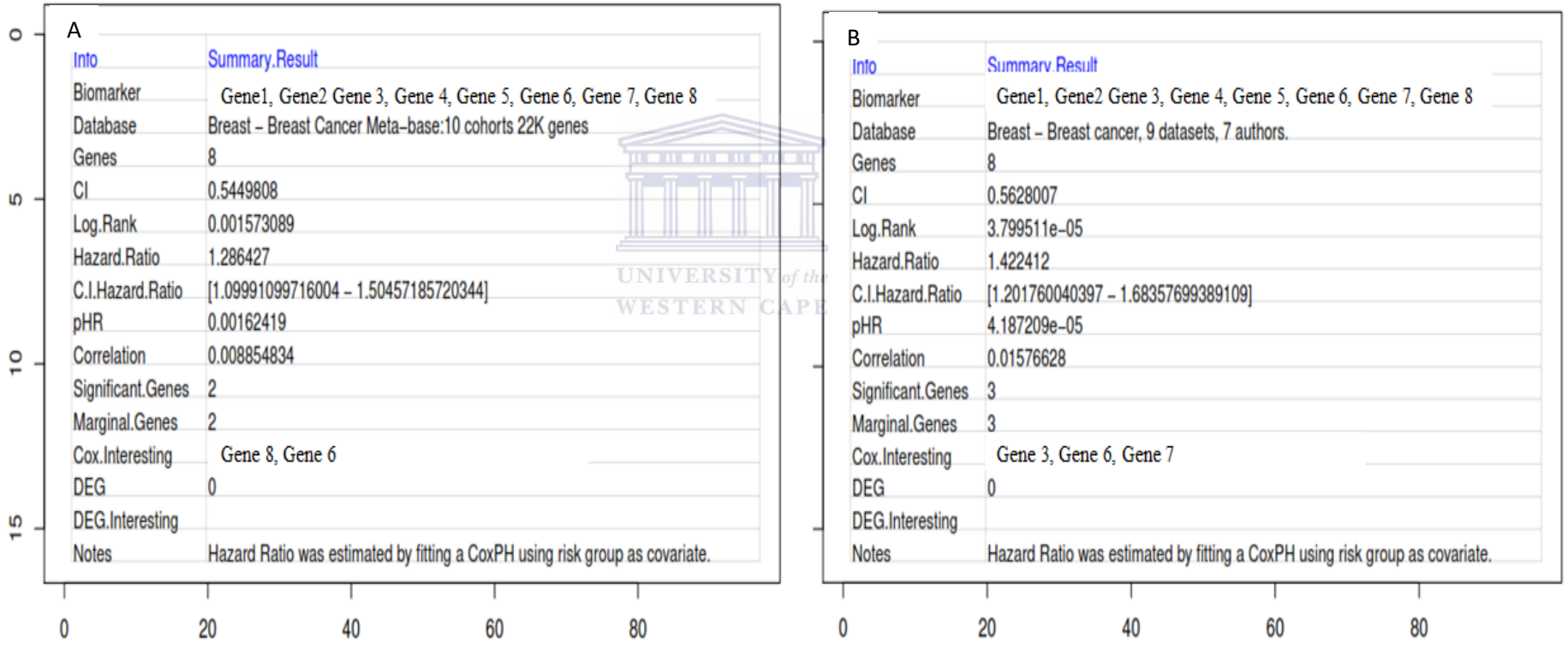
3.6: Results and Discussion

To validate the prognostic and/or predictive value of the identified candidate genes as a set of biomarkers, several prognostic software and databases were used.

3.6.1: SurvExpress

Using the SurvExpress tool, the prognostic ability of this set of candidate biomarkers was determined using two broad datasets; (1) One comprising 9 datasets from the same platform/methods compiled by seven authors comprising 1574 samples (DATA 1) and (2) meta-bases constructed by performing a quantile normalization compiled by SurvExpress comprising 1901 samples (DATA 2). From the results, three genes out of the eight showed promise as good distinguishable prognostic markers based on their p-value ($p < 0.05$) in DATA 1, while two genes were indicated as significant genes in DATA 2 (table 3.1).

Table 3.1: Result summaries of gene expression analysis showing significant genes based on p-values ($p \leq 0.05$) of individual genes from two datasets; (a) DATA 2 and (b) DATA 1 using SurvExpress



From table 3.1, DATA 1 shows 2 genes (Gene 6 and Gene 8) were significantly expressed whilst DATA 2 shows 3 genes (Gene 3, Gene 6 and Gene 7) to be significantly expressed. Gene 6 was common in the output from both datasets. Interestingly this gene was highly differentially expressed in a breast cancer versus normal cell line from the qRT-PCR result (Ngcoza, 2012). Two of the three significant genes from DATA 2 (table 3.1) were also found to be highly differentially expressed in the qRT-PCR experiment using a breast cancer cell line versus a non-cancer breast cell line (Ngcoza, 2012), while the second significant gene from DATA 2 was Gene 8, the additional candidate gene identified from the earlier *in silico* work (Chapter 2).

Figure 3.1 shows the Kaplan-Meier plots by risk group, the log-rank test of differences between risk groups, the hazard-ratio estimate, and the concordance indeces, which estimate the probability that subjects with a higher risk (red) will experience relapse after subjects with a lower risk (green) for both databases. Unlike the variance in significant genes in both datasets, these figures suggest that the genes when combined as biomarkers could significantly predict the prognostic outcome of breast cancer patient based on the difference in outcome between the higher and the lower risk groups. The difference in output from these two datasets showed different individual biomarkers of significance, might be as a result of different approaches used in identifying and compiling these databases. This can be corroborated by a study that showed that the performance of biomarkers may differ in different populations, based on clinical information, probes per gene, the gene expression technology and the conditions used (Aguirre-Gamboa, *et al*., 2013).

However the significant output shown by the Kaplan-Meier plots of the combine genes from the two datasets suggest that our candidate gene list might serve as good prognostic biomarkers for breast cancer. Although a database with the highest

population of patients was selected for this analysis, the gene expression technology and clinical information of individual patient in the datasets may have affected the performance output of the individual candidate genes.

In accordance with our approach of validating these candidate genes using the SurvExpress analytical tool, Cronin, *et al.*, (2007) used this software to analyse 16 biomarker genes to predict the likelihood of breast cancer recurring in early-stage, node-negative, estrogen receptor–positive breast cancer patients. These genes are currently used in a multigene commercial test kit known as Oncotype DX™, a clinically validated, high-complexity, multianalyte reverse transcription–PCR genomic test tool (Cronin *et al.*, 2007).

This kit was used to analyzed four different breast cancer patients datasets, and the result showed that though some of the individual genes may not adequately separate the low and high risk groups, the Kaplan-Meier plots based on the concordance indexes confirmed that the Oncotype DX™ could significantly separate low and high-risk groups in the four datasets tested (Aguirre-Gamboa, *et al.*, 2013). This result is similar to results obtained from this chapter, using the seven candidate genes to monitored breast cancer treatment outcome in the two largest dataset sets using the SurvExpress analytic tool.

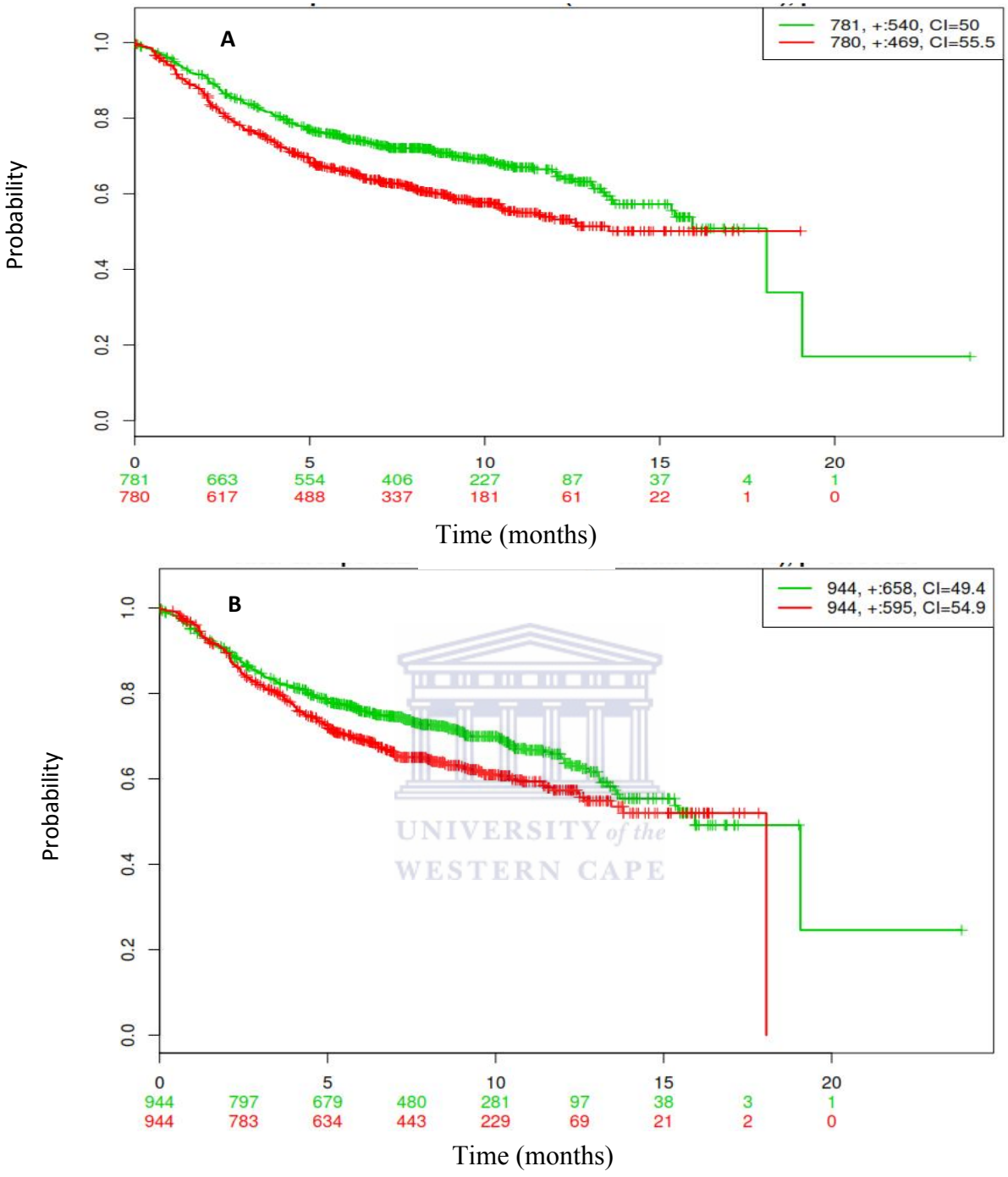


Figure 3.1: Kaplan–Meier analysis for all candidate genes for breast cancer outcomes. By class: treatment/ pre-surgery (overall betas) concordance index = 61.25, log-rank equal curves p=0.02728, r^2=0.043/0.921 risk groups hazard ratio = 1.88 (conf. int. 1.06 ~ 3.32), p=0.02987. a) DATA 1. b) DATA 2.

Two other result output types from SurvExpress analysis were a box plot graph and heat map showing the expression level of each of the candidate genes (Fig, 3.3 and 3.4). The gene expression of each gene is plotted along risk groups obtained in the analysis for the box plot. The box plot indicates whether gene expression levels are different between risk groups based on visual difference (high risk group are represented with red and low risk groups in green) and also the p-value ($p \leq 0.05$). Whilst the heat map shows the expression level (by colour) of each candidate genes against samples ranked by their prognostic index.

Box plots compare the difference of gene expression between risk groups using a t-test. The box plots confirmed the differential expression of most of the candidate genes in both datasets. From DATA 1 seven of the eight candidate genes were differentially expressed whilst five were differentially expressed from DATA 2 ($p < 0.05$).

The heat map ranked the genes based on their prognostic ability. Figure 3.3 showed that most of these genes are highly differentially expressed in the high-risk group ($p < 0.05$). This is an indication that high expression of these genes will present a poor prognosis in breast cancer patients, and the expression signature could be complemented with the repreessed genes in predicting treatment outcome. Figure 3.3 showed two of the highly differentially expressed genes (Gene 4 and Gene 7) from the qRT-PCR analysis (Ngcoza, 2012), were among the three highly ranked genes in the DATA 2, while the third gene (Gene 3) and Gene 8, the gene identified from the earlier *in silico* work (chapter 2), were highly ranked in DATA 1.

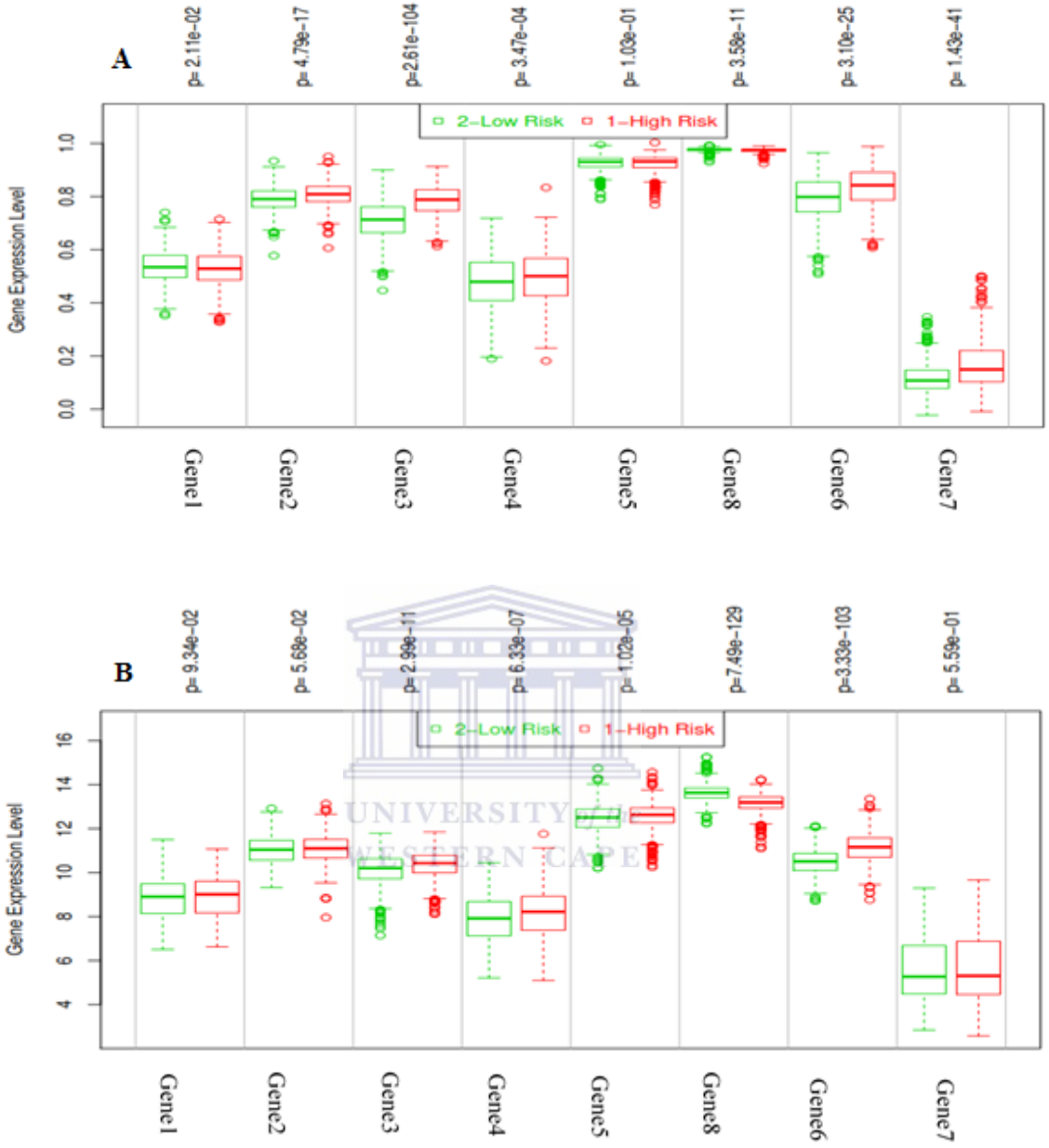


Figure 3.2: Gene expression values across gene groups represented by box plots together with the p-value ($p \leq 0.05$) of the corresponding difference comparing risk groups (a) DATA 1 (b) DATA 2

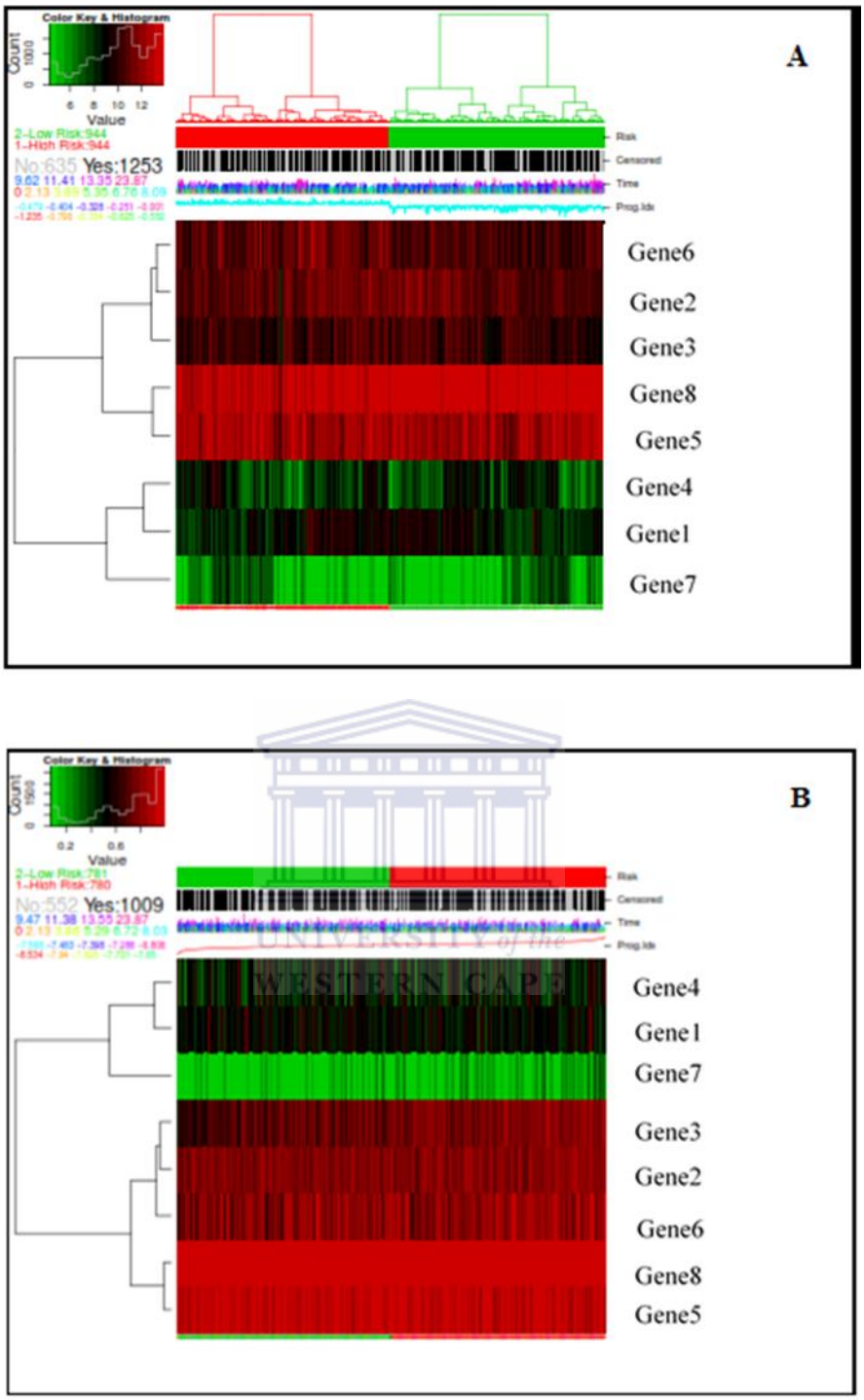


Figure 3.3: Expression profile and gene ranking based on prognostic index. (a) Expression profile from the datasets compiled by SurvExpress comprising 1901 samples. (b) Expression profile from dataset of nine databases compiled by seven authors comprising 1574 samples.

The difference in expression signature comparing the two datasets could be due to variance in cancer types and subtypes. Other major prognostic factors such as, age, tumour size, status of axillary lymph nodes, histologic type of the tumour, pathological grade, and hormone-receptor status could affect the expression pattern of these genes (Fulford, *et al.*, 2007). Vijver *et al.*, 2002 classified patients with primary breast carcinomas using a 70-gene prognosis profile, earlier identified as a gene-expression signature associated with either poor or good prognosis and concluded that the gene-expression profile studied was a more powerful predictor of the outcome of disease in young patients irrespective of the lymph node status. Fan *et al.*, 2006 concluded in their work, that the lack of similarities in performance of different gene expression signatures for breast cancer prognosis, derived from almost similar studies, are not completely known, but may include differences in the patient cohorts, microarray platforms, and mathematical methods of analysis.

3.6.2: Kaplan-Meier plotter

To further determine the prognostic value of the individual biomarker candidates in recurrence of breast cancer, the Kaplan-Meier plotter was used (www.kmplot.com). This software uses gene expression microarray data and survival information of 2,324 patients with breast cancer downloaded from GEO (Affymettrix HGU133A and HGU133+2) to present the expression of an individual gene in patient survival outcome (Romagnoli *et al.*, 2012). High expression level of some of these genes was significantly correlated with poor outcome of breast cancer, while the expression of others did not show a significant distinguishing outcome ($p \leq 0.05$). Genes of major interest were the three genes that were differentially expressed in the qRT-PCR analysis (Ngcoza, 2012).

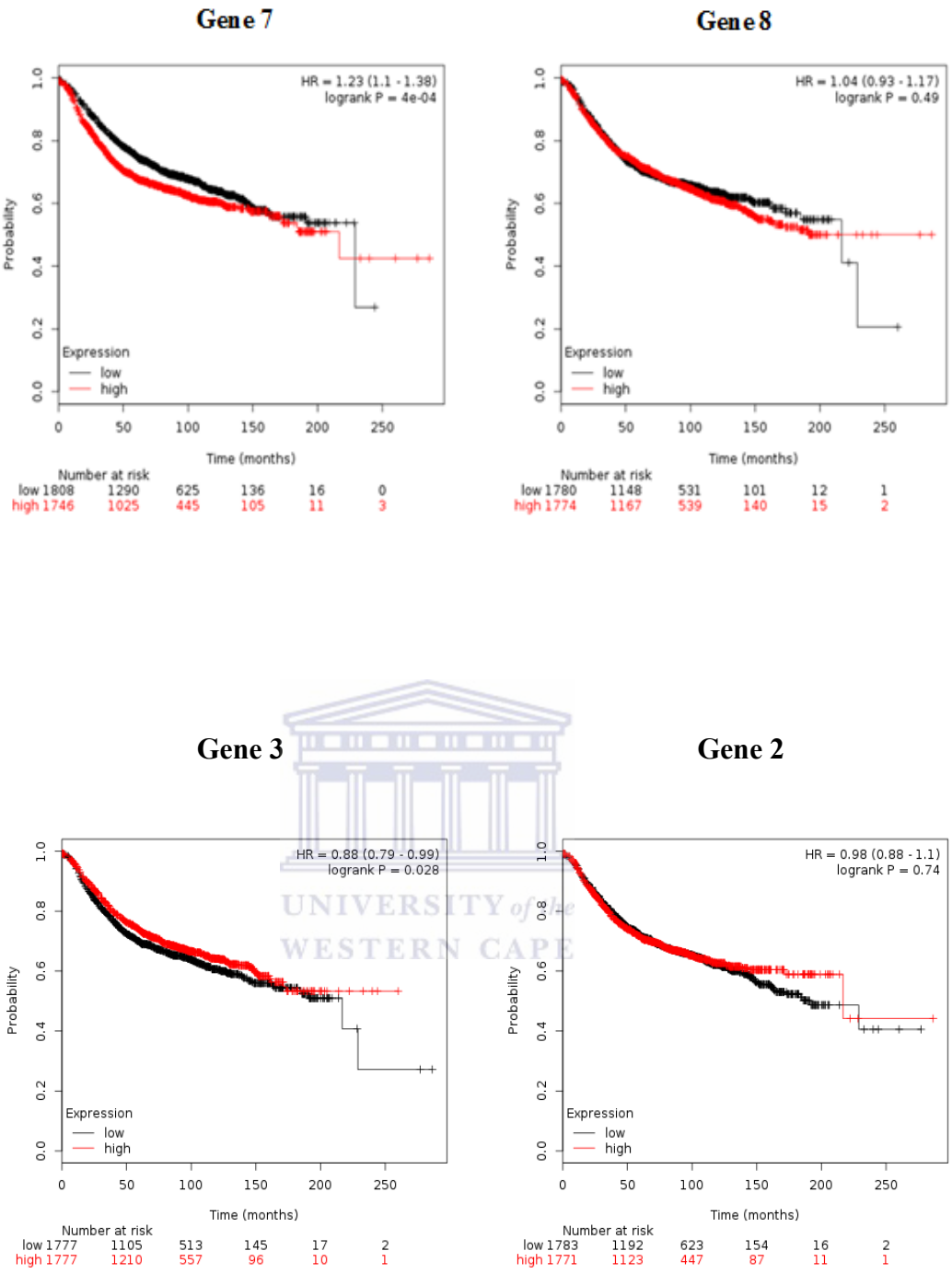


Figure 3.4: Survival risk curves for Gene 2, 3, 7 and 8. Low and high risks are drawn in black and red respectively. The p-value was shown in the right hand corner of each curve ($p \leq 0.05$ is significant)

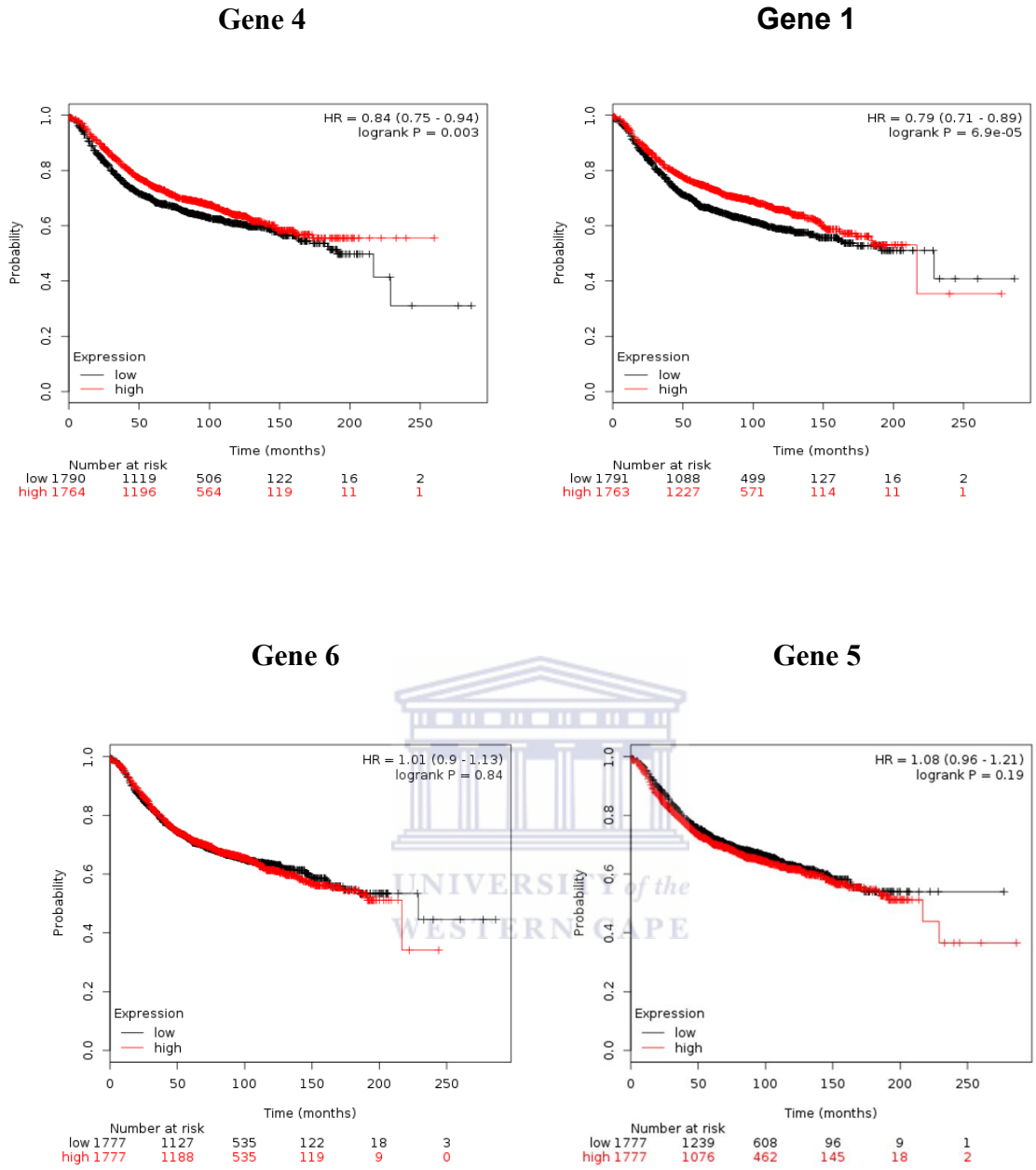


Figure 3.5: Survival risk curves for genes 1, 4, 5 and 6. Low and high risks are drawn in black and respectively. The p-value was shown in the right hand corner of each curve ($p \leq 0.05$ is significant)

Figures 3.4 and 3.5 showed difference in the survival rate of risk groups, based on the expression of the individual genes. From the survival curves, Gene 3, Gene 4 and Gene 7 have a significant p-value in predicting the prognostic outcome based on the differential value of the biomarkers in breast cancer patients (Figure 3.4), however, Gene 8 did not show significant distinguishing value in predicting the disease outcome (fig 3.5). Also in addition to the three genes that showed significant p value, the Kaplan-Meier plots showed Gene 1, potentially have a good prognostic value based on its p-value from this analysis (Figure 3.5). However, this gene was not highly differentially expressed in the qRT-PCR analysis (Ngcoza, 2012) and interestingly, this gene was not listed among the genes of significance in the SurvExpress analysis report from both datasets used (Section, 3.6.1).

The variations observed using different bioinformatics tools might be as a result of the different sample types used in assessing the value of these biomarkers in the various databases. Variation in gene expression has been reported in human mammary epithelial cells and breast cancer tissues, which are two different sample types. A publication suggested that different clusters of genes with coherent expression patterns in cultured cells are different from breast tumour samples and the difference were due to specific features of biological variation among the samples (Buess *et al.*, 2007). Also, the Kaplan-Meier plotter is known to access a database of breast cancer patients, undergoing different types of treatment, whilst the expression analysis using qRT-PCR, involved the use of only two types of breast cell lines (MCF-7 and MCF-12) which are not total representation of the heterogeneous nature of a breast tumour and a normal breast tissue.

Holliday and Speirs (2011) reported that there is a limitation to the extent to which a single cell line will mirror the clinical samples, covering specialised histopathological

types and maintaining the phenotype of a cell line in cell culture just like the breast tumour *in vivo.* However, the tremendous advances in the understanding of the biology of breast cancer were made and still rely on the use of cell lines (Holliday and Speirs 2011).

The approach used in this study to validate the candidate genes as breast cancer biomarkers was similar to that of Romagnoli, *et al.*, (2012). Using the Kaplan Meier plotter, they identified and analyzed the expression pattern of BMP-5 in different cancer types and concluded that the low expression of this gene significantly correlated with breast cancer survival, especially in ERα-negative breast tumours (Romagnoli, *et al.*, 2012).

3.6.3: Gene Expression-Based Outcome for Breast Cancer Online (GOBO)

The effective prognostic ability of this set of candidate biomarkers was determined in different subtypes of breast cancer, using the Gene Expression-Based Outcome for Breast Cancer Online (GOBO) and the results represented with survival curves (Kaplan-Meier plots). Five major subtypes of morphologically similar breast cancer types namely: luminal A, luminal B, normal breast-like, HER2-overexpressing, and basal-like have been identified based on microarray profiling of invasive breast carcinoma (Turashvili, *et al.*, 2011). The basal-like subtype is generally characterized by non-expression of estrogen receptor (ER), progesterone receptor (PR), and HER2 (Bauer *et al.*, 2007).

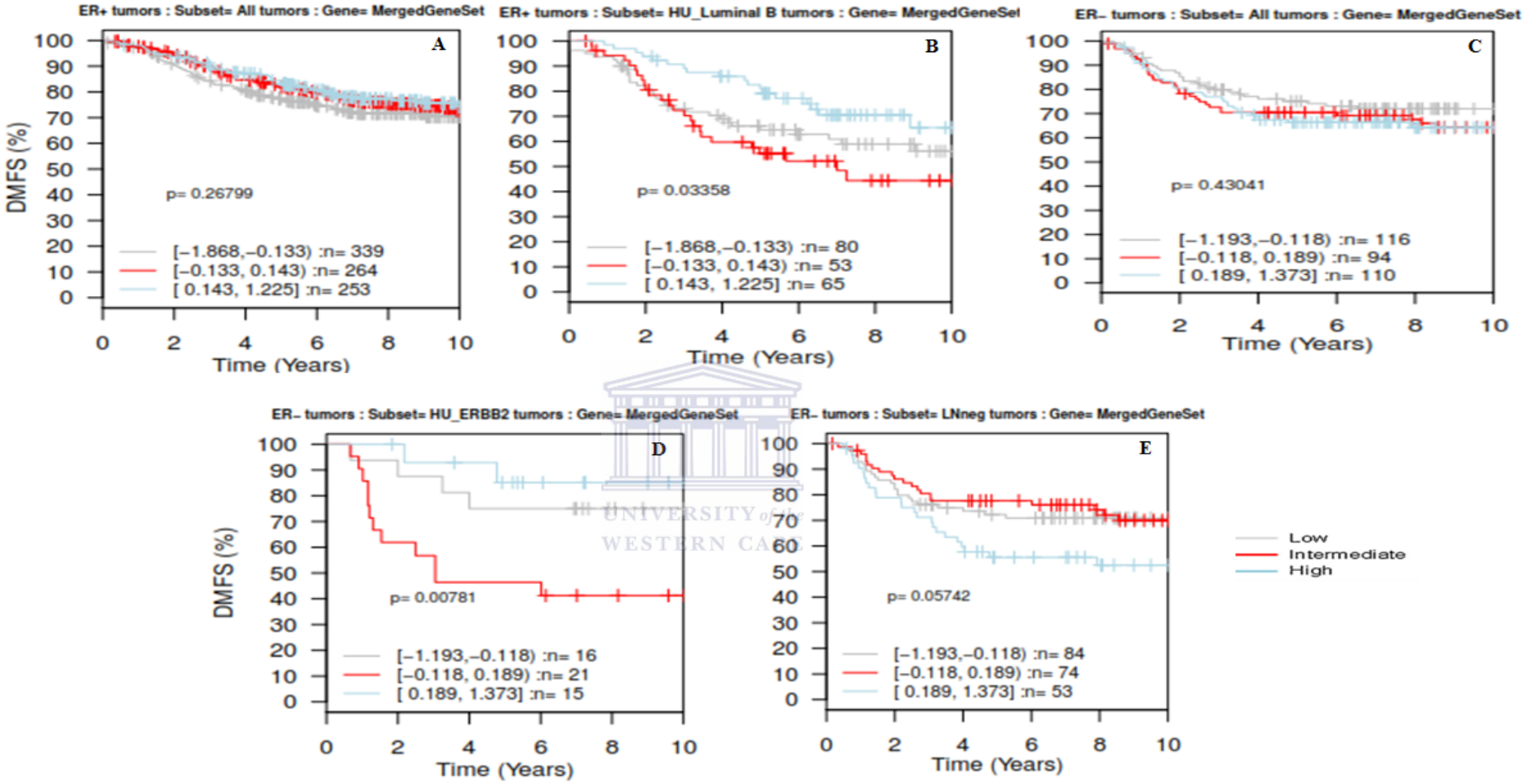


Figure 3.6: Survival curve for different breast cancer subtypes (a) ER+ all tumour subtype (b) ER+ luminal B tumour subtype (c) ER- all tumour subtype (d) ER- ERBB2 tumour subtype (e) ER- luminal negative tumours subtype. ($p \leq 0.05$)

Figures 3.6 represent the survival outcomes in low-risk, intermediate-risk and high-risk patients based on the candidate genes expression pattern in different cancer types. The Kaplan-Meier analysis is performed on different stratified groups based on gene expression in different subtypes and presented as percentage distant metastasis-free survival (DMFS) against time in years. The output from this analysis indicated that this set of candidate biomarkers may not predict effectively all estrogen receptor positive (ER+) breast cancer subtype based on the Kaplan-Meier plot ($p = 0.26799$), however the candidate biomarkers were effective in predicting the outcome of luminal B ER+ breast cancer subtype based on the Kaplan-Meier plot p-value ($p \leq 0.05$) (Figure 3.6).

Also, the GOBO output showed this set of genes could be accurately used to predict the survival outcome of ERBB2 subtype of estrogen receptor negative (ER-) tumours, and lymph node negative subtype of ER- tumours, but not all subtype of ER- tumours based on the individual survival curves outcome for these diseases (Figure 3.6). The percentage DMFS difference between survival groups for a ten year period in ER- tumour patients based on the expression of these genes/proteins was not statistically significant ($p = 0.43041$), while the difference between the risk groups in the ER- human ERBB2 tumour patients ($p = 0.00781$) and the ER- luminal negative patients ($p = 0.05742$), were statically significant ($p \leq 0.05$).

The luminal types of breast cancer exhibit distinct characteristics, features, and difference in response to different drugs. Luminal B breast cancer has been reported to exhibits worse prognosis as compared to luminal A. The luminal B cancer types expresses lower hormone receptors, higher expression of proliferating markers and possess higher histological grade than luminal A (Ades, *et al.*, 2014). The ability of predicting the outcome in luminal B cancer type patients, present a promising set of

biomarkers that could assist in identifying patients with this subtype of breast cancer, monitor their treatments, and predict their treatment outcome, thereby, providing better prognosis for this type of breast cancer patients.

Matboli *et al.*, (2014), used GOBO analytical tool to identify high expression of Histidine-rich glycoprotein RNA (HRG) in breast cancer with relative higher expression in basal and grade 2 subtypes. This was further validated using reverse transciptase PCR (RT-PCR) to monitor the RNA expression level; they concluded that the expression pattern for biomarker purposes was 71.7 % sensitive and 93.3 % specific. Also the HRG serum protein was analysed using ELISA and an outcome with 86.7 % sensitivity and 80 % specificity was reported. It was concluded that HRG tissue RNA and serum protein could serve as markers for breast cancer prediction (Matboli, *et al.*, 2014). Paolicchi *et al.*, (2012) used the GOBO database to analyse the expression of different KDM genes as prognostic biomarker for breast cancer in several datasets after identifying them from Oncomine. The output showed KDM5A expression to be low in different types of breast cancer and correlated its expression pattern to the worse clinical outcome for breast cancer (Paolicchi, *et al.*, 2012).

These approaches used to identified and validate KDM5A protein is similar to the approach used for the search for these candidate genes as breast cancer biomarkers, which also included the identification of these candidate genes from the Oncomine dataset in the earlier study (Ngoza, 2012). This suggest, that this approach is effective in validating biomarkers and the eight candidate genes would probably predict correctly the outcome of breast cancer treatment atleast in the luminal B ER+, ERBB2 ER- and lymph node negative subtype of ER- tumours breast cancer types.

3.7: Conclusion

In recent times, numerous genes that could serve as biomarkers have been identified for various cancer types. Several gene expression studies using high-throughput technologies have identified expression profiles of sets of genes that showed promise as prognostic, predictive biomarkers or both for patients with breast cancer (Fan *et al.*, 2006). One of the most important processes in the discovery of multiple genes/proteins for cancer prognosis is validation of the genes/proteins in clinical outcomes. Often, from these list of identified biomarkers, multiple gene products have been linked to a specific cancer tissue and their subtypes. However, validating the role of these biomarkers in different populations or evaluating competing biomarkers has been a herculean task, even though lots of data for these genes are available (Aguirre-Gamboa, *et al.*, 2013).

Validating these prognostic or predictive candidate genes in different breast cancer cohorts is of great interest (Gyorffy *et al.*, 2009). *In silico* approach to biomarker validation could be a more cost-effective strategy for biomarker development, which is typically a fast and cost effective process compared to the traditional methods of biomarker validation (Chen *et al.*, 2014).

From this study, some significant genes that could independently serve as effective breast cancer diagnosis, prognosis, or predictive biomarkers were identified using the SurvExpress biomarker validation tool. However, the variance in significant genes identified from two different datasets from the SurvExpress biomarker analytical tool, suggest that a panel of multiple genes will be more appropriate as candidate biomarkers for breast cancer than the individual genes. The result confirmed that most of the identified genes are differentially expressed in breast cancer and the expression

signature of these set of genes could serve as prognostic or predictive indicators in breast cancer.

The outcome of this results also suggest that the expression signature(s) of the candidate genes as biomarkers could be specific for different breast cancer types and subtype, and indicated effectiveness in accurately predicting the outcome of estrogen positive luminal B subtype of breast cancer as well as estrogen negative human epidermal growth factor receptor 2 (ERBB2) and the lymph node negative subtypes. The current prevailing view is that the most powerful single biomarker of tumours have already been discovered, thus, multiple markers will serve a more effective purpose as breast cancer markers (Diamandis, 2004). The differential expression and presence of the protein products of these genes in body fluids could provide an easy means of monitoring breast cancer, therefore we proceeded to monitor the expression of these proteins in cancer and non-cancer cell lines using a molecular approach in the subsequent chapter.

Chapter 4: Protein expression analysis of identified candidate genes in a breast cancer cell line

4.1 Introduction

Biomarkers are becoming increasingly important in cancer management because they could serve as effective molecular indicators that can be monitored non-invasively in readily available bodily fluids from humans. These biomarkers will surely benefit cancer patients and clinicians by providing an easy to use point of care device for identifying cancer risk, early detection, prognosis, and monitoring patients' responses to therapeutics (Aebersold *et al.*, 2005). An ideal tumour marker should be easily measurable, and presentable in a reliable, cost-effective and easy to use assay format with high analytical sensitivity and specificity (Kulasingam and Diamandis, 2008). Well characterized biomarkers such as the estrogen receptor (ER), progesterone receptor (PR), and the c-erbB2/ HER-2/neu oncogene are used in clinical analysis of breast cancer by immunohistochemistry and fluorescence *in situ* hybridization (FISH), using formalin-fixed, paraffin-embedded tumour tissue as samples (FFPE) (Zhang *et al.*, 2003). However, these procedures are laborious and time consuming, and present a need for identifying biomarkers that are easily accessible as well as non-invasive for breast cancer monitoring.

Proteins are believed to be a better molecule as biomarkers because they are more diverse than DNA or RNA and therefore carry more information than the nucleic acids (Greenbaum *et al.*, 2003). The gene that produces these proteins could undergo alternative splicing and different post-translational modifications, which may result in multiple forms of a protein product of a gene in a species. Also proteins are dynamic and more reflective of many physiological changes most of which, occur at post-

transcriptional level and may not be revealed at the nucleic acid level. Proteins are also more dynamic and reflective of cellular physiology (Aebersold *et al.*, 2005).

Cancer as well as many other diseases is caused by pathway perturbations that occur as a result of genetic and/ or environmental changes and the effect may manifest in altered expression of mRNA transcripts, microRNA, and subsequently the cellular protein profile. The alterations in microRNA expression ultimately modulates the translation of mRNA transcripts, which subsequently alter the quantity of specific proteins shed into the tissue/organ microenvironment, and ultimately end up in the bodily fluids (Ahn and Simpson. 2007).

The expression of altered proteins either in molecular structure or concentration produced by either diseased tissues or normal adjacent tissues in response to disease stimulus, may get to the peripheral blood circulation for easy sampling through processes as simple as venipuncture (Ralhan *et al.*, 2008). Disease-related proteins may also be present in other bodily fluids such as urine, saliva, cerebrospinal fluid, sweat, cystic fluid, and effusions, most of which can be easily accessed as sample for subsequent analyses in the laboratory (Schrohl *et al.*, 2008). The exciting challenge of modern proteomics seek to identify such protein signature in diseases, with strong believe that if this is met, bodily fluids will become potential bio-specimen sources and windows into health and disease. It is hypothesized that the concentration of these disease signatures will correlate with the extent or type of cancer that influenced their expression and could serve as biomarkers for the early detection of cancer, monitoring of cancer progression, and the check of effect and efficacy of drugs (Ahn and Simpson. 2007).

Though high-throughput technologies like microarray and next generation sequencing analysis provides a means of identifying differentially expressed genes, it is a known fact that not all genes are translated to their protein products (Celis *et al.*, 2000). The basic understanding is that, changes in the synthesis of specific transcripts will lead to alterations in the expression level of protein products of such genes, but the multitude of steps between transcription and translation which provide many different regulatory opportunities have proven otherwise (Taylor *et al.*, 2013). Common practice has been to equate the expression level of mRNA to the concentrations and activities of the corresponding proteins, thereby limiting the analysis of the protein abundance to gene expression analysis. However, advances in proteomic techniques have shown this is not always the case (Vogel and Marcotte, 2012).

Different linked processes which include transcription of mRNA, processing, degradation of mRNA, translation, localization, protein modification and programmed destruction of protein products are involved in the production and maintenance of cellular proteins and protein abundance. These processes are modulated by various factors (Vogel and Marcotte, 2012). Several recent studies have looked at the relationship between gene expression and protein levels. In general, these studies have suggested that there are varied reason for lack of correlation between the absolute levels of transcripts and their protein products. Differences in translational efficiency between transcripts and variations in protein turnover are believed to play prominent roles (Taylor *et al.*, 2013; Ning *et al.*, 2012; Schwanhäusser *et al.*, 2011).

Change in cell growth rates, as well as cellular stress are known to directly affect the translational efficiency and these changes have been shown to change the relationship between mRNA and their corresponding protein abundance in cells (Firczuk *et al.*,

2013). In some cases, translational efficiency was strongly linked to alterations by environmental factors. For example, environmental alterations was reported to change the translation of some mRNA transcripts in yeast, though it was not clear if the changes in translational efficiency or protein turnover were at the level of individual transcripts or general cellular response to alteration in environmental conditions (Taylor *et al.*, 2013).

In human cancer cell lines and tissue samples, there are varying reports correlating protein expression of highly differentially expressed genes to their protein products. Shankavaram *et al.* (2007), reported variations in the level of gene expressed in NCI-60 cell line and their corresponding protein products, reporting that genes involved in maintenance of cellular processes and structural properties showed a stronger correlation than other genes in the cell line. It was also stated that genes with characteristics such as high codon adaptation indices (CAI), and/ or ribosomal occupancy showed higher correlations with the expression of their corresponding proteins than the main population of gene products (Shankavaram *et al.*, 2007). Assessing the correlation of microarray RNA profiles and their corresponding protein products, using immunohistochemistry techniques, of 1066 genes in 23 human cell lines, Gry *et al.* (2009), concluded that only one third of the RNA molecules showed significant correlation with their protein (Gry *et al.*, 2009).

Although gene expression analysis provides an alternative in identifying candidate protein biomarkers from identified genes as protein expression is often known or predicted from gene expression data, differentially highly expressed genes are not always correlated with protein abundance. More so, some proteins are known to undergo modification based on influence of cancer development and cannot be predicted from genomic data (Zhang and Chan 2007). It is therefore necessary to

validate the expression of candidate protein products of putative biomarker genes identified from gene expression data.

Generally protein biomarkers for cancer monitoring are broadly classified into four groups as human tumour antigens. Two groups are majorly patient-specific antigens arising from either somatic mutations in normal gene products (e.g., BRCA1 and BRCA2 gene) or antigens that are generated as a result of mutations related to oncogenic processes. These two groups are reported not suitable for general use in many patients. A third group is an antigen produced in normal tissue but the levels increase in tumours (e.g., PSA in prostate cancer). This class of proteins is common in many patients, but is not necessarily tumour specific, but rather tissue specific. The fourth and most appropriate group is tumour specific, relates to the general oncogenic process and thus is expressed in a wide range of human tumours (Onda *et al.*, 2001).

Based on reports in literature regarding variation between gene and protein expression levels in cancer samples, it is important to analyse the protein expression levels of these identified genes. Thus, the protein expression levels of three of the seven putative biomarkers identified to be highly differentially expressed when these genes were analysed using qRT-PCR in the previous study (Ngcoza, 2012) will be analysed. Furthermore, protein expression of Gene 2 (a member of the seven putative gene) and Gene 8 an additional putative biomarker for breast cancer identified from this study using *in silico* methods (Chapter 2) will also be monitored in different cell lines. Thus the aim of this chapter was to analyse the protein expression levels of these five putative biomarkers.

4.2 Materials and methods

4.2.1 Materials

Table 4.1: List of reagents used and suppliers

Materials	Suppliers
Acetone	Merck
40% acrylamide Bis solution 37:5:1	Serva
Beta Actin antibody-HPR conjugated	Cell Signalling
Ammonium persulphate (APS)	Merck
Bacteriological agar	Merck
Bovine Serum Albumin (BSA)	Roche
Bromophenol blue	Merck
Coomassie Brilliant Blue R-250 (CBB)	Bio-Rad
Cytobuster Protein Extraction Reagent	Novagen
Detection Solution	Bio-Rad
Dithiothreitol (DTT)	Fermentas
Donkey Anti-Mouse antibody	Santa Cruz
Dulbecco's Modified Eagle's Medium (DMEM)	LONZA
Ethylene Diamine Tetra-acetic acid (EDTA)	Merck
Ethanol	Merck

Fetal Bovine Serum (FBS)	Invitrogen
Geneticin (G418)	Sigma
6X Gel Loading Buffer	Fermentas
Glycerol	Merck
Hydrochloric Acid	Merck
Isopropanol	Merck
Metafectene Transfection Reagent	Biontex Laboratories
Methanol	Merck
Nuclease free water	Fermentas
Penicillin-Streptomycin	Invitrogen
Phosphate Buffered Saline (PBS)	Invitrogen
Potassium Acetate	Merck
Sodium Chloride (NaCl)	Merck
Sodium Dodecyl Sulphate (SDS)	Merck
Sodium Hydroxide (NaOH)	Merck
N,N,N',N'-Tetramethylethylenediamine (TEMED)	Sigma Aldrich
Tryptone	Merck
Trypsin	Merck
Tween 20	Merck

4.2.2 Solutions and buffers

Coomasie Brilliant Blue (CBB) staining solution: 0.1 % CBB R-250 in 40 % methanol and 10 % acetic acid made up in distilled water.

Destain Solution: 40 % methanol and 10 % acetic acid made up in distilled water.

DTT buffer: 2 % (w/v) DTT in SDS equilibration buffer.

2X Sample buffer: 100 mM Tris-HCL, pH 6.8, 20 % Glycerol, 0.025 % bromophenol blue, 4 % (w/v) SDS, and DTT added just before use to a final concentration of 200 mM.

10X SDS electrophoresis buffer: 10 g SDS, 30.3 g Tris, 144.1 g Glycine make up to 1.0 Liter in distilled water, pH 8.4

10X TBS-Tween: 24.23 g Tris, 87.66 g NaCl, make up in distilled water, pH 7.4 (0.1 % Tween 20 was added when preparing a 1X solution) solution stored at -20^0 C

10X Transfer Buffer: 30.28 g Tris, 144 g Glycine, in distilled water (20 % methanol was added when making 1X solution) solution stored at -20^0 C

4 % Paraformaldehyde: 16 g Paraformaldehyde was dissolved in 80 ml of distilled water by stirring in fume cupboard on a hot plate, 1 M sodium hydroxide was added dropwise till the solution became clear. The solution was allowed to cool down and the pH was adjusted to 7. 0. The solution was made up to 100 ml and an equal volume of 2X PBS was added.

4.3 Cell culture

All cell lines used for this study were from the American Type Culture Collection (ATCC). All the cells were cultured in filter-cap culture flasks with their respective

recommended media and supplements. The cells were incubated at 37 °C in a 90 % humidified atmosphere of 5 % CO_2 till cells reached confluence.

Table 4.2: Summary of cell lines used and their respective media composition

Cell lines	Tissue	Growth medium
A2780	Cervix	RPMI, 10 % FBS, and Penstrep
CaCo2	Colon	DEMEM, 10 % FBS, and Penstrep
HeLa	Cervix	DEMEM, 10 % FBS, and Penstrep
HepG2	Liver	DEMEM, 10 % FBS, and Penstrep
HT29	Colorectal	DEMEM, 10 % FBS, and Penstrep
Kmst-1	Normal Fibroblast	DMEM, 10 % FBS and Penstrep
MCF7	Breast–pleural effusion	EMEM, 10 % FBS, Penstrep, and insulin
MCF-12A	Normal Breast	DMEM-F12, 10 % FBS, Penstrep, EGF, Hydrocortisone and Insulin

4.3.1 Sub-culturing of cells

Cells were monitored every two to three days and media changed before the cells reached confluency. To split cells, media were poured off, and cells were rinsed with sterile PBS (LONZA) to remove excess media. The cells were subsequently treated with 1X sterile trypsin (LONZA) and incubated for a period of 1-3 minutes for cells to detach from flasks. Trypsin was de-activated by adding fresh complete medium and the cells were pelleted by centrifugation at 3000 X g for 3 min. The pelleted cells were re-suspended in fresh growth medium, split into different flasks at varying densities and incubated under the recommended conditions.

Excess cells were frozen down for storage by suspending cell pellets in 10 % DMSO in complete medium, aliquots of about 1.0-1.5 mL were placed in cryotubes and stored in a bio-freezer at -150^{o} C.

4.4 Protein extraction and quantification

Protein were extracted from cells using the CytoBuster^{TM} Protein Extraction Reagent (Novagen) according to the manufacturer instructions. Cells were cultured to confluency, after which media was removed from the flask and the flask was rinsed with 1X PBS. The recommended volume of CytoBuster^{TM} extraction solution was added and the cells were incubated on a shaker (Belly dancer, Stovall Life Science Inc., USA) for 5 minutes. After incubation, the cells were scraped from the flask using a cell scraper and collected into a 1.5 ml tube. The tubes were centrifuged at 16,000 X g at 4^{0} C for 5 minutes. The supernatant containing the total protein was collected into fresh 1.5 ml tubes and stored at -20^{0} C until required.

4.4.1 Protein quantification

The concentration of total protein extracted from each of the cell lines was determined using the Qubit™ Protein Assay Kits following the manufacturer's instructions. 0.5 ml tubes were labelled for each protein samples plus extra tubes for the standards. Working solutions was made by adding Quant-iT reagent (1 x n) µL with Quant-iT buffer (199 x n) µL (n = number of sample tubes + 3 standards) and vortexed for about 3-5 seconds, while preventing formation of bubbles. The standards were prepared by combining 10 µL of standards 1, 2, 3, provided in the kit with 190 µL of working solution. The protein samples were prepared by adding 5 µL of the individual protein extracts into their labelled tubes combined with 195 µL working solution. Tubes were gently vortexed for 3-5 seconds and incubated for 15 minutes at room temperature. The Qubit™ 3.0 (Invitrogen) was used to read concentration of the samples after taking the readings of the standards.

4.5 1D SDS-Polyacrylamide Gel Electrophoresis

4.5.1 Gel preparation

One dimensional (1D) polyacrylamide gel electrophoresis was used to separate the total protein extracted from each cell line. Glass plates and apparatus (Bio-Rad) used, were first cleaned with 70 % ethanol to ensure it was free of acrylamide. A 12 % separating gel was prepared with the constituents and volumes as stipulated in table 4.3. The solution was poured between the gel plates, overlaid with 1 ml isopropanol, and allowed to solidify.

The stacking gel was prepared as stipulated in table 4.4. Once the separating solution had solidified, the isopropanol was poured off and the gel rinsed with distilled water. TEMED and APS were added to the other components of the stacking gel, a 10 well

Bio-Rad comb was inserted and the gel was allowed to solidify. After the gel solidified, the comb was removed and the gel was set up in a gel electrophoresis tank (Bio-Rad). The tank was filled with 1X SDS running buffer and was pre-run at 100 V for 10 minutes to remove any SDS or liquid acrylamide from the wells of the gel.

Table 4.3: Stock solutions for 12 % separating gel solution

Reagent	Volumes
1.875 M Tris-HCl separating buffer pH 8.8	2 mL
40 % Acrylamide-Bis	3 mL
10 % SDS	100 µL
10 % APS	50 µL
Distilled water	4.84 mL
TEMED	10 µL
Total Volume for two plates	10 mL

Table 4.4: Stock solutions for 4 % stacking gel solution

Reagent	Volumes
0.625 M Tris-HCl stacking buffer pH 6.8	500 µL
40 % Acrylamide-Bis	500 µL
10 % SDS	50 µL
10 % APS	38 µL
Distilled water	4 mL
TEMED	10 µL
Total volume for two plates	5 mL

4.5.2 Electrophoresis

The protein samples were removed from -20^0 C and left to thaw on ice. Equal volume of protein sample with 2X sample buffer was mixed in new tubes and boiled at 95^0 C in a block heater (Stuart Scientific) for 5 minutes. Samples were removed from heater and immediately placed on ice, the samples were centrifuged for about 1 minute using a benchtop Eppendorf centrifuge at 16,000 X g. Approximate concentrations of each sample were loaded onto the polyacrylamide gel and electrophoresed at 110 V for 60-90 minutes.

4.5.3 Gel staining

After electrophoresis, gels were carefully removed from the plates and stained with coomasie staining solution by placing gel in a tray containing the coomasie stain and placed on a shaker (Belly dancer, Stovall Life Science Inc., USA) for 45-90 minutes. The staining solution was then removed, and the gel destained by placing in destaining solution on a shaker for 10-16 hours. Distained gels were viewed and captured with a UVP imager (UVP LLC, Biospectrum).

4.6 Monitoring protein expression using Western blot analysis

4.6.1 Blot transfer of Proteins onto PVDF membrane

The protein samples were electrophoresed with the use of the 1D SDS polyacrylamide gel electrophoresis as described in section 4.5.2. The protein separated on the electrophoresed gels were transferred onto a PVDF membrane (BioTrace^{TM}) using a Trans-Blot Turbo transfer system (Bio-Rad), at 7.0 Amp, for 15 minutes. After blot transfer, the gel was stained in coomasie blue as described in section 4.5.3, to monitor how well the proteins were transferred, while the membrane was further analysed.

4.6.2 Analysis of individual protein

The membrane was blocked in 1 % casein in TBS-Tween for 1 hour at room temperature. After 1 hour, the membrane was washed three times in TBS-Tween for 5 minutes on a shaker. After the third wash, individual membranes were incubated in their respective primary antibodies diluted as stipulated in their different product data sheets, with 1 % casein in TBS-Tween. Membranes were incubated with primary antibodies overnight on a shaker at 4^0 C. The following day, the primary antibody was removed and the membrane washed in TBS-Tween three times for 5 minutes for each wash. The respective membranes were then incubated with their specified secondary

antibodies (diluted 1:500 TBS-Tween) as stipulated in the different product datasheets and incubated on a shaker for 1 hour at room temperature. Following incubation, the membranes were washed with TBS-Tween thrice for 5 minutes at room temperature using a shaker. The membranes were then individually transfered to the UVP-Biospectrum imaging system (USA), incubated in 2 ml Clarity^{TM} western ECL substrate (Bio-Rad), for viewing and the image captured with a UVP-Bospectrum imaging system built in camera.

4.7 Immunofluorescence expression analysis of protein product of the candidate genes

About 1 x 10^5 MCF7 and MCF-12A cells were seeded in six well plates on a sterile cover slip for 24 hours. After 24 hours, the cells on cover slips were wash twice with 1X PBS. Cells were then fixed in 1 ml 4 % paraformaldehyde for 15 minutes. Cells were washed three times for 5 minutes for each wash with 1X PBS, adding 100 mM glycine in the second wash. The cells were permeabilized by incubation in 0.1 % triton X100 (Sigma) in PBS for 2 minutes, after which cells were washed three times for 5 minutes each. The cells were then incubated in blocking solution of 1 % BSA in PBS for 30 minutes to 1 hour. The blocking solution was removed and the cells were incubated for 1 hour with primary antibodies diluted (1:100) in blocking solution. After an hour, cells were washed with PBS three times (30 seconds for each wash) and then incubated with secondary antibodies diluted (1:200) in blocking solution. Cells were then washed three times with PBS before mounting cells with vector shield containing DAPI (Sigma), by placing the cover slips with the cells faced down on a glass slide. The slides were left to solidify at room temperature before storage at 4^0 C, pending viewing using the Axioplan 2 imaging fluorescent microscope (ZEISS).

4.7 Results and Discussion

The different methods available to analyse gene expression patterns in cancer cells in comparison to the patterns in normal cells, such as qRT-PCR, have provided a powerful means of analysing total RNA expression patterns in specific cancers. However, these methods do not necessarily monitor the expression of the protein products of these various RNA's. The emerging field of proteomics provides new tools for the early detection of cancer from human serum, cerebral spinal fluid, urine, saliva, sweat etc., by providing information about the rapid changes that occur regarding the proteome's dynamic as a result of disease conditions (Good *et al.*, 2007).

However, the direct proteomics approach to biomarker discovery is regarded not to be too sensitive. For example, some authors argued that the mass spectrophotometry technology approach, which is a major tool in proteomics, couldn't reliably detect low-abundant cancer-specific proteins in plasma (Diamandis, 2004). Also other major challenges in the proteomics approach in biomarker discovery are reported that include the extreme complexity of the plasma proteome, variation in genomes among individuals in a population, the changes in the plasma proteome as a result of various factors such as sex, age, health status, lifestyle etc., as well as physiological and pathological conditions that could influence the proteomic experimental outcome other than the putative cancer itself (Zhang and Chan 2007). To validate the candidate genes identified as putative breast cancer biomarkers, protein expression of the differentially expressed genes from the qRT-PCR analysis (Ngcoza, 2012) and the additional two genes identified from further *in silico* validation of the candidate genes (Chapter 2) was monitored in different cancer and non-cancer cell lines using western blot analysis.

4.7.1: Protein profile analysis of extracted protein from different cell lines

Figure 4.1 shows the 1D SDS polyacrylamide gel electrophoresis of total protein extracted from different cell lines. From the result, the protein bands are not smeared, thus are of good quality to be used for the western blot analysis.

Figure 4.2 shows Western blot analysis using antibody specific for β-actin to analyse protein expression in different cell lines. The result shows the proteins were well transferred and the protein concentration loaded for MCF-7 and MCF-12A, are approximately equal. Beta-actin (β-actin) is expressed constantly and at high levels in all cell types used in protein research, it is therefore used as loading control in protein expression analysis such as Western blot. β-actin as the control shows that the proteins of interest was loaded equally on a gel, that it is being transferred as well as showing that the reagents are functioning normally (Liao, 2000).

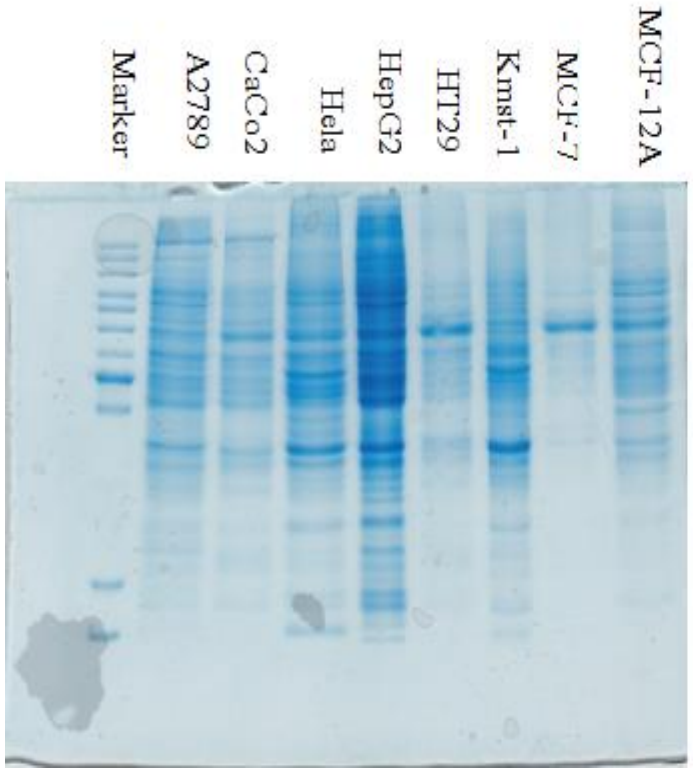


Figure 4.2: 1D SDS PAGE of isolated protein from different cell lines.

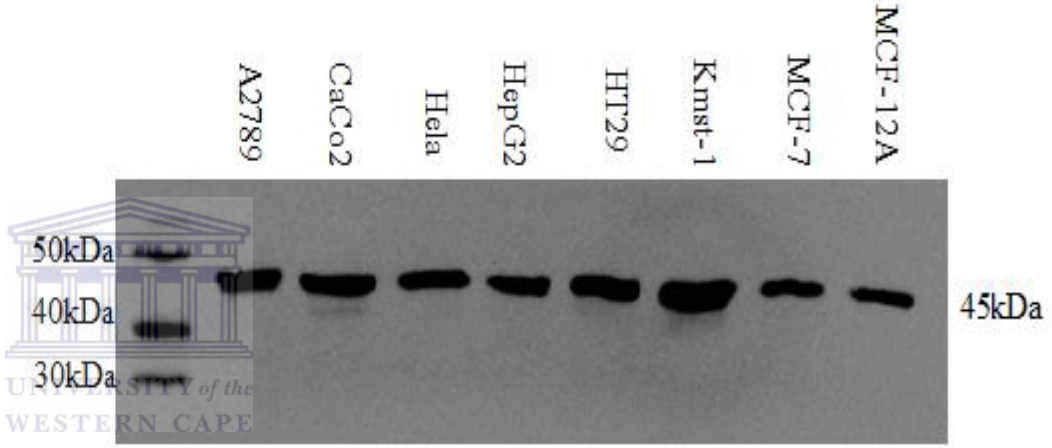


Figure 4.1: β-actin expression analysis in different cell lines suing western blot

4.7.3: Protein expression analysis of Gene 7 (BRG 7)

Figure 4.3 shows the expression of protein product of Gene 7 (BRG 7). Although the protein band was not well pronounced, the result from the western blot analyses confirmed expression of this protein in a breast cancer cell line (MCF-7) was higher when compared to MCF-12A. The expected molecular weight of this protein is 38 KDa (shown by the orange box) although it is known to undergo various post-translational modifications. The result also showed a protein band of higher molecular weight common to both the breast cancer and non-cancer cell lines (Figure 4.3). Another reported member of the NKG2D ligand family of proteins have been described to encode an extra domain whose amino acid sequence is similar to the amino acid sequence of the candidate protein of interest (BRG 7), we postulate that the higher molecular weight band represent that protein. These two proteins have been reported to be co-expressed in cell lines, and are induced by the same stimuli (Ohashi *et al.*, 2010).

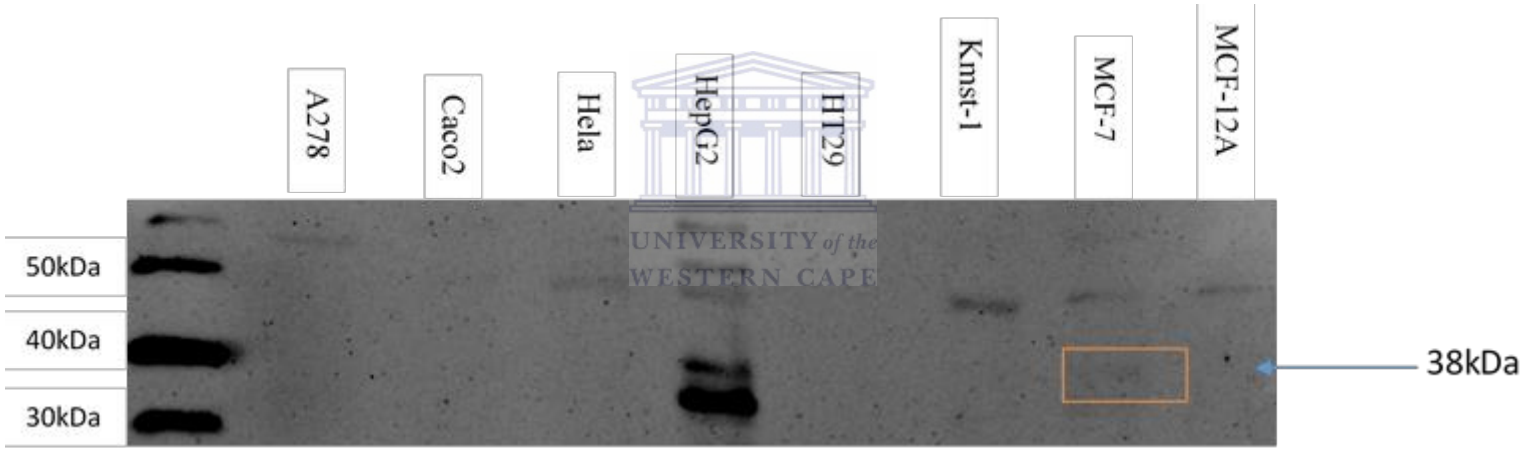


Figure 4.3: Western blot result showing the expression of BRG 7 in different cell lines

Gene 7 was annotated using KEGG (see Section 2.4.2) to be associated with pathways involving natural killer cells. Natural killer (NK) cells serves as the first line of defence against pathogens, cancerous or stressed cells. They expressed specific surface receptors that receive signals from the environment and in response, produce effector molecules that suppresses tumour growth as well as convey information to the immune system for total clearance (Smyth *et al.*, 2002). In disease conditions, the active receptor Natural killer (NK) group 2, member D (NKG2D) is expressed on the cell surface of NK cells but naive cells do not express this NKG2D protein (Skov *et al.*, 2005). The NKG2D receptor recognises some classes of protein ligands that are constitutively expressed on some tumour cells as well as viral/bacterial infected cells (Skov *et al.*, 2005).

Also, members of this protein family are known to undergo different post translation modification. For example, an ortholog of a member of the protein family in mice was reported to undergo post-translation regulation, by ubiquitination on lysine residue (Nice *et al.*, 2009). The protein family has also been reported to undergo N-glycosylation as to get to the membrane surface, and the inhibitor of glycolysis and N-linked glycosylation, 2-deoxy-d-glucose (2DG), potently inhibits surface expression of their ligands. These are alternate mechanisms used by cancer cells to evade immune-detection (Andresen *et al.*, 2012).

The immunofluorescence analysis of this protein in a breast cancer cell line versus a non-cancer breast cell line indicated that this protein is synthesized in the cytoplasm and expected to be translocated to the membrane. Based on fluorescent intensity, the expression of this protein appears less on the membrane of the cancer cell line as compared to the non-cancer cell line (Figure 4.8). The band that is common to both cell lines in Figure 4.3 is also suspected to be the protein of interest (BRG 7) that have

undergone post-translational modification The post-translational forms of this protein could also account for the higher molecular weight band seen in the western blot result (Figure 4.3) as well as the seemingly lower membrane expression level in a breast cancer cell line.

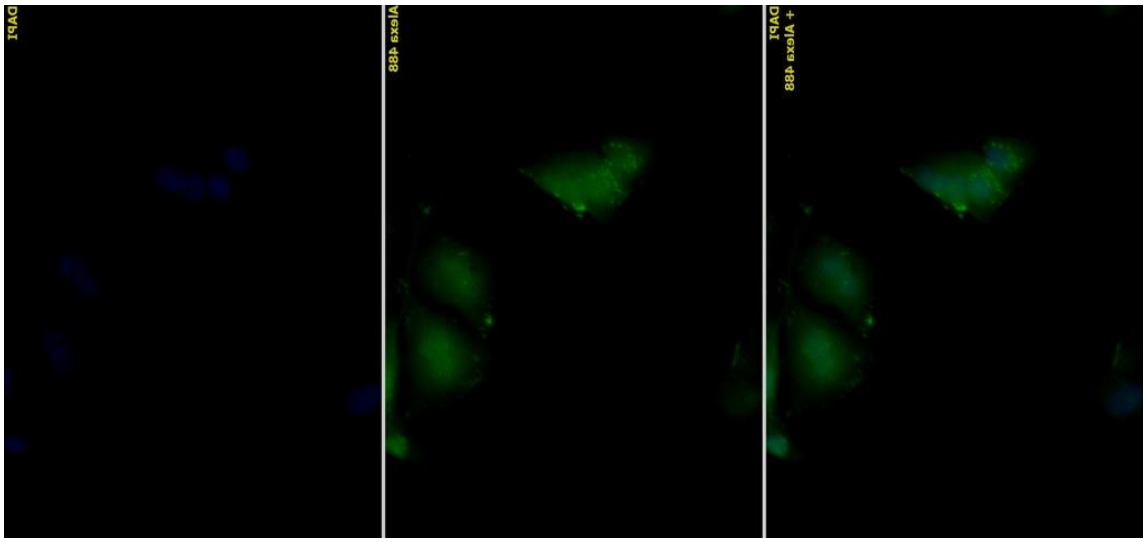
Analyzing the expression of the protein products of Gene 3 and Gene 4 using Western blot analysis was not successful after several trials (result not shown). It was hypothesized that the commercial primary antibodies for these proteins might bind the proteins better in their native forms, thus, leading to analyzing the expression of these proteins using immunofluorescence. It has been reported that, the interaction between a protein and its corresponding antibody may be disrupted by competition with high concentrations of the pure epitope itself, as well as high salt concentrations, extremes pH, and detergents (Janeway *et al.*, 2001).

Furthermore, although use of immunoassays in a clinical setting is believed to provide a cost effective approach to cancer management, the development and production of effective antibodies against identified proteins is known to be a major challenge (Zhang and Chan, 2007). Faça and Hanash, (2009), reported that the analysis of the proteome of ovarian cancer cell lines indicated that some proteins were predominantly in the cell surface fraction, compared with the cell culture medium or to whole cell lysates. This is an indication that these proteins of interest may not necessarily be abundant in the cell lysate that were used for the western blot analysis and may also account for the low intensity of BRG 7 (Figure 4.3) and a need for immunofluorescent analysis.

4.7.4 Immunofluorescence analysis of the protein products of Gene 3, 4 and 7

The expression and localization of the protein products of the three highly differentially expressed genes (Gene 3, Gene 4 and Gene 7) were analyzed using the immunofluorescence analysis. The nuclei of breast cancer and non-cancer cell lines were stained with DAPI (4′, 6-diamidino-2-phenylindole), while the secondary antibodies against the primary antibodies had a green fluorophore (Alexa Fluor®448 from Life Technology) attached to them.

A



B

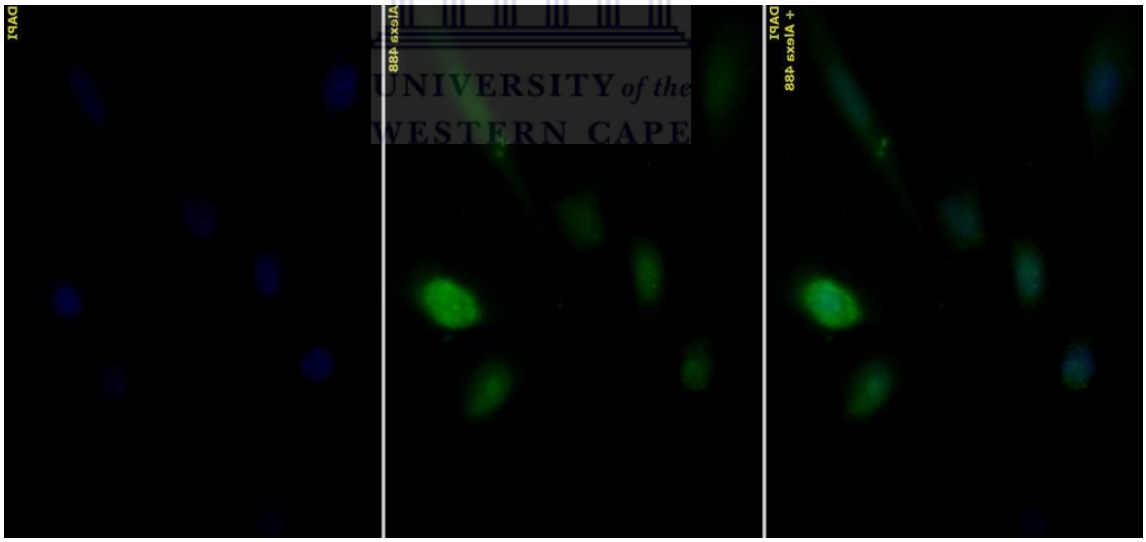
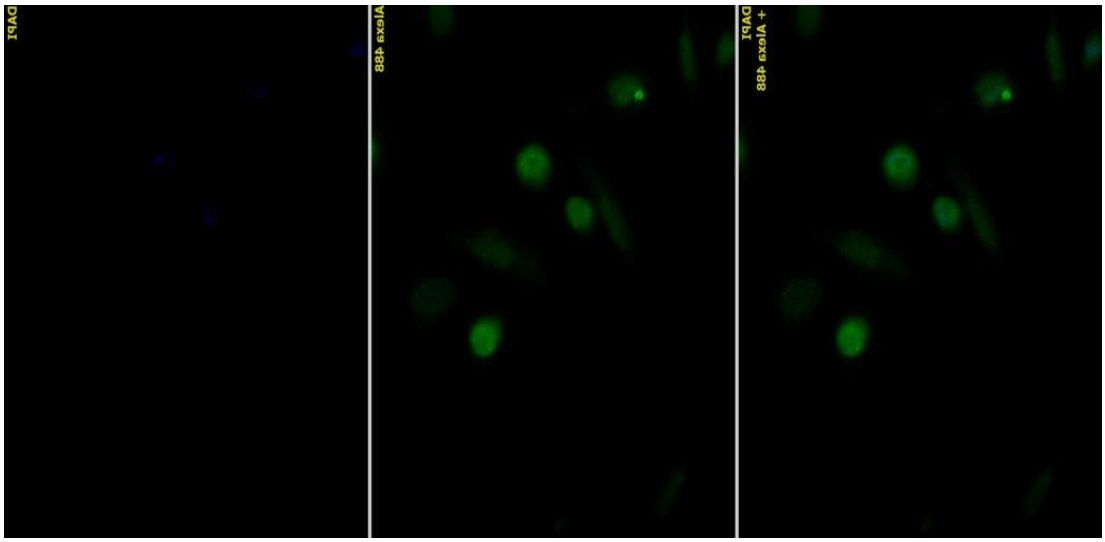


Figure 4. 4: Immunofluorescence analysis of BRG 3 in breast cell lines. (a) MCF-7 (b) MCF-12A (40x magnification)

A



B

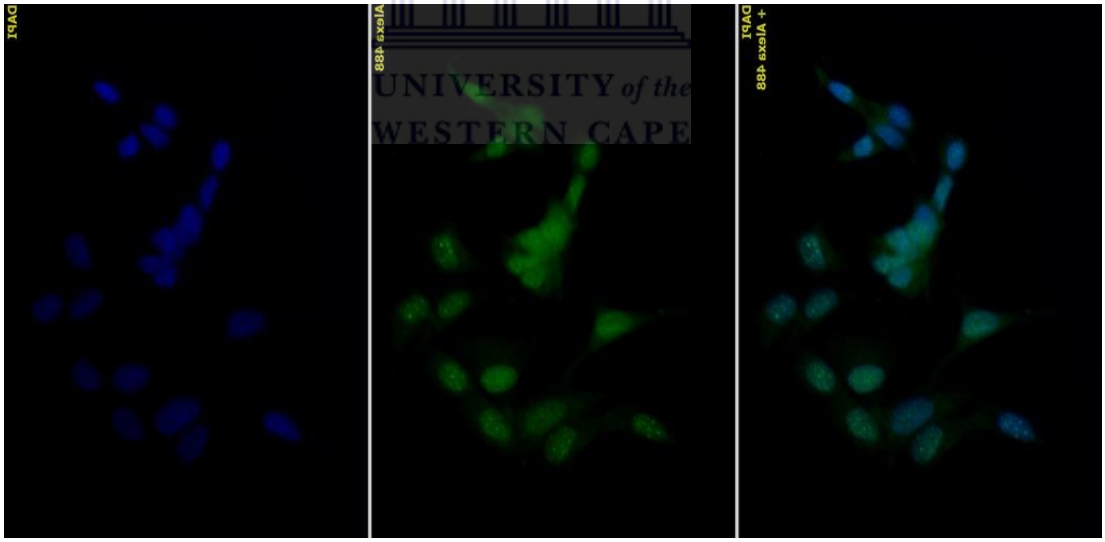
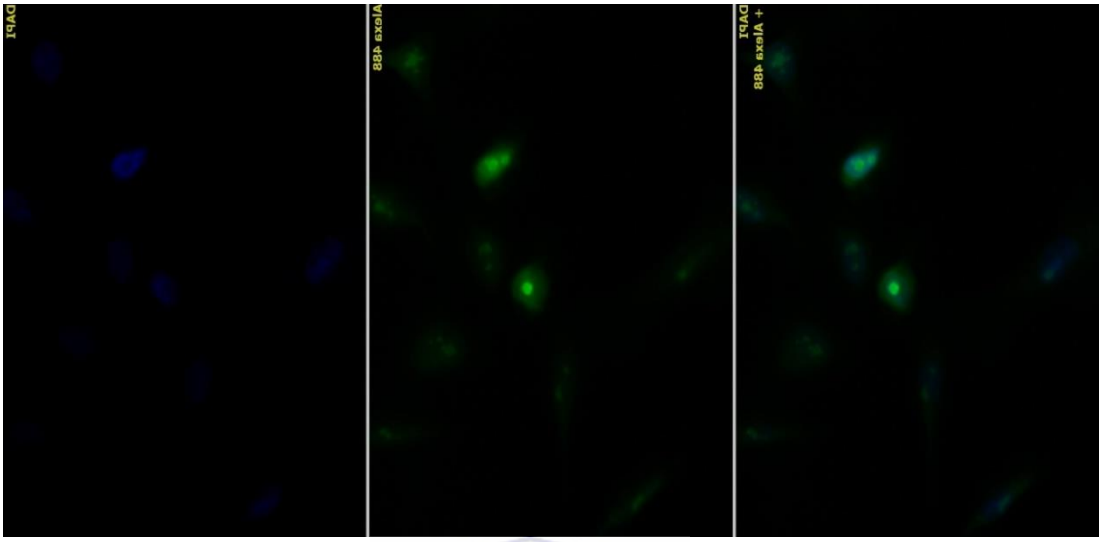


Figure 4.5: Immunofluorescence analysis of BRG 4 in breast cell lines. (a) MCF-7. (b) MCF-12A (40x magnification).

Figures 4.4 and 4.5, showed proteins BRG 3 and 4 are expressed in both cancer and non-cancer cell lines, it would appear the protein expression are more in the breast cancer cell lines than the non-cancer breast cell lines, validating the qRT-PCR analysis result (Ngcoza, 2012). The protein products of Gene 4 (BRG 4), appears to be more localised to the cell membrane of the breast cancer cell line (MCF 7) as compared to the breast non-cancer cell line (MCF-12A) (Figure, 4.5 a, b). Also, BRG 3 appears to be more on the cell membrane of the breast cancer cell line when compared to the non-cancer breast cell line (Figure 4.4 a, b), although a number of the MCF-12A cell population shows an increase concentration of the protein localised to the cell membrane.

The localization of the protein product of Gene 7 was also monitored using the immunofluorescence assay. Figure 4.8 showed the expression and localisation of the protein BRG 7, from the figure, it seems the protein is almost equally expressed in both breast cancer and non-cancer breast cell line, but more are localised on the cell surface of the breast cancer cell line (MCF 7). However, more analysis needs to be done to confirm this.

A



B

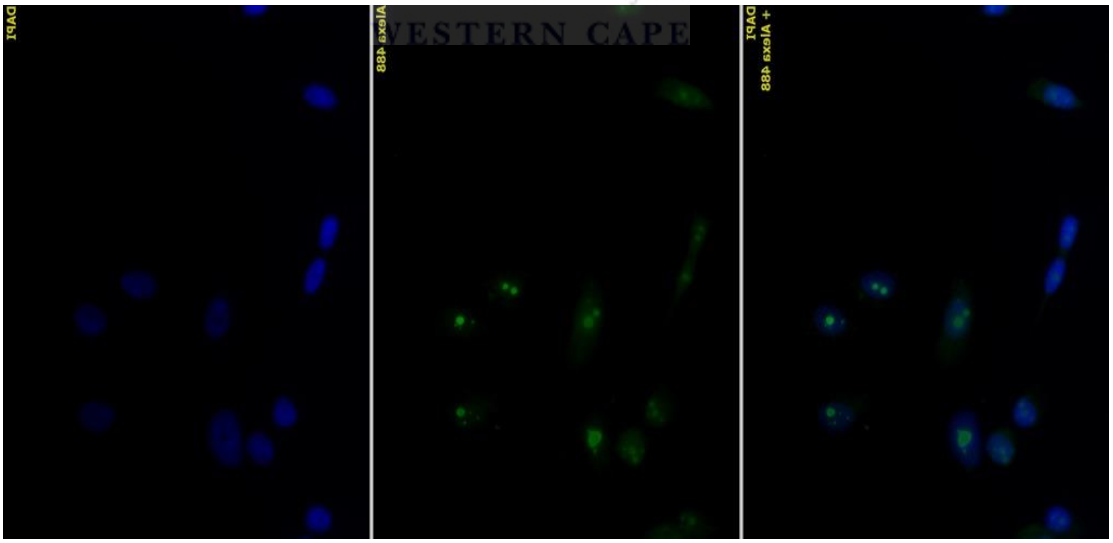


Figure 4.6: Immunofluorescence analysis of BRG 7 in breast cell lines. (a) MCF-7 (b) MCF-12A (40x magnification)

The protein product of the candidate Gene 7 observed with the Western blot analysis (Figure 4.6) is a member of one of the NKG2D groups of ligand found on the membrane of disease cells. Membrane bound proteins are attractive biomarker candidates because of their potential use in cancer imaging, targeted therapeutic strategies, and also they are often shed into the bloodstream, which makes them useful candidate's for early detection (Harsha *et al.*, 2009). BRG 7 is known to be an anchored protein just like most of the NKG2D ligands family memeber. They serve as immune signals in some disease conditions (González *et al.*, 2008).

In evading detection by the immune system, cancer cells are known to shed off some membrane proteins especially immune signaling proteins (Hanahan and Weinberg). Some of the NKG2D ligands have also been reported to be shed off the surface of cancer cells for the cells to evade the immune system. Wang and Sun (2014), reported that, the NK cells stimulates the shedding of some of these ligand proteins in response to cancer treatment (Wang and Sun, 2014) which further validates that a member of this group of protein would be a good candidate biomarker for cancer progression and effective therapeutics monitoring.

4.7.5: Protein expression analysis of Gene 8 (BRG 8)

Figure 4.7 showed the expression level of BRG 8, the protein that was reported to interact with BRG 7 and annotated to regulate ESR (see Section 2.4.6), using an *in silico* method (Chapter 2). Interestingly, the protein was not expressed in the breast and colon cancer cell lines but found to be expressed in a non-cancer breast cell line (Figure 4.7). Annotation of this protein showed it could also be a membrane protein that modulates the expression of estrogen receptors. Estrogen receptor (ERα)-positive primary breast tumours account for more than 70 % of all cancer cases and this protein

is inhibited in breast tumours resistance to hormone therapy drugs and tumour cells involved in breast cancer progression (Kim *et al.*, 2013).

The reason for the breast and colon cancer cell lines not expressing this protein is not known as it is expected to be expressed in breast cancer, however, the differential expression of this protein could be a good candidate biomarker for tamoxifen resistance breast cancer patients. The protein has been reported to be in blood circulation and higher in cancer patient than non-cancer patients (Sievers *et al.*, 2006). Tamoxifen is a synthetic anti-estrogen drug used as standard treatment for patients with ERα-positive breast cancer. The drug serve as a selective ER modulator by directly inhibiting the interactions of estrogen (E2) and the ERα receptor, however, tumours often develop resistance to this drug. Although, the mechanisms by which tumours develop resistance is yet to be defined. The mechanism could include ERα modulating proteins that are co-regulated and the link between the ERα pathway and other growth factors and kinase networks (Yoshimaru *et al.*, 2013).

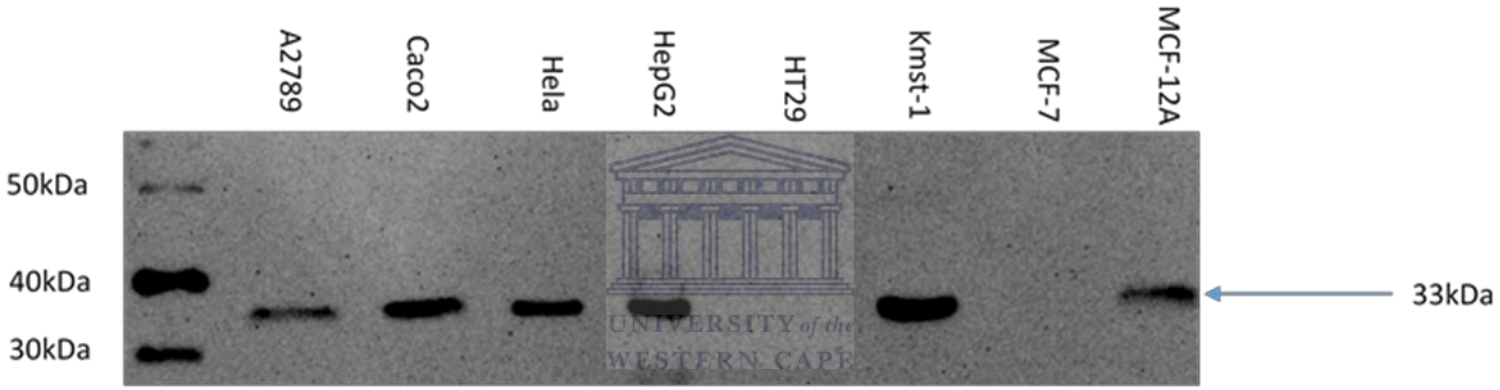


Figure 4.7: BRG 8 expression analysis in different cell lines using western blot analysis

4.7.6 Protein expression analysis of Gene 2 (BRG 2)

Figure 4.8 shows the expression of BRG 2, the protein product of Gene 2, the gene of this protein was amongst the seven candidate genes identified from the earlier work (Ngcoza, 2012). This gene though not highly differentially expressed from the qRT-PCR analysis (Ngcoza, 2012), was found using *in silico* functional analysis (Chapter 2) to be involved in different processes linked to breast cancer in particular and cancer in general.

The protein was expressed in all cancer types used for the western blot analysis. This could be the reason why the gene did not show differential expression in the qRT-PCR analysis (Ngcoza, 2012), however it seems the protein is cleaved in the non-cancerous breast cell line as it is presented as an extra band in the Western blot analysis. This protein is a multifunctional endoplasmic reticulum protein that has been reported to play a major role in intracellular calcium storage and transfer as well as other functions like chaperoning and adhesion (Lwin *et al.*, 2010).

This protein is reported to interact with ERα to prevent cancer cell invasion in breast cancer and also reported as a contributing factor in the role of thrombospondin-1 in invasion (Platet *et al.*, 2000). It is also reported in literature, that molecular chaperones (a function family of protein which BRG 2 belongs), may also induce a humoral immune response and generate autoantibodies that could serve as cancer biomarkers for screening, diagnosis, as well as in immunotherapy (Kabbage *et al.*, 2013). Thus, this protein is expected not to be highly expressed in cancer cells or modulated to prevent it from carrying out its anti-cancer functions. However, the expression pattern needs to be further investigated to validate it for use as a breast cancer biomarker.

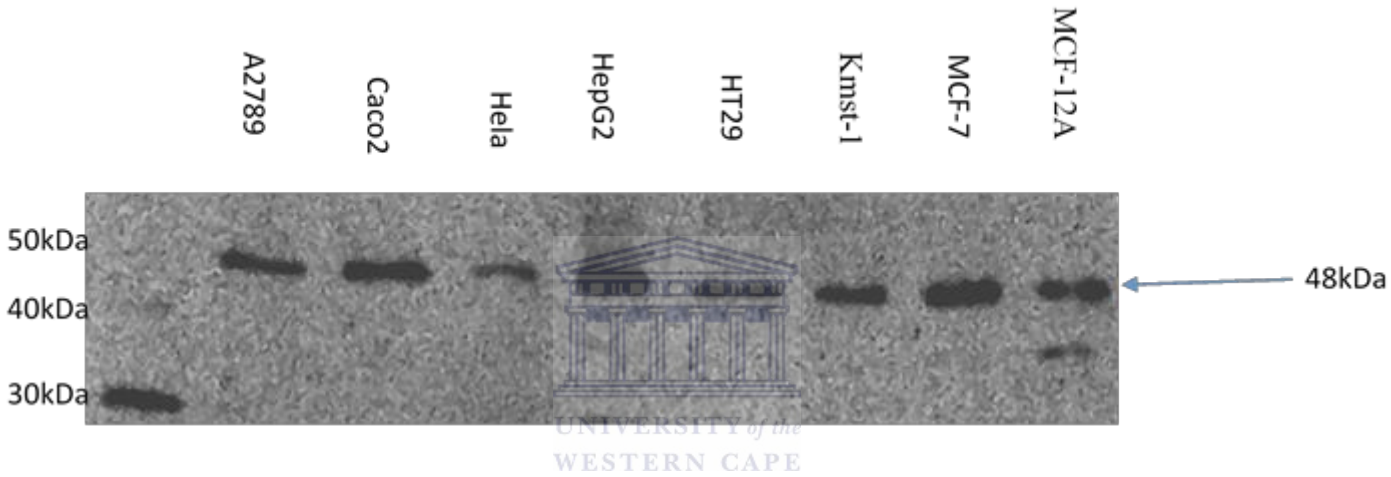


Figure 4. 8: BRG 2-protein expression analysis in different cell lines using western blot analysis.

4.9: Conclusion

From the protein expression analysis using the Western blot procedure, it was demonstrated that the protein products of the genes of interest are expressed and there was an indication that certain proteins may be differentially expressed. BRG 3 and BRG 4 were not successfully demonstrated to be expressed using the Western blot analysis, however the proteins are shown to be expressed using immunofluorescence analysis. The immunofluorescence analysis confirms BRG 3, BRG 4 and BRG 7, the protein products of the three highly differentially expressed genes in the previous study (Ngcoza, 2012) as membrane bound and would probably be shed into the bodily fluids. BRG 8, which was identified from the *in silico* validation study (Chapter 2), was shown to be expressed in other cell lines except the breast and colon cancer cell lines.

It was exciting to observe that this protein which was identified to directly interact with BRG 7 seem not to be highly expressed in the breast cancer cell line. The presence of these proteins however, would have to be looked at in the culture media of these cell lines, this will serve as a strong indication that they are shed off by breast cancer cells at some point of progression and could be present in the serum of breast cancer patients. The detection of cancer biomarkers in plasma and other bodily fluids using immunoassays provides a cost effective approach to cancer diagnosis (Zhang and Chan 2007). Antibodies that target these specific biomarker proteins can be used to develop easy to use point of care devices for early detection as well as tools to complement existing methods for effective management of breast cancer. However, further molecular validation needs to be done to validate these proteins and how well they can serve as breast cancer biomarkers.

Chapter 5: Establishing stable transfected breast cell lines for over-expression and silencing of BRG 7

5.1 Background

The innate immune system is the first line of the body's physiological defense against cancer and the natural killer (NK) cells, are central to mediating this type of immune system which can impede tumour growth through the process of immune surveillance, identifying tumour cells and destroying them (Crane *et al.*, 2014). The NK cell is activated when a specific surface receptor such as the Natural killer group 2, member D (NKG2D) receptor identifies and binds its ligand expressed by tumour cells. This ligation results in the production of cytokines and the release of cytotoxic granules that cause tumour lysis (Orr and Lanier, 2010; Crane *et al.*, 2014). Therefore tumour cells expressing ligands for the NKG2D receptor are susceptible to NK killing

The NKG2D ligands include MHC class I polypeptide-related sequence A (MICA), MHC class I polypeptide-related sequence B (MICB), and UL16-binding proteins 1–6 (ULBP1-6). They are known to be highly expressed by any type of cell under stress as well as cancer cells (Critchley-Thorne *et al.*, 2009). There are various studies and reports highlighting the mechanisms by which NKG2D ligands could be involved in immune evasion in different tumour types and whether the detection of these ligands on circulating monocytes in cancer patients could provide useful biomarkers to either monitor cancer at different stages or monitor relapse (Crane *et al.*, 2014).

Mice that lack the MIC genes but express retinoic acid early inducible-1 (RAE-1) protein, a homologous protein of ULBP, H60 and murine ULBP-like transcript 1, have been used to demonstrate that sufficient high levels of expression of the NKG2D receptors could stimulate cell death. Suggesting that related NKG2D ligands

expression could be expressed even when the expression of the MHC class I is normal (Skov *et al.*, 2005). RAE-1 transfected cancer cells were eliminated *in vivo* by NK cell compared to parent cells that were not expressing RAE-1, also γδ T-cells were reported to kill skin cancer cell by the NKG2D (Skov *et al.*, 2005). These reports strongly suggest a role for NKG2D/NKG2D ligand interaction in tumour rejection and surveillance (Skov *et al.*, 2005).

Soluble NKG2D ligands are believed to be release by tumour cells and some have been identified to serve as markers for tumour prognosis (Wang and Sun, 2014). For example, Kaiser *et al.*, (2007), reported that most epithelial cancer cells express large amounts of MICA on the cell surface, although the late-stage human tumours sustain the surface expression but shed off this NKG2D ligand, which negatively impact on the systemic immune response, thereby promoting tumour immune evasion (Kaiser *et al.*, 2007).

The presence of the soluble MICA could thus serve as biomarker for cancer diagnosis and prognosis. Also, it was reported that pharmacological inhibition of thioreductase and the modulation of endoplasmic reticulum protein (ERp5), a protein that associates with MICA could regulate MICA shedding and ERp5 could serve as a marker and target for therapeutic intervention (Kaiser *et al.*, 2007). Also activation of heat-shock transcription elements in the promoters of corresponding genes is believed to lead to upregulated expression of MICA and MICB in many tumour-cell lines and primary tumours of epithelial origin (Raulet, 2003).

NKG2D ligands are also reported to be involved in modulating the NKG2D-related effector functions. Groh *et al.*, (2002), reported a 60-70 % reduction in expression of NKG2D on $CD8^+$ T cells from MIC positive tumour infiltrating lymphocytes,

extracted from epithelial tumours. Whereas none of the 13 MIC-negative tumour specimens that were examined, had their NKG2D expression altered when compared to normal control peripheral blood monoclonal cells.

Also cisplatin and interleukin 2 (IL-2) have also been demonstrated to enhanced apoptosis and growth inhibition when used as combined therapy in head and neck cancer. The combined therapy showed a synergistic interaction between cisplatin and IL-2, which resulted in NKG2D-based cytotoxicity increasing tumour specific lysis from CTLs and NK cells (Li *et al.*, 2002). Wang and Sun, (2014), showed that NK cell/ apoptosis induced cell surface shedding of ULPB protein ligands through the action of metalloproteinase and concluded that compounds inhibiting NKG2D ligand shedding may offer therapeutic means by reducing excessive pathogenic NK cell activities.

Considering the important role of NKG2D ligands in modulating the effect of NKG2D related effector functions and the great possibility of them being in bodily fluid as a result of cancer cells shedding them to evade detection, it is worthwhile to look at the effect of modulated expression of Gene 7 and its protein product in breast cell lines. From the early study, Gene 7 a NKG2D ligand was detected to be the most highly differentially expressed gene in a breast cancer cell line (Ngcoza, 2012). Also, earlier *in silico* analysis (Chapters 2 and 3) suggest that this protein shows a great promise as a biomarker for monitoring breast cancer.

The success in gene sequencing and the ability to identify, predict and characterize their expression pattern in disease conditions has led to the need to study their role in tissue functions. Major approaches used to study gene function, such as gene knockout in murine subjects, are expensive, time consuming and cannot be applied to human

tissues (Rubinson *et al.*, 2003). The advent of the RNA interference technology (RNAi), have provided the means of analysing gene functions in primary mammalian cells. This approach used to silence a gene of interest, is fast, more cost effective, and can be easily adapted to study homologous gene function in a wide variety of organisms (Stewart et al., 2003).

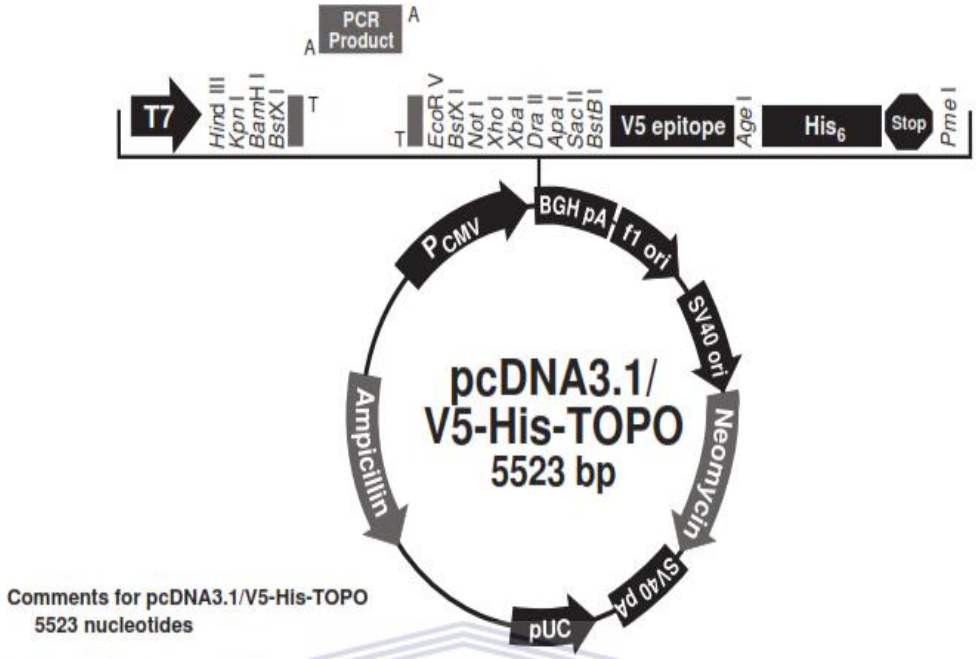
Three mediating methods that are used to achieve an RNAi effect in mammalian cell lines include; small interfering RNA (siRNA), short hairpin RNA (shRNA) and bi-functional shRNA (Rao *et al.*, 2009). The siRNA is the easiest construct to produce and it provides an effective easy means for transient monitoring of silenced genes. However, using optimized shRNA constructs that will be processed by endogenous RNAi processing machinery, allow for high potency, sustainable effects using low concentration, and less off-target effects. The Bi-function shRNA constructs may provide a more enhanced potency of gene silencing, but poses the challenges of tumour selective delivery vehicles and a means for off-target effect evaluation (Rao *et al.*, 2009).

Vector based shRNA delivery has also been reported to provide *in vivo* RNAi effectiveness in mammalian cells (Tong *et al.*, 2005). *In vivo* studies make use of many delivery methods, however lentiviral vectors are increasingly becoming the method of choice for shRNA delivery in mammalian cell transfection. This is because, the virus has the ability to stably integrate into the host cell genome (Cooray *et al.*, 2012). Another advantageous feature of lentivirus vectors is the ability to mediate potent transduction in stable transfected cells in dividing and non-dividing cells both *in vitro* and *in vivo* (Castanotto and Rossi, 2009).

Several studies have shown successful use of lentiviral vectors for the delivery of shRNA constructs in different mammalian systems some of which include: the down-regulation of an activated *Ras* oncogene, which was achieved using a lentiviral-delivered shRNA (Brummelkamp *et al.*, 2002), expression of superoxide dismutase 1 (SOD1) was down-regulated in mouse models of amyotrophic lateral sclerosis using shRNA delivered using a lentirus (Raoul *et al.*, 2005), also a lentiviral vector was used to deliver a shRNA that targets smad3 protein for regeneration of satellite cells and repair of old and injured muscle (Carlson *et al.*, 2008).

Another important approach to study a protein function is to express it in its actual post-translational modified forms, as these contribute to its biological activity (Panavas *et al.*, 2011). Although mammalian proteins have been successfully over-expressed in bacteria, yeast, or insect cells, mammalian cell culture system still remain the system of choice because it allows for specific protein post-translation modification (Varakala 2008; Panavas *et al.*, 2011). Most mammalian secreted and membrane bound receptor proteins often possessed compulsory post-translational modifications that may include disulphide bonds and glycosylation that aid protein proper folding as well as biological activity (Dalton and Barton, 2014).

Generally mammalian cells over-expressing a protein of interest can be constructed to be transient or stable expression system. Transient expression systems allows protein production within a few days after the mammalian cell line have been transfected with expression plasmid, whereas stable cell lines takes a longer time to develop (Andréll and Tate, 2013). However, stable over-expressing mammalian cell lines, provides a source that can be stored and reused over a long period, providing a source for multiple analysis and a medium to study the effect of a specific protein over-expression (Dalton and Barton, 2014)



Comments for pcDNA3.1/V5-His-TOPO
5523 nucleotides

CMV promoter: bases 209-863
T7 promoter/priming site: bases 863-882
Multiple cloning site: bases 902-1019
TOPO® Cloning site: 953-954
V5 epitope: bases 1020-1061
Polyhistidine tag: bases 1071-1088
BGH reverse priming site: bases 1111-1128
BGH polyadenylation signal: bases 1110-1324
f1 origin of replication: bases 1387-1800
SV40 promoter and origin: bases 1865-2190
Neomycin resistance gene: bases 2226-3020
SV40 polyadenylation signal: bases 3039-3277
pUC origin: bases 3709-4382
Ampicillin resistance gene: bases 4527-5387

Figure 5. 1: pcDNA™ 3.1/V5-His TOPO® cloning vector map (Invitrogen)

Table 5.1: TOPO® cloning reaction setup

Reagent	Quantity added
Fresh PCR product (10ng/ml)	2 µl
Salt solution	1 µl
Sterile water	2 µl
pcDNA™ TOPO® vector	1 µl
Total	6 µl

5.3.7.3: Transformation of ligated product

The One Shot® TOP10 chemically competent *E. coli* cells provided in the cloning kit (Invitrogen) as well as freshly prepared LB plates containing 100 µg/ml of ampicillin were used for this step of the cloning process according to the manufacturer's instructions. A vial of the One Shot® TOP10 was thawed on ice, 2 µl of the cloned reaction solution from the previous step was added and mixed gently. This reaction was incubated for 30 minutes on ice, after which, the cells were heat shocked for 30 seconds at 42^0 C without shaking, and the tube was immediately transferred to ice. 250 µl of room temperature SOC medium (Invitrogen) was added, and the tube was incubated while shaking horizontally at 37^0 C for one hour. 100 and 200 µl of the transformation solution was plated onto two different freshly prepared LB plates containing ampicillin, and the plates were incubated overnight at 37^0 C. Different colonies were picked from each plates for further analysis.

5.3.7.4: Colony PCR

The colonies were screened for the presence of the insert with the use of colony PCR. Ten different colonies were randomly picked from the transformed plates and

suspended in 5 µl of nuclease free, double distilled sterile water and this was used as a template for the PCR reaction. The parameters for the PCR were the same as previously described in section 5.3.4, using the 2X KAPA Taq Extra Hotstart ReadyMix with dye. Two positive clones were sent to Inqaba Biotec for sequence analysis.

5.3.8: Gene Knock down analysis

5.3.8.1: shRNA constructs

Short hairpin or small hairpin RNA (shRNA) were used for the knock down study. Glycerol stocks of five constructs of different shRNA sequences that target different parts of Gene 7 was purchased from Sigma Aldrich (MISSION shRNA). Each shRNA is cloned into the pLKO.1 lentivirus puro vector (Figure 5.2).

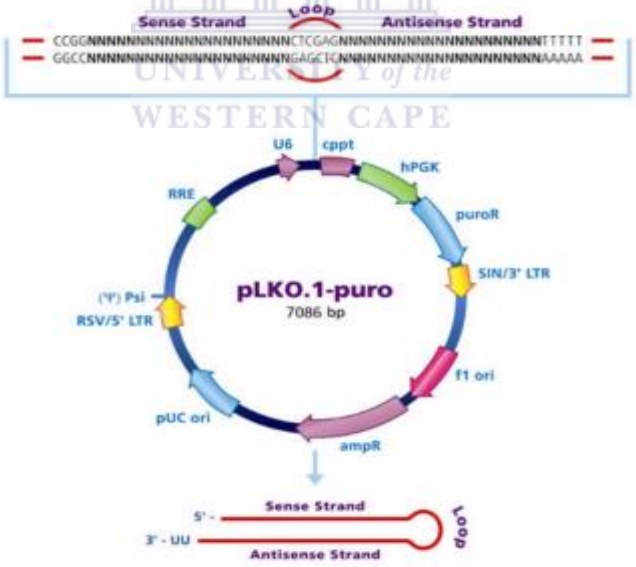


Figure 5.2: TRC1 and TRC1.5 Lentiviral Plasmid Vector pLKO.1-puro (Sigma Aldrich)

5.3.8.2: Isolation and culture of shRNA plasmid from glycerol stocks

The vials of glycerol stock representing the individual shRNA constructs, were gently spin down, and ice splinters of the frozen bacterial glycerol stock was removed from the vials using a sterile loop. The ice splinters was placed into a sterile culture tube containing 0.5 ml of LB without antibiotics, and the culture incubated at 37^0 C on a shaker for 30 minutes. 35 ml of the incubated culture was plated onto freshly prepared Luria broth (LB) plates containing 100 mg/ml ampicillin, and the plates were incubated over night at 37^0 C. Single colonies were picked from the plates and inoculated into a LB containing 100 µg/ml of ampicillin. The plasmid DNA were isolated from the overnight cultures broth.

5.3.8.3: Restriction endonuclease analysis

A double restriction enzyme digestion was used to confirm the isolated plasmid shRNA from the glycerol stocks. An *in silico* digestion was first carried out using Biolabs NEBcutter V2.0 (Vincze, *et al.*, 2003) to predict the number of expected bands and sizes from the digestion. The vector sequence (pLKO. 1-puro) was downloaded from the manufactural site (Sigma Aldrich) https://www.sigmaaldrich.com/ and the shRNA sequences were inserted into the vector sequence. The combined sequence were pasted in the box provided at the software site http://nc2.neb.com/NEBcutter2/, two restriction enzymes (BamH1 and EcoR1) that do not have restriction sequence for the shRNA were selected, and "submit" to carry out the analysis. The output was further analysed by selecting the "virtual image" to get a virtual image of electrophoresed digest, webpage was snapshot and image stored (result not shown).

Restriction enzyme digestion was then carried out in the laboratory using the BamH1 and EcoR1 fast digest from Fermenters according to manufacturer's instruction. The

reaction solutions were set as described in table 5.3.8. The solution was gently mixed, and incubated at 37^0 C and incubated for 60 minutes. The enzymes were deactivated by heating the solution at 80^0 C for 5 minutes.

Table 5.2: Double digestion set up of shRNA plasmid DNA

Reagents	Volume
Nuclease free water	31 µl
10X Fast digest buffer	5 µl
Plasmid DNA (up to 1µg)	10 µl
Fast digest enzyme1 (BamH1)	2 µl
Fast digest enzyme2 (EcoR1)	2 µl
Total	50 µl

5.3.9: Plasmid DNA isolation (pcDNA and shRNA containing plasmid)

The plasmid for both cloned pcDNA and the shRNA glycerol stocks were isolated using the Wizard® Plus SV Minipreps DNA purification system (Promega), according to the manufacturer's instructions. Colonies from the pcDNA clone for expression and shRNA plates were inoculated in 50 ml LB containing 100 µg/ml ampicillin, and cultured overnight at 37^0 C with shaking. The bacterial cultures were harvested by

centrifuging 10 ml of LB culture for 10 minutes at 3500 X g. The supernatants were removed and the tubes were blotted on paper towels. The pelleted cells were re-suspended in 250 µl resuspension solution by gentle pipetting after which, 250 µl of cell lysis solution was added and the solution gently mixed by inverting the tubes four times and the tubes incubated for 5 minutes. 10 µl alkaline protease solution was added to the solution and gently mixed by inverting the tubes four times before incubating for 5 minutes at room temperature.

The action of the alkaline protease was stopped by adding 350 µl neutralization solution and immediately mixed by inverting the tubes 4 times. The bacterial lysate was centrifuge at 14000 X g for 10 minutes at room temperature. A column provided in the extraction kit was placed in a collection tubes and the cleared lysate transferred into the column. The tubes containing the columns were centrifuged for 1 minute at maximum speed, the flow through was discarded, the column re-inserted into the tubes and 750 µl of column wash was added. The tubes were centrifuge at maximum speed for 1 minute, the flow through discarded and the wash repeated with a 250 µl of column wash buffer and centrifuged for 2 minutes at maximum speed. The columns were transferred into new, sterile 1.5 ml micro centrifuge tubes, and the plasmids were eluted with 100 µl nuclease free water, by adding the water to the column and centrifuge the tubes at maximum speed for 2 minutes. The isolated plasmid were electrophoresed on 1 % agarose gel as described in section 5.3.5 and the remaining stored at -20^0 C until required for further analysis.

5.3.10: Transfection of isolated plasmids

The pcDNA cloned plasmid and different constructs of shRNA of the gene of interest were transfected into a breast cancer cell line (MCF-7) and non-cancer breast cell line

(MCF-12A) using the Metafectene® Transfection Reagent for Mammalian Cells (Biontex), according to the manufacturer's instructions.

About 2 x 10^5 cells were seeded into six well plates and left to grow to about 70 % confluency. Two separate solutions comprising the plasmid DNA and the metafectene transfection reagent were made in serum free media. Solution A comprises 1 µg of DNA in 50 µl of serum and antibiotic-free medium and solution B comprises, 3 µl of mectafectene in 50 µl of serum and antibiotic-free medium. The solutions were mixed gently separately by pipetting and combined without any mixing. The mixture was incubated at room temperature for 25 minutes. After the incubation, the mixture was added drop wise to the cells, with gentle swirling, and returned into the incubator for 4 hours, after which complete media was added and the cells incubated for 24 hours.

5.3.11: Selection of stable transfected cells

Media with transfection reagent and plasmid DNA was removed after 24 hours of incubation of the cells, cells were rinse with sterile PBS, and fresh complete media containing antibiotics G418 and puromycin, expressed by the respective vectors. The pcDNA clones were treated with 800 µg/ml G418 (Roche), while the shRNA constructs were treated with 2 µg/ml Puromycin (Gibco®). Media containing these antibiotics were added to the respective transfected cells and the media changed regularly every 48 hours until all the cells in the control wells (cells not transfected with plasmid) were all confirmed dead. The selective media were then replaced with complete media into all wells including control wells, and media were changed every 48 hours until transfected cells grow to 70-90 % confluence and no cell growth was seen in the control wells. Once wells were confluent, the cells were trypsinized and sub-cultured in a 25 cm^2 flask for further analysis. Transfected cells were monitored

using a light microscope (Nikkon TMS) and images captured using a Leica EC3 camera (Leica Microsystem Switzerland). The total protein was extracted and western blot analysis of BRG 7 was carried out as described earlier in section 4.6, to confirm over-expression and “knock down”.

5.4: Results and Discussion

5.4.1: Polymerase Chain Reaction (PCR) of cDNA using the two sets of primers

Figure 5.3 shows the PCR products of the two sets of primers (A and B) used to amplify cDNA's from different breast cell lines. The figure showed the expected size of the gene (741bp). PCR amplification was then repeated using primer A and cDNA from MCF-12A, the PCR product was electrophoresed (result not shown). The DNA (PCR product) was extracted from the gel and cloned into the pcDNA3.5 TOPO vector. The ligated products was used to transformed TOP10 bacterial competent cells (invitrogen).

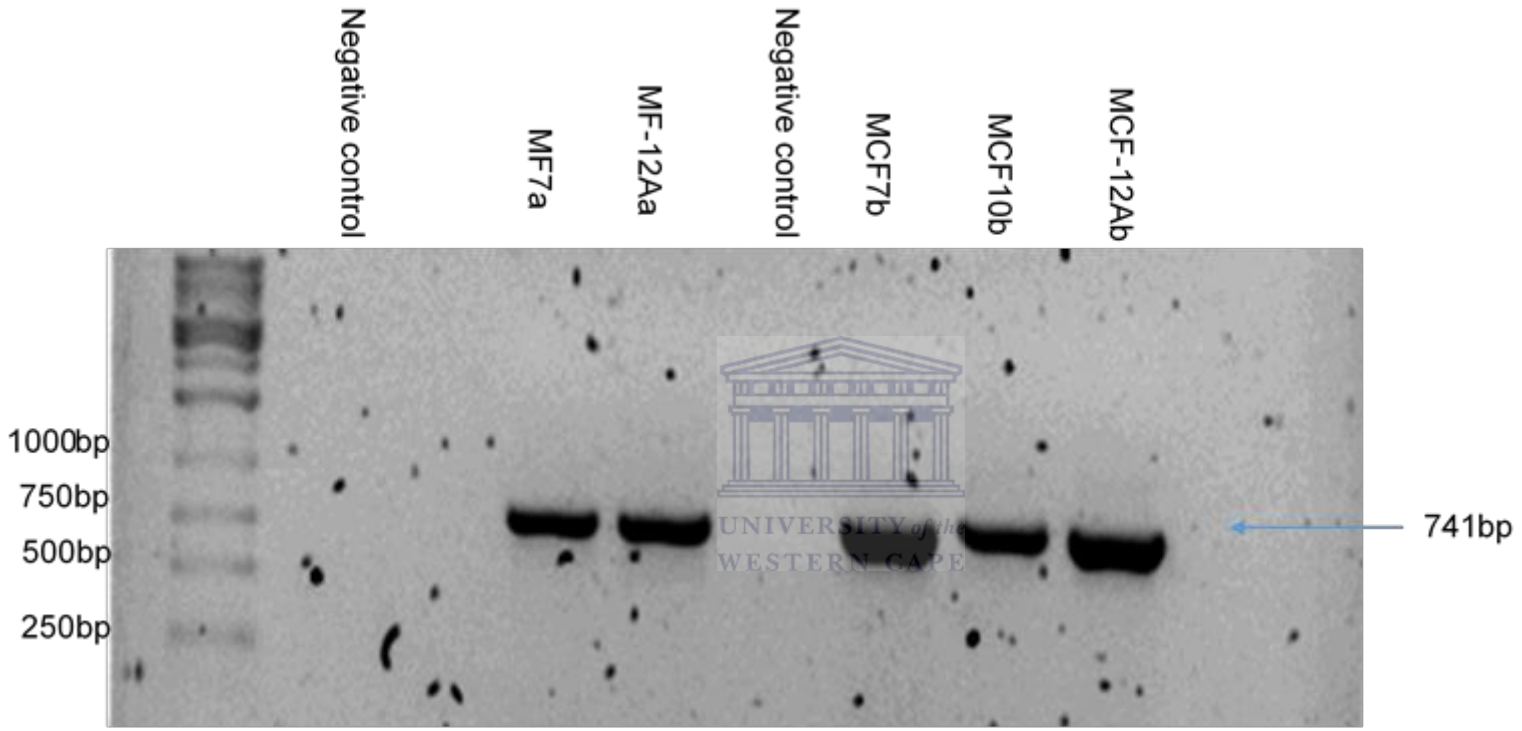


Figure 5.3: PCR amplification of Gene 7 with cDNA from different breast cell lines using two sets of primers.

5.4.2: Colony PCR

Transformed bacterial cells, following cloning, were streaked onto two LB plates containing ampicillin antibiotics. The plates were incubated for 24 hours at 37^0 C and five individual colonies from each plate were selected and used for colony PCR.

Figure 5.4 showed the electrophoresed PCR products of the five different colonies taken from the respective LB plates. Columns represent PCR products of five colonies taken from two different plates representing transformed bacteria cells of clones of Gene 7.

Plasmids that express the BRG 7 protein were successfully cloned and confirmed using the colony PCR (Figure 5.4). Colonies that were successfully transformed with the gene of interest were screened using the gene specific primers thus expecting a PCR product of 714 bp. First column represents the negative control without any template (bacterial cells) but water. The columns that showed specific bands of the expected product size of Gene 7. Colonies 6, 7, 8 and 10 were sent for sequencing at Inqaba Biotec to confirm that the gene of interest was successfully cloned into the vector. Products of colony 7 and 8 showed 99 % similarity in sequence to the gene of interest using the NCBI nucleotide blast tool (BLASTn). The DNA sequence result received from Inqaba Biotec was translated into the protein sequence using the ExPASy nucleotide-translating tool (www.expasy.org) and was confirmed to be the protein of interest by blasting the protein sequence against sequences in the NCBI database using BLASTp tool and colony 8 were used for further analysis after isolating plasmid DNA.

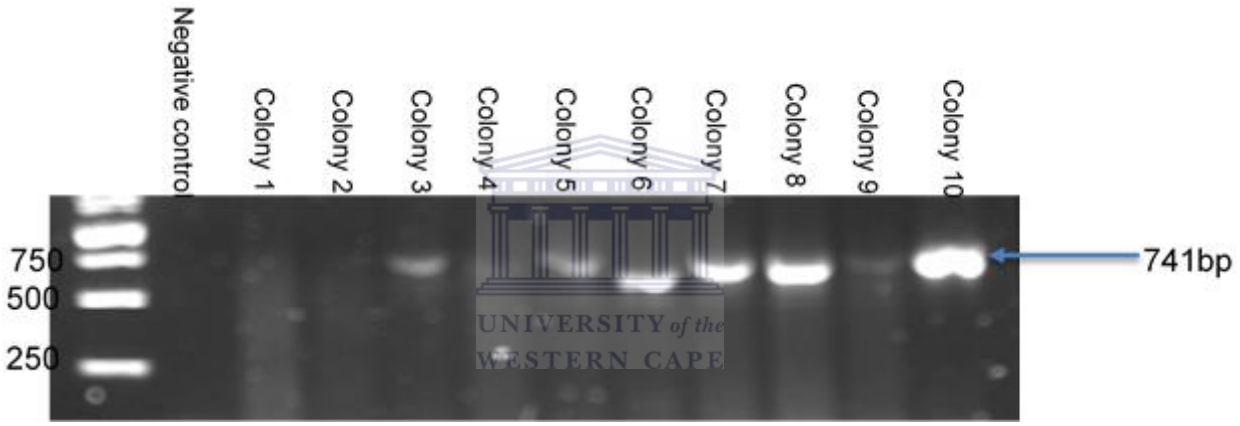


Figure 5.4: Colony PCR products of transformed bacteria cells electrophoresed on a 1 % agarose gel

5.4.3: Isolation of plasmid DNA from shRNA glycerol stocks

Figure 5.5 shows the isolated plasmid DNA carrying shRNA constructs against the gene of interest. The plasmids were isolated from glycerol stocks of the PLKO.1 lentivirus clones of three different shRNAs that target different regions of Gene 7 (Sigma Aldrich). Lentivirus vectors have been reported to be an efficient delivery tool of short hairpin RNAs (shRNAs) in transformed and primary cells for effective silencing of target genes (Rubinson, *et al.*, 2003).

The figure confirmed plasmids were successfully isolated from the glycerol stock and the plasmid DNA are of the expected molecular weight according to the manufacturer product sheet (Sigma Aldrich).

To further confirm the isolated plasmids as the PLKO.1 vector with inserted shRNAs, endonuclease digestion was performed after an *in silico* digestion analysis using the NEBcutter software (Biolabs), using BamH1 and EcoR1 restriction endonucleases.

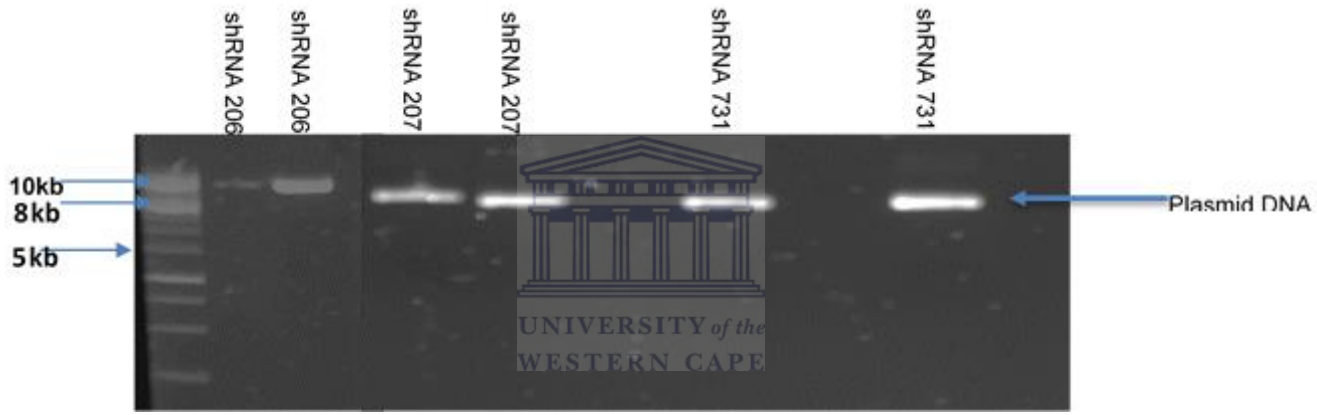


Figure 5. 5: Electrophoresed plasmid DNA of five different shRNA construct that targets Gene 7

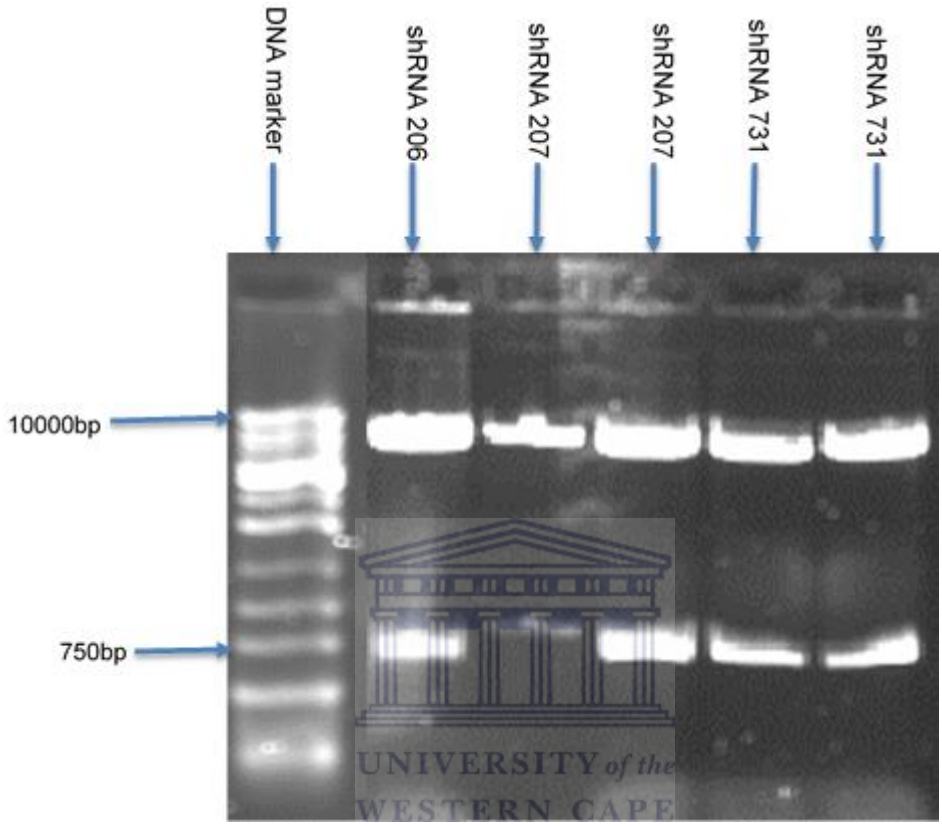


Figure 5.6: Endonuclease digestion of plasmid shRNA digested using BamH1 and EcoR1. Digested reactions electrophoresed on 1% agarose gel.

Figure 5.6 shows the electrophoresed products of double digestion of restriction endonucleases BamH1 and EcoR1 respectively. The number of bands and their approximate sizes of the virtual gels generated from the *in silico* analysis using the NEBcutter software (results not shown) were similar to the double digestion analysis using BamHI and ECORI endonucleases. This confirms the glycerol stocks contain samples of the pLKO.1 vector constructs with the different shRNA targets and could be used for further analysis.

5.4.4: Creating stable transfected cell lines

The breast cancer cell line MCF-7 and breast non-cancer cell line MCF-12A were transfected with the different constructs for protein over-expression (pcDNA clone) and knockdown (shRNAs constructs). Figures 5.7 and 5.8 shows transfected MCF-12A and MCF-7 transfected cells respectively after selection with appropriate antibiotics. The breast non-cancer cells (MCF-12A) showed a change in cell morphology and inhibition in growth in cells transfected with the clone for protein over-expression (figure 5.7). However, the over-expression of the BRG 7 protein seem not to affect the morphology of the breast cancer cell line (MCF-7) (figure 5.8). This suggests the protein plays an important role in cell growth and it's over expression may lead to cell senescence in normal cells.

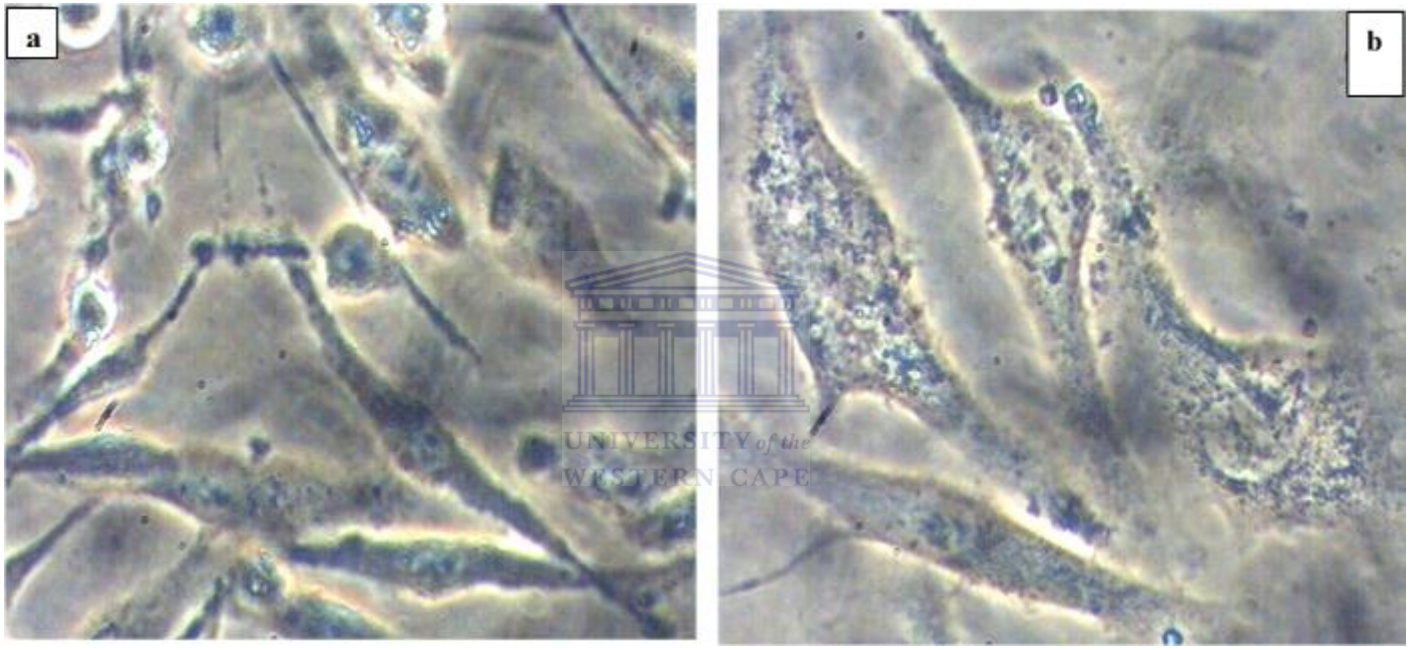


Figure 5.7: Breast non-cancer (MCF-12A) cell lines (a) normal MCF-12A un-transfected cell line (b) stable transfected MCF-12A cell line over-expressing BRG 7 protein.

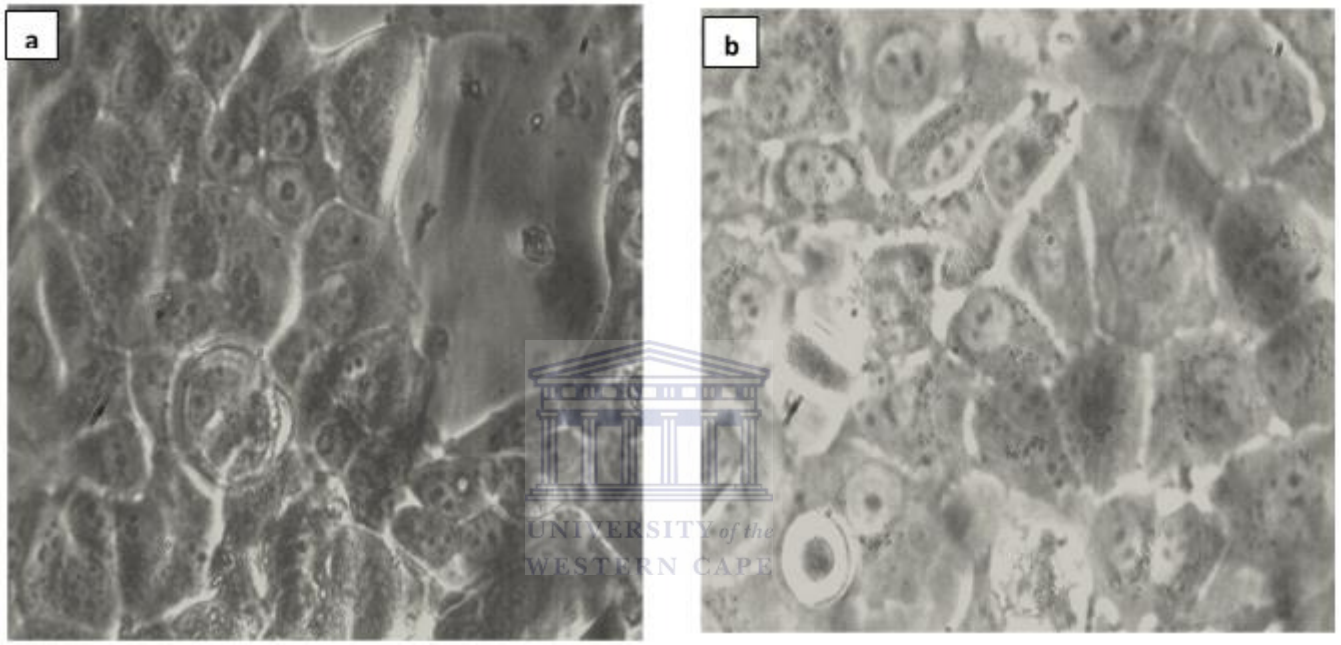


Figure 5.8: Breast cancer (MCF-7) cell lines (a) normal MCF-7 un-transfected cell line (b) stable transfected MCF-7 cell line over-expressing BRG 7 protein

Increased NKG2D ligand expression is regulated by activation of DNA damage. DNA damage leads to cell apoptosis and a senescence-like phenotype cell in response to most of the chemotherapeutic agents commonly used in clinical trials (Soriani *et al.*, 2009). A major feature of senescent cells that distinguishes them from most quiescent cells, is a large flattened morphology and high expression of senescent-associated β-galactosidase activity. They are also known to down-regulate genes involved in proliferation and extracellular matrix production and up-regulate inflammatory cytokines as well as other molecules known to modulate the immune response (Krizhanovsky *et al.*, 2008).

The MCF-12A cells ceased growing after the first passage, and the total protein extracted was very low. This could be as a result premature and spontaneous aging of the cell due to BRG 7 over-expression. Like most cells undergoing senescence, the transfected MCF-12A cells showed a striking change in morphology. They became flat, and multinucleated. A flat cell phenotype is commonly seen in cells undergoing H-ras induced senescence (Kuilman *et al.*, 2010).

Abnormal expression of H-ras in normal fibroblast has been shown to result in a cell senescence type that is very similar to replicative senescence (Mason *et al.*, 2004). Cellular senescence leads to permanent cell cycle arrest that limits the uncontrolled proliferation and tumourigenesis of cells (Reddel, 2000). Senescence was first observed in replicative cultured human fibroblast cells that exhausted their replicative potential after a number of cell division and this was termed replicative senescence (Mason, *et al.*, 2004). However, studies have shown that senescence can be triggered not necessarily by number of cell division or length of telomere, but also by various types of stress induced by various types of compounds and molecules and this type of senescence is referred to as premature senescence (Shay and Roninson, 2004).

Figures 5.9 and 5.10 shows immunoblots of BRG 7 protein and β-actin respectively in stable transfected MCF-7 and MCF-12A cell lines. The β-actin immunoblot result (figure 5.10) showed the proteins are well transferred and the protein concentration loaded are adequate. Figure 5.9 immunoblot showed the expression of the protein to be higher in pcDNA cloned transfected MCF-7 cell line compared to the MCF-12A cell line. Also, the blot showed the gene expression was successfully "knocked down" in both MCF-7 and MCF-12A transfected with the shRNA constructs.

The over-expression of the BRG 7 protein in cancer cell line and non-cancer breast cell line transfected with pcDNA vector clone and those transfected for gene "knock-down" were confirmed using the western blot analysis. Unlike the non-cancer breast cell line (MCF-12A), that showed the cells becoming senescent as a result of over-expression of the protein, the MCF-7 cells did not present any visible physiological change after transfecting them with both clones for over-expression as well as those for "knockdown". However, a longer period of time was required for some of the shRNA construct transfected cells to grow back after transfection. This suggests that, though the over-expression of this protein may initiates cell senescence; its basal expression might be needed for normal cell growth.

One of the characteristics of cancer cells is their ability to overcome senescence, and one of the mechanisms suggested by which cancer cells overcome growth restriction posed by senescence is by overcoming the cell autonomous blockade placed to cause senescence in oncogenic damaged or stressed cells, though it is not so clear how tumour cells overcome this regulatory phenomenon (Rodier and Campisi, 2011).

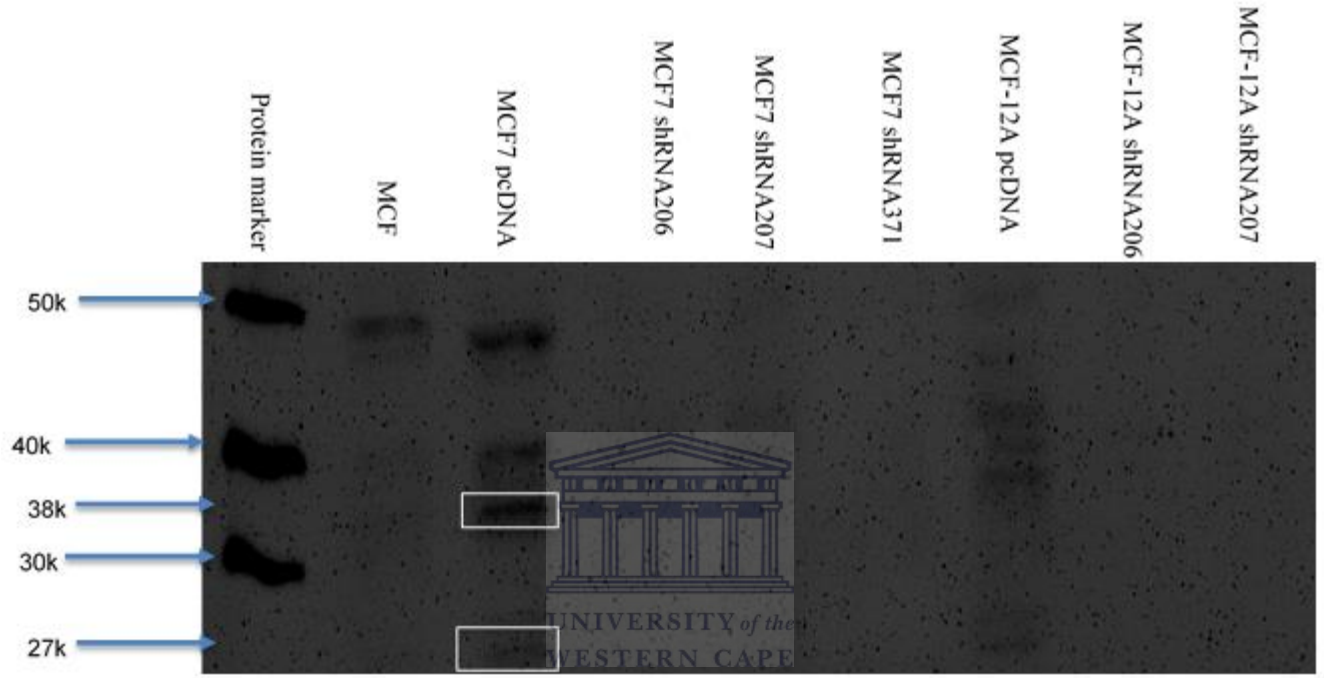


Figure 5.9: Western blot analysis of transfected cells for gene "knock down" and protein over-expressing BRG 7 in MCF-7 and MCF-12A cell lines.

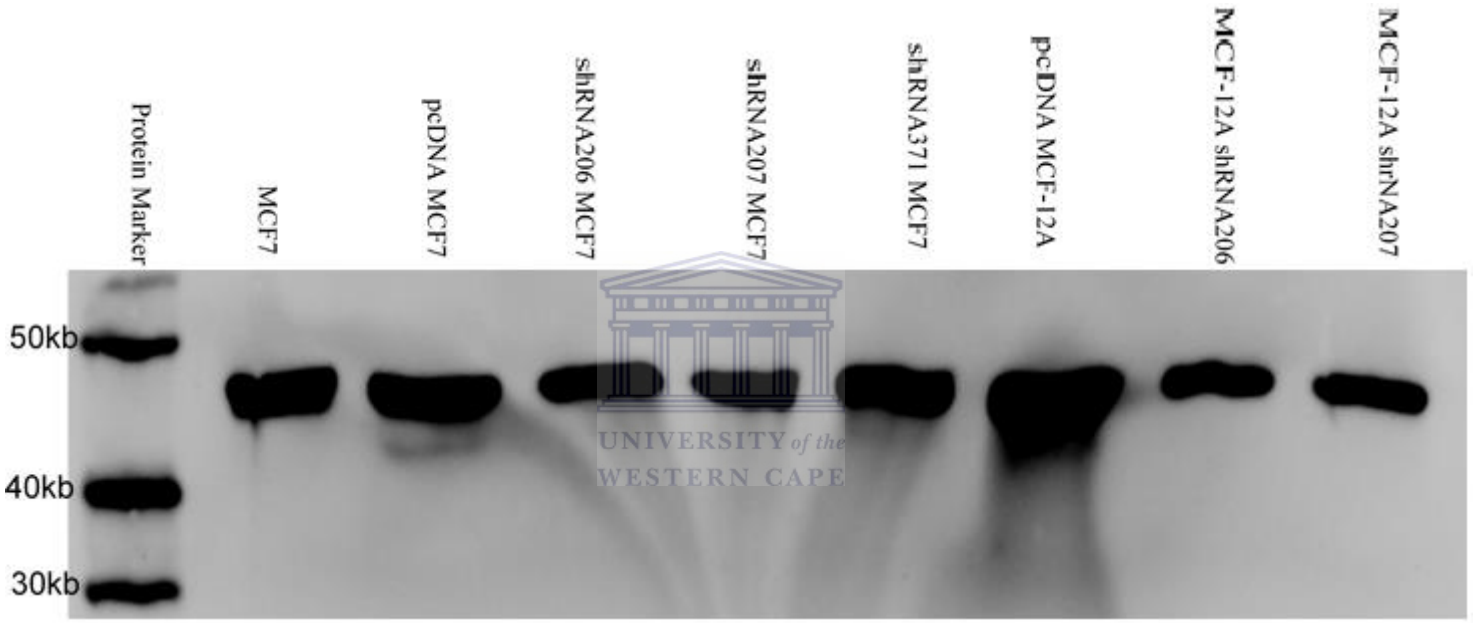


Figure 5.10: β-actin expression analysis in non-transfected and transfected cell lines.

5.6: Conclusion

Clones that express BRG 7 protein of interest was successfully generated, the plasmid isolated and transfected into breast cancer cell line (MCF-7), and breast non-cancer cell line (MCF-12A). Also plasmid DNA representing different shRNAs that target BRG 7 were isolated from commercially available glycerol stocks from Sigma Aldrich, and were successfully transfected into MCF-7 and MCF-12A cell lines.

Western blot analysis confirmed higher expression of this protein in the breast cancer cell line (MCF-7) and non-cancer cell line (MCF-12A). The over-expression in MCF-12A cell line resulted in the cell going into senescent, characterized by stop of cell growth and change in cell morphology. It seems the breast cancer cell line may have a way to overcome the senescence induced by the over-expression of BRG 7. However, further analysis is required to confirm this.

Thus, further effect of over-expression of BRG 7 and its repression in enhancing chemotherapy induced apoptosis, was monitored in the breast cancer cell line in the next chapter.

Chapter 6: Effect of the NKG2D ligand on apoptosis induction in breast cancer

6.1: Background

Programmed cell death or apoptosis is essential for the development of metazoans and the development and maintenance of cellular homeostasis of the immune system in mammals (Henson and Hume, 2006). It is a genetic process of cell death that responds to both normal and pathological stimuli (Opferman and Korsmeyer, 2003). One of the mechanism by which apoptosis is exerted on the target cell is by exocytosis. The granules from the cytoplasm of the NK cell are transferred into the target cells. Also the interaction of the Fas ligand (FasL) or Tumour Necrosis Factor (TNF) on the Natural Killer (NK) cell with Fas (CD95) or TNF receptor on the tumour target cells could be employed. The NK cells themselves can also undergo apoptosis following the interaction of certain NK receptors with the target cells (Snow *et al.*, 2010).

The innate immune cells, which include the natural killer (NK) cells, play an important role in tumour regression (Textor *et al.*, 2011). NK cells are activated by the interaction of receptor ligands expressed by tumour cells and the specific activating receptors expressed by the NK cells. These NK cell receptors include cytotoxic receptors, ADAM-1 and NKG2D. Also the NK cells expresses the MHC class I receptor, an inhibitory receptor whose expression level must be low in NK cells for it to be activated (Raulet and Guerra, 2009).

NKG2D ligands are distant family members of the MHC class I molecules and generally, they are of two major types; MICA/MICB and RAET1/ULBP (ULBP1-10). The MIC gene family members are mainly found in the human major histocompatibility complex,

whereas the ten members of the ULBP family can be found in humans and mice. In mice, the Raet1 genes are subdivided into Rae1, H60 and Mult1 gene subfamilies, and are relatively distinct in amino acid sequence to each other (Gasser and Raulet, 2006).

The NK ligands are normally not expressed on the cell surface of healthy cells. Tumour cells frequently express high levels of activating NK cell receptor ligands and low levels of MHC class I molecules, rendering these cells highly susceptible to NK cell–mediated lysis (Textor *et al.*, 2011). Soluble NKG2D ligands have been reported in various cancer types in humans, and many studies have suggested that the elevation of the concentration of these ligands in the serum of cancer patient could represent an immune escape mechanism that evolve in cancer cells to evade being detected in their host (Gasser and Raulet, 2006).

Tumour cells are subdued by different anti-tumour effector lymphocytes. However, tumour cells are known to evade the immune surveillance by several means. For example, cytotoxic T-lymphocyte (CTL)-mediated surveillance can be escaped by tumour cells by the repression of the expression of the HLA class I, which will affect the recognition of tumour cells (Poggi *et al.*, 2005). Various types of cancer have been linked with different immune cancer cells (Karin and Greten, 2005). NK cells infiltration has been found in human gastric and colorectal carcinoma, whilst innate immune cell types like macrophages are seen in human breast cancer and mast cells in human lung cancer types (de Visser *et al.*, 2006).

The NKG2D ligands have shown great importance in shaping NK cell effector function, as a result of cytotoxicity or cytokine production (Wang and Sun, 2014). The molecular

basis for the ability of NK cells to recognize stressed cells is not well understood. However, studies suggest that cells exposed to chemotherapeutic agents, genotoxic stimuli, or inhibited DNA replication cycles, upregulates the stimulatory NKG2D ligands (Gasser and Raulet, 2006). Different genotoxic compounds (e.g., aphidicolin, cisplatin, fluorouracil, γ-irradiation, and UV radiation), response to heat shock, antigen-mediated activation of T cells, and dysregulation of Dicer expression, have been shown to stimulate upregulation of the expression of NKG2D ligands (Fine *et al.*, 2010).

Soluble NKG2D ligands have been reported in the sera of patients of different cancer types, and their increased protein concentrations correlated with poor cancer prognosis (Paschen *et al.*, 2009). Wang and Sun, (2014) reported a lost in expression of ULBP2 on the surface of tumour cells and a shedding of this NKG2D ligand was linked to target cell apoptosis in contrast to spontaneous shedding, though, both the spontaneous shedding and shedding due to apoptosis were due to metalloproteinase activity (Wang and Sun, 2014).

Also tumour suppressors have been reported to modulate the expression of NKG2D ligands. Heinemann *et al.*, (2012) reported that tumour suppressing miRNA suppresses the expression of the NKG2D ligands by binding at the 3'-untranslated region. The study specifically showed that the member of the miR-34 family (miR-34a and miR-34c) directly regulates ULBP2 by degradation and affecting its translation in a melanoma cell line (Heinemann *et al.*, 2012). This class of miRNA are reported to induce cell cycle arrest, senescence as well as apoptosis and generally regulate cellular differentiation (Heinemann and Paschen, 2012).

Apart from the intrinsic factors and the natural features of cancer cells, the host immune response could serve major prognostic and predictive roles in determining the fate of cancer patients treated with conventional or targeted chemotherapies (Zitvogel *et al.*, 2011). The NK cells are reported to present the most efficient immune mechanism that detects and destroy tumour cells, thus, playing a significant role in host defence against carcinogenesis (Pende *et al.*, 2002). The NKG2D protein is a major receptor of the NK cells that triggers a cytotoxic response when activated by the interaction of tumour cells through the NKG2D ligands (Moretta and Moretta, 2004).

Considering the important role of NKG2D and its ligands in predicting cancer outcomes in different cancer types (Bui *et al.*, 2007), there is a need to further monitor the role of BRG 7 (a NKG2D ligand) in the modulation of chemotherapy induce apoptosis in breast cancer for it to serve as a prognostic or predictive biomarker in breast cancer disease. Range of compounds has been reported to activate apoptosis through different pathways (Marks and Xu, 2009). Due to this consideration three different established apoptotic inducers namely: Camptothecin, Doxorubicin and DMSO were used to monitor the role of BRG 7 in apoptosis induction in breast cancer.

6.1.1: Camptothecin

Camptothecin (CPT) is a cytotoxic alkaloid reputed for its potent apoptosis inducing property. It was originally isolated from the *Camptotheca acuminate* plants native to China and Tibet and have been extensively used in Chinese medicine (Thomas *et a*l., 2004). CPT and its analogues are reported to cause cells to enter into apoptosis by inhibiting the function of the enzyme topoisomerase I (topo I) (Liu *et al.*, 2000). DNA topo I relaxes super coiled DNA through the cleavage of the double stranded DNA and

the formation of a phosphoryl bond between the cleaved DNA and the active site tyrosine of the enzyme (Holden, 2001). The activity of topo I is inhibited by camptothecin by preventing the re-ligation of the cleaved DNA strands. Camptothecin reversibly stabilizes the topo I-DNA complexes and the replication forks during the S-phase of the cell cycle, which activates stress-associated signalling that ultimately induces apoptosis (Thomas *et a*l., 2004).

6.1.2: Doxorubicin (Adriamycin)

Doxorubicin (DOX) is an anthracycline antibiotic from the bacterium *Streptomyces*, considered to be one of the most effective agents in the treatment of breast cancer. DOX is used to treat early-stage or node-positive breast cancer, HER2-positive breast cancer and metastatic disease and resistance to it leads to an unsuccessful outcome in many breast cancer patients (Smith *et al.*, 2006). Induction of apoptosis by DOX is majorly believed to be through its action of inhibiting topoisomerase II (topo II), thereby preventing the breaking-religation of this enzyme, as well as through the formation of reaction oxygen species (ROS) (Döhrmann *et al.*, 2012). Free radical formation and DNA damage by inhibiting the action of topo II may primarily be responsible for the cytotoxic effect of DOX. This drug is also known to induce apoptosis by activating caspases and disruption of mitochondrial potential (Eom *et al.*, 2005).

6.1.3: Dimethyl sulfoxide

Dimethyl sulfoxide (DMSO) is a polar solvent that readily permeates the cell membrane, majorly use in cell biology as solvent for chemical compounds designated for treatment in cell culture, for cryopreservation of many cell types and for improving primary cells viability in cell culture (Banič *et al.*, 2011) . DMSO has been reported to be effective in

initiating apoptosis in disease cells by inducing the release of Cytochrome c in the mitochondria, collapse of the mitochondrial membrane potential, activation of caspase-9 and -3 as well as induces apoptotic changes in a murine lymphoma cell line and down-regulation of Bcl-2 (Santos *et al.*, 2003).

Stably transfected MCF-7 cell lines for BRG 7 protein over-expression was henceforth referred to as pcDNABRG 7, while BRG7shRNA206 and BRG7shRNA207 were used to represent the stably transfected cells for Gene 7 "knock-down" in this study

6.2: Aim

The aim of this chapter was to monitor the role of the BRG 7 (a member of the NKG2D ligand proteins) in modulating apoptosis induced in breast cancer cell lines using three pro-apoptotic compounds.

6.2.1: Objectives

- Investigate the role of BRG 7 over-expression in camptothecin, doxorubicin and DMSO induced apoptosis in stably transfected breast cancer cell line (pcDNABRG7)

- Investigate the role of Gene 7 "knock down" in camptothecin, doxorubicin and DMSO induced apoptosis in stably transfected breast cancer cell line (BRG7shRNA206 and BRG7shRNA207).

6.3: Materials and methods

6.3.1: Cell culture and cell treatment

pcDNABRG7, BRG7shRNA206 and BRG7shRNA207 were seeded into 6 well plates using the MEM reconstituted media comprising 1 % penicillin/streptomycin and 10 % fetal serum bovine (FBS) solution. The cells were cultured as described earlier in section 4.3, at 37^0 C in a 5 % CO_2 aerated incubator to between 70-80 % confluency. Apoptosis was induced in cells using camptothecin (100 µg/ml) dissolved in DMSO and diluted further in complete media to a final concentration of 2 µM, an effective concentration established by Abedin *et al*., (2007), 5 % v/v Dimethyl sulfoxide (DMSO) in complete media, and 1 µg/ml of doxorubicin (Cuvillier *et al*., 2001) in different plates for 24 hours. After 24 hours, cells were harvested and the apoptotic cells measured using the flow cytometer (BD accuri 6) after performing the Apo-percentage assay.

6.3.2: Apo percentage assay

Seeded cells were treated with various concentration of different pro-apoptotic compounds stated above in complete media for 24 hours. After 24 hours, the media were removed into 15 ml tubes. Wells were washed with 1 ml PBS and added to the individual 15 ml tubes. Cells were trypsinized by the addition of 0.5 ml 1X trypsin to each well and monitored until they were about to detach and the trypsinization process stopped by adding 0.5 ml complete medium. The cells were gently washed off the plates and added to the individual 15 ml tubes. The cells were pelleted by centrifuging the tubes at 3000 rpm for 10 minutes, and the supernatant discarded. Excess supernatant was removed as much as possible and the pellets re-suspended in 250 µl diluted (1:160) 3, 4, 5, 6-tetrachloro-2', 4', 5', 7'-tetraiodofluorescein (TCTF). Cells were incubated in dye at 37^0

C for 30 minutes, after which 4 ml of PBS was added and the tubes centrifuged for 5 minutes at 3000 rpm, and supernatant discarded. Cells were re-suspended in 60 µl PBS and analyzed on a flow cytometer.

6.3.3: Flow Cytometry Analysis

Stained apoptotic cells were analysed using the FL2 filter on the BD Accuri™ C6 flow cytometer (BD Biosciences, San Jose, CA). Acquisitions were done in linear mode and 10,000 cells per sample were acquired for analysis.

6.3.4: Statistical analysis

All assays were repeated twice to ensure reproducibility and three replicates of each group were performed in each test. The significance of difference in apoptosis between un-transfected cells and transfected cells induced using the different effector agents (camptothecin, DMSO and doxorubicin) were analyzed by one way analysis of variance (ANOVA) followed by Benferroni's post-hoc comparison using Prism 5.0 (Graph pad, USA). The number of cells undergoing apoptosis was represented as bar charts depicting the mean ± standard error percentage of triplicate samples.

6.4: Results and Discussion

The effect of chemotherapy on cancer can be influenced by the hosts' immune system by either stimulating immune response to the disease or by increasing the susceptibility of the cancer cells to immune attack. Also, chemotherapy may directly stimulate anticancer immune effectors or repress immunosuppressive mechanisms (Zitvogel *et al.*, 2011).

Soluble NKG2D ligands of which BRG 7 is a member of, are reported to be present in sera of some cancer patients and are thus regarded as good predictive markers for the prognosis of these cancers (Wang and Sun, 2014). However, better understanding of the effector role BRG 7 protein play in stimulating apoptosis to repress breast cancer onset/progression could facilitate a point of care tool for breast cancer diagnosis, prognosis or use to monitor the efficacy of anti-tumour drugs.

Most studies carried out to validate the NKG2D ligands as biomarkers, majorly focused on the spontaneous release of these ligands in their bid to evade NK cell tumour surveillance. However, better understanding of the mechanism of shedding of BRG 7 may facilitate the development of BRG 7 based diagnosis methods for breast cancer (Wang and Sun, 2014; Paschen *et al.*, 2009; Guerra *et al.*, 2008). This study seek to understand the role of BRG 7 in modulating apoptosis induced by chemotherapeutics in a breast cancer cell line.

6.4.1: The effect of BRG 7 over-expression and silencing on camptothecin induced apoptosis in a breast cancer cell line

The sensitivity of cells to CPT is not only based on its action of DNA damage through the inhibition of topo I, but also on the ability of the cells to activate checkpoints in the S

and G2 phases of the cell cycle through the action of p53 protein. Thus tumour cells that lack functional p53 protein lose their sensitivity to a wide range of DNA-damaging agents including CPT (Jones *et al.*, 2000).

Expression of p53 in breast cancer cell lines has been reported to be repressed in order to promote cell proliferation (Chang *et al.*, 2007). A variety of molecules are known to override the DNA damage-dependent G2-checkpoint and enhance drug-induced cytotoxicity (Jones *et al.*, 2000).

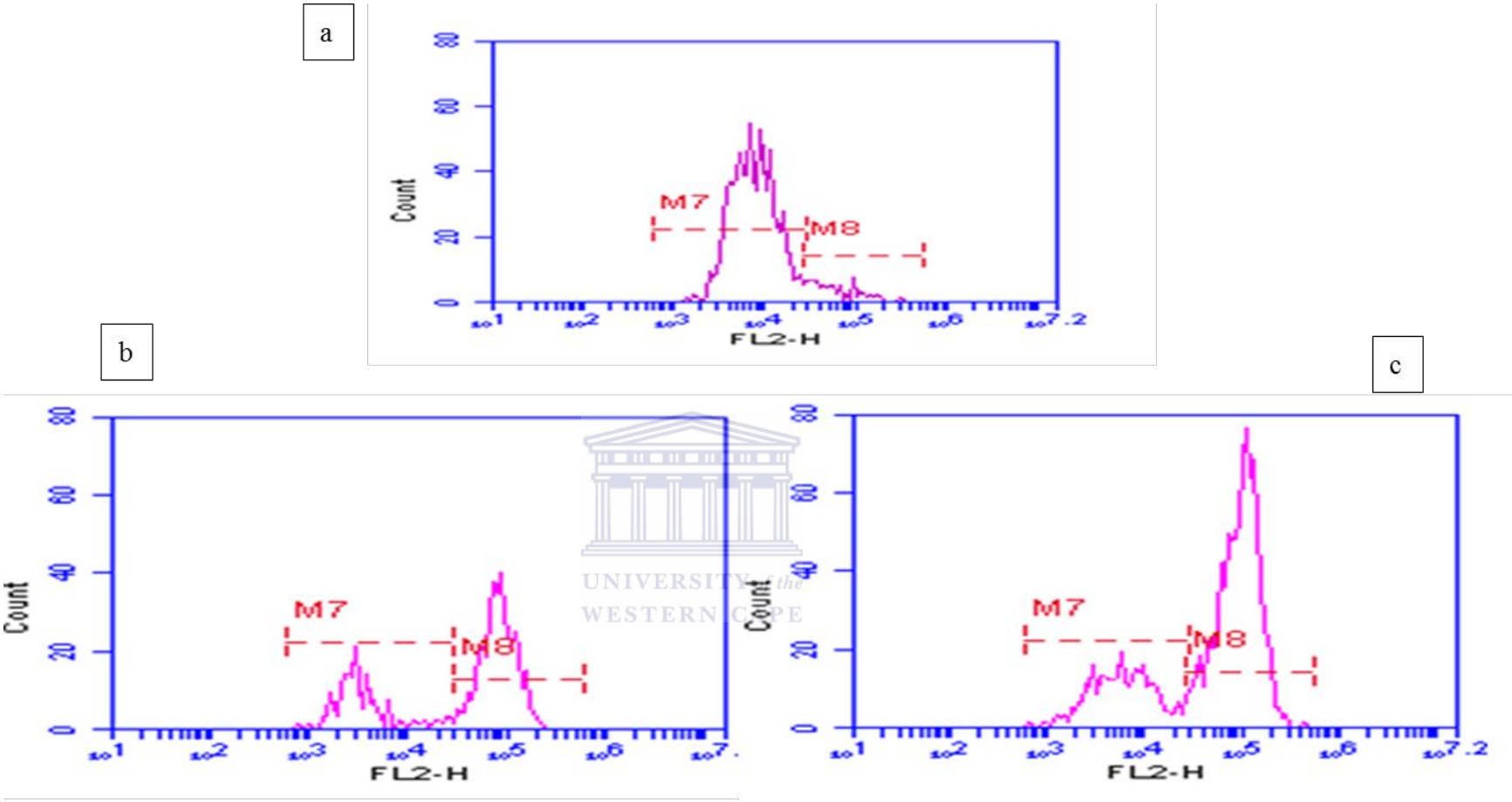


Figure 6.1: Quantification of apoptosis in MCF-7 cell lines using the flow cytometer (a) a representation of the MCF-7 untreated cell line (b) a representation of camptothecin treated MCF-7 cell line (c) a representation of pcDNA transfected, CPT treated MCF-7 cell line

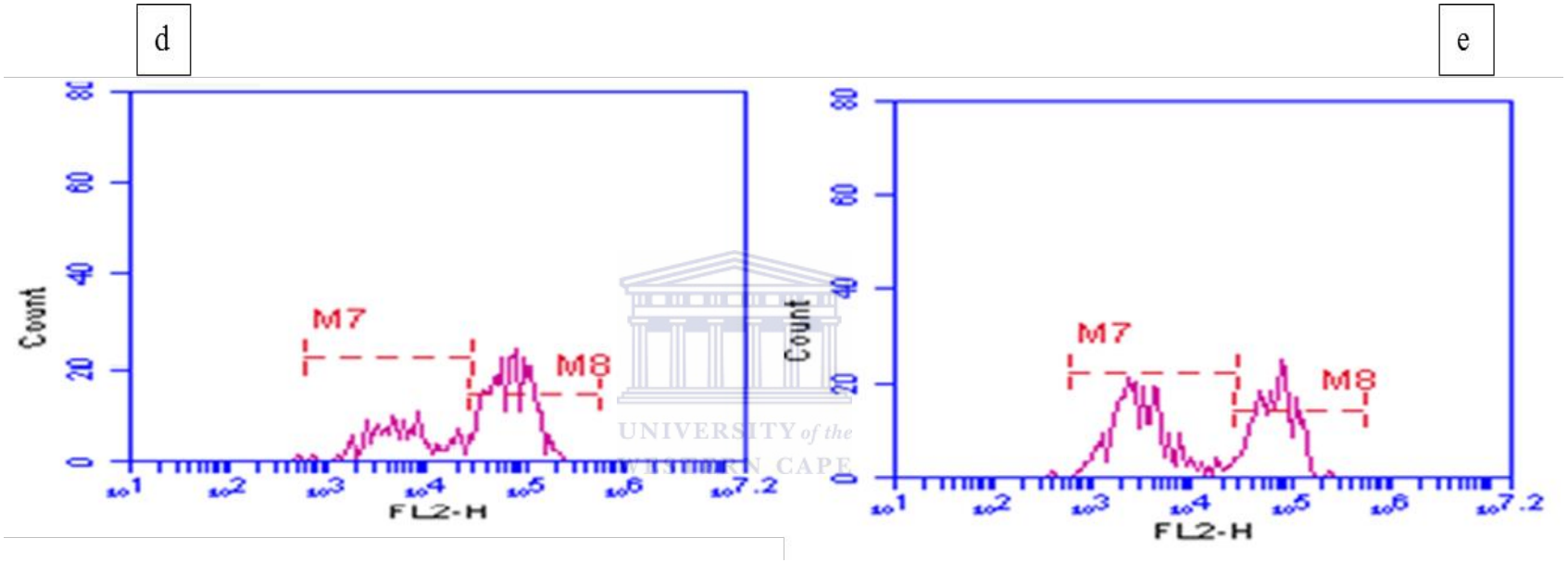


Figure 6. 2: Quantification of apoptosis in BRG7 silenced MCF-7 cell lines using flow cytometer (d) a representation of shRNA206 transfected and CPT treated MCF-7 cell line (e) a representation of shRNA207 transfected and CPT treated MCF-7 cell line

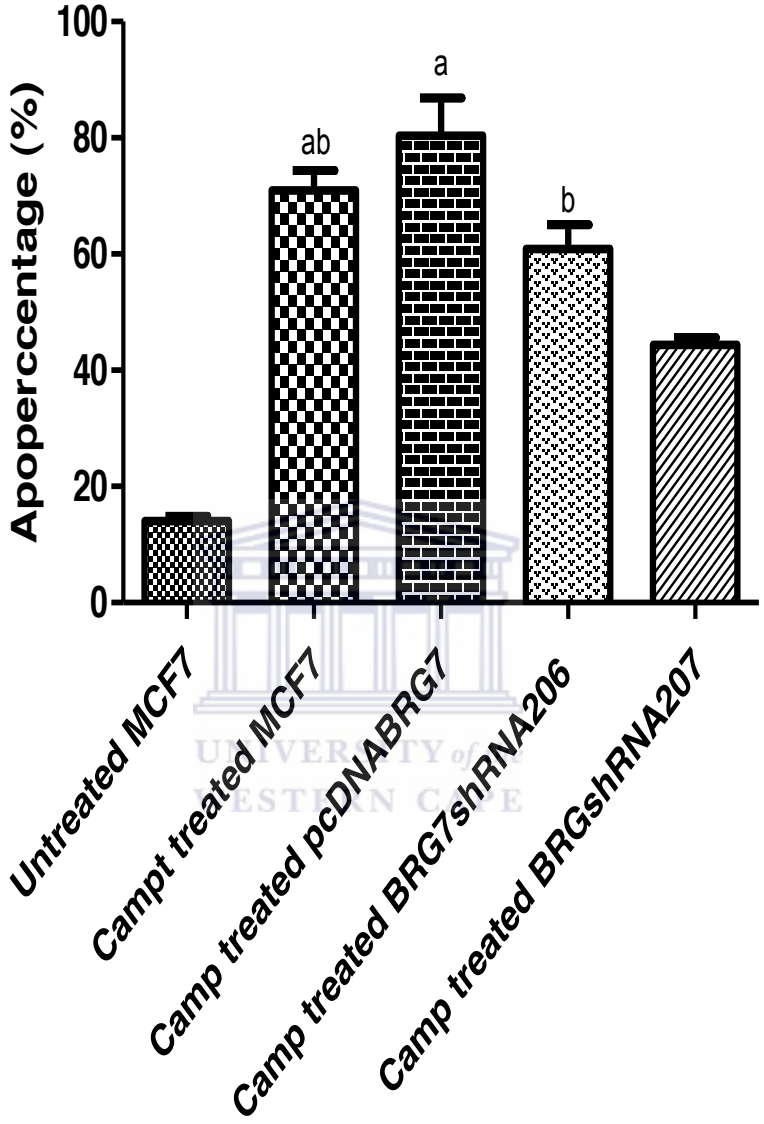


Figure 6.3: Percentage apoptotic cells of camptothecin treated MCF-7 cell lines; bar chart represents the mean and standard error of triplicate Apo-percentage values of cells after 24 hours of incubation with- or without camptothecin. Bars with same superscript are statically not significant ($p \leq 0.05$).

Figure 6.1 and 6.2 shows the flow cytometry percentage apoptotic cells (raw data) in a breast cancer cell line (MCF-7), transfected to over-express the BRG 7 (pcDNABRG7) protein and those transfected to "knock down" Gene 7 expression (BRG7shRNA206 and BRG7shRNA207) treated with camptothecin. Figure 6.3 (analysed data) showed the effect of camptothecin-induced apoptosis is higher in pcDNABRG7 compared to the un-transfected MCF-7 cell lines as well as in BRG7shRNA206 and BRG7shRNA207 cell lines.

Although the difference in percentage of apoptotic cells in the pcDNABRG7 cell is not statistically significant when compared to the un-transfected MCF-7 cell, the increase in apoptosis could be an indication that the NKG2D ligand, positively enhance the effect of CPT in stimulating apoptosis in breast cancer cells. Also, lower effect of this drug in MCF-7 cell lines transfected for the BRG 7 gene repression, lend more credence to the probability of the BRG 7 having a positive modulating effect in enhancing CPT action in breast cancer treatment, and the shedding and repression of this protein could be another means by which cancer cells resist the effect of cytotoxic drugs.

MCF-7 is an estrogen positive, tumourigenic, non-invasive, primary human mammary epithelial cell line that does not express caspase-3, the stimulatory protein of the hallmark caspase cascade. MCF-7 responds to variety of apoptotic stimuli such as UV radiation, paclitaxel, tamoxifen, just to mention a few. Caspase-3 activity is characteristic of the apoptotic pathway, thus, the route of apoptosis induction in MCF-7 is not too clear (Abedin *et al.*, 2007). Also the induction of apoptosis in MCF-7 by camptothecin is delayed (Abedin *et al.*, 2007).

Both innate and adaptive immune systems have been reported to play crucial roles in the anti-tumour effects of chemotherapy and radiotherapy based cancer treatment (Schreiber *et al.*, 2011). Based on the biochemical homogeneous process of apoptosis, it is generally assumed that pro-apoptotic chemotherapy would not elicit immunogenic cell death, thus cancer cells treated with mitomycin C, a drug that induces classical apoptosis, was shown not to stimulate an immune response when injected subcutaneously into histocompatible hosts in the absence of adjuvant therapy (Zitvogel *et al.*, 2008). However, DNA-damaging agents such as the topoisomerase inhibitors are reported to induce the activation of tumour-suppressor proteins including the transcription factor p53 and the DNA-damage response induces the expression of NKG2D ligands on tumour cells (Gasser, *et al.*, 2005). Thus, over-expression of BRG 7 a NKG2D ligand could possibly enhance the function or stimulation of p53 to help enhance drug-induced cytotoxicity. However p53 was not observed to directly interact with BRG 7 in the earlier *in silico* analysis study (Chapter 2).

Eukaryotic topoisomerase I (topo I) is an enzyme required for DNA replication to remove supercoils by introducing breaks and nick the DNA helical structure (Liu *et al.*, 2000). Camptothecin (CPT) antitumour activity was reported to include inhibiting the action of topo I by preventing its activity and causing an accumulation of topo I-DNA adducts reported to collide with the replication fork during DNA replication, resulting in double-strand break which leads to cell death (Staker *et al.*, 2002). NKG2D ligands are reported to be upregulated in cells as a result of chemotherapeutic agents, genotoxic stimuli, or stalled DNA replication and the upregulation of these ligands has been reported to be as a result of diverse genotoxic stimuli (Gasser and Raulet, 2006), thus some of these ligands

may be required for the effectiveness or to enhance the potency of the drugs. This could probably explain, the increase in apoptosis in cells transfected to express an increase level of the BRG 7 protein, as compared to those whose BRG 7 gene expression is knocked down both compared to the breast cancer cell line (MCF-7).

6.4.2: The effect of BRG 7 over-expression and silencing on doxorubicin induced apoptosis in breast cancer cell line

DOX is reported to be very effective in a wide variety of breast cancer types including both metastatic and early disease, but because of its reported cardiotoxicity activity, its use is limited (Lao *et al.*, 2013). Thus, the needs to identify biomarkers that can predict treatment response, enhance the drug efficacy, or help design therapy tailored to individual patients.

Doxorubicin like most chemotherapeutics, have been reported to stimulate the immune system to complement their role of clearing tumour cells by preventing an immunosuppressive effect in the tumour microenvironment or enhance the infiltration of immune effector cells into tumour cells to sensitize them for immune attack (Hu *et al.*, 2013). Doxorubicin and interleukin-12 have been reported to increase infiltration of $CD8^+$ T cells into cancer cells and also induces NKG2D expression on the $CD8^+$ T cells which enhances the antitumour immune surveillance of the cells (Hu *et al.*, 2014). The aim of this section was to monitor the effect of the BRG 7 protein in modulating the apoptosis inducing effect of DOX in a breast cancer cell line.

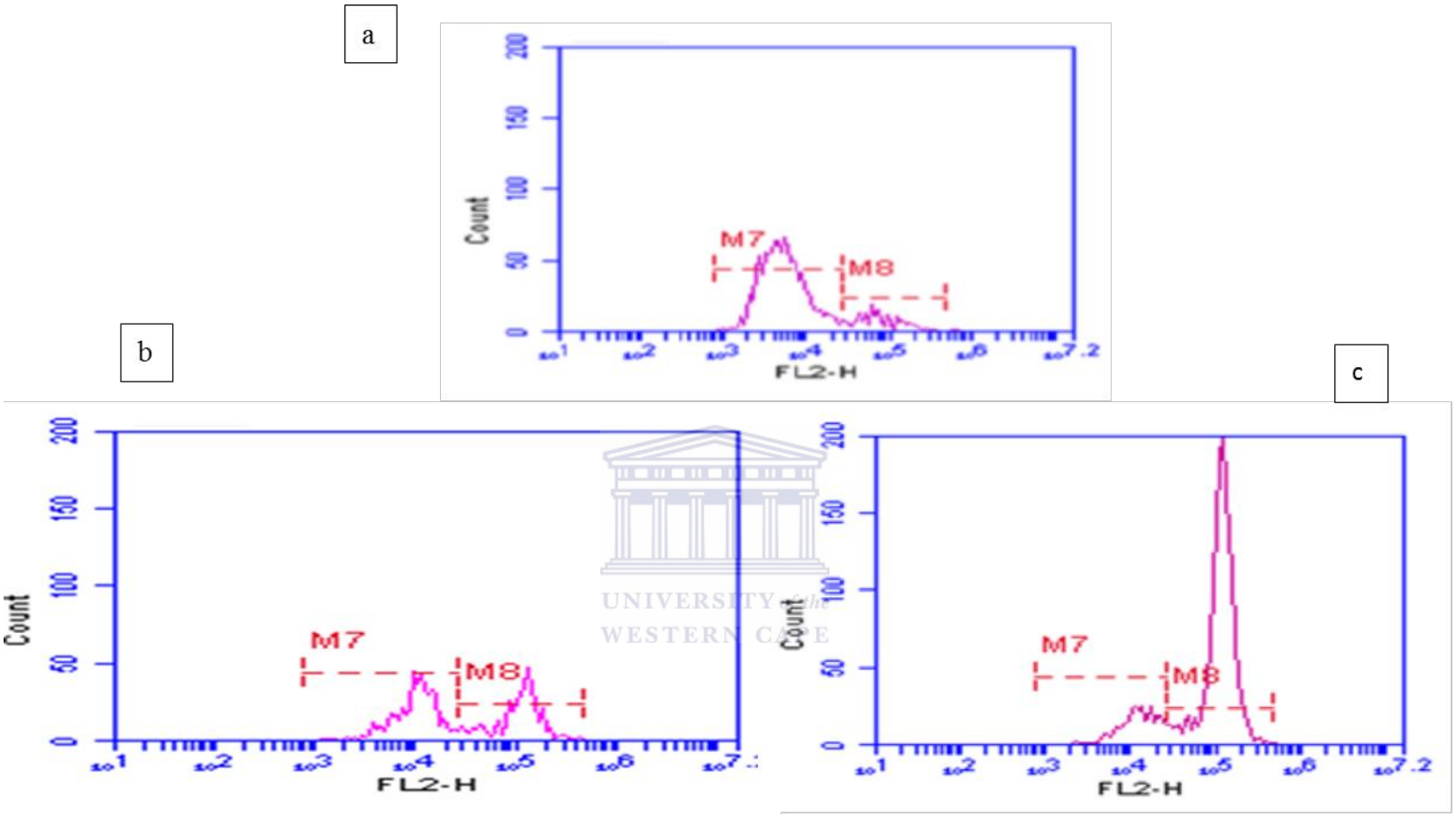


Figure 6.4: Quantification of apoptosis in MCF-7 cell lines using flow cytometer (a) a representation of the MCF-7 untreated cell line (b) a representation of DOX treated MCF-7 cell line (c) a representation of pcDNA transfected, DOX treated MCF-7 cell line

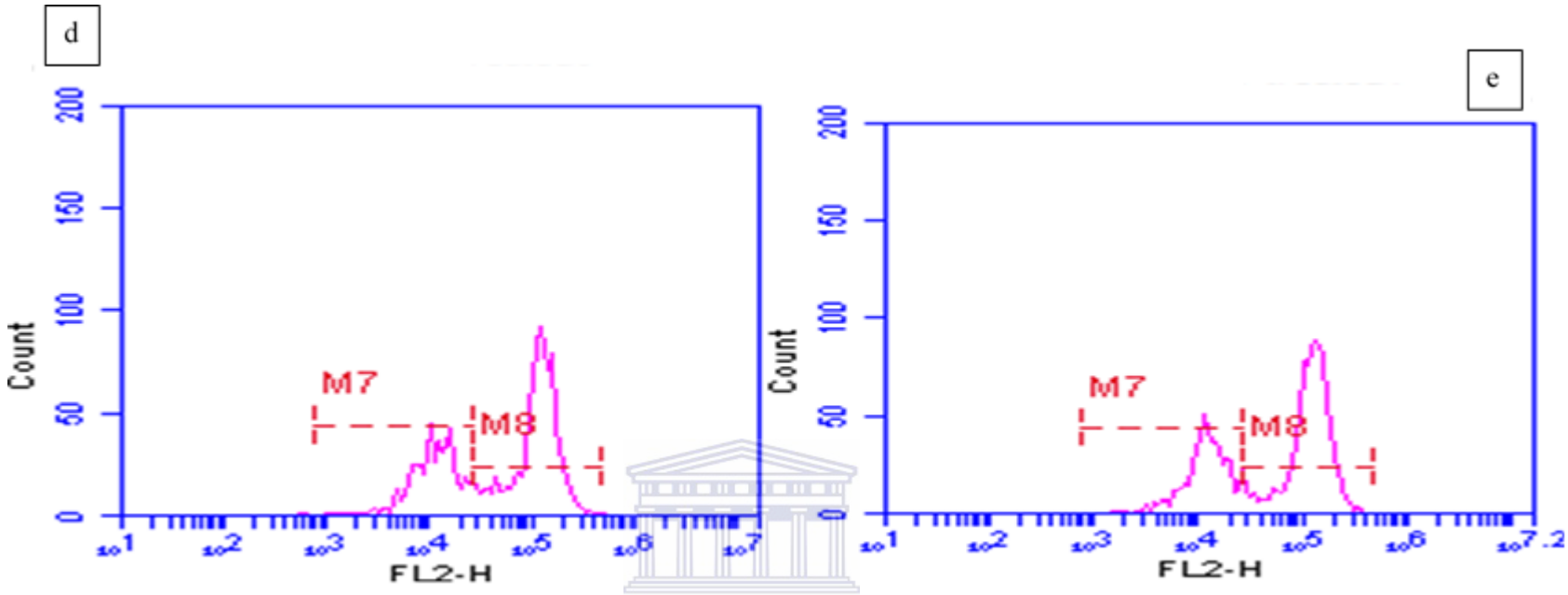


Figure 6.5: Quantification of apoptosis in BRG 7 silenced MCF-7 cell lines using flow cytometer (d) a representation of shRNA206 transfected and DOX treated MCF-7 cell line (e) a representation of shRNA207 transfected and DOX treated MCF-7 cell line.

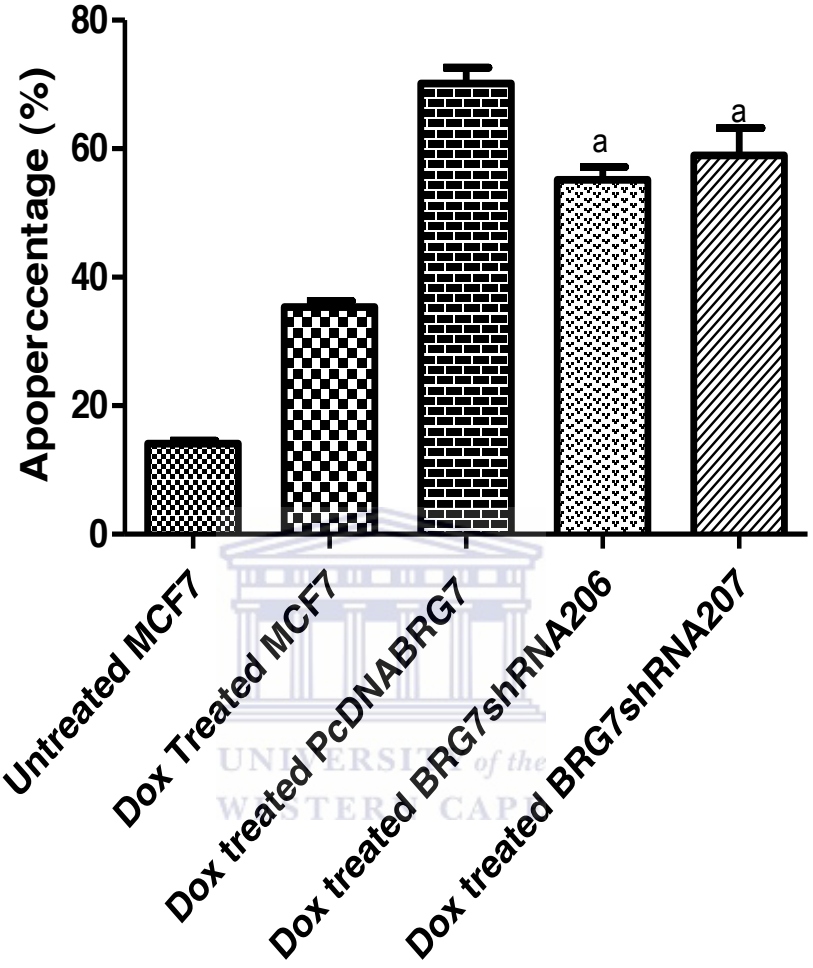


Figure 6.6: Doxorubicin treated MCF-7 cell lines; bar chart represents the mean and standard error of triplicate apo-percentage values of cells after 24 hours of incubation with/without Doxorubicin. Bars with same superscript are statically not significant ($p \leq 0.05$).

Figure 6.6, shows the difference in the level of apoptosis in an epithelial cancer cell line (MCF-7), MCF-7 transfected to over-express BRG 7 protein, a ligand to the NKG2D receptor and MCF-7 cells transfected to repress the expression of the protein by two shRNA constructs that targets two different parts of Gene 7 after 24 hrs treatment with DOX. The result showed the pcDNABRG7 cell lines are undergoing more apoptosis than the MCF-7, BRG7shRNA206 and BRG7shRNA cell lines. The percentage apoptosis difference when the three categories of cell lines were compared were statically significant ($p \leq 0.05$). There was an indication that BRG7shRNA206 and BRG7shrNA207 were undergoing apoptosis faster than the un-transfected MCF-7 cell line.

Having more apoptotic cells in MCF-7 with knocked down NKG2D ligand gene than the un-transfected MCF-7 cell line cannot be readily explained. However, this may suggest that, although the NKG2D ligand may enhance the efficacy of DOX, complete "knock-down" of the gene may also be detrimental to the cell. Knocking down this gene could stress the cell and stimulate the expression of other NKG2D ligands or other immune signals, which can enhance the effect of the cytotoxic drug.

The engagement between NKG2D and its ligands enhances cell-mediated cytotoxicity and cytokine production against transformed cells. Some of these ligands have been described as stress-related proteins that can be induced by heat shock and are regulated by heat shock transcription elements, which are activated during cell transformation (Okita *et al.*, 2012). The result suggest that either the over-expression of BRG 7 or its repression could assist breast cancer treatment using DOX. Also the over-expression or

repression could probably be used to determine the effective concentration of DOX, thus, the potential of this ligand protein serving as biomarker for therapeutic outcome of DOX.

Breast cancer cells developing resistance to DOX remain a major obstacle in the successful treatment of breast cancer patients. Biomarkers that enhances and correlates with treatment response, would allow treatment to be tailored to an individual on the basis of biomarker expression (Smith *et al.*, 2006). Although breast tumours are regarded as one of the most chemo-sensitive solid tumours, there are reports that the initially responsive tumours could relapse and develop resistance to a wide spectrum of drugs making them refractory to cytotoxic drugs which is typically incurable (Harris *et al.*, 2012).

The most frequently identified molecular changes observed in DOX resistance breast tumours are reduced intracellular concentration of the drug as a result of lots of membrane efflux pumps, increase in enzymes involve in drug metabolism and failure of the apoptotic pathways (AbuHammad and Zihlif, 2013). Result from this study suggest that increase in the level of BRG 7 protein could be used to enhance the efficacy of DOX in treating breast cancer, as well as serve as a biomarker to determine the appropriate drug dose, however the actual concentration of this NKG2D ligand that correlates to the effective dose of drug, needs to be determined.

6.4.3: The effect of BRG 7 over-expression and silencing on DMSO induced apoptosis in breast cancer cell line

DMSO has been reported to stimulate apoptosis in different cancer cells (Wang *et al.*, 2012). The apoptotic initiating effect of DMSO was reported to vary based on the cell types (Aita *et al.*, 2005). Deng *et al.*, (2014) reported a DMSO dose dependent tumour growth inhibition in mouse 4T1 breast cancer cell line that lasted for at least three days in breast cancer mice (Deng *et al.*, 2014). Also 2 % DMSO was reported to increase the expression levels of MDR-I gene and protein and increase the chemo sensitivity of MCF-7 after 48 hours of incubation (Zhang, 2001). A MTT chemosensitivity analysis for DMSO was carried out to determine the IC_{50} in an earlier study, and it was observed that about 50 % of the cells died when cells were incubated with 5 % DMSO for 24 hours (result not shown).

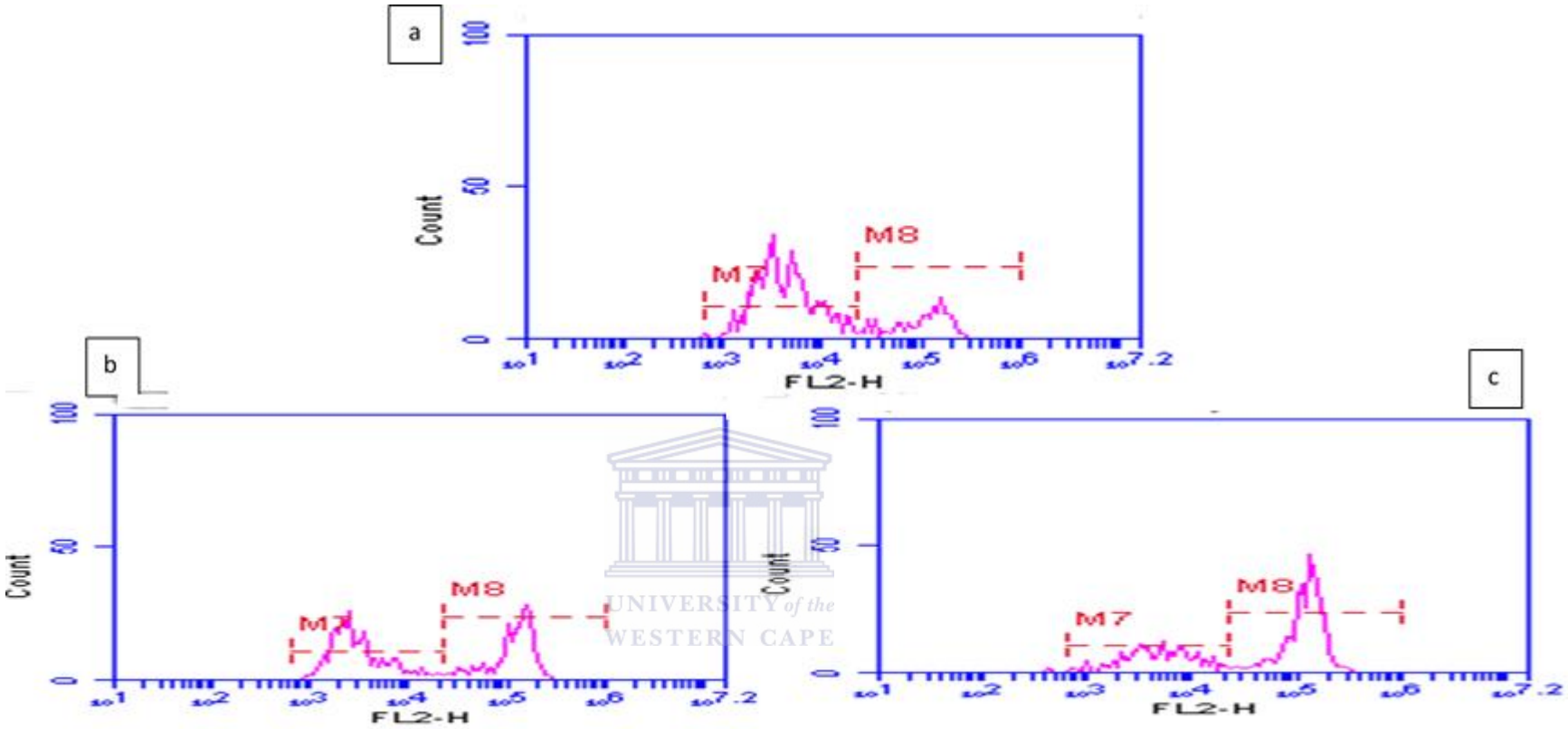


Figure 6.7: Quantification of apoptosis in MCF-7 cell lines using flow cytometer (a) a representation of the MCF-7 untreated cell line (b) a representation of DMSO treated MCF-7 cell line (c) a representation of pcDNA transfected, DMSO treated MCF-7 cell line

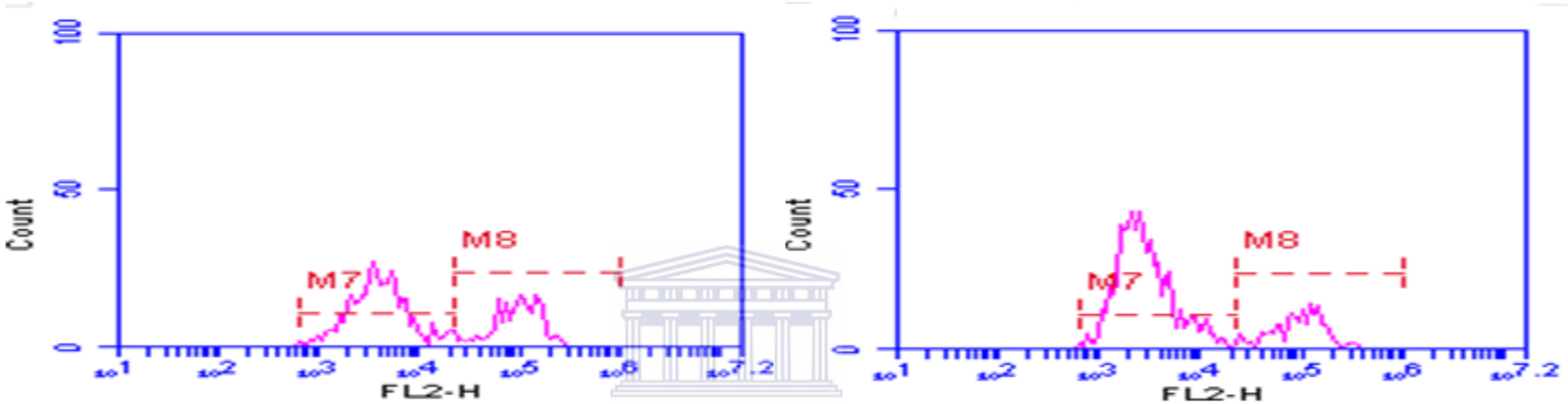


Figure 6.8: Quantification of apoptosis in BRG 7 silenced MCF-7 cell lines using flow cytometer (d) a representation of shRNA206 transfected and DMSO treated MCF-7 cell line (e) a representation of shRNA207 transfected and DMSO treated MCF-7 cell line.

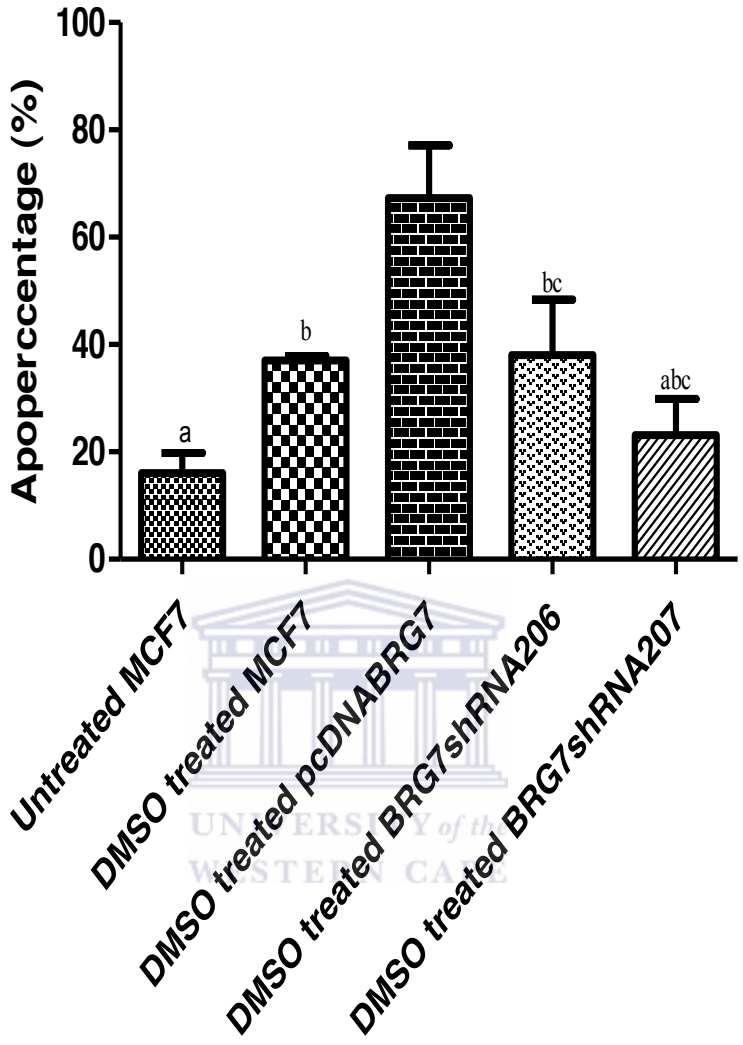


Figure 6.9: DMSO treated MCF-7 cell lines; bar chart represents the mean and standard error of triplicate apo-percentage values of cells after 24 hours of incubation with- or without DMSO. Bars with same superscript are statically not significant ($p \leq 0.05$).

Figure 6.9, shows the difference in the level of apoptosis in an epithelial cancer cell line (MCF-7), transfected to over-express (pcDNABRG7) or repress BRG 7 (BRG7shRNA206 and BRG7shRNA207) treated with DMSO for 24 hours. The result just like the CPT and DOX treated cells showed that the pcDNABRG7 cell line responded more positively to the cytotoxic effect of the DMSO as compared to MCF-7, BRG7shRNA206 and BRG7shRNA207. Also the BRG7shRNA cells showed more resistance to the cytotoxic effect of DMSO when compared to MCF-7, although the difference was not statistically significant.

DMSO has been reported to be effective as an anticancer drug against many types of cancer. Wang *et al.*, (2012), showed the prospect of DMSO as an anticancer drug for lung adenocarcinoma by stimulating tumour suppressor proteins. DMSO was reported to present a synergistic effect when used in combination with antineoplastic agents against ovarian cancer (Pommler *et al.*, 1988). Its anti-tumour activity has also been linked to the induction of the tumour necrosis factor of the p-53 mitochondrial pathway (Koiri and Trigun, 2011). DMSO has also been used to treat refractory cancer pain (Hoang *et al.*, 2011).

DMSO is regarded generally as an inert compound, with little or no physiological effect in biological system. However, studies have shown the amphipathic compound could stimulate the immune system. DMSO have been shown to activate Natural Killer T-cells and NK cells *in vivo* in hepatic tissue, increasing the intracellular levels of the cytotoxic effector molecules, interferon (IFN)-γ and granzyme B protein (Masson *et al.*, 2008). The increased level of the NK cells as a result of the stimulating effect of the DMSO could increase apoptosis in the transfected cells expressing more of the BRG 7, a protein ligand of NKG2D receptor on the NK cells.

Lower expression of the ligand in the BRG7shRNA206 and BRG7shRNA207 cell lines could be the reason for the reduction in apoptosis in these cells. However, DMSO also possesses an anti-oxidative property, which could be used to selectively induced cell death of tumour cells and extending the half-life of the NKT cells (Masson *et al.*, 2008). BRG 7 was modulated by AP-2 a member of the activator protein (AP) transcription factor family (Chapter, 2) and the effect of DMSO in stimulating apoptosis could also be, its influence on the AP transcription factor families. DMSO was reported to modulate the activity of AP-1, a member of the activator protein transcription factor family and the effect suppresses the expression of ICAM-1 (Intercellular Adhesion Molecule 1) also known as CD54 in the respiratory syncytial virus (RSV)-induced production of IL-8 to promote tumour suppression in A549 epithelial cells (Wang *et al.*, 2012).

6.4.4: Effect of different shRNA targets on Gene 7 "knock-down" on apoptosis induction

The BRG7shRNA206 and BRG7shRNA207 are both stable constructs of MCF-7 cell lines with Gene 7 "knock down" using different shRNA constructs. The effects of the various pro-apoptotic compounds on these stably transfected cell lines varied, although BRG 7 expression was repressed in both cell line (Chapter 5). Apoptosis was more induced in the BRG7shRNA206 compared to the BRG7shRNA207 when the cells were treated with camptothecin, although the difference was not significant (figure, 6.2), also more cells were apoptotic in the BRG7shRNA206 cells when DMSO was used (figure, 6.6). Comparing percentage of apoptotic cells in BRG7shRNA207 and the MCF-7 untreated cells in DMSO treated study; the difference was not statistically significant. However, apoptosis was more induced in the BRG7shRNA207 cell lines compared to the BRG7shRNA206 when cells were

treated with DOX (figure, 6.4). This variance of drug effect in shRNA Gene 7 "knocked down" cells is probably due to the shRNA targeting different regions of the gene, thereby modulating the expression of protein and effect of drugs differently.

Several studies have reported variance to pro-apoptotic drug in cells using different short RNA constructs to target different region of a candidate gene for gene "knock down". Holen *et al.*, (2002), showed that siRNA synthesized against different region of a particular mRNA demonstrated difference in silencing. They carried out limited tiling of different siRNA across a region of the human coagulation trigger factor gene (TF) and reported that by shifting the siRNA target site by only three nucleotides at a time resulted in different degree of silencing. Also Pretorius *at al.*, (2013), showed difference in the level of RBBP6 gene knock down using two different siRNA constructs that targets different region of the gene. They also showed that the level of resistance of the two different siRNA stably transfected NIH 3T3 cells treated with camptothecin to induce apoptosis differed as cell that showed higher "knock down" of RBBP6 presented more resistance to apoptosis (Pretorius *et al.*, 2013).

6.5 Conclusion

MCF-7 is an estrogen receptor positive breast cancer cell line, which accounts for about 65 % of all detectable breast cancer cases in human (Keen, 2011). Estrogen receptor positive cancers present the most favourable prognosis of all subtypes and are usually believed to respond normally to hormone therapy (Speirs and Kerin, 2000). Based on the numerous biological and clinical findings that ER positive patients don't derive equal benefit from chemotherapy and hormonal based treatment, Paik *et al.*, (2006), suggested the use of chemotherapy as well as hormone based treatments and their outcomes, should not only be based on baseline prognostic information but also prediction of degree of benefit from chemotherapy (Paik *et al.*, 2006). Identifying BRG 7 as a prime candidate in a set of identified candidate genes as breast cancer biomarkers suggest a high prospect for this gene and it protein product in particular and the others in general in breast cancer management. The findings from this chapter confirmed that cancer cells over-expressing the BRG 7 protein responded faster and better to the three apoptotic inducing drugs.

ER positive breast cancer patients have been reported to have a slow response to camptothecin (Abedin *et al.*, 2007). A Study carried out to determine the underlying mechanisms for the delayed death response of a breast cancer cell line to camptothecin, showed that the slow response of MCF-7 to cytotoxic apoptosis stimulatory compounds may be due to lack of caspase-3 activity, and possible defects in the formation of apoptosome/ caspase-9 activation (Abedin *et al.*, 2007). The effect of CPT was reduced and probably delayed in breast cancer cell lines not expressing the BRG 7 protein compared with the un-transfected cells, whereas the effect of CPT was enhanced in cells transfected for expression of BRG 7, suggesting that this protein could serve as a biomarker for predicting how a specific breast cancer type would

respond to CPT treatment. The result also suggest that BRG 7 could serve as a biomarker in determining the appropriate concentration and drug combination in designing drug specific individual treatment for breast cancer patients. There was an indication from this result that, apart from BRG 7 enhancing the apoptotic stimulating effect of CPT, it could probably serve as a biomarker to monitor the prognosis and therapeutic outcome of CPT and other related apoptotic chemo-therapeutics.

Doxorubicin is regarded as one of the most effective substances for the effective treatment of breast cancer patients. However, resistance to this drug is reported to be common and present a major challenge in breast cancer management (Smith *et al.*, 2006). From the result, there is an indication that the NKG2D ligand (BRG 7) could probably be used to modulate the outcome of doxorubicin treatment in breast cancer patients. Designing treatment tailored on an individual patient basis, patients expressing a high level of the BRG 7 protein would indicate a good outcome when treated with DOX. Also, the result showed that a complete repression of BRG 7 could enhances apoptosis in DOX treated breast cancer cell line.

Dimethyl sulfoxide (DMSO) is used as a reagent in the field of bioscience as a universal water solvent of insoluble compounds, as a protectant in the process of cryopreservation of cells, as well as used as an apoptosis regulator. However, the mode of action of the substance is not well established (Aita *et al.*, 2005). The trend the DMSO treated transfected cells showed, was better defined as compared to camptothecin and doxorubicin treated transfected cells for BRG 7 over-expression and repression. The result suggest that breast cancer cells over-expressing BRG 7 will better respond to DMSO induced apoptosis whilst cells having this protein repressed, have apoptosis induction reduced in them. The mode of DMSO induction of apoptosis in cells in not well defined (Hanslick *et al.*, 2009), however the result observed in the

DMSO treated transfected cells is similar to the CPT treated transfected cells. This suggests that DMSO could probably induce apoptosis following the same route as CPT.

The difference in the response of BRG7shRNA206 and BRG7shRNA207 induction of apoptosis by different pro-apoptotic compounds, further proves that BRG 7 could play a significant role in the treatment and management of breast cancer using pro-apoptotic compounds. The variance in the effect of the pro-apoptotic compounds on BRG7shRNA206 and BRG7shRNA207 suggest targeting the gene at different regions, for gene knock down will present different outcome when pro-apoptotic compounds are used for treatment. Hence, BRG 7 could probably serve as a good biomarker for the effective management of ER positive breast cancer, however more studies is required to confirm this.

Chapter 7: General Discussion and Conclusion

7.1 General discussion

Breast cancer remains the major type of cancer amongst women and a global health problem (AbuHammad and Zihlif, 2013). Based on the challenges posed by breast cancer, there is an urgent need for biomarkers to improve the diagnosis, prognosis, guided therapy and management, as well as monitor patient response to therapeutics. The advent of high throughput analysis technology have provided a fast means of genomic analysis which have provided differential expression level of gene in disease condition and whose protein products have been predicted and published as candidate cancer biomarkers (Kulasingam and Diamandis, 2008). Using different bioinformatics tools and processes, Ngcoza, (2012) identified seven genes, whose protein products are membrane bound, from different databases e.g. Oncomine. Most of the differentially expressed genes, identified from these various databases were obtained from microarray analysis (Ngcoza, 2012).

Differential expression analysis of these seven genes in different cancer cell lines and non-cancerous cell lines including breast cell lines was carried out using semi-quantitative PCR and qRT-PCR. The results showed that three of these genes; Gene 3, Gene 4 and Gene 7 were highly expressed in a breast cancer cell line as compared to the non-cancerous breast cell line and other cancer cell lines (Ngcoza, 2012). Although finding a single biomarker has been the target for most biomarker discovery research, and some have been identified in some disease types, there is growing consensus that multiple (e.g., three to five) markers used individually or as part of an integrated panel will be more appropriate for cancer management. The performance

and advantages of a panel of biomarkers in cancer management have been confirmed in several publications (Rifai *et al.*, 2006; Hanash *et al.*, 2008; Hu *et al.*, 2008).

One of the most important processes in the discovery of multiple genes/proteins for cancer biomarkers is validation of the genes/proteins in clinical outcomes. In recent times, numerous genes have been identified as biomarkers for various cancer types based on gene expression analysis, using high throughput techniques such as microarray profiling and more recently the RNA-Seq technologies and qRT-PCR. Often, from these list of identified biomarkers multiple gene products have been linked to a specific cancer tissue and their subtypes. However, validating the role of these biomarkers in different populations or evaluating competing biomarkers has been a herculean task, even though lots of data for these genes are available (Aguirre-Gamboa *et al.*, 2013).

In silico approach to biomarker validation could be a more effective strategy for biomarker development, which is typically a fast and cost effective process compared to the traditional methods of biomarker validation (Chen *et al.*, 2014). To validate the diagnostic, prognostic and/or predictive value of the identified candidate genes as a set of biomarkers, several software and databases were used. Most of the genes were confirmed as having good prospects as a biomarker, based on their association with different cancer related processes using the *in silico* validation procedure.

Two additional genes were hypothesized as having a good prospect as breast cancer biomarkers after several *in silico* functional analysis. These genes included another member of the seven putative breast cancer biomarker gene list (Gene 2), and Gene 8, whose protein product was observed to have a direct interaction with the protein product of the most highly expressed Gene 7 from the qRT-PCR study (Ngcoza,

2012). Gene 2 was not shown to be highly expressed from the qRT-PCR analysis, however the gene was found to be functionally involved in many cancer processes in general, using in silico annotation tools, and could potentially be differentially expressed in breast cancer. The protein product of Gene 8 was also annotated as a modulator of expression of the estrogen receptor protein (ESR). ESR is a prognostic factor in a large group of breast cancer multiple cohorts (Hartman *et al.*, 2009). Also the *in silico* functional analysis (Chapter 2) showed some of the other candidate genes were linked to at least one process linked to cancer progression. It was postulated that this set of candidate biomarkers could complement one another as breast cancer prognosis or predictive biomarkers, thus, an *in silico* validation of these candidate genes was carried out (Chapter 3) to determine their prognostic and predictive values.

Several genes have been identified as candidate biomarkers especially as prognostic, predictive or both for patients with breast cancer (Fan *et al.*, 2006). Validating these prognostic or predictive candidate genes in different breast cancer cohorts is now of great interest in cancer management (Gyorffy *et al.*, 2009). These candidate genes are usually validated using several molecular techniques, which could be cumbersome and time consuming. To validate the prognostics and/or predictive values of the identified candidate genes as a set of biomarkers, some prognostics software and databases were used.

Using the SurvExpress tool, the prognostic ability of this set of candidate biomarkers were analysed in two datasets namely DATA 1 and DATA 2. DATA 1 is a dataset comprising 9 datasets from the same platform compiled by seven authors using 1574 samples, while DATA 2 is a meta-bases constructed by performing a quantile normalization compiled by SurvExpress with 1901 samples. The two datasets showed variance in number and the genes that are relevant in predicting breast cancer

treatment outcome based on their differential expression ($p \leq 0.05$). Although one of the seven genes was common to both datasets, three genes were listed as being significant in DATA 1 and two listed in DATA 2, with one of the genes common to both. The variance in the significance of the set of genes listed as being significant, suggest this set of genes will be more effective when combined than used as individual biomarkers for breast cancer. The effective differentiation of risk groups, based on combined gene expression, add credence to this suggestion. The combined set of genes showed positive prognostic and predictive value using the two different datasets from SurvExpress.

The Kaplan Meier plotter was used to analyze the prognostic ability of the individual candidate biomarkers based on them differentiating between the low and the high risk patients groups based on their expression levels. From the survival curves, Gene 3, Gene 4 and Gene 7 showed significant P-value in predicting the prognostic outcome of breast cancer based on the differential value of the biomarkers. These genes were also highly expressed in the qRT-PCR analysis (Ngcoza, 2012), validating their significance as putative breast cancer biomarkers. In addition to these genes, Gene 1 also showed significance in differentiating the two risk groups based on its expression. However, this gene was not differentially expressed in the qRT-PCR analysis (Ngcoza, 2012).

The difference in outcomes of the different bioinformatics tools was postulated to be as a result of different cancer types and techniques used in analyzing the differential gene expression. The ability of this bioinformatics tool to confirm the same genes from the qRT-PCR study as having value as potential biomarkers is an indication that the tool is effective for *in silico* analysis of prospective biomarkers. The corroboration

of the Kaplan Meier plotter with previously observed results (Ngcoza, 2012) also give more credence to further analyse these genes as breast cancer biomarkers.

The effective prognostic ability of this set of candidate biomarkers was determined in different subtypes of breast cancer, using the Gene Expression-Based Outcome for Breast Cancer Online (GOBO). GOBO present Kaplan-Meier plots performed on different stratified groups based on gene expression in different subtypes and presented as percentage distant metastasis-free survival (DMFS) against time in years. The output from this analysis indicated that, this set of candidate biomarkers may not predict effectively all estrogen receptor positive (ER+) breast cancer subtypes, however the candidate biomarkers were effective in predicting the outcome of luminal B ER+ breast cancer subtype. The GOBO results also showed the candidate genes could probably be used to predict the survival outcome of ERBB2 subtype of estrogen receptor negative (ER-) tumours, and lymph node negative subtype of ER- tumours, but not all subtype of ER- tumours based on the individual survival curves outcome for these diseases (Chapter 3).

The specific differential expression signature of the candidate genes in different cancer types, suggest that a single candidate gene or protein may not be very specific as breast cancer biomarker, but combining these genes/proteins base on their expression pattern could be more appropriate. The current prevailing view is that the most powerful single biomarker of tumours have already been discovered, thus, multiple markers will serve a more effective purpose as breast cancer markers (Diamandis, 2004). However some genes within the list of putative biomarkers showed more promise as effective biomarkers and needed to be further analyzed for their effective use as breast cancer biomarkers.

Based on the Kaplan Meier plotter results, box plots and the GEA analysis, Gene 3, Gene 4 and Gene 7 were validated to be differentially expressed in breast cancer. These three genes were the genes differentially expressed in the qRT-PCR (Ngcoza, 2012). In addition to these genes, Gene 2, one of the seven genes that was shown to be involved in almost all the biochemical processes associated to cancer initiation and progression from the GO and protein interaction analysis (Chapter 2) was also considered a likely candidate biomarker. Although, this gene was not observed to be differentially expressed in breast cancer from the *in silico* analysis (GEA analysis) in chapter 2 and the qRT-PCR analysis (Ngcoza, 2012). Gene 8 was also postulated to be a good biomarker for breast cancer management based on its direct association with the most differentially expressed gene from the qRT-PCR analysis Gene 7 (Ncogza, 2012). BRG 8 was also directly linked to estrogen receptor (ESR) protein (Chapter 2). Estrogen receptor (ER) status of invasive carcinoma has been reported to be a useful prognostic and predictive factor and has become a common practice in the management of breast carcinomas. ER positivity predicts for response to endocrine therapy such as anti-estrogen (tamoxifen) administration (Manavathi *et al*., 2014).

The protein expressions of these five candidate genes were monitored in different cancer cell lines using Western blot analysis. Western blot analysis confirms the expression of three of the five protein products of these genes and also there was an indication from the Western blot analysis that certain proteins may be differentially expressed when the expression level of these proteins in MCF-7 was compared with expression in other cell lines. However, BRG 3 and BRG 4 were not successfully demonstrated to be expressed using Western blot analysis. This was thought to be as a result of the commercial primary antibody to these proteins unable to bind the proteins in their native forms. It has been reported that, the interaction between a

protein and its corresponding antibody may be disrupted by competition with high concentrations of the pure epitope itself, as well as high salt concentrations, extreme pH, and detergents (Janeway *et al.*, 2001). Thus, immunofluorescence analysis was used to analyze the expression of BRG 3, BRG 4 and BRG 7 in whole cell lines, comparing the expression in MCF-7 (a breast cancer line) to MCF-12A (a breast non-cancerous cell line).

From the immunofluorescence analysis results, the proteins were shown to be expressed in both cell lines. The results also suggest these proteins were more expressed in the breast cancer cell line as compared to the non-cancerous breast cancer cell line, however, a quantitative immunofluorescence analysis needs to be done to confirm this. The immunofluorescence protein expression analysis also validated BRG 3, BRG 4 and BRG 7 to be membrane bound, thus, would probably be shed into the bodily fluids. Membrane bound proteins are attractive biomarker candidates because of their potential use in cancer imaging, targeted therapeutic strategies, and also they are often shed into the bloodstream, which makes them useful candidate's for early disease detection (Harsha *et al.*, 2009).

The expression analysis of BRG 8 using Western blot analysis showed this protein was not expressed in breast cancer as well as the colon cancer cell lines. Gene 8 was shown to directly interact with BRG 7 (Chapter 2) and its differential expression in breast cancer compared to breast non-cancerous cell line was concluded, could serve or complement BRG 7 as breast cancer biomarker. Also the expression of these proteins could modulate chemotherapeutics effect in treating cancer and hence can be use as biomarker to monitor the prognosis and therapeutic outcome in breast cancer based on its direct association with an estrogen receptor protein, a prognostic and predictive factor in breast cancer types.

BRG 7 was consistently shown through the *in silico* and the protein expression analysis as a probable effective biomarker for breast cancer management. BRG 7 is a member of the NKG2D ligand family, reported to be anchored on the cell surface of diseased cells to serve as immune signal, that signal the NK cells against cancer cells and/or virus infection cells (González *et al.*, 2008). Some of the NKG2D ligands have also been reported to be shed off the surface of cancer cells for these cells to evade the immune system.

Wang and Sun (2014), reported that, most studies looked at the spontaneous shedding of NKG2D ligands in order to evade the immune system and thus reported that, NK cells stimulates the shedding of some of these ligand proteins in response to pro-apoptotic cancer treatment (Wang and Sun, 2014). We postulated that, BRG 7, a NKG2D ligand will enhance the effectiveness of pro-apoptotic compounds in breast cancer treatment, and the expression level could be use as biomarker to monitor the therapeutic outcome of pro-apoptotic drugs.

In recent times, discovery of biomarkers for cancer management for therapeutics has been of great interest (Kuppusamy *et al.*, 2014), thus, the effect of over-expression of this protein and the "knock down" of its gene in enhancing apoptosis, a major hallmark of cancer progression was monitored in breast cancer cell line (Chapters 5 and 6).

To confirm the effect of the over-expression of BRG 7 in enhancing the initiation of apoptosis in breast cancer, MCF-7 (an estrogen positive cancer cell line) and MCF-12A (a non-cancerous breast cell line), were stably transfected to over-express BRG 7 using a cloned plasmid and repressed by "knocking down" the gene using three commercial shRNA constructs. The cloned plasmid and the commercial shRNA constructs were confirmed using colony PCR and endonuclease digestion respectively

(Chapter 5). Also, the isolated plasmid was sent for sequencing and the result confirms the plasmid insert having 99 % homology to Gene 7 using the NCBI nucleotide blast tool (BLASTn).

The over-expression of BRG 7 in non-cancerous cell line (MCF-12A) lead the cell to undergo senescence as the cell presented a changed in morphology to flattened, elongated and nucleated cells characteristic of senescence. However, the over-expression of the BRG 7 protein seems not to affect the morphology of the breast cancer cell line (MCF-7) (Chapter 5). A major feature of senescent cells that distinguishes them from most quiescent cells is a large flattened morphology. They are also known to down-regulate genes involved extracellular matrix production and upregulate inflammatory cytokines as well as other molecules known to modulate the immune response (Krizhanovsky *et al.*, 2008). Increased NKG2D ligand expression is regulated by activation of DNA damage. DNA damage leads to cell apoptosis and a senescence-like phenotype cell in response to most of the chemotherapeutic agents commonly used in clinical trials (Soriani *et al.*, 2009). This result suggest that the over-expression of this protein in breast cells might be a defense mechanism against cancer, while it seems the breast cancer cells, have a mechanism to overcome the induced senescence by over-expression of the protein.

Stably transfected MCF-7 cell lines for BRG 7 protein over-expression was referred to as pcDNABRG 7, whilst BRG7shRNA206 and BRG7shRNA207 were used to represent the stably transfected cells for Gene 7 “knock-down” in this study.

The effect of BRG 7 in modulating camptothecin, doxorubicin and DMSO in MCF-7, an ER positive cell line was invstigated (Chapter 6). Estrogen receptor positive breast cancer accounts for about 65 % of all detectable breast cancer cases in human

(Keen, 2011). However, ER positive breast cancer patients have been reported to have a slow response to Campotethecin (Abedin *et al.*, 2007). Doxorubicin is regarded as one of the most effective substances for the effective treatment of breast cancer patients. However, resistance to this drug is reported to be common and present a major challenge in breast cancer management (Smith *et al.*, 2006). DMSO has been reported to stimulate apoptosis in different cancer cells (Wang *et al.*, 2012).

The result from this study showed that BRG 7 enhances the effect of CPT by increasing the level of apoptosis in pcDNABRG 7, whilst the effect of CPT in inducing apoptosis in MCF-7, BRG7shRNA206 and BRG7shRNA207 seems delayed and reduced, although, the difference in percentage apoptotic cells between pcDNABRG 7 and MCF-7 was not statistically significant. Camptothecin antitumour activity is majorly by preventing DNA replication through the inhibition of topo I activity (Staker *et al.*, 2002). NKG2D ligands are reported to be upregulated in cells as a result of chemotherapeutics, which stall DNA replication, and the upregulation of these ligands has been reported to be as a result of diverse genotoxic stimuli (Gasser and Raulet, 2006). This suggest that NKG2D ligands might be required for effective genotoxic effect of CPT in inducing apoptosis and the over-expression of the BRG 7 could help in improving the delayed response of ER positive tumours to CPT.

Also pcDNABRG 7 cells showed a more effective response to apoptosis induced by DOX. Also there was an indication that BRG7shRNA206 and BRG7shrNA207 were undergoing apoptosis faster than the un-transfected MCF-7 cell line. Having more apoptotic cells in MCF-7 with knocked down NKG2D ligand gene expression than the un-transfected MCF-7 cell line, cannot be readily explained, however, this suggest that, although the NKG2D ligand may enhance the efficacy of DOX, complete "knockdown" of the gene may also be detrimental to the cell.

Although breast tumours are regarded as one of the most chemo-sensitive solid tumours, there are reports that the initially responsive tumours could relapse and develop resistance to a wide spectrum of drugs making them refractory to cytotoxic drugs which is typically incurable (Harris *et al.*, 2012). These results suggest that the mode of BRG 7 expression in breast cancer could aid the design of treatment tailored on an individual basis. Patients expressing high level of the BRG 7 protein would indicate a good outcome when treated with DOX, while those not expressing or whose expression is below an established baseline could indicate the breast cancer patient may not respond well to DOX. Also combined therapy that repressed BRG 7 could be employed with DOX to get a more effective cancer treatment therapy strategy using DOX.

BRG 7 showed an effective enhancement of apoptosis induction by DMSO. pcDNABRG 7 cells were shown to experience more apoptosis when treated with DMSO as compared to MCF-7, BRG7shRNA206 and BRG7shrNA207 and the differences when compared were statistically significant. Although the mode of action of DMSO in initiation apoptosis in cancer cells is not well established (Aita, *et al.*, 2005). The trend in apoptosis undergone by DMSO treated transfected and un-transfected cells showed some similarities to that seen with the camptothecin treated cells. This suggest that mode of action of DMSO inducing apoptosis in breast cancer cells expressing high levels of BRG 7 protein is similar to that of CPT.

Studies have reported variance to pro-apoptotic drug in cells using different short RNA constructs to target different region of a candidate gene for gene "knock down". Holen *et al.*, (2002). The effect of "knocked down" of BRG 7 in modulating effect of the three pro-apoptotic drugs varied comparing BRG7shRNA206 to BRG7shRNA207. This could be as result of these shRNA transfected constructs

binding the gene at different region, thereby generating different effects. The level of resistance of the two different siRNA stably transfected NIH 3T3 cells treated with camptothecin to induce apoptosis were shown to differ as cell showed variation in gene “knock down” and cell that showed higher “knock down” of RBBP6 presented more resistance to apoptosis (Pretorius *et al.*, 2013).

7.2: Conclusion

The outcome of the functional analysis of the seven candidate genes using an *in silico* techniques;

- The three genes (Gene 3, Gene 4 and Gene 7) that were differentially expressed from the qRT-PCR analysis were confirmed to be highly expressed in breast cancer, as shown by the *in silico* analysis in chapter 2.
- An additional gene (Gene 2) from the seven candidate genes was found to be involved in most of the cancer processes, however not differentially expressed in the qRT-PCR analysis (Ngcoza, 2012), was considered to be a good biomarker candidate for breast cancer for further investigation.
- Also another protein candidate (BRG 8) was discovered from the protein interaction analysis based on its direct interaction with ESR protein and BRG 7 (Chapter 2).
- The individual genes were shown to be involved in one cancer process or the other and we concluded these genes will serve well as a panel of biomarkers for breast cancer management based on the *in silico* functional analysis (Chapter 2).
- The prognostic/predictive validation of these individual genes also showed variations as effective individual breast cancer biomarkers.
- A panel of these biomarkers were shown to be more effective in predicting the outcomes of breast cancer patients based on the analysis using two different databases from SurvExpress.
- Gene 3, Gene 4 and Gene 7 showed consistency as breast cancer biomarkers in predicting breast cancer treatment outcome, based on their individual survival analysis (Kaplan Meier plotter), also Gene 1 was also shown to be

significant as a predictive biomarker based on the Kaplan Meier plotter result (Chapeter 3).

- The *in silico* prognosis analysis of these genes also suggest that the expression signature of the candidate genes as biomarkers could be specific for different breast cancer types and subtype, and indicated effectiveness in accurately predicting the outcome of estrogen positive luminal B subtype of breast cancer, estrogen negative human epidermal growth factor receptor 2 (ERBB2) and the lymph node negative subtypes.

The protein expression of five of the candidate genes were monitored in breast cancer cell line compared to non-cancerous breast cell line and other cancer cell lines. These genes are: Gene 3, Gene 4 and Gene 7, the genes that showed differential expression based on the qRT-PCR analysis and based on their consistency as effective biomarker from the *in silico* analysis. In addition Gene 2; the gene that was found to be involved with most cancer processes based on the *in silico* analysis, although not differentially expressed in the qRT-PCR analysis and Gene 8; the gene whose protein product directly interact with BRG 7 and ESR from the STRING analysis.

- Western blot analysis was used to show BRG 7 is expressed in breast cancer cell lysate, although the expression was not well pronounced. It was concluded the protein might have been shed off the cell membrane.
- BRG 3, BRG 4 and BRG 7 were successfully shown to be differentially highly expressed in MCF-7 (breast cancer cell line) as compared to MCF-12A (non-cancerous breast cell line) using immunoassay analysis and may serve as breast cancer biomarkers.

- BRG 8 was not expressed in MCF-7 but expressed in MCF-12A and was concluded could be an effective biomarker.
- BRG 2 was expressed in all cancer cell lines analyzed but not differentially expressed. It was concluded this gene will not effectively serve as an individual breast cancer biomarker.

The modulating effect of BRG 7 in apoptosis induction was monitored in breast cancer cell lines stably transfected for BRG 7 overe-expression and repression:

- The findings of the modulating role of BRG 7 in drug induced apoptosis, suggest it could probably play an important role in personalised medicine and could serve as a biomarker to monitor the prognosis and/or therapeutic outcome of camptothecin, doxorubicin and DMSO treated breast cancer patients.
- There was an indication that targeting the gene at different region for repression resulted to variance modulating effect on pro-apoptotic compounds.

7.3: Future work

Over-expression of BRG 7 in non-cancerous breast cell line resulted in the cell changing morphology to flat nucleated cell, which is a major feature of cell in senescence. However this needs to be further validated by monitoring the level of β-galactosidase. Also the effect of over-expression on the p53 protein a major modulating protein of cell senescence, apoptosis and a tumour suppressor protein need to be evaluated as well as other apoptotic proteins. In addition to the modulating effect of BRG 7 on apoptosis, the effect of over-expression and knock down on cell cycle needs to be monitored in breast cancer and non-cancerous cell lines. Also, the modulating effect of the protein needs to be monitor in other breast cancer cell line types.

Some of the proteins were qualitatively shown to be differentially expressed in breast cancer cell line using the immunofluorescence assay, however the expression level of the protein needs to be quantified using quantitative immunofluorescence analysis or enzyme linked immunosorbent assay (ELISA). As these candidate proteins are annotated to be cell surface proteins and are expected to be shed into bodily fluids, the presence of these proteins need to be validated in various bodily fluids and tissue samples of breast cancer patients.

References

Abedin, M. J., Wang, D., McDonnell, M. A., Lehmann, U and Kelekar, A. (2007). Autophagy delays apoptotic death in breast cancer cells following DNA damage. *Cell Death & Differentiation*, *14*(3), 500-510.

Abbott, R. G., Forrest, S., & Pienta, K. J. (2006). Simulating the hallmarks of cancer. *Artificial life*, *12*(4), 617-634.

AbuHammad, S and Zihlif, M. (2013). Gene expression alterations in doxorubicin resistant MCF7 breast cancer cell line. *Genomics*, *101*(4), 213-220.

Adam, B. L., Qu, Y., Davis, J. W., Ward, M. D., Clements, M. A., Cazares, L. H., Semmes, O.J, Schellhammer, P.F, Yasui, Y, Feng Z and Wright, G. L. (2002). Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Research*, *62*(13), 3609-3614.

Adams, J.M. and Cory, S. (2007). The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene* 26, 1324–1337.

Ades, F., Zardavas, D., Bozovic-Spasojevic, I., Pugliano, L., Fumagalli, D., de Azambuja, E., Viale, G., Sotiriou, C. and Piccart, M. (2014). Luminal B breast cancer: molecular characterization, clinical management, and future perspectives. *Journal of Clinical Oncology*, *32*(25), 2794-2803.

Aebersold, R., Anderson, L., Caprioli, R., Druker, B., Hartwell, L. and Smith, R. (2005). Perspective: a program to improve protein biomarker discovery for cancer. *Journal of Proteome Research*, *4*(4), 1104-1109.

Aguirre-Gamboa, R., Gomez-Rueda, H., Martínez-Ledesma, E., Martínez-Torteya, A., Chacolla-Huaringa, R., Rodriguez-Barrientos, A., Tamez-Pen, G. and Trevino, V. (2013). SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis. *PloS one*, *8*(9), e74250.

Ahn, S. M. and Simpson, R. J. (2007). Body fluid proteomics: Prospects for biomarker discovery. *PROTEOMICS-Clinical Applications*, *1*(9), 1004-1015.

Aita, K., Irie, H., Tanuma, Y., Toida, S., Okuma, Y., Mori, S. and Shiga, J. (2005). Apoptosis in murine lymphoid organs following intraperitoneal administration of dimethyl sulfoxide (DMSO). *Experimental and Molecular Pathology*, *79*(3), 265-271.

Aittokallio, T. and Schwikowski, B. (2006). Graph-based methods for analysing networks in cell biology. *Briefings in Bioinformatics*, *7*(3), 243-255.

Alaiya, A. A., Franzén, B., Auer, G. and Linder, S. (2000). Cancer proteomics: from identification of novel markers to creation of artifical learning models for tumour classification. *Electrophoresis*, *21*(6), 1210-1217.

Allinen, M., Beroukhim, R., Cai, L., Brennan, C., Lahti-Domenici, J., Huang, H., Porter D, Hu M, Chin L, Richardson A, Schnitt S, Sellers W.R. and Polyak, K. (2004). Molecular characterization of the tumour microenvironment in breast cancer. *Cancer cell*, *6*(1), 17-32.

Allred, D. C., Mohsin, S. K. and Fuqua, S. A. (2001). Histological and biological evolution of human premalignant breast disease. *Endocrine-related cancer*, *8*(1), 47-61.

Amir, E., Freedman, O. C., Seruga, B. and Evans, D. G. (2010). Assessing women at high risk of breast cancer: a review of risk assessment models. *Journal of the National Cancer Institute*, *102*(10), 680-691.

Andréll, J. and Tate, C. G. (2013). Overexpression of membrane proteins in mammalian cells for structural studies. *Molecular Membrane Biology*, *30*(1), 52-63.

Andresen, L., Skovbakke, S. L., Persson, G., Hagemann-Jensen, M., Hansen, K. A., Jensen, H. and Skov, S. (2012). 2-deoxy D-glucose prevents cell surface expression of NKG2D ligands through inhibition of N-linked glycosylation. *The Journal of Immunology*, *188*(4), 1847-1855.

Avraham, R. and Yarden, Y. (2011). Feedback regulation of EGFR signalling: decision making by early and delayed loops. *Nature Reviews Molecular Cell Biology*, *12*(2), 104-117.

Baeriswyl, V. and Christofori, G. (2009). The angiogenic switch in carcinogenesis. Semin. *Cancer Biol*. 19, 329–337.

Banič, B., Nipič, D., Šuput, D. and Milisav, I. (2011). DMSO modulates the pathway of apoptosis triggering. *Cellular & Molecular Biology Letters*, *16*(2), 328-341.

Barabasi, A. L. and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, *5*(2), 101-113.

Barabási, A. L., Gulbahce, N. and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, *12*(1), 56-68.

Barh, D., Zambare, V., and Azevedo, V. (Eds.). (2013). *Omics: Applications in Biomedical, Agricultural, And Environmental Sciences*. CRC Press. Pg 276

Bauer, K. R., Brown, M., Cress, R. D., Parise, C. A. and Caggiano, V. (2007). Descriptive analysis of estrogen receptor (ER) -negative, progesterone receptor (PR) -negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype. *Cancer*, *109*(9), 1721-1728.

Baylin, S. B. and Jones, P. A. (2011). A decade of exploring the cancer epigenome—biological and translational implications. *Nature Reviews Cancer*, *11*(10), 726-734.

Bean, R. G., Ganesan, Y.T., Dong, Y., Takeda, S., Liu, H., Chan, P.M., Huang, Y., Chodosh, L.A., Zambetti, G.P., Hsieh, J.J. and Cheng, E.H.Y. (2013). PUMA and BIM are required for oncogene inactivation-induced apoptosis. *Sci. Signal*. 6, ra20

Ben-Hur, A. and Noble, W. S. (2005). Kernel methods for predicting protein–protein interactions. *Bioinformatics*, *21*(suppl 1), i38-i46.

Bertucci, F., Finetti, P., Rougemont, J., Charafe-Jauffret, E., Cervera, N., Tarpin, C., Nguyen, C., Xerri, L., Houlgatte, R., Jacquemier, J., Viens, P. and Birnbaum, D. (2005). Gene expression profiling identifies molecular subtypes of inflammatory breast cancer. *Cancer Research*, *65*(6), 2170-2178.

Berx, G. and van Roy, F. (2009). Involvement of members of the cadherin superfamily in cancer. Cold Spring Harb. *Perspect. Biol*. 1, a003129.

Bhatt, A. N., Mathur, R., Farooque, A., Verma, A. and Dwarakanath, B. S. (2010). Cancer biomarkers-Current perspectives. *Indian J Med Res*, 132:129-49

Bhowmick, N.A., Neilson, E.G., and Moses, H.L. (2004). Stromal fibroblasts in cancer initiation and progression. *Nature,* 432: 332–337.

Bishop, J.M., and Weinberg, R.A., eds. (1996). Molecular Oncology. (New York: Scientific American, Inc.).

Böcker, W., Moll, R., Poremba, C., Holland, R., van Diest, P. J., Dervan, P., Bürger, H., Wai, D., Ina Diallo, R., Brandt, B., Herbst, H., Schmidt, A., Lerch, M.M. and Buchwallow, I. B. (2002). Common adult stem cells in the human breast give rise to glandular and myoepithelial cell lineages: a new cell biological concept. *Laboratory Investigation*, *82*(6), 737-746.

Bosher, J. M., Totty, N. F., Hsuan, J. J., Williams, T. and Hurst, H. C. (1996). A family of AP-2 proteins regulates c-erbB-2 expression in mammary carcinoma. *Oncogene*, *13*(8), 1701-1707

Boyd, N.F., Guo, H., Martin, J.L., Sun, L., Stone, J., Fishell, E., Jong, R.A., Hislop, G., Chiarelli, A., Minkin, S. and Yaffe, M.J. (2007) Mammographic density and the risk and detection of breast cancer. *N Engl J Med.*; 356(3):227–236

Brummelkamp, T. R., Bernards, R. and Agami, R. (2002). Stable suppression of tumourigenicity by virus-mediated RNA interference. *Cancer cell*, *2*(3), 243-247.

Buess, M., Nuyten, D. S., Hastie, T., Nielsen, T., Pesich, R. and Brown, P. O. (2007). Characterization of heterotypic interaction effects in vitro to deconvolute global gene expression profiles in cancer. *Genome Biol*, *8*(9), R191.

Bui, J. D. and Schreiber, R. D. (2007). Cancer immunosurveillance, immunoediting and inflammation: independent or interdependent processes? *Current opinion in immunology*, *19*(2), 203-208.

Burkhart, D.L. and Sage, J. (2008). Cellular mechanisms of tumour suppression by the retinoblastoma gene. Nat. Rev. Cancer 8, 671–682.

Cairns, R. A., Harris, I. S. and Mak, T. W. (2011). Regulation of cancer cell metabolism. *Nature Reviews Cancer*, *11*(2), 85-95.

Cantor, S. B., Bell, D. W., Ganesan, S., Kass, E. M., Drapkin, R., Grossman, S., Wahrer, D.C., Sgroi, D.C., Lane, W.S., Haber, D.A. and Livingston, D. M. (2001). BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell*, *105*(1), 149-160.

Cantor, S., Drapkin, R., Zhang, F., Lin, Y., Han, J., Pamidi, S. and Livingston, D. M. (2004). The BRCA1-associated protein BACH1 is a DNA helicase targeted by clinically relevant inactivating mutations. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(8), 2357-2362.

Carlson, M. E., Hsu, M. and Conboy, I. M. (2008). Imbalance between pSmad3 and Notch induces CDK inhibitors in old muscle stem cells. *Nature*, *454*(7203), 528-532.

Carmeliet, P. (2005). VEGF as a key mediator of angiogenesis in cancer. *Oncology*, *69* (Suppl. 3), 4-10.

Castanotto, D. and Rossi, J. J. (2009). The promises and pitfalls of RNA-interference-based therapeutics. *Nature*, *457*(7228), 426-433.

Cavallaro, U. and Christofori, G. (2004). Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. Nat. Rev. Cancer 4, 118–132.

Celis, J. E., Kruhøffer, M., Gromova, I., Frederiksen, C., Østergaard, M., Thykjaer, T., Gromov, P., Yu, J., Pálsdóttir, H., Magnusson, N. and Orntoft, T.F. (2000). Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS letters*, *480*(1), 2-16.

Chaffer, C. L. and Weinberg, R. A. (2011). A perspective on cancer cell metastasis. *Science*, *331*(6024), 1559-1564.

Chang, T. C., Wentzel, E. A., Kent, O. A., Ramachandran, K., Mullendore, M., Lee, K. H., Feldmann, G., Yamakuchi, M., Ferlito, M., Lowenstein, C.J., Arking, D.E., Beer, M.A., Maitra, A. and Mendell, J. T. (2007). Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Molecular cell*, *26*(5), 745-752.

Chatterjee, S. K. and Zetter, B. R. (2005). Cancer biomarkers: knowing the present and predicting the future. Future Oncology; 1(1), 37-50

Chen, X., Sun, X. and Hoshida, Y. (2014). Survival analysis tools in genomics research. *Human Genomics*, *8*(1), 21.

Cheng, N., Chytil, A., Shyr, Y., Joly, A., and Moses, H.L. (2008). Transforming growth factor-beta signaling-deficient fibroblasts enhance hepatocyte growth factor signaling in mammary carcinoma cells to promote scattering and invasion. Mol. Cancer Res. 6, 1521–1533.

Cho, W. C. (2007). Contribution of oncoproteomics to cancer biomarker discovery. *Mol Cancer*, *6*(25), 25-27.

Clark, H. F., Gurney, A. L., Abaya, E., Baker, K., Baldwin, D., Brush, J., Chen, J., Chow, B., Chui, C., Crowley, C., Currell, B., Deuel, B., Dowd, P., Eaton, D., Foster, J., Grimaldi, C., Gu, Q., Hass, P.E., Heldens, S., Huang, A., Kim, H.S., Klimowski, L., Jin, Y., Johnson, S., Lee, J., Lewis, L., Liao, D., Mark, M., Robbie, E., Sanchez, C., Schoenfeld, J., Seshagiri, S., Simmons, L., Singh, J., Smith, V., Stinson, J., Vagts, A., Vandlen, R., Watanabe, C., Wieand, D., Woods, K., Xie, M.H., Yansura, D., Yi,

S., Yu, G., Yuan, J., Zhang, M., Zhang, Z., Goddard, A., Wood, W.I., Godowski, P. and Gray, A. (2003). The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: a bioinformatics assessment. *Genome research*, *13*(10), 2265-2270.

Coccia, M. (2012). Evolutionary growth of knowledge in path–breaking targeted therapies for lung cancer: radical innovations and structure of the new technological paradigm. *International Journal of Behavioural and Healthcare Research*, *3*(3), 273-290.

Cooray, S., Howe, S. J. and Thrasher, A. J. (2012). 3 Retrovirus and Lentivirus Vector Design and Methods of Cell Conditioning. *Methods in Enzymology*, *507*, 29.

Coulibaly, I. and Page, G. P. (2008). Bioinformatic tools for inferring functional information from plant microarray data II: analysis beyond single gene. *International Journal of Plant Genomics*, *2008*.

Crane, C. A., Austgen, K., Haberthur, K., Hofmann, C., Moyes, K. W., Avanesyan, L., Fong, L., Campbell, M.J., Cooper, S., Oakes, S.A., Parsa, A.T. and Lanier, L. L. (2014). Immune evasion mediated by tumour-derived lactate dehydrogenase induction of NKG2D ligands on myeloid cells in glioblastoma patients. *Proceedings of the National Academy of Sciences*, *111*(35), 12823-12828.

Critchley-Thorne, R. J., Miller, S. M., Taylor, D. L. and Lingle, W. L. (2009). Applications of cellular systems biology in breast cancer patient stratification and diagnostics. *Combinatorial Chemistry & High Throughput Screening*, *12*(9), 860-869.

Cronin, M., Sangli, C., Liu, M. L., Pho, M., Dutta, D., Nguyen, A., Jeong, J., Wu, J., Langone, K. C., and Watson, D. (2007). Analytical validation of the Oncotype DX

genomic diagnostic test for recurrence prognosis and therapeutic response prediction in node-negative, estrogen receptor–positive breast cancer. *Clinical Chemistry*, *53*(6), 1084-1091.

Crusio, W. E. and Gerlai, R. T. (Eds.). (1999). *Handbook of molecular-genetic techniques for brain and behavior research*. Elsevier.

Cuvillier, O., Nava, V. E., Murthy, S. K., Edsall, L. C., Levade, T., Milstien, S. and Spiegel, S. (2001). Sphingosine generation, cytochrome c release, and activation of caspase-7 in doxorubicin-induced apoptosis of MCF7 breast adenocarcinoma cells. *Cell Death and Differentiation*, *8*(2), 162-171.

Czene, K., Lichtenstein, P. and Hemminki, K. (2002). Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish family-cancer database. *International Journal of Cancer*, *99*(2), 260-266.

Dalton, A. C., & Barton, W. A. (2014). Over-expression of secreted proteins from mammalian cell lines. *Protein Science*, *23*(5), 517-525.

Damsky, C. H. and Werb, Z. (1992). Signal transduction by integrin receptors for extracellular matrix: cooperative processing of extracellular information. *Current Opinion in Cell Biology*, *4*(5), 772-781.

de Visser, K. E., Eichten, A. and Coussens, L. M. (2006). Paradoxical roles of the immune system during cancer development. *Nature Reviews Cancer*, *6*(1), 24-37.

DeBerardinis, R. J., Lum, J. J., Hatzivassiliou, G. and Thompson, C. B. (2008). The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. *Cell Metabolism*, *7*(1), 11-20.

Denault, J. B. and Salvesen, G. S. (2002). Caspases: keys in the ignition of cell death. *Chemical Reviews*, *102*(12), 4489-4500.

Deng, R., Wang, S. M., Yin, T., Ye, T. H., Shen, G. B., Li, L., Zhao, J.Y., Sang, Y.X., Duan, X.G. and Wei, Y. Q. (2014). Dimethyl sulfoxide suppresses mouse 4T1 breast cancer growth by modulating tumour-associated macrophage differentiation. *Journal of Breast Cancer*, *17*(1), 25-32.

Dennis Jr, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C. and Lempicki, R. A. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol*, *4*(5), P3.

Deshpande, A., Sicinski, P. and Hinds, P.W. (2005). Cyclins and cdks in development and cancer: a perspective. *Oncogene* 24, 2909–2915.

Devadoss, A. V., Sheeba, J. J. and Victor, S. M. (2013). Analysis of Role of Protein in different Stages of Cancer using Weighted Multi Expert Neural Networks System. *International Journal of Computing Algorithm*, *2*, 399-409.

Diamandis, E. P. (2004). Mass spectrometry as a diagnostic and a cancer biomarker discovery tool opportunities and potential limitations. *Molecular & Cellular Proteomics*, *3*(4), 367-378.

Döhrmann, M., Schorr, A., Hülsenbeck, J., Rasmussen, T. L., Issinger, O. G. and Fritz, G. (2012). Comparative analysis of stress responses of H9c2 rat cardiomyoblasts following treatment with doxorubicin and tBOOH. *Experimental Cell Research*, *318*(6), 779-788.

Draghici, S., Khatri, P., Tarca, A. L., Amin, K., Done, A., Voichita, C., Georgescu, C. and Romero, R. (2007). A systems biology approach for pathway level analysis. *Genome Research*, *17*(10), 1537-1545.

Duffy, M. J. (2001). Biochemical markers in breast cancer: which ones are clinically useful? *Clinical Biochemistry*, *34*(5), 347-352.

DuPrie, M. (2010). The Role of EZH2 in Genomic Stability and Tumourigenesis in Breast Cancer.

Eom, Y. W., Kim, M. A., Park, S. S., Goo, M. J., Kwon, H. J., Sohn, S., Kim, W.H., Yoon, G. and Choi, K. S. (2005). Two distinct modes of cell death induced by doxorubicin: apoptosis and cell death through mitotic catastrophe accompanied by senescence-like phenotype. *Oncogene*, *24*(30), 4765-4777.

Essaghir, A. and Demoulin, J.B. (2012). A Minimal Connected Network of Transcription Factors Regulated in Human Tumours and Its Application to the Quest for Universal Cancer Biomarkers. PLoS ONE 7(6): e39666. doi:10.1371/journal.pone.0039666

Essaghir, A., Toffalini, F., Knoops, L., Kallin, A., van Helden, J. and Demoulin, J. B. (2010). Transcription factor regulation can be accurately predicted from the presence of target gene signatures in microarray gene expression data. *Nucleic Acids Research*, *38*(11), e120-e120.

Faça, V. M. and Hanash, S. M. (2009). In-depth proteomics to define the cell surface and secretome of ovarian cancer cells and processes of protein shedding. *Cancer Research*, *69*(3), 728-730.

Fan, C., Oh, D. S., Wessels, L., Weigelt, B., Nuyten, D. S., Nobel, A. B., van't Veer, L.J and Perou, C. M. (2006). Concordance among gene-expression–based predictors for breast cancer. *New England Journal of Medicine*, *355*(6), 560-569.

Ferlay, J., Héry, C., Autier, P. and Sankaranarayanan, R. (2010). Global burden of breast cancer. In *Breast Cancer Epidemiology* (pp. 1-19). Springer New York.

Ferlay, J., Shin, H. R., Bray, F., Forman, D., Mathers, C. and Parkin, D. M. (2010). An Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer*, *127*(12), 2893-2917.

Fernald, K. and Kurokawa, M. (2013). Evading apoptosis in cancer. Trends in Cell Biology. 23(12); 620-633

Ferrara, N. (2010). Pathways mediating VEGF-independent tumour angiogenesis. Cytokine Growth Factor Rev. 21, 21–26.

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

Fiebig, A. A., Zhu, W., Hollerbach, C., Leber, B. and Andrews, D. W. (2006). Bcl-XL is qualitatively different from and ten times more effective than Bcl-2 when expressed in a breast cancer cell line. *BMC cancer*, *6*(1), 213.

Fine, J. H., Chen, P., Mesci, A., Allan, D. S., Gasser, S., Raulet, D. H. and Carlyle, J. R. (2010). Chemotherapy-induced genotoxic stress promotes sensitivity to natural killer cell cytotoxicity by enabling missing-self recognition. *Cancer research*, *70*(18), 7102-7113.

Firczuk, H., Kannambath, S., Pahle, J., Claydon, A., Beynon, R., Duncan, J., Westerhoff, H., Mendes, P. and McCarthy, J. E. (2013). An in vivo control map for the eukaryotic mRNA translation machinery. *Molecular Systems Biology*, *9*(1).

Fisher, L. W., Jain, A., Tayback, M. and Fedarko, N. S. (2004). Small integrin binding ligand N-linked glycoprotein gene family expression in different cancers. *Clinical Cancer Research*, *10*(24), 8501-8511.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C. and Jensen, L. J. (2013). STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research*, *41*(D1), D808-D815.

Frank, R., & Hargreaves, R. (2003). Clinical biomarkers in drug discovery and development. *Nature Reviews Drug Discovery*, *2*(7), 566-580.

Fredlund, E., Staaf, J., Rantala, J. K., Kallioniemi, O., Borg, A. and Ringnér, M. (2012). The gene expression landscape of breast cancer is shaped by tumour protein p53 status and epithelial-mesenchymal transition. *Breast Cancer Res*, *14*(4), R113.

Fregene, A., & Newman, L. A. (2005). Breast cancer in sub-Saharan Africa: How does it relate to breast cancer in African-American women? *Cancer*, *103*(8), 1540-1550.

Fu, W., Lobocki, C. A., Silberberg, B. K., Chelladurai, M. and Young, S. C. (2001). Molecular markers in Paget disease of the breast. *Journal of Surgical Oncology*, *77*(3), 171-178.

Fulda, S. and Debatin, K. M. (2006). Extrinsic versus intrinsic apoptosis pathways in anticancer chemotherapy. *Oncogene*, *25*(34), 4798-4811.

Fulford, L. G., Reis-Filho, J. S., Ryder, K., Jones, C., Gillett, C. E., Hanby, A., Easton, D. and Lakhani, S. R. (2007). Basal-like grade III invasive ductal carcinoma of the breast: patterns of metastasis and long-term survival. *Breast Cancer Research*, *9*(1), R4.

Gasser, S. and Raulet, D. H. (2006). The DNA damage response arouses the immune system. *Cancer Research*, *66*(8), 3959-3962.

Gasser, S., Orsulic, S., Brown, E. J. and Raulet, D. H. (2005). The DNA damage pathway regulates innate immune system ligands of the NKG2D receptor. *Nature*, *436*(7054), 1186-1190.

Ghebranious, N. and Donehower, L.A. (1998). Mouse models in tumour suppression. Oncogene 17, 3385–3400.

Gibbons, F. D. and Roth, F. P. (2002). Judging the quality of gene expression-based clustering methods using gene annotation. *Genome Research*, *12*(10), 1574-1581.

Goel, M. K., Khanna, P. and Kishore, J. (2010). Understanding survival analysis: Kaplan-Meier estimate. *International Journal of Ayurveda Research*, *1*(4), 274.

González, S., López-Soto, A., Suarez-Alvarez, B., López-Vázquez, A. and López-Larrea, C. (2008). NKG2D ligands: key targets of the immune response. *Trends in Immunology*, *29*(8), 397-403.

Greenbaum, D., Colangelo, C., Williams, K. and Gerstein, M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol*, *4*(9), 117.

Greenhough, A., Smartt, H. J., Moore, A. E., Roberts, H. R., Williams, A. C., Paraskeva, C. and Kaidi, A. (2009). The COX-2/PGE2 pathway: key roles in the hallmarks of cancer and adaptation to the tumour microenvironment. *Carcinogenesis*, *30*(3), 377-386.

Groh, V., Wu, J., Yee, C. and Spies, T. (2002). Tumour-derived soluble MIC ligands impair expression of NKG2D and T-cell activation. *Nature*, *419*(6908), 734-738.

Gry, M., Rimini, R., Strömberg, S., Asplund, A., Pontén, F., Uhlén, M. and Nilsson, P. (2009). Correlations between RNA and protein expression profiles in 23 human cell lines. *BMC Genomics*, *10*(1), 365.

Guerra, N., Tan, Y. X., Joncker, N. T., Choy, A., Gallardo, F., Xiong, N., Knoblaugh, S., Cado, D., Greenberg, N.M. and Raulet, D. H. (2008). NKG2D-deficient mice are defective in tumour surveillance in models of spontaneous malignancy. *Immunity*, *28*(4), 571-580.

Győrffy, B., Lánczky, A. and Szállási, Z. (2012). Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients. *Endocrine-Related Cancer*, *19*(2), 197-208.

Györffy, B., Lanczky, A., Eklund, A. C., Denkert, C., Budczies, J., Li, Q. and Szallasi, Z. (2010). An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Research and Treatment*, *123*(3), 725-731.

Győrffy, B., Surowiak, P., Budczies, J. and Lánczky, A. (2013). Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One*, *8*(12), e82241.

Hahn, W. C. and Weinberg, R. A. (2002). Modelling the molecular circuitry of cancer. *Nature Reviews Cancer*, *2*(5), 331-341.

Hanahan, D and Weinberg, R.A. (2000). The Hallmarks of Cancer, *Cell*, 100, 57–70

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, *144*(5), 646-674.

Hanahan, D. and Folkman, J. (1996). Patterns and emerging mechanisms of the angiogenic switch during tumourigenesis. *Cell* 86, 353–364.

Hanash, S. M., Pitteri, S. J. and Faca, V. M. (2008). Mining the plasma proteome for cancer biomarkers. *Nature*, *452*(7187), 571-579.

Hanslick, J. L., Lau, K., Noguchi, K. K., Olney, J. W., Zorumski, C. F., Mennerick, S. and Farber, N. B. (2009). Dimethyl sulfoxide (DMSO) produces widespread apoptosis in the developing central nervous system. *Neurobiology of Disease*, *34*(1), 1-10.

Harris, J. R., Lippman, M. E., Osborne, C. K. and Morrow, M. (2012). *Diseases of the Breast*. Lippincott Williams & Wilkins.

Harris, L., Fritsche, H., Mennel, R., Norton, L., Ravdin, P., Taube, S., Somerfield, M.R., Hayes, D.F. and Bast, R. C. (2007). American Society of Clinical Oncology 2007 update of recommendations for the use of tumour markers in breast cancer. *Journal of Clinical Oncology*, *25*(33), 5287-5312.

Harsha, H. C., Kandasamy, K., Ranganathan, P., Rani, S., Ramabadran, S., Gollapudi, S., Balakrishnan, L., Dwivedi, S.B., Telikicherla, D., Selvan, L.D., Goel, R., Mathivanan, S., Marimuthu, A., Kashyap, M., Vizza, R.F., Mayer, R.J., Decaprio,

J.A., Srivastava, S., Hanash, S.M., Hruban, R.H, and Pandey, A. (2009). A compendium of potential biomarkers of pancreatic cancer. *PLoS medicine*, *6*(4), e1000046.

Hartman, J. L., Garvik, B. and Hartwell, L. (2001). Principles for the buffering of genetic variation. *Science*, *291*(5506), 1001-1004.

Hartman, J., Ström, A. and Gustafsson, J. A. (2009). Estrogen receptor beta in breast cancer diagnostic and therapeutic implications. *Steroids*, *74*(8), 635-641.

Hayes, D. F., Thor, A. D., Dressler, L. G., Weaver, D., Edgerton, S., Cowan, D., Broadwater, G., Goldstein, L.J., Martino, S., Ingle, J.N., Henderson, I.C., Norton, L., Winer, E.P., Hudis, C.A., Ellis, M.J. and Berry, D. A. (2007). HER2 and response to paclitaxel in node-positive breast cancer. *New England Journal of Medicine*, *357*(15), 1496-1506.

Heinemann, A. and Paschen, A. (2012). Tumour suppressors control ULBP2, an innate surface ligand of the lymphocyte immune receptor NKG2D. *Oncoimmunology*, *1*(4), 535-536.

Heinemann, A., Zhao, F., Pechlivanis, S., Eberle, J., Steinle, A., Diederichs, S., Schadendorf, D and Paschen, A. (2012). Tumour suppressive microRNAs miR-34a/c control cancer cell expression of ULBP2, a stress-induced ligand of the natural killer cell receptor NKG2D. *Cancer Research*, *72*(2), 460-471.

Henson, P. M. and Hume, D. A. (2006). Apoptotic cell removal in development and tissue homeostasis. *Trends in Immunology*, *27*(5), 244-250.

Hess, K. R., Anderson, K., Symmans, W. F., Valero, V., Ibrahim, N., Mejia, J. A., Booser, D., Theriault, R.L., Buzdar, A.U., Dempsey, P.J., Rouzier, R., Sneige, N.,

Ross, J.S., Vidaurre, T., Gómez, H.L., Hortobagyi, G.N, Pusztai, L. (2006). Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *Journal of Clinical Oncology*, *24*(26), 4236-4244.

Hoang, B. X., Tran, D. M., Tran, H. Q., Nguyen, P. T., Pham, T. D., Dang, H. V., Ha, T.V., Tran, H.D., Hoang, C., Luong, K.N. and Shaw D.G. (2011). Dimethyl sulfoxide and sodium bicarbonate in the treatment of refractory cancer pain. *Journal of Pain and Palliative Care Pharmacotherapy*, *25*(1), 19-24.

Holden, J. A. (2001). DNA Topoisomerases as anticancer drug targets from the laboratory to the clinic. *Current Medicinal Chemistry-Anti-Cancer Agents*, *1*(1), 1-25.

Holliday, D. L. and Speirs, V. (2011). Choosing the right cell line for breast cancer research. *Breast Cancer Res*, *13*(4), 215.

Hollstein, M., Sidransky, D., Vogelstein, B. and Harris, C. C. (1991). p53 mutations in human cancers. *Science*, *253*(5015), 49-53.

Høyer, A. P., Gerdes, A. M., Jørgensen, T., Rank, F. and Hartvig, H. B. (2002). Organochlorines, p53 mutations in relation to breast cancer risk and survival. A Danish cohort-nested case-controls study. *Breast Cancer Research and Treatment*, *71*(1), 59-65.

Hu, J., Zhu, S., Xia, X., Kleinerman, E. and Li, S. (2013). Co-stimulation of IL-12 gene delivery plus doxorubicin induces NKG2D-dependent CD8+ T cell infiltration to tumours and inhibition of tumour progression. *Cancer Research*, *73*(8 Supplement), 4996-4996.

Hu, J., Zhu, S., Xia, X., Zhang, L., Kleinerman, E. S. and Li, S. (2014). CD8 T cell-specific induction of NKG2D receptor by doxorubicin plus interleukin-12 and its contribution to CD8 T cell accumulation in tumours. *Molecular Cancer*, *13*, 34.

Hu, S., Arellano, M., Boontheung, P., Wang, J., Zhou, H., Jiang, J., Elashoff, D., Wei R., Loo, J.A. and Wong, D. T. (2008). Salivary proteomics for oral cancer biomarker discovery. *Clinical Cancer Research*, *14*(19), 6246-6252.

Hu, Y., Wu, G., Rusch, M., Lukes, L., Buetow, K. H., Zhang, J. and Hunter, K. W. (2012). Integrated cross-species transcriptional network analysis of metastatic susceptibility. *Proceedings of the National Academy of Sciences*, *109*(8), 3184-3189.

Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44-57.

Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, *37*(1), 1-13.

Huang, H. C., Zheng, S., VanBuren, V. and Zhao, Z. (2010). Discovering disease-specific biomarker genes for cancer diagnosis and prognosis. *Technology in Cancer Research & Treatment*, *9*(3), 219-229.

Hubner, A., Barnett, T., Flavell, R. A. and Davis. R.J. (2008). Multisite phosphorylation regulates Bim stability and apoptotic activity. Mol. Cell 30, 415–425

Hunt, K. K., Green, M. C. and Buchholz, T. A. (2012). Diseases of the breast. *Townsend CM, Beauchamp RD, Evers BM, Mattox KL. Sabiston Textbook of Surgery, 19th ed. St. Louis, Mo: WB Saunders*.

Hunter, T. (2000). Signaling—2000 and beyond. *Cell*, *100*(1), 113-127.

Iorio, M. V., Ferracin, M., Liu, C. G., Veronese, A., Spizzo, R., Sabbioni, S., Magri E., Pedriali, M., Fabbri, M., Campiglio, M., Ménard, S., Palazzo, J.P., Rosenberg, A., Musiani, P., Volinia, S., Nenci, I., Calin, G.A., Querzoli, P., Negrini, M. and Croce, C. M. (2005). MicroRNA gene expression deregulation in human breast cancer. *Cancer Research*, *65*(16), 7065-7070.

Janeway, C. A., Travers, P., Walport, M. and Shlomchik, M. J. (2001). The interaction of the antibody molecule with specific antigen. The Immune System in Health and Disease. 5th edition. New York: Garland Science.

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E. and Forman, D. (2011). Global cancer statistics. *CA: A Cancer Journal for Clinicians*, *61*(2), 69-90.

Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, and von Mering C. (2009). STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Research*, *37*(suppl 1), D412-D416.

Jones, C. B., Clements, M. K., Wasi, S. and Daoud, S. S. (2000). Enhancement of camptothecin-induced cytotoxicity with UCN-01 in breast cancer cells: abrogation of S/G2 arrest. *Cancer Chemotherapy and Pharmacology*, *45*(3), 252-258.

Junttila, M.R., and Evan, G.I. (2009). p53—a Jack of all trades but master of none. *Nat. Rev. Cancer* 9, 821–829.

Kabbage, M., Trimeche, M., Bergaoui, S., Hammann, P., Kuhn, L., Hamrita, B., ben Nasr, H., Chaieb, A., Chouchane, L. and Chahed, K. (2013). Calreticulin expression

in infiltrating ductal breast carcinomas: relationships with disease progression and humoral immune responses. *Tumour Biology*, *34*(2), 1177-1188.

Kaiser, B. K., Yim, D., Chow, I. T., Gonzalez, S., Dai, Z., Mann, H. H., Strong, R.K., Groh, V. and Spies, T. (2007). Disulphide-isomerase-enabled shedding of tumour-associated NKG2D ligands. *Nature*, *447*(7143), 482-486.

Kanagathara, N. and kavitha, K. (2014). Evaluation of HER2/NEU over expression in breast cancer. *International Journal of Pharmacy & Pharmaceutical Sciences*. *6*(2), 898-900.

Kanehisa, M., Goto, S., Kawashima, S. and Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic Acids Research*, *30*(1), 42-46.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research*, *42*(D1), D199-D205.

Kapushesky, M., Adamusiak, T., Burdett, T., Culhane, A., Farne, A., Filippov, A., Holloway, E., Klebanov, A., Kryvych, N., Kurbatova, N., Kurnosov, P., Malone, J., Melnichuk, O., Petryszak, R., Pultsin, N., Rustici, G., Tikhonov, A., Travillian, R.S., Williams, E., Zorin, A., Parkinson, H. and Brazma, A. (2012). Gene Expression Atlas update—a value-added database of microarray and sequencing-based functional genomics experiments. *Nucleic Acids Research*, *40*(D1), D1077-D1081.

Kapushesky, M., Emam, I., Holloway, E., Kurnosov, P., Zorin, A., Malone, J., Rustici, G., Williams, E., Parkinson, H. and Brazma, A. (2009). Gene expression atlas at the European bioinformatics institute. *Nucleic Acids Research*, gkp936.

Karam, J. A., Lotan, Y., Karakiewicz, P. I., Ashfaq, R., Sagalowsky, A. I., Roehrborn, C. G. and Shariat, S. F. (2007). Use of combined apoptosis biomarkers for prediction of bladder cancer recurrence and mortality after radical cystectomy. *The Lancet Oncology*, *8*(2), 128-136.

Karin, M. and Greten, F. R. (2005). NF-κB: linking inflammation and immunity to cancer development and progression. *Nature Reviews Immunology*, *5*(10), 749-759.

Karley, D., Gupta, D. and Tiwari, A. (2011). Biomarker for cancer: a great promise for future. *World Journal of Oncology*, *2*(4), 151-157.

Kartal, M., Ozcakar, N., Hatipoglu, S., Tan, M. N. and Guldal, A. D. (2014). Breast cancer risk perceptions of Turkish women attending primary care: a cross-sectional study. *BMC women's health*, *14*(1), 152.

Kazerounian, S., Yee, K.O. and Lawler, J. (2008). Thrombospondins in cancer. *Cell. Mol. Life Sci.* 65, 700–712.

Keen, J. C. (2011). *Breast Cancer Cell Line Development and Authentication*. INTECH Open Access Publisher.

Kessenbrock, K., Plaks, V. and Werb, Z. (2010). Matrix metalloproteinases: Regulators of the tumour microenvironment. *Cell* 141, 52–67.

Khaleel, S. S., Andrews, E. H., Ung, M., DiRenzo, J. and Cheng, C. (2014). E2F4 regulatory program predicts patient survival prognosis in breast cancer. *Breast Cancer Research*, *16*(6), 486.

Khatri, P., Sirota, M. and Butte, A. J. (2012). Ten years of pathway analysis: current approaches and outstanding challenges. *Plos Computational Biology*, *8*(2), e1002375.

Kim, J. W., Akiyama, M., Park, J. H., Lin, M. L., Shimo, A., Ueki, T., Daigo, Y., Tsunoda, T., Nishidate, T., Nakamura, Y. and Katagiri, T. (2009). Activation of an estrogen/estrogen receptor signaling by BIG3 through its inhibitory effect on nuclear transport of PHB2/REA in breast cancer. *Cancer Science*, *100*(8), 1468-1478.

Kim, T. K. and Eberwine, J. H. (2010). Mammalian cell transfection: the present and the future. *Analytical and Bioanalytical Chemistry*, *397*(8), 3173-3178.

Kirmiz, C., Li, B., An, H. J., Clowers, B. H., Chew, H. K., Lam, K. S., Ferrige, A., Alecio, R., Borowsky, A.D., Sulaimon, S., Lebrilla, C.B. and Miyamoto, S. (2007). A serum glycomics approach to breast cancer biomarkers. *Molecular & Cellular Proteomics*, *6*(1), 43-55.

Knudson, A.G. (1985). Hereditary cancer, Oncogenes, and antioncogenes, *Cancer Res*, 45, 1437-1443.

Koiri, R. K. and Trigun, S. K. (2011). Dimethyl sulfoxide activates tumour necrosis factorα-p53 mediated apoptosis and down regulates d-fructose-6-phosphate-2-kinase and lactate dehydrogenase-5 in Dalton's lymphoma in vivo. *Leukemia Research*, *35*(7), 950-956.

Kolonel, L. N., Altshuler, D. and Henderson, B. E. (2004). The multiethnic cohort study: exploring genes, lifestyle and cancer risk. *Nature Reviews Cancer*, *4*(7), 519-527.

Kouros-Mehr, H., Slorach, E. M., Sternlicht, M. D. and Werb, Z. (2006). GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell*, *127*(5), 1041-1055.

Kriege, M., Brekelmans, C. T., Boetes, C., Besnard, P. E., Zonderland, H. M., Obdeijn, I. M., Manoliu, R.A., Kok, T., Peterse, H., Tilanus-Linthorst, M.M., Muller, S.H., Meijer, S., Oosterwijk, J.C., Beex, L.V., Tollenaar, R.A., de Koning, H.J., Rutgers, E.J. and Klijn, J. G. (2004). Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. *New England Journal of Medicine*, *351*(5), 427-437.

Krizhanovsky, V., Yon, M., Dickins, R. A., Hearn, S., Simon, J., Miething, C., Yee, H., Zender, L. and Lowe, S. W. (2008). Senescence of activated stellate cells limits liver fibrosis. *Cell*, *134*(4), 657-667.

Kuilman, T., Michaloglou, C., Mooi, W. J. and Peeper, D. S. (2010). The essence of senescence. *Genes & Development*, *24*(22), 2463-2479.

Kulasingam, V. (2008). *Identification and Validation of Candidate Breast Cancer Biomarkers: A Mass Spectrometric Approach* (Doctoral dissertation, University of Toronto).

Kulasingam, V. and Diamandis, E. P. (2008). Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nature Clinical Practice Oncology*, *5*(10), 588-599.

Kuppusamy, P., Govindan, N., Yusoff, M. M. and Ichwan, S. J. (2014). Proteins are potent biomarkers to detect colon cancer progression. *Saudi Journal of Biological Sciences*.

Lacey, J. V., Brinton, L. A., Leitzmann, M. F., Mouw, T., Hollenbeck, A., Schatzkin, A. and Hartge, P. (2006). Menopausal hormone therapy and ovarian cancer risk in the

National Institutes of Health–AARP Diet and Health Study Cohort. *Journal of the National Cancer Institute*, *98*(19), 1397-1405.

Lao, J., Madani, J., Puértolas, T., Álvarez, M., Hernández, A., Pazo-Cid, R., Artal A. and Antón Torres, A. (2013). Liposomal doxorubicin in the treatment of breast cancer patients: a review. *Journal of Drug Delivery*, *2013*.

Lee, A. H. S. (2008). Recent developments in the histological diagnosis of spindle cell carcinoma, fibromatosis and phyllodes tumour of the breast. *Histopathology*, *52*(1), 45-57.

Lee, J. M. and Sonnhammer, E. L. (2003). Genomic gene clustering analysis of pathways in eukaryotes. *Genome Research*, *13*(5), 875-882.

Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., Chen, L., Chen, L., Chen, T.M., Chin, M.C., Chong, J., Crook, B.E., Czaplinska, A., Dang, C.N., Datta, S., Dee, N.R., Desaki, A.L., Desta, T., Diep, E., Dolbeare, T.A., Donelan, M.J., Dong, H.W., Dougherty, J.G., Duncan, B.J., Ebbert, A.J., Eichele, G., Estin, L.K., Faber, C., Facer, B.A., Fields, R., Fischer, S.R., Fliss, T.P., Frensley, C., Gates, S.N., Glattfelder, K.J., Halverson, K.R., Hart, M.R., Hohmann, J.G., Howell, M.P., Jeung, D.P., Johnson, R.A., Karr, P.T., Kawal, R., Kidney, J.M., Knapik, R.H and Kuan, C. L. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, *445*(7124), 168-176.

Leist, M. and Jäättelä, M. (2001). Four deaths and a funeral: from caspases to alternative mechanisms. *Nature reviews Molecular cell biology*, *2*(8), 589-598.

Li, D., Ronson, B., Guo, M., Liu, S., Bishop, J. S., Van Echo, D. A. and O'Malley, B. W. (2002). Interleukin 2 gene transfer prevents NKG2D suppression and enhances antitumour efficacy in combination with cisplatin for head and neck squamous cell cancer. *Cancer Research*, *62*(14), 4023-4028.

Li, J., Di, C., Mattox, A. K., Wu, L. and Adamson, D. C. (2010). The future role of personalized medicine in the treatment of glioblastoma multiforme. *Pharmacogenomics and personalized medicine*, *3*, 111.

Li, J., Zhang, Z., Rosenzweig, J., Wang, Y. Y. and Chan, D. W. (2002). Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clinical Chemistry*, *48*(8), 1296-1304.

Liao, J., Xu, X. I. N. G. Z. H. I. and Wargovich, M. J. (2000). Direct reprobing with anti-beta-actin antibody as an internal control for western blotting analysis. *Biotechniques*, *28*(2), 216-218.

Lindor, N. M., McMaster, M. L., Lindor, C. J. and Greene, M. H. (2008). Concise handbook of familial cancer susceptibility syndromes. *Journal of the National Cancer Institute. Monographs*, (38), 1-93.

Lipinski, M.M., and Jacks, T. (1999). The retinoblastoma gene family in differentiation and development. *Oncogene*, 18, 7873–7882.

Liu, L. F., Desai, S. D., LI, T. K., Mao, Y., Sun, M. E. I. and SIM, S. P. (2000). Mechanism of action of camptothecin. *Annals of the New York Academy of Sciences*, *922*(1), 1-10.

Liu, R., Wang, X., Chen, G. Y., Dalerba, P., Gurney, A., Hoey, T., Sherlock, G., Lewicki, J., Shedden, K. and Clarke, M. F. (2007). The prognostic role of a gene

signature from tumourigenic breast-cancer cells. *New England Journal of Medicine*, *356*(3), 217-226.

Livasy, C. A., Karaca, G., Nanda, R., Tretiakova, M. S., Olopade, O. I., Moore, D. T. and Perou, C. M. (2006). Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Modern Pathology*, *19*(2), 264-271.

Loeb, K. R. and Loeb, L. A. (2000). Significance of multiple mutations in cancer. *Carcinogenesis*, *21*(3), 379-385.

López-Árias, E., Aguilar-Lemarroy, A., Felipe Jave-Suárez, L., Morgan-Villela, G., Mariscal-Ramírez, I., Martínez-Velázquez, M., Alvarez, A.H., Gutiérrez-Ortega, A. and Hernández-Gutiérrez, R. (2012). Alpha 1-antitrypsin: A novel tumour-associated antigen identified in patients with early-stage breast cancer. *Electrophoresis*, *33*(14), 2130-2137.

López-Otín, C. and Matrisian, L. M. (2007). Emerging roles of proteases in tumour suppression. *Nature reviews cancer*, *7*(10), 800-808.

López-Soto, A., Quiñones-Lombraña, A., López-Arbesú, R., López-Larrea, C. and González, S. (2006). Transcriptional regulation of ULBP1, a human ligand of the NKG2D receptor. *Journal of Biological Chemistry*, *281*(41), 30419-30430. Lowe, S.W., Cepero, E., and Evan, G. (2004). Intrinsic tumour suppression. *Nature* 432, 307–315.

Ludwig, J. A. and Weinstein, J. N. (2005). Biomarkers in cancer staging, prognosis and treatment selection. *Nature Reviews Cancer*, *5*(11), 845-856.

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. *Modern Pathology*, *23*(12), 1559-1566.

Ma, I., Dueck, A., Gray, R., Wasif, N., Giurescu, M., Lorans, R., Pizzitola, V. and Pockaj, B. (2012). Clinical and self-breast examination remain important in the era of modern screening. *Annals of Surgical Oncology*, *19*(5), 1484-1490.

Manavathi, B., Samanthapudi, V. S. and Gajulapalli, V. N. R. (2014). Estrogen receptor coregulators and pioneer factors: the orchestrators of mammary gland cell fate and development. *Frontiers in Cell and Developmental Biology*, *2*.34

Marks, P. A. and Xu, W. S. (2009). Histone deacetylase inhibitors: Potential in cancer therapy. *Journal of Cellular Biochemistry*, *107*(4), 600-608.

Martin, S.G. (2003). Cell Signaling and Cancer. *Cancer Cell*, 4, 164-174

Martinou, J.C. and Youle, R.J. (2011) Mitochondria in apoptosis: Bcl-2 family members and mitochondrial dynamics. *Dev. Cell* 21, 92–101

Mason, D. X., Jackson, T. J. and Lin, A. W. (2004). Molecular signature of oncogenic ras-induced senescence. *Oncogene*, *23*(57), 9238-9246.

Masson, M. J., Carpenter, L. D., Graf, M. L. and Pohl, L. R. (2008). Pathogenic role of natural killer T and natural killer cells in acetaminophen-induced liver injury in mice is dependent on the presence of dimethyl sulfoxide. *Hepatology*, *48*(3), 889-897.

Matboli, M., Eissa, S. and Said, H. (2014). Evaluation of histidine-rich glycoprotein tissue RNA and serum protein as novel markers for breast cancer. *Medical Oncology*, *31*(4), 1-7.

Mathers, C. and Bonita, R. (2009). Current global health status. *Global Public Health: A New Era*, 23-61.

McPherson, K., Steel, C. M. and Dixon, J. M. (2006). 5 Breast cancer-epidemiology, risk factors, and genetics. *ABC of Breast Diseases*, *572*, 24.

McPherson, K., Steel, C. and Dixon, J. M. (2000). ABC of breast diseases: breast cancer-epidemiology, risk factors, and genetics. *BMJ: British Medical Journal*, *321*(7261), 624.

McShane, L. M., Altman, D. G., Sauerbrei, W., Taube, S. E., Gion, M. and Clark, G. M. (2005). Reporting recommendations for tumour marker prognostic studies. *Journal of Clinical Oncology*, *23*(36), 9067-9072.

Michl, P., Ramjaun, A. R., Pardo, O. E., Warne, P. H., Wagner, M., Poulsom, R., D'Arrigo, C., Ryder, K., Menke, A., Gress, T. and Downward, J. (2005). CUTL1 is a target of TGFβ signaling that enhances cancer cell motility and invasiveness. *Cancer Cell*, *7*(6), 521-532.

Moore, J. H., Asselbergs, F. W. and Williams, S. M. (2010). Bioinformatics challenges for genome-wide association studies. *Bioinformatics*, *26*(4), 445-455.

Moretta, L. and Moretta, A. (2004). Unravelling natural killer cell function: triggering and inhibitory human NK receptors. *The EMBO Journal*, *23*(2), 255-259.

Morita, S., Kojima, T. and Kitamura, T. (2000). Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Therapy*, *7*(12), 1063-1066.

Mucci, L. A., Wedren, S., Tamimi, R. M., Trichopoulos, D. and Adami, H. O. (2001). The role of gene–environment interaction in the aetiology of human cancer: examples

from cancers of the large bowel, lung and breast. *Journal of Internal Medicine*, *249*(6), 477-493.

Narayanan, R. (2007). Bioinformatics approaches to cancer gene discovery. In *Target Discovery and Validation Reviews and Protocols* (pp. 13-31). Humana Press.

Neagu, M., Constantin, C., & Tanase, C. (2010). Immune-related biomarkers for diagnosis/prognosis and therapy monitoring of cutaneous melanoma. *Expert Review of Molecular Diagnostics*, *10*(7), 897-919.

Ngcoza, N. (2012). Identification of Biomarkers in Breast Cancer as Potential Diagnostic and Therapeutic agents using a combined In-silico and Molecular approach. Unpublished Master's thesis. University of the Western Cape.

Niaid, N. (2006). Database for Annotation, Visualization and Integrated Discovery (DAVID).

Nice, T. J., Coscoy, L. and Raulet, D. H. (2009). Posttranslational regulation of the NKG2D ligand Mult1 in response to cell stress. *The Journal of Experimental Medicine*, *206*(2), 287-298.

Ning, K., Fermin, D. and Nesvizhskii, A. I. (2012). Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data. *Journal of Proteome Research*, *11*(4), 2261-2271.

Ohashi, M., Eagle, R. A. and Trowsdale, J. (2010). Post-translational modification of the NKG2D ligand RAET1G leads to cell surface expression of a glycosylphosphatidylinositol-linked isoform. *Journal of Biological Chemistry*, *285*(22), 16408-16415.

Okita, R., Mougiakakos, D., Ando, T., Mao, Y., Sarhan, D., Wennerberg, E., Seliger, B., Lundqvist, A., Mimura K and Kiessling, R. (2012). HER2/HER3 signaling regulates NK cell-mediated cytotoxicity via MHC class I chain-related molecule A and B expression in human breast cancer cell lines. *The Journal of Immunology*, *188*(5), 2136-2145.

Onda, H., Ohkubo, S., Shintani, Y., Ogi, K., Kikuchi, K., Tanaka, H., Yamamoto, K., Tsuji, I., Ishibashi, Y., Yamada, T., Kitada, C., Suzuki, N., Sawada, H., Nishimura, O. and Fujino, M. (2001). A novel secreted tumour antigen with a glycosylphosphatidylinositol-anchored structure ubiquitously expressed in human cancers. *Biochemical and Biophysical Research Communications*, *285*(2), 235-243.

Opferman, J. T. and Korsmeyer, S. J. (2003). Apoptosis in the development and maintenance of the immune system. *Nature Immunology*, *4*(5), 410-415.

Orr, M. T. and Lanier, L. L. (2010). Natural killer cell education and tolerance. *Cell*, *142*(6), 847-856.

Packer, N. H., Gooley, A. A. and Wilkins, M. R. (2007). One gene, many proteins. In *Proteome Research* (pp. 95-121). Springer Berlin Heidelberg.

Paik, S., Tang, G., Shak, S., Kim, C., Baker, J., Kim, W., Cronin, M., Baehner, F.L., Watson, D., Bryant, J., Costantino, J.P., Geyer, C.E. Jr., Wickerham, D.L. and Wolmark, N. (2006). Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor–positive breast cancer. *Journal of Clinical Oncology*, *24*(23), 3726-3734.

Panavas, T., Lu, J., Liu, X., Winkis, A. M., Powers, G., Naso, M. F. and Amegadzie, B. (2011). Generation and evaluation of mammalian secreted and membrane protein

expression libraries for high-throughput target discovery. *Protein Expression and Purification*, *79*(1), 7-15.

Paolicchi, E., Crea, F., Farrar, W. L., Green, J. E. and Danesi, R. (2013). Histone lysine demethylases in breast cancer. *Critical Reviews in Oncology/Hematology*, *86*(2), 97-103.

Park, S. M., Gaur, A. B., Lengyel, E. and Peter, M. E. (2008). The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes & Development*, *22*(7), 894-907.

Parsons, M.J., McCormick, L., Janke, L., Howard, A., Bouchier-Hayes, L. and Green, D.R. (2013). Genetic deletion of caspase-2 accelerates MMTV/c-neu-driven mammary carcinogenesis in mice. *Cell Death and Differentiation*. 20, 1174–1182

Paschen, A., Sucker, A., Hill, B., Moll, I., Zapatka, M., Nguyen, X. D., Sim, G.C., Gutmann, I., Hassel, J., Becker, J.C., Steinle, A., Schadendorf, D. and Ugurel, S. (2009). Differential clinical significance of individual NKG2D ligands in melanoma: soluble ULBP2 as an indicator of poor prognosis superior to S100B. *Clinical Cancer Research*, *15*(16), 5208-5215.

Paschen, A., Sucker, A., Hill, B., Moll, I., Zapatka, M., Nguyen, X. D., Sim, G.C., Gutmann, I., Hassel, J., Becker, J.C., Steinle, A., Schadendorf, D. and Ugurel, S. (2009). Differential clinical significance of individual NKG2D ligands in melanoma: soluble ULBP2 as an indicator of poor prognosis superior to S100B. *Clinical Cancer Research*, *15*(16), 5208-5215.

Pattabiraman, D. R. and Weinberg, R. A. (2014). Tackling the cancer stem cells—what challenges do they pose?. *Nature Reviews Drug Discovery*, *13*(7), 497-512.

Patterson, S. D. and Aebersold, R. H. (2003). Proteomics: the first decade and beyond. *Nature Genetics*, *33*, 311-323.

Pende, D., Rivera, P., Marcenaro, S., Chang, C. C., Biassoni, R., Conte, R., Kubin, M., Cosman, D., Ferrone, S., Moretta, L. and Moretta, A. (2002). Major Histocompatibility Complex Class I-related Chain A and UL16-Binding Protein Expression on Tumour Cell Lines of Different Histotypes Analysis of Tumour Susceptibility to NKG2D-dependent Natural Killer Cell Cytotoxicity. *Cancer Research*, *62*(21), 6178-6186.

Perissi, V., Menini, N., Cottone, E., Capello, D., Sacco, M., Montaldo, F. and De Bortoli, M. (2000). AP-2 transcription factors in the regulation of ERBB2 gene transcription by oestrogen. *Oncogene*, *19*(2), 280-288.

Pisano, E. D., Gatsonis, C., Hendrick, E., Yaffe, M., Baum, J. K., Acharyya, S., Conant, E.F., Fajardo, L.L., Bassett, L., D'Orsi, C. and Jong, R. (2006). Diagnostic performance of digital versus film mammography for breast-cancer screening. *Obstetrical & Gynecological Survey*, *61*(3), 178.

Platet, N., Cunat, S., Chalbos, D., Rochefort, H. and Garcia, M. (2000). Unliganded and liganded estrogen receptors protect against cancer invasion via different mechanisms. *Molecular Endocrinology*, *14*(7), 999-1009.

Poggi, A., Massaro, A. M., Negrini, S., Contini, P. and Zocchi, M. R. (2005). Tumour-induced apoptosis of human IL-2-activated NK cells: role of natural cytotoxicity receptors. *The Journal of Immunology*, *174*(5), 2653-2660.

Pommler, R. F., Woltering, E. A., Milo, G. and Fletcher, W. S. (1988). Synergistic cytotoxicity between dimethyl sulfoxide and antineoplastic agents against ovarian cancer in vitro. *American Journal of Obstetrics and Gynecology*, *159*(4), 848-852.

Pretorius, A., Bankole, H.A., Meyer, M., February, F. and Rees, D. J. G. (2013). Silencing of Mouse RBBP6 Using Interference RNA Implicates It in Apoptosis and the Cell Cycle. *Journal of BioSciences (JBio)*, *2*(2).

Rakha, E. A., Reis-Filho, J. S. and Ellis, I. O. (2008). Basal-like breast cancer: a critical review. *Journal of Clinical Oncology*, *26*(15), 2568-2581.

Ralhan, R., DeSouza, L. V., Matta, A., Tripathi, S. C., Ghanny, S., Gupta, S. D., Bahadur, S. and Siu, K. M. (2008). Discovery and verification of head-and-neck cancer biomarkers by differential protein expression analysis using iTRAQ labeling, multidimensional liquid chromatography, and tandem mass spectrometry. *Molecular & Cellular Proteomics*, *7*(6), 1162-1173.

Ramakrishnan, R., Agrawal, R., Freytag, J. C., Bollinger, T., Clifton, C. W., Dzeroski, S., Hipp, J., Keim, D., Kramer, S., Kriegel, H., Leser, U., Liu, B., Mannila, H., Meo, R., Morishita, S., Ng, R., Pei, J., Raghavan, P., Spiliopoulou, M., Srivastava, J. and Torra, V. (2005). *Data mining: The next generation*. R. Agrawal, J. C. Freytag, & R. Ramakrishnan (Eds.). Internat. Begegnungs-und Forschungszentrum für Informatik.

Rangiah, K., Tippornwong, M., Sangar, V., Austin, D., Tétreault, M. P., Rustgi, A. K., Blair, I. A. and Yu, K. H. (2009). Differential secreted proteome approach in murine model for candidate biomarker discovery in colon cancer. *Journal of Proteome Research*, *8*(11), 5153-5164.

Rao, D. D., Vorhies, J. S., Senzer, N. and Nemunaitis, J. (2009). siRNA vs. shRNA: similarities and differences. *Advanced Drug Delivery Reviews*, *61*(9), 746-759.

Raoul, C., Abbas-Terki, T., Bensadoun, J. C., Guillot, S., Haase, G., Szulc, J., Henderson, C.E. and Aebischer, P. (2005). Lentiviral-mediated silencing of SOD1 through RNA interference retards disease onset and progression in a mouse model of ALS. *Nature Medicine*, *11*(4), 423-428.

Rashidi, H. H. and Buehler, L. K. (1999). *Bioinformatics basics: Applications in Biological Science and Medicine*. CRC press.

Raulet, D. H. (2003). Roles of the NKG2D immunoreceptor and its ligands. *Nature Reviews Immunology*, *3*(10), 781-790.

Raulet, D. H. and Guerra, N. (2009). Oncogenic stress sensed by the immune system: role of natural killer cell receptors. *Nature Reviews Immunology*, *9*(8), 568-580.

Reddel, R. R. (2000). The role of senescence and immortalization in carcinogenesis. *Carcinogenesis*, *21*(3), 477-484.

Reece, W. O. (2009). *Functional anatomy and physiology of domestic animals*. John Wiley & Sons.

Reis-Filho, J. S. and Pusztai, L. (2011). Gene expression profiling in breast cancer: classification, prognostication, and prediction. *The Lancet*, *378*(9805), 1812-1823.

Rhee, S. Y., Wood, V., Dolinski, K. and Draghici, S. (2008). Use and misuse of the gene ontology annotations. *Nature Reviews Genetics*, *9*(7), 509-515.

Rhodes, D. R., Barrette, T. R., Rubin, M. A., Ghosh, D. and Chinnaiyan, A. M. (2002). Meta-analysis of microarrays interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Research*, *62*(15), 4427-4433.

Rifai, N., Gillette, M. A. and Carr, S. A. (2006). Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature Biotechnology*, *24*(8), 971-983.

Rigel, D. S., Robinson, J. K., Ross, M. I., Friedman, R., Cockerell, C. J., Lim, H. and Stockfleth, E. (2011). *Cancer of the Skin: Expert Consult*. Elsevier Health Sciences.

Ringnér, M., Fredlund, E., Hakkinen, J., Borg, A. and Staaf, J. (2011). GOBO: gene expression-based outcome for breast cancer online. *PloS one*, *6*(3), e17911.

Rodier, F. and Campisi, J. (2011). Four faces of cellular senescence. *The Journal of Cell Biology*, *192*(4), 547-556.

Rölle, A., Mousavi-Jazi, M., Eriksson, M., Odeberg, J., Söderberg-Nauclér, C., Cosman, D., Kärre K. and Cerboni, C. (2003). Effects of human cytomegalovirus infection on ligands for the activating NKG2D receptor of NK cells: up-regulation of UL16-binding protein (ULBP) 1 and ULBP2 is counteracted by the viral UL16 protein. *The Journal of Immunology*, *171*(2), 902-908.

Romagnoli, M., Belguise, K., Yu, Z., Wang, X., Landesman-Bollag, E., Seldin, D. C., Barille-Nion, S., Jezequel, P., Seldin, M. L. and Sonenshein, G. E. (2012). Epithelial-to-mesenchymal transition induced by TGF-β1 is mediated by Blimp-1–dependent repression of BMP-5. *Cancer Research*, *72*(23), 6268-6278.

Rosen, P. P. (Ed.). (2001). *Rosen's breast pathology*. Lippincott Williams & Wilkins.

Rubinson, D. A., Dillon, C. P., Kwiatkowski, A. V., Sievers, C., Yang, L., Kopinja, J., Rooney, D.L., Zhang, M., Ihrig, M.M., McManus, M.T., Gertler, F.B., Scott, M.L. and Van Parijs, L. (2003). A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nature Genetics*, *33*(3), 401-406.

Safarpour, D., Pakneshan, S., and Tavassoli, F. A. (2014). Androgen receptor (AR) expression in 400 breast carcinomas: is routine AR assessment justified? *American Journal of Cancer Research*, *4*(4), 353.

Safran, M., Dalah, I., Alexander, J., Rosen, N., Stein, T. I., Shmoish, M., Nativ N., Bahir, I., Doniger, T., Krug, H. and Lancet, D. (2010). GeneCards Version 3: the human gene integrator. *Database*, *2010*, baq020.

Safran, M., Dalah, I., Alexander, J., Rosen, N., Stein, T. I., Shmoish, M., Nativ, N., Bahir, I., Doniger, T., Krug, H., Sirota-Madi, A., Olender, T., Golan, Y., Stelzer, G., Harel, A. and Lancet, D. (2010). GeneCards Version 3: the human gene integrator. *Database*, *2010*, baq020.

Sahoo, C. K., Rao, S. R., Sudhakar, M., Sahoo, T. K. and Sahoo, P. K. (2015). A review on breast cancer. *International Journal of Pharmaceutical Development & Technology*. 5(1), 35-51.

Sakamoto, K. M. and Frank, D. A. (2009). CREB in the pathophysiology of cancer: implications for targeting transcription factors for cancer therapy. *Clinical Cancer Research*, *15*(8), 2583-2587.

Santos, N. C., Figueira-Coelho, J., Martins-Silva, J. and Saldanha, C. (2003). Multidisciplinary utilization of dimethyl sulfoxide: pharmacological, cellular, and molecular aspects. *Biochemical Pharmacology*, *65*(7), 1035-1041.

Saslow, D., Boetes, C., Burke, W., Harms, S., Leach, M. O., Lehman, C. D., Morris, E., Pisano, E., Schnall, M., Sener, S., Smith, R.A., Warner, E., Yaffe, M., Andrews, K.S., Russell, C.A. and Russell, C. A. (2007). American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA: A Cancer Journal for Clinicians*, *57*(2), 75-89.

Schiess, R., Wollscheid, B. and Aebersold, R. (2009). Targeted proteomic strategy for clinical biomarker discovery. *Molecular Oncology*, *3*(1), 33-44.

Schneider, B. P., Winer, E. P., Foulkes, W. D., Garber, J., Perou, C. M., Richardson, A., Sledge, G.W. and Carey, L. A. (2008). Triple-negative breast cancer: risk factors to potential targets. *Clinical Cancer Research*, 14(24), 8010-8018.

Schreiber, R. D., Old, L. J. and Smyth, M. J. (2011). Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science*, *331*(6024), 1565-1570.

Schrohl, A. S., Würtz, S., Kohn, E., Banks, R. E., Nielsen, H. J., Sweep, F. C. and Brünner, N. (2008). Banking of biological fluids for studies of disease-associated protein biomarkers. *Molecular & Cellular Proteomics*, *7*(10), 2061-2066.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, *473*(7347), 337-342.

Sestak, I., Cuzick, J. and Evans, G. (2012). Breast Cancer: Epidemiology, Risk Factors and Genetics. *ABC of Breast Diseases*, *100*, 41.

Seth, A. and Watson, D. K. (2005). ETS transcription factors and their emerging roles in human cancer. *European Journal of Cancer*, *41*(16), 2462-2478.

Shankavaram, U. T., Reinhold, W. C., Nishizuka, S., Major, S., Morita, D., Chary, K. K., Reimers, M.A., Scherf, U., Kahn, A., Dolginow, D., Cossman, J., Kaldjian, E.P., Scudiero, D.A., Petricoin, E., Liotta, L., Lee, J.K. and Weinstein, J. N. (2007). Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study. *Molecular Cancer Therapeutics*, *6*(3), 820-832.

Sharma, S., Kelly, T. K. and Jones, P. A. (2010). Epigenetics in cancer. *Carcinogenesis*, *31*(1), 27-36.

Shaulian, E. and Karin, M. (2001). AP-1 in cell proliferation and survival. *Oncogene*, *20*(19), 2390-2400.

Shay, J. W. and Roninson, I. B. (2004). Hallmarks of senescence in carcinogenesis and cancer therapy. *Oncogene*, *23*(16), 2919-2933.

Shay, J.W., Zou, Y., Hiyama, E. and Wright, W.E. (2001). Telomerase and Cancer. *Human Molecular Genetics*. 10 (7): 677-685

Sherr, C.J. and McCormick, F. (2002). The RB and p53 pathways in cancer. *Cancer Cell* 2, 103–112.

Sievers, C., Billig, G., Gottschalk, K. and Rudel, T. (2010). Prohibitins are required for cancer cell proliferation and adhesion. *PLoS One*, *5*(9), e12735.

Simpson, R. J., Bernhard, O. K., Greening, D. W. and Moritz, R. L. (2008). Proteomics-driven cancer biomarker discovery: looking to the future. *Current Opinion in Chemical Biology*, *12*(1), 72-77.

Skov, S., Pedersen, M. T., Andresen, L., Straten, P. T., Woetmann, A. and Ødum, N. (2005). Cancer cells become susceptible to natural killer cell killing after exposure to histone deacetylase inhibitors due to glycogen synthase kinase-3–dependent expression of MHC class I–related chain A and B. *Cancer Research*, *65*(23), 11136-11145.

Smith, L. M., Wise, S. C., Hendricks, D. T., Sabichi, A. L., Bos, T., Reddy, P., Brown, P.H. and Birrer, M. J. (1999). cJun overexpression in MCF-7 breast cancer cells produces a tumourigenic, invasive and hormone resistant phenotype. *Oncogene*, *18*(44), 6063-6070.

Smith, L., Watson, M. B., O'Kane, S. L., Drew, P. J., Lind, M. J. and Cawkwell, L. (2006). The analysis of doxorubicin resistance in human breast cancer cells using antibody microarrays. *Molecular Cancer Therapeutics*, *5*(8), 2115-2120.

Smith, R. A., Saslow, D., Sawyer, K. A., Burke, W., Costanza, M. E., Evans, W. P. I. I. I., Foster, R.S., Eyre, E.H.H.J. and Sener, S. (2003). American Cancer Society guidelines for breast cancer screening: update 2003. *CA: A Cancer Journal for Clinicians*, *53*(3), 141-169.

Smyth, M. J., Hayakawa, Y., Takeda, K. and Yagita, H. (2002). New aspects of natural-killer-cell surveillance and therapy of cancer. *Nature Reviews Cancer*, *2*(11), 850-861.

Snow, A. L., Pandiyan, P., Zheng, L., Krummey, S. M. and Lenardo, M. J. (2010). The power and the promise of restimulation-induced cell death in human immune diseases. *Immunological Reviews*, *236*(1), 68-82.

Soriani, A., Zingoni, A., Cerboni, C., Iannitto, M. L., Ricciardi, M. R., Di Gialleonardo, V., Cippitelli, M., Fionda, C., Petrucci, M.T., Guarini, A., Foà, R. and Santoni, A. (2009). ATM-ATR–dependent up-regulation of DNAM-1 and NKG2D ligands on multiple myeloma cells by therapeutic agents results in enhanced NK-cell susceptibility and is associated with a senescent phenotype. *Blood*, *113*(15), 3503-3511.

Sotiriou, C., Neo, S. Y., McShane, L. M., Korn, E. L., Long, P. M., Jazaeri, A., Martiat, P., Fox, S.B., Harris, A.L. and Liu, E. T. (2003). Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proceedings of the National Academy of Sciences*, *100*(18), 10393-10398.

Speirs, V. and Kerin, M. J. (2000). Prognostic significance of oestrogen receptor β in breast cancer. *British Journal of Surgery*, *87*(4), 405-409.

Sporn, M.B. (1996). The war on cancer. *Lancet 347*, 1377–1381.

Staker, B. L., Hjerrild, K., Feese, M. D., Behnke, C. A., Burgin, A. B. and Stewart, L. (2002). The mechanism of topoisomerase I poisoning by a camptothecin analog. *Proceedings of the National Academy of Sciences*, *99*(24), 15387-15392.

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. *Journal of the Society for Integrative Oncology*, *6*(4), 158.

Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksöz, E, Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H. and Wanker, E. E. (2005). A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, *122*(6), 957-968.

Stewart, S. A., Dykxhoorn, D. M., Palliser, D., Mizuno, H., Yu, E. Y., An, D. S., Sabatini, D.M., Chen, I.S., Hahn, W.C., Sharp, P.A., Weinberg, R.A. and Novina, C. D. (2003). Lentivirus-delivered stable gene silencing by RNAi in primary cells. *Rna*, *9*(4), 493-501.

Strasser, A., Cory, S. and Adams, J.A. (2011). Deciphering the rules of programmed cell death to improve therapy of cancer and other diseases. *EMBO Journal* 30(18); 3667-3683

Sugimoto, M., Kawakami, M., Robert, M., Soga, T. and Tomita, M. (2012). Bioinformatics tools for mass spectroscopy-based metabolomic data processing and analysis. *Current Bioinformatics*, *7*(1), 96.

Tait, S. W. and Green, D. R. (2010). Mitochondria and cell death: outer membrane permeabilization and beyond. *Nature Reviews Molecular Cell Biology*, *11*(9), 621-632.

Talmadge, J.E. and Fidler, I.J. (2010). AACR centennial series: the biology of cancer metastasis: historical perspective. *Cancer Res*. 70, 5649–5669.

Taylor, R. C., Robertson, B. J. M. W., Markillie, L. M., Serres, M. H., Linggi, B. E., Aldrich, J. T., Hill E.A., Romine, M.F., Lipton, M.S. and Wiley, H. S. (2013).

Changes in translational efficiency is a dominant regulatory mechanism in the environmental response of bacteria. *Integrative Biology*, *5*(11), 1393-1406.

Textor, S., Fiegler, N., Arnold, A., Porgador, A., Hofmann, T. G. and Cerwenka, A. (2011). Human NK cells are alerted to induction of p53 in cancer cells by upregulation of the NKG2D ligands ULBP1 and ULBP2. *Cancer Research*, *71*(18), 5998-6009.

Thomas, C. J., Rahier, N. J. and Hecht, S. M. (2004). Camptothecin: current perspectives. *Bioorganic & Medicinal Chemistry*, *12*(7), 1585-1604.

Thompson, E. W. and Newgreen, D. F. (2005). Carcinoma invasion and metastasis: a role for epithelial-mesenchymal transition? *Cancer Research*, *65*(14), 5991-5995.

Thorsteinsdottir, U., Krosl, J., Kroon, E., Haman, A., Hoang, T. and Sauvageau, G. (1999). The oncoprotein E2A-Pbx1a collaborates with Hoxa9 to acutely transform primary bone marrow cells. *Molecular and Cellular Biology*, *19*(9), 6355-6366.

Tong, A. W., Zhang, Y. A. and Nemunaitis, J. (2005). Small interfering RNA for experimental cancer therapy. *Current Opinion In Molecular Therapeutics*, *7*(2), 114-124.

Turashvili, G., McKinney, S. E., Goktepe, O., Leung, S. C., Huntsman, D. G., Gelmon, K. A., Los, G., Rejto, P.A. and Aparicio, S. A. (2011). P-cadherin expression as a prognostic biomarker in a 3992 case tissue microarray series of breast cancer. *Modern Pathology*, *24*(1), 64-81.

Van De Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E.T., Friend, S.H. and Bernards, R. (2002). A gene-expression signature as a

predictor of survival in breast cancer. *New England Journal of Medicine*, *347*(25), 1999-2009.

Varakala, R. (2008). Proteome based improvements for immobilized metal affinity chromatography in an Esherichia coli expression system. ProQuest.

Verma, S. (2007). Bioinformatics Approaches to Biomarker Discovery.

Vesely, M. D., Kershaw, M. H., Schreiber, R. D. and Smyth, M. J. (2011). Natural innate and adaptive immunity to cancer. *Annual Review of Immunology*, *29*, 235-271.

Vincze, T., Posfai, J. and Roberts, R. J. (2003). NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Research*, *31*(13), 3688-3691.

Visvader, J. E. (2011). Cells of origin in cancer. *Nature*, *469*(7330), 314-322.

Vogel, C. L., Cobleigh, M. A., Tripathy, D., Gutheil, J. C., Harris, L. N., Fehrenbacher, L., Slamon, D.J., Murphy, M., Novotny, W.F., Burchmore, M., Shak, S., Stewart, S.J. and Press, M. (2002). Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *Journal of Clinical Oncology*, *20*(3), 719-726.

Vogel, C. and Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, *13*(4), 227-232.

Vogelstein B, Lane D, Levine AJ. (2000). Surfing the p53 network. *Nature* 408: 307–310.

Von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M.A. and Bork, P. (2005). STRING: known and predicted

protein–protein associations, integrated and transferred across organisms. *Nucleic Acids Research*, *33*(suppl 1), D433-D437.

Von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S. and Bork, P. (2002). Comparative assessment of large-scale data sets of protein–protein interactions. *Nature*, *417*(6887), 399-403.

Vorobiof, D.A., Sitas, F. and Vorobiof G. (2001). Breast cancer incidence in South Africa. *J Clin Oncol*; 19(18 Suppl):125S–127S

Vuong, L., Conley, S. M. and Al-Ubaidi, M. R. (2012). Expression and role of p53 in the retina. *Investigative Ophthalmology & Visual Science*, *53*(3), 1362-1371.

Waldner, M., Schimanski, C. C. and Neurath, M. F. (2006). Colon cancer and the immune system: the role of tumour invading T cells. *World Journal of Gastroenterology*, *12*(45), 7233.

Walsh, T. and King M.C. (2007). Ten genes for inherited breast cancer. *Cancer Cell,* 11(2):103–105.

Wang, C. C., Lin, S. Y., Lai, Y. H., Liu, Y. J., Hsu, Y. L. and Chen, J. J. (2012). Dimethyl sulfoxide promotes the multiple functions of the tumour suppressor HLJ1 through activator protein-1 activation in NSCLC cells. *PloS one*, *7*(4), e33772.

Wang, R. and Sun, P. D. (2014). Natural Killer Cell-Mediated Shedding of ULBP2. *PloS one*, *9*(3), e91133.

Wang, S., Konorev, E. A., Kotamraju, S., Joseph, J., Kalivendi, S. and Kalyanaraman, B. (2004). Doxorubicin induces apoptosis in normal and tumour cells via distinctly

different mechanisms intermediacy of H2O2-and p53-dependent pathways. *Journal of Biological Chemistry*, *279*(24), 25535-25543.

Wang, T., Niu, G., Kortylewski, M., Burdelya, L., Shain, K., Zhang, S., Bhattacharya, R., Gabrilovich, D., Heller, R., Coppola, D., Dalton, W., Jove, R., Pardoll, D. and Yu, H. (2004). Regulation of the innate and adaptive immune responses by Stat-3 signaling in tumour cells. *Nature Medicine*, *10*(1), 48-54.

Wehland, M., Bauer, J., Magnusson, N. E., Infanger, M. and Grimm, D. (2013). Biomarkers for anti-angiogenic therapy in cancer. *International Journal of Molecular Sciences*, *14*(5), 9338-9364.

Weigel, M. T. and Dowsett, M. (2010). Current and emerging biomarkers in breast cancer: prognosis and prediction. *Endocrine-related cancer*, *17*(4), R245-R262.

Wong, R.S.J (2011). Apoptosis in cancer: from pathogenesis to Treatment. *Journal of Experimental & Clinical Cancer Research*, 30:87

Yang, W. T., Le-Petross, H. T., Macapinlac, H., Carkaci, S., Gonzalez-Angulo, A. M., Dawood, S., Resetkova, E., Hortobagyi, G.N. and Cristofanilli, M. (2008). Inflammatory breast cancer: PET/CT, MRI, mammography, and sonography findings. *Breast cancer research and treatment*, *109*(3), 417-426.

Yi, J. K., Chang, J. W., Han, W., Lee, J. W., Ko, E., Kim, D. H., Bae, J.Y., Yu, J., Lee, C., Yu, M.H. and Noh, D. Y. (2009). Autoantibody to tumour antigen, alpha 2-HS glycoprotein: a novel biomarker of breast cancer screening and diagnosis. *Cancer Epidemiology Biomarkers & Prevention*, *18*(5), 1357-1364.

Yoshimaru, T., Komatsu, M., Matsuo, T., Chen, Y. A., Murakami, Y., Mizuguchi, K., Mizohata, E., Inoue, T., Akiyama, M., Yamaguchi, R., Imoto, S., Miyano, S., Miyoshi,

Y., Sasa, M., Nakamura, Y. and Katagiri, T. (2013). Targeting BIG3–PHB2 interaction to overcome tamoxifen resistance in breast cancer cells. *Nature communications*, *4*, 1-13

Youle, R. J. and Strasser, A. (2008). The BCL-2 protein family: opposing activities that mediate cell death. *Nature Reviews Molecular Cell Biology*, *9*(1), 47-59.

Zhang, D., Salto-Tellez, M., Putti, T. C., Do, E. and Koay, E. S. C. (2003). Reliability of tissue microarrays in detecting protein expression and gene amplification in breast cancer. *Modern Pathology*, *16*(1), 79-85.

Zhang, H. and Chan, D. W. (2007). Cancer biomarker discovery in plasma using a tissue-targeted proteomic approach. *Cancer Epidemiology Biomarkers & Prevention*, *16*(10), 1915-1917.

Zhang, R. H. and He, S. G. (2001). Modulation of MDR-1 gene in human breast cancer cells by sodium butyrate and DMSO. *Chinese Journal of Cancer Research*, *13*(1), 27-30.

Zitvogel, L., Apetoh, L., Ghiringhelli, F. and Kroemer, G. (2008). Immunological aspects of cancer chemotherapy. *Nature Reviews Immunology*, *8*(1), 59-73.

Zitvogel, L., Kepp, O. and Kroemer, G. (2011). Immune parameters affecting the efficacy of chemotherapeutic regimens. *Nature Reviews Clinical Oncology*, *8*(3), 151-160.